



UiT The Arctic University of Norway

Fakultet for naturvitenskap og teknologi

Institutt for Informatikk

Epidemiology Network Analysis

Rafael Adolfo Nozal Cañadas

A dissertation for the degree of Philosophiae Doctor

2024 August 01



Preface

Social influence has been a notion nesting in human brains since we started living in a community. In the Poetic Edda, the poems of Fáfnismál [1] [2] and Völuspá [3] describe the existence of beings known as the “Nornir”, who are in charge of weaving the threads of destiny for gods and humans alike. Völuspá describes the Nornir as three individuals: Urðr (“Fate”), Verðandi (“What Is To Come”), and Skuld (“What Needs To Occur”). In Germanic folklore, fate is not fixed nor unique. Because of the influence of others, and because of our own past actions, humans do not have complete freedom to do what they want, but we are not enslaved by divine determinism either. The Nornir are constantly knitting the tapestry of life, in which each person’s string is tugging with those of others, always changing and defining who we are or will be.

A thousand years later, Jean Jacques Rousseau argues that we surrender our freedom to the "community". Rousseau synthesized the transition from natural freedom to civil liberty with the phrase: "*Freedom consists not so much in doing one's will as in not being subjected to that of others; it still consists in not submitting the will of others to ours*" [4]. Rousseau argued that one must surrender his freedom and act in the best interest of the "general will" because by definition the general will can never be wrong. This "general will" has been used as justification for preserving liberty and to build the foundations for 20th-century totalitarianism enforcing oppression. In contrast, John Locke argued that there are rights undeniable to each individual, and no government, not even the general will, has the right to take them away.

This begs the question, if other people always influence us and we are vulnerable to others' ideas, is freedom an illusion? In my thesis, I argue that your health is not different and will never be alone. Your social network has been estimated to weigh between 15% to 40% [5] of the relative contributions from health determinants (such as genetics, environment, or medical care) to health outcomes. You are forced to be part of a community and you need to understand the trade-off of your options. People give you the risk of being infected by pathogens, but isolation makes you more likely to have

alcohol addiction or suffer from depression. Friends will influence you to skip diets and drinks at the bar, advertisers will constantly push you to consume junk food. Do you eat what you want or do you eat what is enforced upon you? Do you choose your health or is it imposed by others?

You are locked in the tapestry with other people and they are always influencing your fate, but you are still in charge of it. Being part of a group doesn't excuse you from being responsible for your actions. Being aware of that influence allows you to act with freedom and not necessarily by obeying the will of others. I invite you to read this work and how we measure such influences and their consequences, how to recognize them, and how to avoid them, in the hope that you use this knowledge to preserve the elemental health rights of individuals.

This thesis is the product of my work at the Department of Computer Science, UiT The Arctic University of Norway, Tromsø, to obtain the Philosophiae Doctor of Science.

Rafael Adolfo Nozal Cañadas

2024 January 10

Abstract

Research questions: The primary objective of this doctoral dissertation is an explorative investigation into the social network dynamics within eight high schools, located in Tromsø and Balsfjord (North Norway), and the extent to which these dynamics contribute to the overall health and well-being of the students, such as in the context of infectious disease spread and the transmission of negative or positive health effects, and also in comparison with non-social host factors such as sports or recreational drug frequencies. Secondarily, we aim to develop new analytical methods and provide a framework for enabling agnostic evaluation of social networks in epidemiological studies and faster iterations of developing scripts for general statistical research.

Methodology: Using the Fit Futures gathered data on friendship, we used simulations, homophily, X^2 tables, logistic regression, and random forests as the main methods to analyze social influence in our topics of interest. We applied classical database normalization and data cleaning to the original data and developed scripts for automatic analysis in R and Python exporting results directly in plain text, Latex, and HTML.

Results: We found that the social network influences significantly the spread of *Staphylococcus aureus* (*S. aureus*). Students close in the network tend to have similar inflammatory biomarkers, 25-hydroxyvitamin D (25(OH)D), and Body mass index (BMI) levels. Some high schools tend to consume similar levels of over-the-counter medicines and tend to share the same brand of prescribed medicines. There is also a bias on recreational drug usage by high school.

Conclusions: Social influence is shown to be significant in every analysis. These findings emphasize the importance of considering social network dynamics in understanding and addressing health and well-being issues among students. Further research and interventions targeting social network influences can contribute to developing more effective health strategies.

Originality: Use of non-parametric simulation and machine learning methods to estimate social influence. We are measuring social influence on 25(OH)D, in an inflammatory proteomic assay.

Significance: Social influence, whether from virtual friends or physical ones, is a growing area of interest in many fields. In Epidemiology in particular we saw a boost in popularity after the Sars-Cov-2 pandemic.

Keywords: Social Networks, *S. aureus*, statistics, epidemiology, vitamin D, obesity, inflammation, random forests, prescriptions, drugs.

Acknowledgements

A fair listing of all the positive qualities and contributions of everyone connected to this thesis would render another 200 pages, and would still be the short version of it. Instead, I hope that the people named here would forgive me for describing only one or two of the many good characteristics that helped in this PhD journey.

First, I want to express my gratitude to Befolkningsundersøkelser i nord (BiN) and the UiT The Arctic University of Norway for providing financial support, and The Regional Committee of Medical and Health Research Ethics (REK) and the Norwegian Data Protection Authority for approval of the data collection (reference: 2018/1975/REK Nord).

I would like to express my sincere gratitude to my main supervisor Lars-Ailo Bongo for his invaluable contributions to this project. His writing feedback has been incredibly thorough and detailed, providing me with the guidance and support I needed to refine my work and take it to the next level. In addition, his high level of expertise and knowledge in many fields has been instrumental in shaping the direction and scope of this project. His insights and recommendations have been invaluable, and I feel fortunate to have had the opportunity to work with someone of such exceptional talent and dedication.

Co-supervisors Anne-Sofie Furberg and Anne Merethe Hanssen have provided valuable insights and expertise in the areas of general medicine and microbiology, which have helped to shape this project, and been a source of moral support throughout this project, providing encouragement and motivation. They have also introduced me to other individuals, listed below, who have later made significant contributions to this project.

Christopher Nilsen for his invaluable contributions to this doctoral thesis. Christopher played a pivotal role in the conceptualization of Fit Futures. In particular, he also insisted on the inclusion of the social network data that made this PhD research possible. His foresight was critical in shaping the direction of this research and many others as he is persistently securing funding for new projects.

To my department leader Anders Andersen and his efforts to improve the department's resources thanks to which I had the means to do this PhD.

Other co-authors in this project are Dina Benedicte Berg Stensen, Mohsen Askar. I would like to express my deep appreciation to them for their invaluable contributions to the scientific articles we have worked on together. Their expertise in the fields of endocrinology and pharmacy has been essential in shaping the direction and scope of our research, and their collaboration has been instrumental in bringing these papers to fruition. Their dedication, hard work, and attention to detail have been truly inspiring, and I feel grateful to have had the opportunity to work with such talented and knowledgeable individuals.

Other co-authors in other projects, Rocío Raya Miranda and María Luz Gamiz Perez. I did my first statistical course ever with Rocío and I will be forever grateful for her introducing me to this world. María Luz continued teaching me biostatistics and survival analysis later on in my studies. Hopefully, I have become a productive member of this community and thanks to them I have grown to the point that I could teach them a thing or two and we share co-authorship in several articles.

Other collaborators Rocío Bonillo León, Svendsen Kristian, Lars Småbrekke, Guri Grimmes. I don't have a background in medicine, but thanks to them I gained a deeper understanding of many topics, including nutrition, pharmacology, immunology, endocrinology, and physiology. For me is important to get the best mental image possible of how things work rather than crunching the numbers and hoping that someone else deals with the results.

All other group coworkers for their continuous feedback, encouragement, and dedication to showing me and others how to push the boundaries of knowledge. Also, to the cleaning crew in the UiT, for their tireless efforts, their hard work and dedication often go unnoticed, but their contributions are invaluable to the smooth operation of the university. Without their daily efforts, the university would not be able to function effectively. And finally to the personnel in charge of making food in the cafeteria. Their commitment to providing nutritious and delicious meals to the university community is greatly appreciated and kept me running many endless days in the office.

All their valuable wisdom and efforts encourage me and others to reach our full potential.

List of papers

This thesis is based on the following papers, available in the appendix section A.

Paper A: Stensen DB, Cañas RAN, Småbrekke L, Olsen K, Nielsen CS, Svendsen K, Hanssen AM, Sollid JUE, Simonsen GS, Bongo LA, Furberg AS. Social network analysis of *Staphylococcus aureus* carriage in a general youth population. International Journal of Infectious Diseases Volume 123, October 2022, Pages 200-209
<https://doi.org/10.1016/j.ijid.2022.08.018>

Paper B: The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels. Cañas RAN, Nielsen CS, Furberg AS, Hanssen AM, Bongo LA
<https://www.medrxiv.org/content/10.1101/2023.11.29.23299188v1.full.pdf>

Paper C: Askar M, Cañas RAN, Svendsen K. An introduction to network analysis for studies of medication use. Research in Social and Administrative Pharmacy Volume 17, Issue 12, December 2021, Pages 2054-2061 <https://doi.org/10.1016/j.sapharm.2021.06.021>

In addition, the following results are a summary of another four manuscripts that have not yet been published, which can be found in the appendix section B.

Result I: Social network influences on obesity in a general youth population

Result II: Social network influences on inflammatory response in a general youth population

Result III: Measuring social influence with random forest regression and artificial neural networks.

Result IV: Frequency consumption of medication and social network influence in a general youth population.

Abbreviations

S. aureus *Staphylococcus aureus*. iii, iv, 2, 3, 86, 87, 95, 97, 98, 105, 106, 156, 161, 173–175

1,25(OH)2D 1,25-dihydroxyvitamin D. 107

25(OH)D 25-hydroxyvitamin D. iii, iv, 107, 157

Ag Antigens. 54

AI Artificial Intelligence. 38

ALA alpha-linolenic acid. 69

ANNs Artificial Neural Networks. 41, 344

AP-1 activator protein-1. 80

APCs Antigen presenting cells. 54

APL Average Path Length. 29, 30

APR Acute phase reactants. 70, 166

BCNF Boyce - Codd normal form. 171

BDNF Brain-derived Neurotrophic factor. 165

BiN Befolkningsundersøkelser i nord. v

BMI Body mass index. iii, 156, 166, 175

C3a Complement component 3 A. 63

C3b Complement component 3 B. 96

- CAMs** cell adhesion molecules. 63
- CD** cluster of differentiation. 54
- CIfA** Clumping factor proteins A. 95
- CIfB** Clumping factor proteins B. 95
- CNA** Collagen adhesin. 96
- COX** cyclooxygenase. 125
- CRP** C-reactive protein. 71, 165
- CSF1** Macrophage colony-stimulating factor 1. 165
- CVD** Cardiovascular diseases. 119
- DAMPs** Internal Damage Associated Molecular Patterns. 83
- DDIs** drug-drug interactions. 158, 164
- DHA** docosahexaenoic acid. 69
- DPN** Drug Prescription Networks. 158
- DrL** Distributed Recursive Layout. xxiv, 17, 18
- DXM** Dextromethorphan. 125
- EPA** eicosapentaenoic acid. 69
- EPS** extracellular polymeric substances. 93
- ERGM** Exponential Random Graph Models. 32
- ESR** erythrocyte sedimentation rate. 71, 166
- FEST** Norwegian Electronic Prescription Support System. 158
- FF** Fit Futures. 130, 173
- FF1** Fit Futures 1. xxiv, 20, 156

FF2 Fit Futures 2. 130, 344

FGF fibroblast growth factors. 55

FnBPA Fibronectin binding protein A. 94, 173

GAS group A streptococcus. 92

GBS group B streptococcus. 92

GIT Gastrointestinal Track. 97

GMM Generalized Method of Moments. 51

HACEK Haemophilus, Aggregatibacter Cardiobacterium, Eikenella, and Kingella organisms group. 100

HC Hormonal contraceptives. 133

HDL High-Density Lipoprotein. 119

HIV Human Immunodeficiency Virus. 2, 54, 161

HVEM herpes virus entry mediator. 81

ICD-10 International Statistical Classification of Diseases and Related Health Problems. 138

ICE Individual Conditional Expectation. 44

ICP intracranial pressure. 100

IDE Integrated Development Environment. 172

IE infective endocarditis. 95

IFN- α Interferon alpha. 79

IFN- β Interferon beta. 79

IFN- γ Interferon gamma. 62, 67, 80, 98

Ig Immunoglobulin. 59

IgE Immunoglobulin E. 123

IgG Immunoglobulin G. 94, 96, 98

IL-1 Interleukin 1. 72, 86, 89, 98

IL-2 Interleukin 2. 73, 98

IL-2R interleukin-2 receptor. 73

IL-6 Interleukin 6. 86, 165

IRF Interferon regulatory factor. 79

IsdA Iron-regulated surface protein A. 97

IsdB Iron-regulated surface protein B. 97

ISGs Interferon-stimulated genes. 80

JAK-STAT Janus kinase signal transducers and activators of the transcription proteins.
67

KNN K-Nearest Neighbor. 42

LDL Low-Density Lipoprotein. 119

LIME Local interpretable model-agnostic explanations. 44

LOD Limit of Detection. 359

LPS lipopolysaccharides. 59, 75, 83, 89

LT Leukotrienes. 68

LT α Lymphotoxin- α . 81

LT β Lymphotoxin- β . 81

LT β R Lymphotoxin- β receptor. 81

LZRSA *Linezolid resistance in Staphylococcus aureus*. 103

M-BMs memory-based dietary assessment methods. 157, 163

MAE Mean Absolute Error. 345

MAOIs Monoamine oxidase inhibitors. 126

MDI Mean Decrease in Impurity. 45, 159, 166, 167, 344

MDS Multidimensional Scaling. xxiv, 17, 20

MEDAS 14-Item Mediterranean Diet Adherence Screener. 167

MET Metabolic Equivalent of Task. 157, 163

MHC major histocompatibility complex. 54

MHC2 major histocompatibility complex 2. 98

ML Machine Learning. 38

MLE Maximum Likelihood Estimation. 51

MRSA *Methicillin-resistant Staphylococcus aureus*. 86, 103, 105, 161

MS Multiple sclerosis. 120

MSCRAMM Microbial Surface Component Recognizing Adhesive Matrix Molecules. 94, 100

MSSA *Methicillin-sensitive Staphylococcus aureus*. 103

NA Network Analysis. 158

NAPQI N-acetyl-p-benzoquinone imine. 123

NETs neutrophil extracellular traps. 96

NF-κB Nuclear factor kappa-light-chain-enhancer of activated B cells. 56

NK Natural Killer. 59

NMOSD Neuromyelitis optica spectrum disorders. 74

NorPD Norwegian Prescription Database. 158

NPX Normalized Protein eXpression. 359

NSAIDs Nonsteroidal anti-inflammatory drug. 68, 102, 123, 124

OTC Over-the-counter. 122, 125, 128, 175

PA Physical Activity. 156, 157, 163

PAMPs Pathogen Associated Molecular Patterns. 74, 83

PBP penicillin-binding protein. 90, 103

PCA Principal Component Analysis. 39

PDP Partial Dependence Plot. 44

PEA Proximity Extension Assay. 359

PG prostaglandins. 67, 125

PLS Partial Linear Regression. 39

PRRs Patter Recognition Receptors. 79, 84

PTH Parathyroid hormone. 107

PVL Panton–Valentine leukocidin. 97, 103

RAAS Renin-Angiotensin-Aldosterone system. 119

RANKL Receptor activator of nuclear factor kappa-B ligand. 109, 378

REK The Regional Committee of Medical and Health Research Ethics. v

RF Random Forests. 42, 344

RLRs RIG-I-like receptors. 84

ROS Reactive oxygen species. 53

SasC Staphylococcal surface protein C. 94

- SasG** Staphylococcal surface protein G. 94
- SasX** Staphylococcal surface protein X. 94
- SEA** Staphylococcal enterotoxin A. 102
- SEB** Staphylococcal enterotoxin B. 102
- SHAP** SHapley Additive exPlanations. 44, 159, 166, 167, 344
- SNA** Social Network Analysis. 1–4, 164, 175
- Spa** Staphylococcal protein A. 94, 135, 156
- SPMs** Specialized pro-resolving mediators. 85, 86, 173
- SSSS** Staphylococcal scalded skin syndrome. 101, 102
- SVM** Support-vector machines. 42
- TCR** T cell receptor. 54
- Th** Helper T cell. 61
- TLRs** Toll-like receptors. 80, 84
- TNF α** Tumor Necrosis Factor α . 62, 86, 89, 98
- TSST-1** Toxic Shock Syndrome Type-1. 98, 102
- UiT** UiT: The Arctic university of Norway. 137
- UNN** University Hospital of North Norway. 131, 137
- UVB** Ultraviolet B. 157, 163
- VDR** Vitamin D receptor. 107
- VDSP** Vitamin D Standardization Program. 113, 137
- vgs** viridans group streptococci. 92
- VRSA** *Vancomycin-resistant Staphylococcus aureus*. 103

vWbp von Willebrand factor-binding protein. 95

vWD von Willebrand disease. 62

vWF von Willebrand factor. 62, 84

WHO World Health Organization. 137

Contents

Preface	i
Abstract	iii
Acknowledgements	v
List of papers	vii
List of abbreviations	vii
List of Figures	xxiii
List of Tables	xxxi
1 Introduction	1
1.1 Social Network Analysis in health	1
1.2 Social network interventions	2
1.3 Thesis impact	2
2 Aims of the thesis	3
3 Background	5
3.1 Graph theory	5
3.1.1 Historical Overview	5
3.1.2 Definition	6
3.1.3 Nodes	7
3.1.4 Edges	8
3.1.5 Network	12
3.1.6 Layout	14
3.1.7 Networks metrics	22
3.2 Machine learning	38

3.2.1	Introduction	38
3.2.2	Basic concepts	38
3.2.3	Algorithms	41
3.2.4	Interpretability	43
3.3	Statistical Analysis	46
3.3.1	Introduction	46
3.3.2	Contingency tables and Pearson X^2 test	46
3.3.3	Logistic regression	48
3.3.4	Autocorrelation models	49
3.4	Inflammation	52
3.4.1	Introduction	52
3.4.2	Immunology basic concepts	53
3.4.3	Immunology advance concepts	66
3.4.4	Inflammation cycle	82
3.5	<i>Staphylococcus aureus</i>	86
3.5.1	Introduction	86
3.5.2	Microbiology basic concepts	87
3.5.3	<i>S. aureus</i> main characteristics	93
3.5.4	<i>Staphylococcus aureus</i> diseases	98
3.5.5	Antibiotics resistance strains	103
3.5.6	Vaccines	103
3.5.7	Epidemiology	104
3.6	Vitamin D	106
3.6.1	Introduction	106
3.6.2	Sources and metabolism	107
3.6.3	VDR	107
3.6.4	Calcium	109
3.6.5	Phosphate	111
3.6.6	Deficiency and toxicity	113
3.6.7	Epidemiology	116
3.7	Over-the-counter medicine	122
3.7.1	Introduction	122
3.7.2	Principles and direct adverse effects	123
3.7.3	Adverse drug interactions	126
3.7.4	Other ethical considerations	127

4.1	Fit Futures	130
4.1.1	Social network assessment	131
4.1.2	Host risk factors	133
4.1.3	Hormonal contraceptive	133
4.1.4	<i>S. aureus</i> assessment	135
4.1.5	SPA typing	135
4.1.6	Anthropometry assessment	137
4.1.7	Vitamin D assessment	137
4.1.8	OLINK Target 96 Inflammation	138
4.2	Data cleaning	138
4.2.1	Naming	138
4.2.2	Dates	139
4.2.3	General tables	139
4.2.4	Relational Normalization	140
4.2.5	IDs transformation	141
4.2.6	Mapping	142
4.2.7	Diseases	144
4.2.8	Medicines	146
4.3	Statistical and Machine Learning	147
4.3.1	Software	147
4.3.2	Simulations	147
4.3.3	Friendship ratio	153
4.3.4	Statistical methods	155
4.4	Ethical considerations	155
5	Summary of main papers	156
5.1	Paper A	156
5.2	Paper B	157
5.3	Paper C	158
5.4	Result I	158
5.5	Result II	158
5.6	Result III	159
5.7	Result IV	159
6	Discussion	161
6.1	Papers and results	161
6.1.1	Paper A	161

6.1.2	Paper B	162
6.1.3	Paper C	163
6.1.4	Result I	164
6.1.5	Result II	165
6.1.6	Result III	166
6.1.7	Result IV	167
6.2	Network modelization	168
6.3	Direct influence vs common environment	169
6.4	Optimization of resources in public interventions	170
6.5	Challenges in privacy	171
6.6	Challenges in reproducibility	171
6.7	Challenges in framework developing	172
6.8	Future projects	173
7	Conclusions	175
Bibliography		175
A	Main papers	242
A.1	Paper A	242
A.2	Paper B	270
A.3	Paper C	322
B	Preliminary Results	331
B.1	Result I	331
B.2	Result II	334
B.3	Result III	344
B.4	Result IV	354
C	Olink panel	359
C.1	Introduction	359
C.2	Biomarkers	360
C.2.1	ADA - Adenosine Deaminase	360
C.2.2	ART -Artemin	360
C.2.3	AXIN1 - Axin-1	360
C.2.4	BDNF - Brain-derived neurotrophic factor	360
C.2.5	BNGF - Beta-nerve growth factor	361
C.2.6	CASP8 - Caspase-8	361

C.2.7	CCL2 / MCP1 - Monocyte chemotactic protein 1	361
C.2.8	CCL4 - C-C motif chemokine 4	362
C.2.9	CCL7 / MCP3 - Monocyte chemotactic protein 3	362
C.2.10	CCL8 / MCP2 - Monocyte chemotactic protein 2	362
C.2.11	CCL11 - Eotaxin	362
C.2.12	CCL13 / MCP4 - Monocyte chemotactic protein 4	363
C.2.13	CCL19 - C-C motif chemokine 19	363
C.2.14	CCL20 - C-C motif chemokine 20	363
C.2.15	CCL23 - C-C motif chemokine 23	363
C.2.16	CCL25 - C-C motif chemokine 25	363
C.2.17	CCL28 - C-C motif chemokine 28	364
C.2.18	CD244 - Natural killer cell receptor 2B4	364
C.2.19	CD5 - T-cell surface glycoprotein CD5	364
C.2.20	CD6 - T-cell differentiation antigen CD6	364
C.2.21	CD40 / TNFSF5R - Tumor necrosis factor receptor superfamily member 5	365
C.2.22	CDCP1 - CUB domain-containing protein 1	365
C.2.23	CSF1 - Macrophage colony-stimulating factor 1	365
C.2.24	CST5 - Cystatin D	365
C.2.25	CX3CL1 - Fractalkine	366
C.2.26	CXCL1 - C-X-C motif chemokine 1	366
C.2.27	CXCL5 - C-X-C motif chemokine 5	366
C.2.28	CXCL6 - C-X-C motif chemokine 6	366
C.2.29	CXCL9 - C-X-C motif chemokine 9	367
C.2.30	CXCL10 - C-X-C motif chemokine 10	367
C.2.31	CSCL11 - C-X-C motif chemokine 11	367
C.2.32	DNER - Delta and Notch-like epidermal growth factor-related receptor	367
C.2.33	EIF4EBP1 - Eukaryotic translation initiation factor 4E-binding protein 1	367
C.2.34	FGF5 - Fibroblast growth factor 5	368
C.2.35	FGF19 - Fibroblast growth factor 19	368
C.2.36	FGF21 - Fibroblast growth factor 21	368
C.2.37	FGF23 - Fibroblast growth factor 23	368
C.2.38	FLT3LG - Fms-related tyrosine kinase 3 ligand	369
C.2.39	GDNF - Glial cell line-derived neurotrophic factor	369

C.2.40	HGF - Hepatocyte growth factor	369
C.2.41	IFNG - Interferon gamma	369
C.2.42	IL1A - Interleukin-1 alpha	370
C.2.43	IL2 - Interleukin-2	370
C.2.44	IL2RB - Interleukin-2 receptor subunit beta	370
C.2.45	IL4 - Interleukin-4	370
C.2.46	IL5 - Interleukin-5	370
C.2.47	IL6 - Interleukin-6	370
C.2.48	IL7 - Interleukin-7	370
C.2.49	IL8 / CXCL8 - Interleukin-8	371
C.2.50	IL10 - Interleukin-10	371
C.2.51	IL10RA - Interleukin-10 receptor subunit alpha	371
C.2.52	IL10RB - Interleukin-10 receptor subunit beta	371
C.2.53	IL12B - Interleukin-12 subunit beta	371
C.2.54	IL13 - Interleukin-13	371
C.2.55	IL15RA - Interleukin-15 receptor subunit alpha	371
C.2.56	IL17A - Interleukin-17A	372
C.2.57	IL17C - Interleukin-17C	372
C.2.58	IL18 - Interleukin-18	372
C.2.59	IL18R1 - Interleukin-18 receptor 1	372
C.2.60	IL20 - Interleukin-20	372
C.2.61	IL20RA - Interleukin-20 receptor subunit alpha	372
C.2.62	IL22RA1 - Interleukin-22 receptor subunit alpha-1	372
C.2.63	IL24 - Interleukin-24	373
C.2.64	IL33 - Interleukin-33	373
C.2.65	KITLG / SCF - Stem cell factor	373
C.2.66	LIF - Leukemia inhibitory factor	373
C.2.67	LIFR - Leukemia inhibitory factor receptor	373
C.2.68	MMP1 - Matrix metalloproteinase-1	374
C.2.69	MMP10 - Matrix metalloproteinase-10	374
C.2.70	NRTN - Neurturin	374
C.2.71	NT3 - Neurotrophin-3	374
C.2.72	OSM - Oncostatin-M	375
C.2.73	PD-L1 - Programmed cell death 1 ligand 1	375
C.2.74	S100A12 / ENRAGE - Protein S100-A12	375
C.2.75	SIRT2 - SIR2-like protein 2	376

C.2.76	SLAM / SLAMF1 - Signaling lymphocytic activation molecule	376
C.2.77	SULT1A1 - Sulfotransferase 1A1	376
C.2.78	STAMBP - STAM-binding protein	376
C.2.79	TGFA - Transforming growth factor alpha	377
C.2.80	LAP TGF- β 1 / TGFB1 - Latency-associated peptide transforming growth factor beta-1	377
C.2.81	TNFSF1 / LTA / TNFB - Tumor Necrosis Factor Beta	377
C.2.82	TNFSF2 / TNF / TNFA - Tumor necrosis factor	377
C.2.83	TNFSF9 - Tumor necrosis factor receptor superfamily member 9 .	377
C.2.84	TNFSF10 / TRAIL - TNF-related apoptosis-inducing ligand	378
C.2.85	TNFSF11 / RANKL / TRANCE - TNF-related activation-induced cytokine	378
C.2.86	TNFRSF11A / RANK / TSLP - Thymic stromal lymphopoietin .	379
C.2.87	TNFRSF11B / OPG - Osteoprotegerin	379
C.2.88	TNFSF12 / TWEAK - TNF-related weak inducer of apoptosis	380
C.2.89	TNFSF14 / LIGHT - Tumor necrosis factor ligand superfamily member 14	380
C.2.90	UPA - Urokinase-type plasminogen activator	380
C.2.91	VEGFA - Vascular endothelial growth factor A	381

List of Figures

1.1 Proportion of papers published about networks on the topic of health across the years, compared with the number of papers published on obesity. Image reproduced with permissions from "International Encyclopedia of the Social and Behavioral Sciences" [6].	1
3.1 Königsberg according to an engraving by Joachim Bering from 1613 [7]. Blue tones for the river Pregel and orange tones for the bridges have been added for better visualization. The original names in the legend have also been added for context.	5
3.2 An example of a network with 10 nodes labeled from A to J. Each node has a color attribute that can be either red or blue. Nodes are connected via directed relationships. Nodes B and C, and nodes G and F have a reciprocal relationship.	7
3.3 An example of a network with 10 nodes and weighted undirected relationships. A spring layout of this network can be seen in figure 3.8	9
3.4 A histogram with the total of undirected friends per student in the overall network. Some students are popular with up to 12 connections, while others are isolated with 0 connections.	9
3.5 An example of a path. Node A can reach node J following the red highlighted edges and passing through the green highlighted nodes. All the edges weigh the same value (1), so the path is equal to 4.	12
3.6 An example of two layouts. On the left, the relationship between E and F is not clear. On the right, is obvious that E and F are disconnected from all other nodes	14
3.7 Overall network placing the nodes randomly in a 2D grid. This is useful when the amount of connections is very low, but otherwise useless for visualizing patterns in this network.	15

3.8	An example of a network with 10 nodes and weighted undirected relationships with a spring layout. Nodes with smaller weights are brought closer together in the 2D space as if the edges were acting as springs with greater or lower spring force equilibrium alike to Hooke's law.	16
3.9	Our recurrent example of a network layout in a hierarchical pattern. In this example, it is highlighted that D is a very important node in the network, and is placed on top.	16
3.10	Overall network with a DrL algorithm.	18
3.11	Overall network with a Fruchterman - Reingold algorithm.	19
3.12	Overall network with a graphopt layout	19
3.13	Overall network with a Kamada-Kawai algorithm.	20
3.14	FF1 overall network with MDS layout.	20
3.15	Overall network sorting nodes into circles.	21
3.16	Overall network placing nodes according to high-school geographical location.	21
3.17	Overall network coloring nodes by their closeness. Disconnected nodes and subnetworks have outlier values and have been removed for better color scale	23
3.18	Overall network coloring nodes by their betweenness. Node size is proportional to betweenness.	24
3.19	Overall network coloring nodes by their eigencentrality. Node size is also proportional to eigencentrality. A few nodes are popular people also connecting to other popular people.	25
3.20	A network with two components. Nodes are labeled with the number of degree centrality in each. On the left component, we have an average degree of 1.3, while in the right component, we have an average degree of 1.71	26
3.21	A network with two components. Black edges represent existing edges, and red thin, and dashed edges represent possible edges. On the left component, the 3 possible edges are shown and labeled. On the right component, no edge is labeled. The left component has a density of 2/3 (66%), while the right component has a density of 6/49 (12%). Notice that the right component, despite having higher average connectivity, also has a lower density.	26
3.22	An example of coverage. For a given path of distance equal to 3, node A can reach all the green highlighted nodes using the red highlighted paths.	30

3.23 A reachability plot in the overall network. The x-axis represents the number of steps taken. The y-axis represents the proportion of network coverage. After 4 steps, we see that about 25% of the nodes manage to cover 25% of the network. In the next step, nearly 50% of the nodes, cover 50% of the network. In the next step, we see that the median surpasses the 75% coverage, so the node reach grows exponentially in this network. It never reaches 100% because some of the nodes are isolated, we see this better in the outliers that start showing from step 7 at the bottom of each box.	31
3.24 Simulation of the spread of an unknown disease in the network. In the x-axis, we have steps which is an arbitrary representation of time. In the y-axis the proportion of the network that each category has at any given time. Each red, blue, and red, line represents one of the 200 simulations. Red represents infected, Blue represents immunized, and Black represents dead.	35
3.25 Simulation of the spread of an unknown disease in the network. In the x-axis, we have the steps in time. In the y-axis the proportion of the network that each category has at any given time. Each red, blue, and red, line represents one of the 200 simulations. Red represents infected, Blue represents immunized, and Black represents dead. A vaccine is introduced at $t = 10$ which suddenly spikes the immunization ratio and lowers dead rate	36
3.26 a	37
3.27 b	37
3.28 c	37
3.29 d	37
3.30 A simulation of a disease advancing through the School Network using an MDS layout. Each sub-figure represents the disease advance after (a) 0 steps with 5 random initials infected (b) 30 steps (c) 60 steps (d) 90 steps. Infected people are highlighted in violet, and each school is highlighted in a different color. The disease starts in H1 and spreads to closer contacts. H5 and all isolated components are spared because is not connected and thus impossible to reach in this particular model. The complete video of the simulation can be found at https://github.com/uit-hdl/mimisbrunnr/blob/main/results/network/simulation.mp4	37

3.31 Overview of the formation of blood cellular components (hematopoiesis). All cellular blood components are derived from pluripotent stem cell roots. Image from Wikimedia	57
3.32 Overview of the T-cells activation and differentiation. Image from "Medical Gallery of Mikael Häggström 2014". WikiJournal of Medicine	61
3.33 Overview of the histamine/thrombin cascade signaling which activates PS19 (P-selectin) to migrate up to the cell wall. This mechanism also activates the side chains VE-cadherin to open allowing for leukocytes to abandon the blood stream. Source: Wikimedia.	64
3.34 Overview of the roles of lipid mediators in vascular and leukocyte response, from the initiation of inflammation to the resolution of the event. Image from "Pro-resolving lipid mediators are leads for resolution physiology"	65
3.35 Overview of the prostaglandin pathways. Image from Wikipedia	69
3.36 Overview of EPA and DHA conversion into resolvins, protectins, and maresins. Image from "Pro-resolving lipid mediators are leads for resolution physiology"	70
3.37 Overview of all interleukins affecting tumoral growth. Image reproduced from "PRO-AND ANTITUMOR ROLE OF THE INTERLEUKINS 1 TO 41" [8]	71
3.38 Overview of factors that activate human TLR. Reproduced from https://wikimania.org	81
3.39 Overview of cell migration from a blood vessel to an inflammation site. Image reproduced with permissions from "Deep insight into neutrophil trafficking in various organs" [9].	85
3.40 Overview of a rod-shaped Gram-positive cell with a capsule. Reproduced with permissions from Pearson Education.	88
3.41 Differences between Gram-positive and Gram-negative cell wall and cell membranes. Reproduced with permission from https://www.lecturio.com/	88
3.42 Different types of bacteria morphologies. Source: Wikimedia.	90
3.43 Overview of <i>S. Aureus</i> identification and other bacteria. Self-made figure.	92
3.44 Conceptual framework of the functions of microbial extracellular polymeric substances (EPS) in soil. Reproduced from https://doi.org/10.3389/fmicb.2018.01636	94

3.45 An artist's drawing of osteomyelitis progression, subacute to chronic. A) Initial site of osteomyelitis involving the medial aspect of the long bone metaphysis/intraosseous abscess (green). B) The abscess extends through the cortex into the subperiosteal space forming a subperiosteal abscess (shades of green). C) Chronic osteomyelitis with a detached central necrotic bone fragment/sequestrum (brown) within the intraosseous abscess (green) with peripheral new bone formation/involucrum and a cortical and periosteal opening in the involucrum/cloaca allowing the pus to escape. Reproduced with permission from https://doi.org/10.5114/pjr.2022.113825	101
3.46 Vitamin D effects on the immune system. Figure reproduced under CC4.0 license from "Recent advances in vitamin D implications in chronic respiratory diseases" [10].	108
3.47 Overview of vitamin D, calcium, and phosphate interactions. Reproduced with permission from: Vitamin D for Health: A Global Perspective	112
4.1 Geographical location of all eight high schools included in the Fit Futures study. "A" refers to the area around Tromsøya, and "B" is the area in the town of Nordkjosbotn (Balsfjord). "C" shows the distance between "A" and "B".	131
4.2 All networks in FF1 following a MDS layout. Each student is represented by one node in the network. Each relationship is represented by an undirected edge, i.e., a line, in the network.	132
4.3 Histogram with all the answers to the question: " <i>To what degree does this table of friends give an overview of your social network? Please indicate on a scale from 0 (small degree) to 10 (high degree).</i> "	134
4.4 All the S. aureus combinations. A PDF version of this image can be found at the GitHub repository: https://github.com/rafanozal/PhDThesis/blob/main/Images/carrierDefinition.pdf	136
4.5 Overview of the data cleaning process. To the left, each step that transforms the raw data into a more practical version. This process is done only once. To the right, the subsequent analyses are performed as many times as needed.	138
4.6 BCNF transformation from the original data. The database schema for "Person" can be obtained upon request.	141

4.7	In this figure we see the overall process of simulating new networks. On the left, we have a representation of a real network that has 60% green nodes and 40% magenta nodes as attributes. The edges connect nodes, and 8 edges connect nodes with the same attributes (red) while 3 edges connect nodes with different attributes (black). On the center, we strip the nodes of the attributes but we keep the topology of the network, keeping nodes and edges as they originally were. On the right, we have 3 simulations (top, center, bottom) that use the center topology, but distribute the attributes randomly as described on the original network on the left. The simulated network will also have a total of edges connecting nodes sharing their attributes, in this case, the top has 7, the center has 2 and the bottom has 5; the average of the simulated network connecting nodes with the same attributes would be 5.7. This example has only 3 simulations, which is very limited, but seems to indicate that the real network (left) has some sort of bias connecting attributes as the 8 original same-to-same relationships are bigger than the 5.7 average of the simulations using the same attributes' distribution.	148
B.1	An overview of the evolution of BMI from Fit Futures 1 (2010-2011) to Fit Futures 2 (2012-2013) for every student with valid BMI data in both studies ($n = 692$). On the left column, we have the FF1 BMI categories. In the right column, we have the FF2 categories two years later. The threads from left to right indicate the evolution of each person with respect to the BMI category.	333
B.2	The influence of friends from FF1 have with respect to FF2 BMI. The X-axis is the number of friends in FF1 with $\text{BMI} > 25 \text{ kg/m}^2$ The Y-axis is the relative number of students belonging to each BMI category in FF2.334	
B.3	Heatmap with the normalized mean $ \text{SHAP values} $ for the model with respect to each dataset (top names) against each variable (left names). Values closer to 1 indicate strong importance in the model.	346
B.4	Barplots using MDI in RF for each of the datasets. The x-axis represents the MDI value, the greater the value the greater the importance. The y-axis represents each of the studied variables.	348

B.9	Partial dependencies plots for datasets E regarding recreational drugs and sport frequency for the RF models. On the X-axis, the total number of healthy friends, with a light grey histogram in the background. On the Y-axis, how this variable is expected to modify the model output (BMI) as the variable changes.	353
B.10	Absolute frequency of medicine consumption sorted by ATC group. Hormonal contraceptives, anti-inflammatories, painkillers, antiasthmatic, and antihistaminics dominate the frequency of use.	355
B.11	Absolute frequency of medicine consumption group by ATC group and divided by sex, which seems to indicate a higher consumption among females	356
B.12	Absolute frequency of self-reported diseases divided by sex. To avoid visual cluttering, diseases with 5 or fewer instances are not included in this figure. Females also lead in dermatological and psychological conditions, plus migraines. Males seem to have a short lead in respiratory conditions.	357

List of Tables

3.1 Example of a Laplacian matrix corresponding to figure 3.2. Summation by row represents the out-degree. Summation by columns represents the in-degree. Cells that are symmetrical from the main diagonal, represent reciprocal relationships.	11
3.2 Overview of all basic statistics of the six networks. From left to right, the name of the network, the total number of edges in the network, average out-degree, average in-degree, average reciprocity, the average number of connections per node, and the average distance between non-isolated nodes. Notice that average out-degree and average in-degree must have the same value, but the standard deviation varies. Reciprocity in "Overall" is increased because nodes can nominate in one network but receive the nomination from the same friend from another network.	11
3.3 The average centralities measures for the whole population, each high school, and also men and women. Notice that in H5 we have a higher amount of disconnected triads, which causes outliers in the average value for closeness.	27
3.4 Comparison between the different simulation results. The first two cases can be seen in figure 3.24, and figure 3.25. Death rate and immunity rate are calculated at the final step of the simulation. Peak infection is calculated with respect to the maximum infection coverage in each simulation, and in which step did happen.	34
3.5 Table with example data for contingency analysis.	47
3.6 Overview of all interleukins explained in this document	78
3.7 Vitamin D deficiency ($25(\text{OH})\text{D} < 50\text{nmol/L}$) prevalence around the world.	113
3.8 Vitamin D daily recommendation.	115
4.1 Information with the school names, study program, the total number of students in FF1, astronomical season, and whether a significant amount of students' blood samples were taken during the polar night.	133

4.2	Summary of lost connections at data cleaning.	133
4.3	Types of hormonal contraceptives classification and their respective brands.	134
4.4	Original values for the sleeping habits.	144
4.5	Table with the common 7 chronic diseases asked in a subsection of the questionnaire.	145
4.6	Table with the up to 5 self-reported chronic diseases asked during the interview.	145
B.1	Bias with respect to the total number of relationships between each BMI category. The top column shows the absolute and relative frequencies of the population ($n = 1034$). People with unknown BMI are excluded from the analysis ($n = 4$). Each combination of categories represents people who nominate a friend (rows) and people who are nominated (columns). Each cell is divided into three parts, the left-most is the total number of relationships in this combination, the center one is the expected number of relationships (only given if a bias was found), the right part contains an arrow indicating over (up ↑) or underrepresented (down ↓) using a two-sided binomial test with at least p-value < 0.1. The table χ^2 test is < 10^{-10} . In the bottom and to the left, we have the marginal absolute and relative frequencies for each combination.	332
B.2	Results of the spread of BMI across all networks simulations. The first column is the name of the network. Second is how many relationships are in that network. The third column is how many of those relationships share the same BMI category. The next 7 columns are the details of the simulation results, with the important one being the "Average" and "SD (Standard Deviation)", which shows how many same-to-same relationships in average we had in the 1000 simulations, using a network with the same topology but randomizing the BMI according to the BMI probability density data of the original network. The last column is the p-value, rounded to 3 decimals, which shows if there is a significant difference between the averaged same-to-same simulated relationship and the real same-to-same relationships.	332
B.3	Overview of all biomarkers with respect to each high school ($n=8$) for men. All biomarkers with either $R^2 < 0.2$ and p-value > 0.05, are hidden from the tables. P-values are expressed in GP Prism 5.04/d format.	337

B.4 Overview of all biomarkers with respect to each high school (n=8) for women. All biomarkers with either R ² < 0.2 and p-value > 0.05, are hidden from the tables. P-values are expressed in GP Prism 5.04/d format.	338
B.5 Ratio of average square distances between each person's biomarker level and friend's biomarker levels, and average square distance between each person's biomarker levels and non-friends biomarker levels. Values greater than 1 suggest that clusters of friends have similar biomarkers levels in comparison with the rest of the non-friend population. Relevant values are highlighted in bold, with green for >1.1 and red for <0.9. Values are rounded to two decimals but highlighted according to the original values. (Table 1 of 4)	339
B.6 Ratio of average square distances (Table 2 of 4)	340
B.7 Ratio of average square distances (Table 3 of 4)	341
B.8 Ratio of average square distances (Table 4 of 4)	342
B.9 Anthropometric variables with respect to biomarkers levels in men. Each cell is a p-value in GP Prism 5.04/d format. Biomarkers rows with no statistically significant p-values are hidden. All p-values are corrected for Bonferroni.	342
B.10 Anthropometric variables with respect to biomarkers levels in women. Each cell is a p-value in GP Prism 5.04/d format. Biomarkers rows with no statistically significant p-values are hidden. All p-values are corrected for Bonferroni.	343
B.11 Summary of all models and datasets used. From left to right, the ID of each dataset is represented by a letter, the short name of the dataset, the total samples in the dataset, Mean Absolute Error (MAE) for each model (RF or ANN) performing in this dataset, mean BMI in FF1 and mean BMI in FF2 in this dataset.	345
B.12 χ^2 significant bias of each medicine type.	357

Chapter 1: Introduction

1.1 Social Network Analysis in health

In recent years there has been a growing recognition of the profound impact that social relationships and networks of friends have on health outcomes (figure 1.1). This includes common and well-known topics such as the spread of obesity [11, 12], recreational drugs usage such as smoking [13, 14] alcohol [15, 16] or cannabis [17], and depression [15].

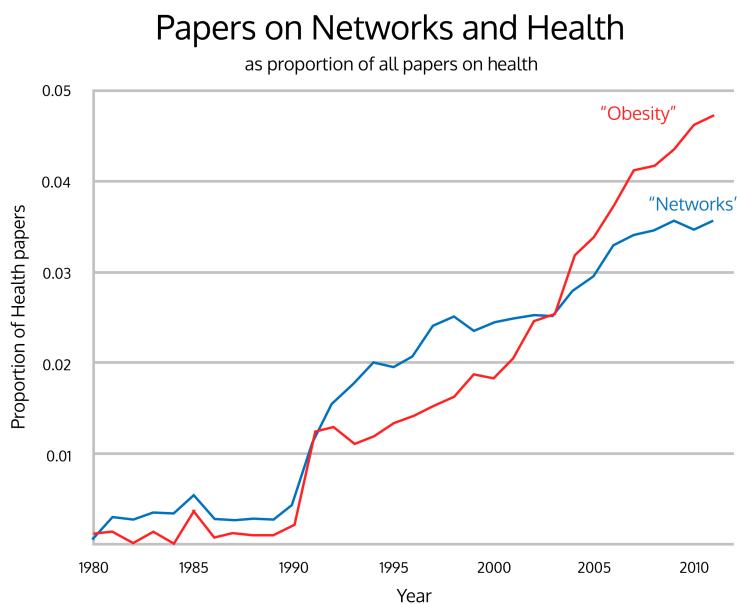


Figure 1.1: Proportion of papers published about networks on the topic of health across the years, compared with the number of papers published on obesity. Image reproduced with permissions from "International Encyclopedia of the Social and Behavioral Sciences" [6].

Social Network Analysis (SNA) is a powerful tool for understanding how social connections influence health behaviors, disease transmission, and healthcare access. This thesis leverages SNA techniques to explore the dynamic interplay between social networks and health, aiming to provide valuable insights that can inform interventions, policies, and practices for improving health outcomes.

1.2 Social network interventions

Within the realm of health research, SNA has demonstrated its versatility and applicability across a wide range of topics. It is possible to detect early outbreaks of influenza [18]. To prevent the spread of Human Immunodeficiency Virus (HIV) infections [19]. To improve the life of chronic illnesses patients as well as their families and community [20]. It has been shown that information on social networks can increase the prediction of health-related models at around 50% [21]. It has been used to design interventions and strategies to enhance communication, collaboration, and knowledge sharing among health professionals, ultimately improving the overall performance and effectiveness of the health organization [22, 23]. A systematic review of 37 studies suggests that social network interventions are associated with positive health behaviors and outcomes [24]. And of course, social network interventions have been used to better the outcome of obesity [25–28], mental health [29, 30], overcoming tobacco addiction [31, 32], transmittable diseases in humans [33–36] and in cattle such as cows, sheep, and pigs [37, 38], and recently we experienced these interventions first hand with COVID-19 [39, 40].

These few examples show that SNA has proven practical value in understanding the diffusion of diseases, as well as tracking the spread of infectious agents through clusters or interconnected social groups. It has been found to play a crucial role in improving individuals' health and preventing further deterioration of their well-being. Different authors have evaluated that social relationships influence a person's health between 15% to 40% [5], putting it ahead of the environment and even medical care. And yet it remains a vastly under-utilized and underrated technique.

1.3 Thesis impact

In this thesis, we employ SNA techniques to shed light on the complex relationships between social networks and health outcomes within a specific population. We have shown how *Staphylococcus aureus* (*S. aureus*), vitamin D, inflammation, medication usage, and obesity are influenced by social networks in a general youth population in Tromsø. We also expanded non-parametric methods for group comparisons in graphs using simulations and applied machine learning models to measure the influence of peers on obesity. Finally, we lay down the basics for a framework to obtain a more efficient analysis framework. We hope this leads to a significant contribution to public interventions and policies that ultimately lead to improved health outcomes in Norway and beyond.

Chapter 2: Aims of the thesis

The overarching objective of this research is to develop methodologies and conduct exploratory studies to evaluate the impact of social influence on a range of health-related topics. This work targets several topics and is done with the help of an interdisciplinary team of health professional researchers. In parallel, this work aims to provide a framework that enables researchers to facilitate faster analysis and produce improved visualizations.

The results of these studies must be a quantifiable measure of how much measuring or modifying the social networks could benefit the studied population. Secondarily, speculate how these results could affect the general population and the advantages and disadvantages of influencing and changing their social network.

The first step of our research is to study how infections and the immune system behave in the population. First, by measuring the spread of *S. aureus* (Paper A) and investigating the possibility that inflammatory processes may be similar across individuals or schools (Result II). The second step is to inspect the social aspect of obesity with classical methods (Result I) and using machine learning models (Result III). Lastly, we look into other variables of interest such as how friends influence vitamin D levels (Paper B) or medication usage (Result IV).

For our secondary objectives, we want to present tutorials and proof of concept on how to apply SNA techniques. For this, we choose prescriptions and drug interactions (Paper C). We also want to present a user-friendly wrapper library to abstract away the complexity and details of SNA and other statistic and machine learning models, where biases are checked automatically, fundamental analysis reports are generated with minimal intervention from the programmer, and plots or figures follow basic design rules for good visualization.

Starting with infectious diseases is a good starting point as it is the classical topic for understanding the network structure, and testing and developing new methods. It also has the advantage of updating targeted interventions in similar populations in the

future. Later on, obesity in particular is a good topic for SNA due to having a substantial impact on individual lifestyle choices, including diet, exercise, and weight management; all of which are influenced by friends. Lastly, topics such as vitamin D levels are purely explorative and we want to determine if there was a connection with the social network.

Regarding our programming objective, wrapper libraries serve as a layer of abstraction that simplifies the usage of lower-level functionality, allowing programmers to build applications more efficiently and with less effort. We also want to extend these advantages and make a high-level abstraction in the statistical context.

Chapter 3: Background

3.1 Graph theory

3.1.1 Historical Overview

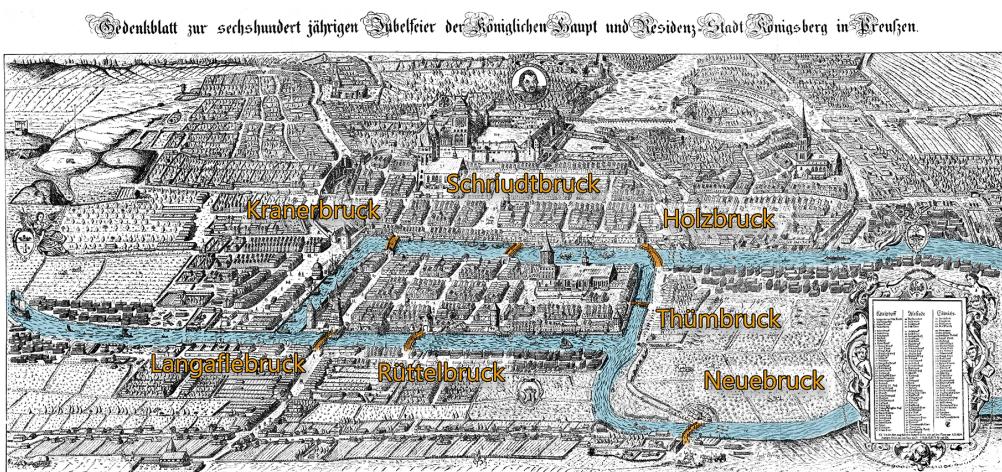


Figure 3.1: Königsberg according to an engraving by Joachim Bering from 1613 [7]. Blue tones for the river Pregel and orange tones for the bridges have been added for better visualization. The original names in the legend have also been added for context.

A famous historical puzzle was raised by Carl Leonhard Gottlieb Ehler, the mayor of Danzig (now Gdańsk), to Leonhard Euler in 1736. The problem proposal was like this. In the city of Königsberg in Prussia (now Kaliningrad, Russia) there are seven bridges crossing the Pregel river (figure 3.1). Starting anywhere you want in the city, how to transverse the 7 bridges crossing each bridge only once. Reaching an island or mainland bank other than via bridges, or accessing a bridge without crossing it is not permitted.

Euler was annoyed by this proposal and expressed so in a letter to the mayor: "...Thus you see, most noble Sir, how this type of solution bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution, is based on reason alone, and its discovery does not depend

on any mathematical principle. Because of this, I do not know why even questions that bear so little relationship to mathematics are solved more quickly by mathematicians than by others."

Even though Euler found the problem very easy, he was curious about the mathematical foundation behind it. He wrote to the Italian mathematician Giovanni Marinoni: "*This question is so banal, but seemed to me worthy of attention in that geometry, nor algebra, nor even the art of counting was sufficient to solve it. In view of this, it occurred to me to wonder whether it belonged to the geometry of position which Leibniz had once so much longed for. And so, after some deliberation, I obtained a simple, yet completely established, rule with whose help one can immediately decide for all examples of this kind, with any number of bridges in any arrangement, whether such a round trip is possible, or not..."*"

Nevertheless, Euler proved that the problem had no solution, and by doing so he invented a new branch of mathematics now known as graph theory. Euler realized that the details of Königsberg were irrelevant, and only the masses of land and the bridges matter. Turning a complicated structure of a city into a very simple graph. A temporal solution was found in April 1942 when two of the bridges were blown up during World War II.

3.1.2 Definition

Graph theory is a branch of mathematics that studies the properties and applications of graphs. A graph is a mathematical structure consisting of a set of vertices (also referred to as nodes) and a set of edges of these vertices [41]. It is a fascinating subject with numerous practical applications in several fields including physical network connections in computer science, optimizing logistics in transportation, or in our case measuring the influence of peers in social networks.

Graphs provide an abstract representation of the relationships between objects, which may be used to find paths, network flow, connectivity, and more. In this context, we will talk about graphs and networks interchangeably.

A graph has the following mathematical definition:

$$G = (V, E) \quad (3.1)$$

Where G is the graph, V is the set of vertices, and E is the set of edges. In figure 3.2, we define G as $V = \{A, B, C, D, E, F, G, H, I, J\}$ and $E = \{ \{A,D\}, \{B,C\}, \{C,A\}, \{C,B\}, \{D,G\}, \{E,D\}, \{F,G\}, \{F,I\}, \{G,F\}, \{H,G\}, \{J,F\} \}$

3.1.3 Nodes

A node or vertex is the fundamental element of a network. It is usually represented as a point with lines, known as edges, coming out of it, which connect it with other nodes in the network. Each node represents one elemental object in the network, which in our case is a total of 1038 students. In different contexts, nodes can be cities, computers, or any other concept.

Mathematically, we notate all nodes as V, with each individual node in a graph with lowercase variables, such as x, y, or z. The total number of nodes is notated with $|V|$. In figure 3.2, we have 10 nodes.

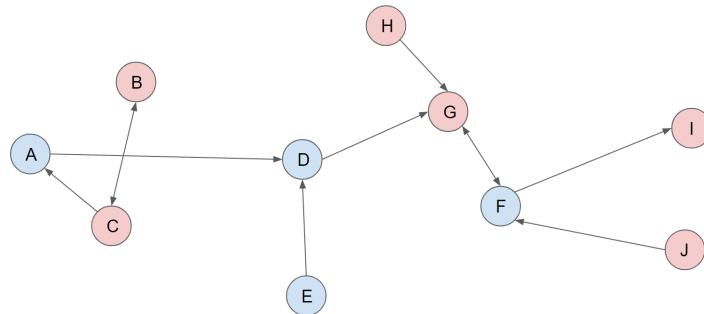


Figure 3.2: An example of a network with 10 nodes labeled from A to J. Each node has a color attribute that can be either red or blue. Nodes are connected via directed relationships. Nodes B and C, and nodes G and F have a reciprocal relationship.

Attributes

Each node may have different variables, such as sex, BMI, or any other intrinsic variable proper to the object the node is representing. Each of these variables is known as the attributes of a node. Each attribute of a node is notated with subindexes, such as $x_1, x_2, \dots, x_i, \dots, x_n$. In figure 3.2, $A_{color} = \text{blue}$

3.1.4 Edges

An edge represents a relationship between two nodes. Our network represents a relationship of undirected friendship. The assessment of friendship is formally introduced in section 4.1.1. For now, let's just comment that we have 5 types of relationships that represent the social interactions that happen between students. Two nodes can have multiple edges with the same or different weights between them, or have none. In our case, all weights are equal to one as all relationships are considered of equal value. If all edges in the graph have at least one connection to every other node, then it is called a complete graph.

The mathematical definition of all edges in a graph is as follows:

$$E \subseteq \{(x, y) | (x, y) \in V^2 \wedge x \neq y\} \quad (3.2)$$

We denote the total number of edges as $|E|$. A particular edge between two nodes is simply (x,y) . In figure 3.2, we have 11 edges that are directed, and in figure 3.3 we have 9 edges that are undirected.

Directionality

Directionality refers to whether the edges are directed or undirected. In the case of friendship, a student may nominate others as friends while none of them nominate him back. Or they might have a strong friendship between them so everyone reciprocates everybody. Nodes can be referred to as "*Ego*" or "*Alter*". An Ego node is the node from which the relationship is emanating, and an Alter node is the node that is receiving the relationship. A phrase such as "*Ego-perceived friends*" means that we are focusing on a node (*ego*) and this node has nominated some friends.

In our network, edges are undirected, meaning that it doesn't matter which student nominates whom as a friend, in either case, there is a relationship between them. In figure 3.4 we can see the number of undirected edges by the number of total nominations.

Reciprocity ratio

The reciprocity ratio is the number of edges that have a reciprocal relationship. This can only be measured in networks that are directed.

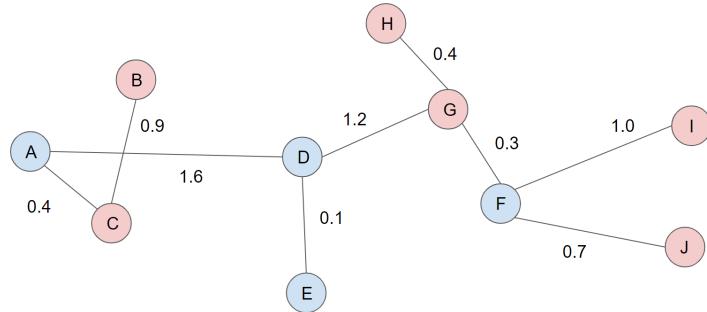


Figure 3.3: An example of a network with 10 nodes and weighted undirected relationships. A spring layout of this network can be seen in figure 3.8

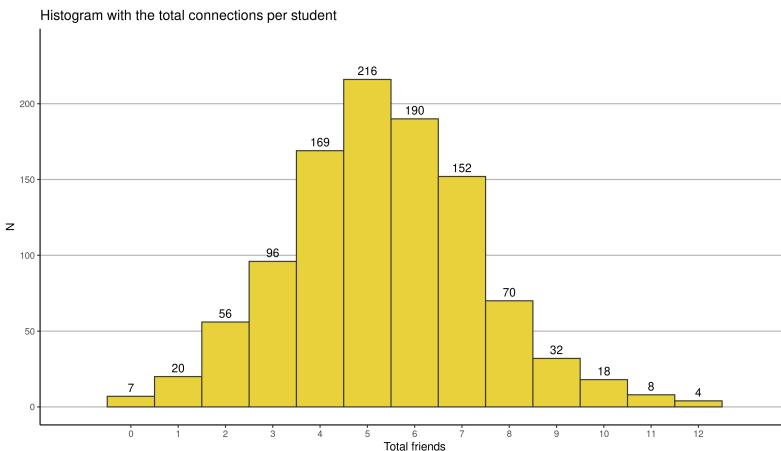


Figure 3.4: A histogram with the total of undirected friends per student in the overall network. Some students are popular with up to 12 connections, while others are isolated with 0 connections.

Weight

All edges have an additional property called weight. This is a variable, typically a real number, that indicates some quantifying property of the relationship. For example the distance between two cities, or how many hours per day you spend in the company of a person. By default, all edges have a weight of 1, meaning all relationships have the same importance; this is our case study.

Nodes that are not connected can be defined to either have a weight of zero or a weight of infinity, depending on the wording of your definition.

Loop

A relationship of a node with itself is allowed in graph theory and is called a loop. Be aware that this may not make sense in a context such as ours. Philosophically a person might be friends with himself if we define it as being happy with his life, but mathematically we don't have any use for such a definition. But it can make sense if we talk about a network of collaborations and we define a loop as authors' self-promotion. In a transportation network, a loop can represent a roundabout where a vehicle can circle around the same point.

Connectivity

Connectivity is the summation of all edges' weight. When all the nodes in the network have at least 1 connectivity is it said that the network is fully connected. A summary of the connectivity in our network can be found in table 3.1.4. There are four main methods of measuring connectivity:

- **In-degree:** Summation of all the weights coming into the node.
- **Out-degree:** Summation of all the weights going outside the node.
- **Reciprocity:** Summation of weights in reciprocal edges.
- **Undirected:** Summation of in and out connections.

Laplacian matrix

A graph can be represented as a matrix, in which each cell combination of row and column represents an edge between two nodes. This is known as the Laplacian matrix (table 3.1.4) of a graph, or simply the matrix.

Distance

A distance is the summation of the edge's weights between two nodes, whether they are directly connected or have some intermediary nodes between them. Distance is not necessarily unique, a couple of nodes can have several different distances between them. If two nodes have no connection, the distance is equal to infinity. If your network allows for edges forming loops then you will also have an infinite combination of nodes and thus infinite possible distances.

Table 3.1: Example of a Laplacian matrix corresponding to figure 3.2. Summation by row represents the out-degree. Summation by columns represents the in-degree. Cells that are symmetrical from the main diagonal, represent reciprocal relationships.

	Node receiving the edge									
	A	B	C	D	E	F	G	H	I	J
A	0	0	0	1	0	0	0	0	0	0
B	0	0	1	0	0	0	0	0	0	0
C	1	1	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	1	0	0	0
E	0	0	0	1	0	0	0	0	0	0
F	0	0	0	0	0	0	1	0	1	0
G	0	0	0	0	0	1	0	0	0	0
H	0	0	0	0	0	0	1	0	0	0
I	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	1	0	0	0	0

Table 3.2: Overview of all basic statistics of the six networks. From left to right, the name of the network, the total number of edges in the network, average out-degree, average in-degree, average reciprocity, the average number of connections per node, and the average distance between non-isolated nodes. Notice that average out-degree and average in-degree must have the same value, but the standard deviation varies. Reciprocity in "Overall" is increased because nodes can nominate in one network but receive the nomination from the same friend from another network.

Name	Edges	Out	In	Reciprocity	Connections	Distance
Overall	3767	3.63 ± 1.32	3.63 ± 2.11	1.92 ± 1.27	5.34 ± 2.04	8.00
Physical	2823	2.72 ± 1.71	2.72 ± 1.96	1.25 ± 1.20	4.18 ± 2.19	9.91
School	2979	1.20 ± 1.35	1.20 ± 1.25	0.49 ± 0.75	1.92 ± 1.66	6.90
Sport	598	2.87 ± 1.59	2.87 ± 1.85	1.42 ± 1.20	4.32 ± 1.92	13.73
Home	1247	0.58 ± 1.11	0.58 ± 1.04	0.18 ± 0.56	0.97 ± 1.48	3.58
Other	1095	1.05 ± 1.47	1.05 ± 1.11	0.26 ± 0.58	1.85 ± 1.68	7.08

Path

A path is the minimum weighted distance between two nodes. The path to the same node can be 0, or another value specified in a loop. A path can be negative if the graph is allowed to have negative distances. A path can be infinite if, for example, two nodes are unreachable from one to another. A directed graph can have different path lengths from any given pair of two nodes. An example of a path is shown in figure 3.5.

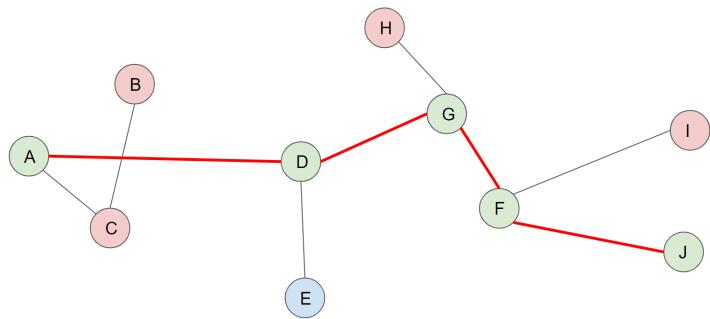


Figure 3.5: An example of a path. Node A can reach node J following the red highlighted edges and passing through the green highlighted nodes. All the edges weigh the same value (1), so the path is equal to 4.

3.1.5 Network

Isomorphic

Two graphs are said to be isomorphic if they have the same edges and nodes if we remove the attributes of all nodes in both graphs. This is also known as network topology.

Network motifs

A set of two nodes is called a diad. A set of three nodes is called a triad. A set of four nodes is called a tetrad. An undirected diad can have only two states, either the two nodes share an edge, or they don't. A directed diad can have 4 states: no common edge, the edge from A to B, the edge from B to A, reciprocal edges from A to B, and B to A. Similarly, triads and tetrads have a limited amount of combinations of edges and no edges between their nodes. All of these examples are called network motifs.

Networks that have the same context, for example, how animals share spaces in the wilderness, and how followers are distributed in internet social networks, tend to have the same distribution of motifs. Sometimes it is useful to compare the motif count with other

network motif counts to check if their structure is similar. Other times it is interesting to simulate a network by giving a set of probabilities for each network motif to force a specific structure.

Network motifs are an important concept in social science studies to describe the structure of the network [42], but there are better metrics on how to describe a network [43].

Components

If no path exists that connects all nodes with all other nodes in the network it means that you have a network divided into components. A network can have isolated nodes, isolated diads, triads, or be divided into several sub-networks. Each of these cases is known as a component of a network. In figure 3.20 we have one network with two components.

Clusters and Communities

Community detection is the process of discovering groups or clusters of nodes in a social network that have a high degree of connectivity within themselves, or low with other groups. There is no best algorithm and the use of each depends on the needs of the researcher:

- **Modularity:** This method compares the density of different groups and is useful for comparing hierarchical and overlapping communities.
- **Clustering:** This is a family of algorithms that consists of dividing groups with similar attributes. Different approaches are k-means, hierarchical clustering, or spectral clustering
- **Link-based:** This method uses the edges in the graph and measures how far you can walk from one node to another. Communities tend to have shorter walking distances between them.
- **Latent variable:** These methods are somewhat similar to PCA analysis, in which we decide on several hidden latent variables, and the algorithm groups automatically nodes based on those latent variables.

In our case, we already have the data divided in communities by high schools, down to the same class levels. We have several attributes in each node to compare

different values. As we have not done any study regarding the intricacies of each high school individually, no clustering has been applied so far other than calculating homophily (section 3.1.7).

Multimodal networks

The network can be multimodal. This means that a node can connect to other nodes that are not of the same class. For example, a researcher can collaborate with other researchers, which are nodes of the same class. They publish a paper together which is another class of nodes that nevertheless connect to researchers. The paper is published in a journal which is another class. Multimodal networks are characterized as being hard to analyze due to the different meanings of the different relationships.

3.1.6 Layout

Introduction

Before we continue exploring other important aspects of the network, it would be helpful to visualize such concepts within the network. However, this is not straightforward. For visualization purposes, there are several possibilities as to how we can spread the nodes in a typically 2D or 3D image [44–48]. An example of the same network with two different layouts can be seen in figure 3.6

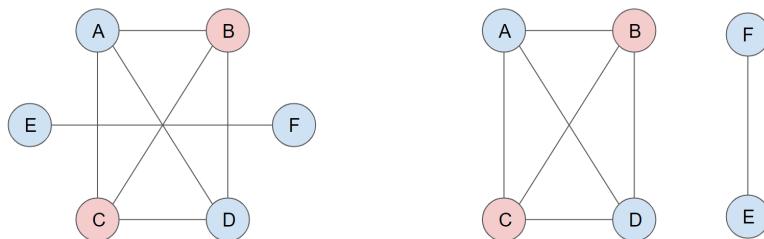


Figure 3.6: An example of two layouts. On the left, the relationship between E and F is not clear. On the right, it is obvious that E and F are disconnected from all other nodes

Proper node placement can help in creating a clear and intuitive picture making it easier to understand the relationships between the nodes. On the other hand, the emerging pattern in the network can be hidden if a poor option is chosen. This is called network layout. There are several algorithms for vertex placement. The basic categories are:

Random

Random placement of the nodes. Just spread the nodes around in no particular order and hope for the best. An example can be seen in figure 3.7.

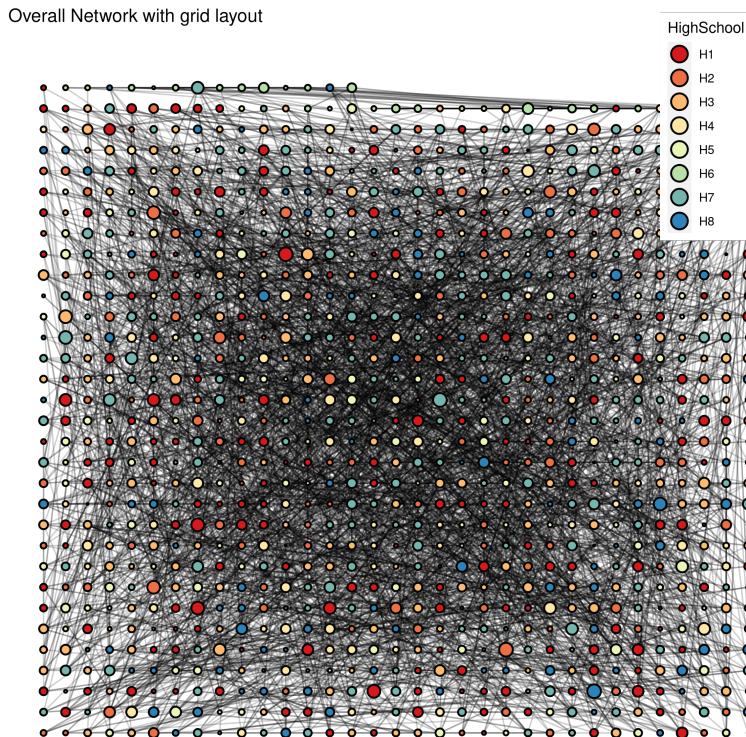


Figure 3.7: Overall network placing the nodes randomly in a 2D grid. This is useful when the amount of connections is very low, but otherwise useless for visualizing patterns in this network.

Force directed

This uses physics-based simulations to determine the position of nodes based on the attractive and repulsive forces between them. For example, we can imagine the nodes as electrical charges based on node properties, or the edges as springs with an elastic force proportional to the edge value.

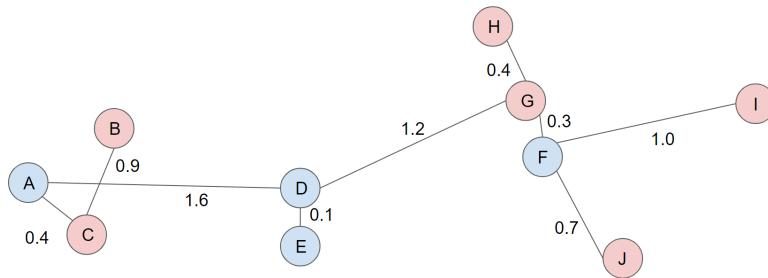


Figure 3.8: An example of a network with 10 nodes and weighted undirected relationships with a spring layout. Nodes with smaller weights are brought closer together in the 2D space as if the edges were acting as springs with greater or lower spring force equilibrium alike to Hooke's law.

Hierarchical

Hierarchical layouts tend to put important nodes on top, and less important nodes on the bottom (figure 3.9). This type of layout only makes sense when you actually have a hierarchical relationship. Common in multi-modal networks, or where there are many triads with only two nodes connecting. Hierarchy would also mean that nodes' placement in the y-axis has some meaning of importance.

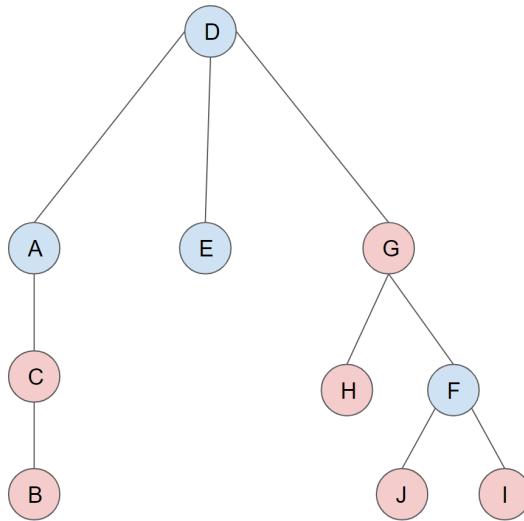


Figure 3.9: Our recurrent example of a network layout in a hierarchical pattern. In this example, it is highlighted that D is a very important node in the network, and is placed on top.

Spectral

A way that minimizes a particular energy function derived from the graph's Laplacian matrix. The spectral placement technique is designed to produce graph layouts that not only look aesthetically pleasing but also provide some semantic information about the underlying network. An example of spectral layouts are:

- **DrL.** The Distributed Recursive Layout (DrL) algorithm [46] is hierarchical and is well-suited for presenting large-scale graphs with a clear structure. In our case, it clusters high schools somewhat together but is not the best choice for our network as seen in figure 3.10.
- **Fruchterman - Reingold.** The Fruchterman - Reingold algorithm [44] keeps related vertices together. We can see an example of this in figure 3.11 we have some isolated nodes that are too far away, while the main network is clumped in one little space.
- **Graphopt.** The Graphopt (graph optimization) layout [47] uses physical analogies for defining attracting and repelling forces among the vertices and then the physical system is simulated until it reaches an equilibrium. It aimed to minimize the number of edge crossings to avoid visual cluttering, but in our case, as seen in figure 3.12 we have too many edges everywhere and an emerging pattern is not clear.
- **Kamada-Kawai.** The Kamada-Kawai algorithm [45] emphasizes the distance as information. In figure 3.13 we see that more isolated nodes are on the outside, and more dense connected areas in the middle. Again, not an ideal choice for us as the core of the network is just a mess of black lines.
- **MDS.** The Multidimensional Scaling (MDS) layout [48], is a popular dimension reduction technique that tries to maintain similar nodes close to each other. For our work, we use the MDS as default for almost every figure as is the one that best displays our network, as shown in figure 3.14.
- **Grouping.** This does not refer to a particular algorithm, but rather aligning manually the nodes with a common particular attribute and showing the relationships. In figure 3.15 we see an example with the overall network sorting nodes into circles, where each circle is a high school. Here is hard to see relationships inside high schools but is easier to see relationships between every two pairs of high schools.

- **Geographical.** Another popular option is to organize nodes based on their geographical location (if such information exists). In figure 3.16, we see that schools are placed according to their geographical location. A background outlined map of Tromsø from OpenStreetMaps has been added. Note that H5 is about 120km away from the rest and is brought closer for better visualization.

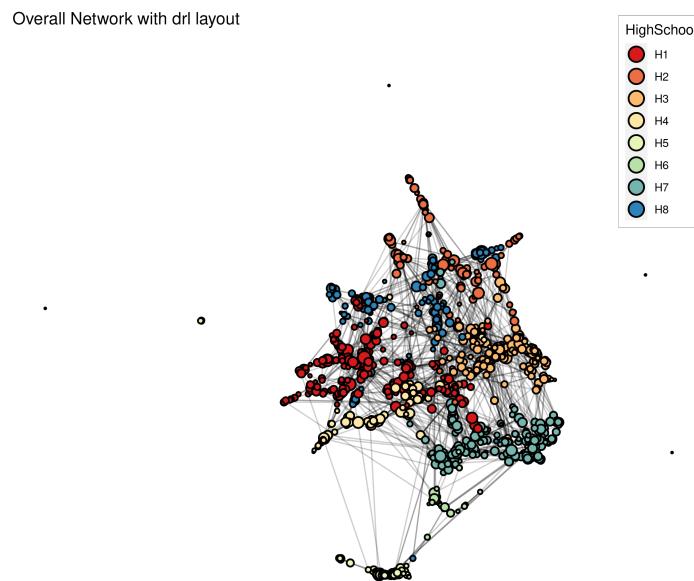


Figure 3.10: Overall network with a DrL algorithm.

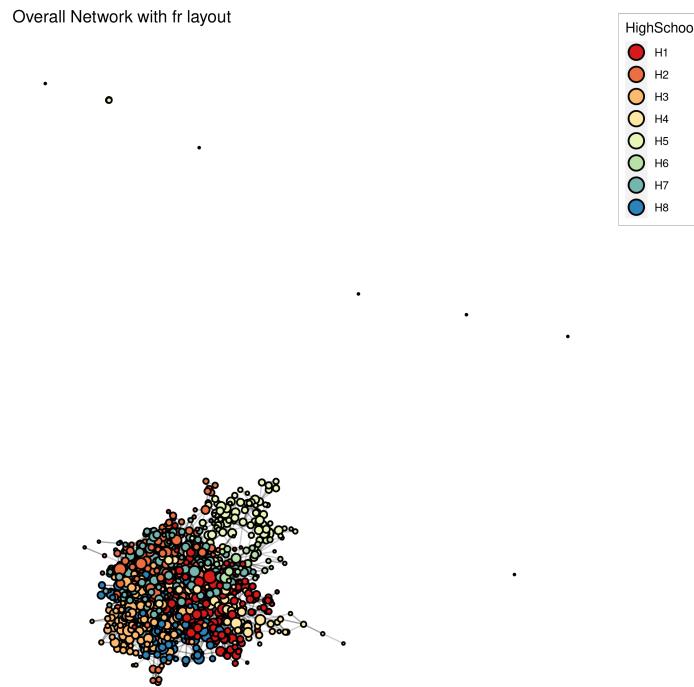


Figure 3.11: Overall network with a Fruchterman - Reingold algorithm.

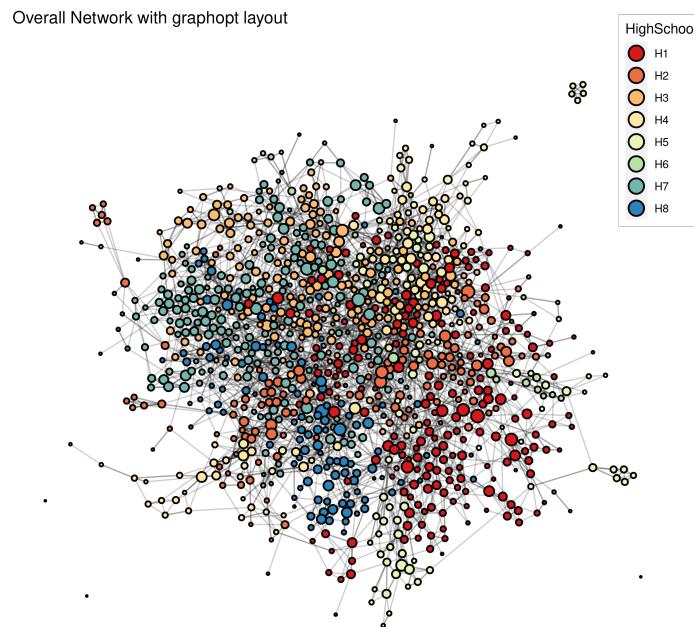


Figure 3.12: Overall network with a graphopt layout

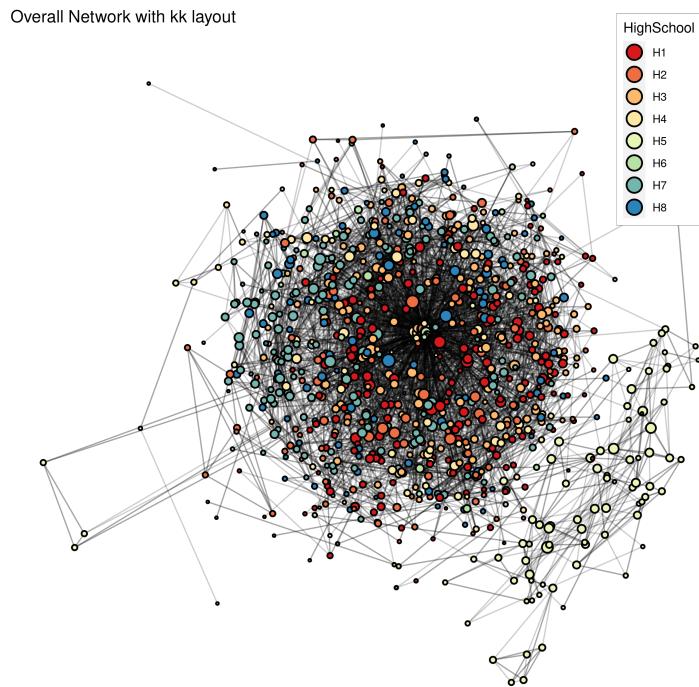


Figure 3.13: Overall network with a Kamada-Kawai algorithm.

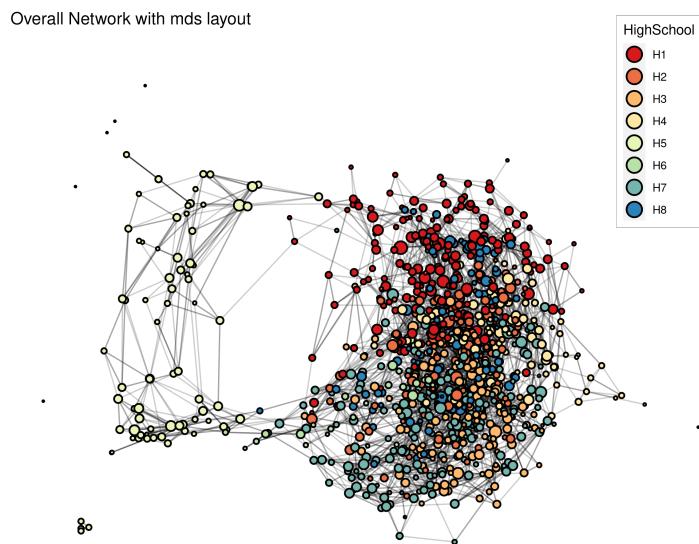


Figure 3.14: Fit Futures 1 (FF1) overall network with MDS layout.

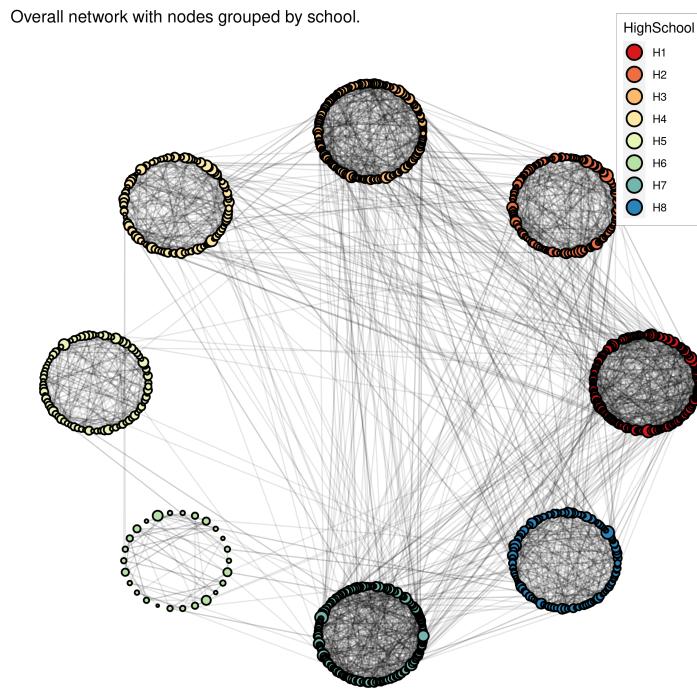


Figure 3.15: Overall network sorting nodes into circles.

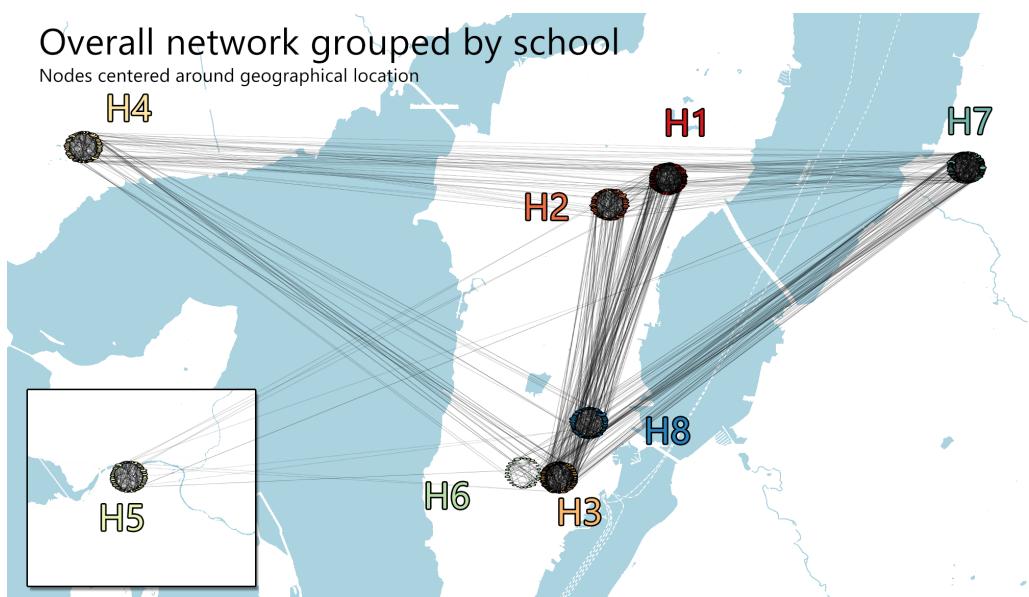


Figure 3.16: Overall network placing nodes according to high-school geographical location.

3.1.7 Networks metrics

How important is a node in a network?

Nodes are not only important because they have many connections as we saw in section 3.1.4. The following metrics can help you identify nodes in the network that might also be relevant. But none of them is a better or worse method, it depends on what you consider to be important in your context:

Degree centrality

The most obvious one is to measure the number of connections [49]. In social networks, well-connected nodes are people with high popularity, which tends to be indicative of importance. In medical interventions, we like to target high connectivity nodes to become healthy because they tend to influence their peers to become healthy as well. In figure 3.14 we can see the network with the node size proportional to the number of connections each student has.

Closeness centrality

In this context, we measure the importance of the nodes proportionally to their path with any other given node [49]. The idea is that an important node is capable of reaching other nodes very quickly. The importance of this is more obvious in topology networks with weighted edges, where a town can be well connected to other towns, but using semi-destroyed and badly maintained roads, making it a less attractive node to put a central hub. In comparison, a town that is less connected but with a newly built 4-lane highway might be a better option. When a closeness centrality number is high, it means that the node is central in the network and is close to other nodes in the network. This indicates that the node has better accessibility to other nodes and can easily reach these nodes. An example of this metric can be seen in figure 3.17.

Betweenness centrality

Central nodes are nodes that connect several subnetworks [49]. In epidemiology identifying these nodes is critical because we can stop the progress of the disease by targeting these nodes alone, and thus isolating each of the individual clusters. An example of this metric can be seen in figure 3.18.

Overall Network with nodes proportional to normalized closeness centrality
How fast can the node reach other nodes in the network? (higher is faster)

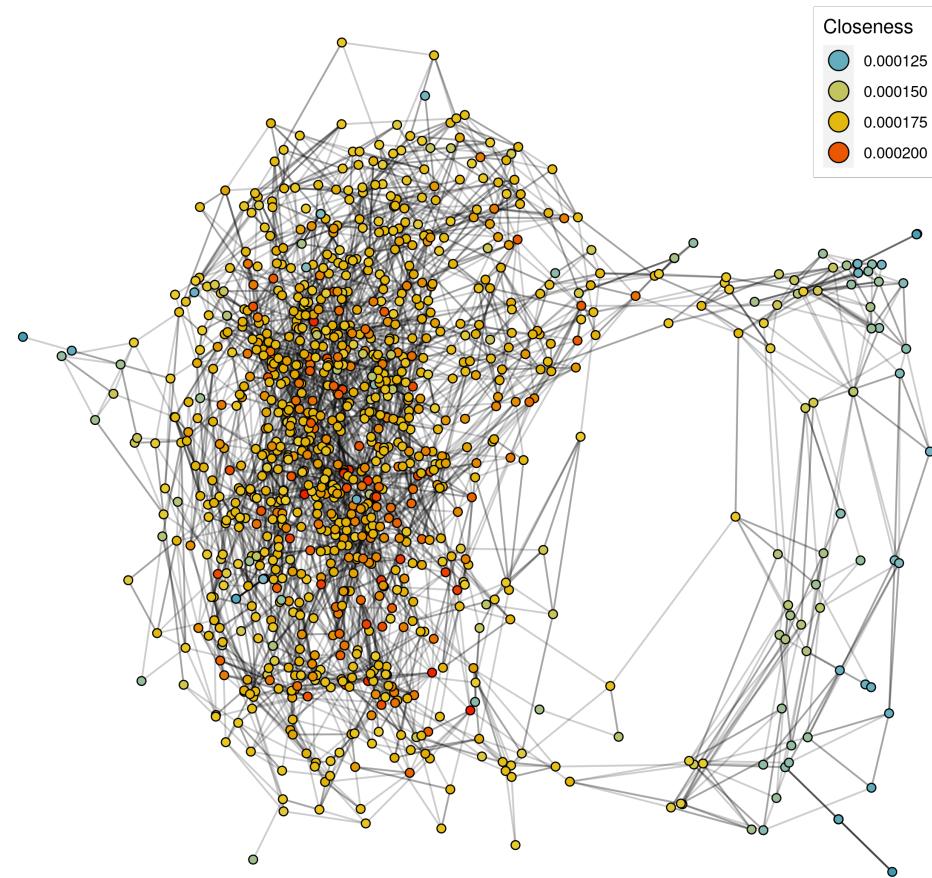


Figure 3.17: Overall network coloring nodes by their closeness. Disconnected nodes and subnetworks have outlier values and have been removed for better color scale

Eigencentrality

When you don't care about the number of connections, but about the quality of the connections, then eigencentrality is what is used to measure the important nodes [50]. In general, vertices with high eigenvector centralities are those that are connected to many other vertices which are, in turn, connected to many others, and so on. An example of this metric can be seen in figure 3.19.

Bonacich Centrality

Is the same as Eigencentrality, but you get a decay factor the further away you are from well-connected people [50].

Overall Network with nodes proportional to normalized between centrality
Is the node connecting subnetworks together? (higher is yes)

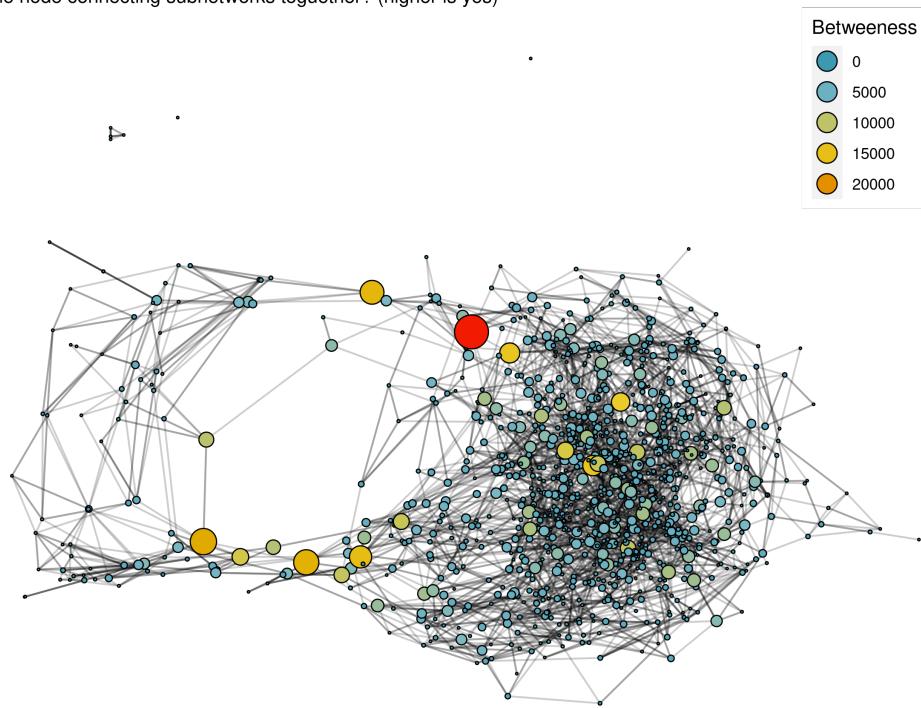


Figure 3.18: Overall network coloring nodes by their betweenness. Node size is proportional to betweenness.

Connectivity average

Connectivity is the average chosen centrality per node in the network. It gives you an idea of how well the network is interconnected. Higher degree centrality numbers mean that information or diseases travel around much easier. In table 3.3 we can see that connectivity is not constant in the network, for example, some schools display higher connectivity than others, and the same goes for sex dynamics.

Density

Density is another overview of the number of relationships in a network. Is the amount of total connections divided by the amount of possible connections. Unlike connectivity, this value is bounded between 0 and 1. An example is shown in figure 3.21. Density

Overall Network with nodes proportional to eigencentrality
How 'well connected' is this node? (higher is better)

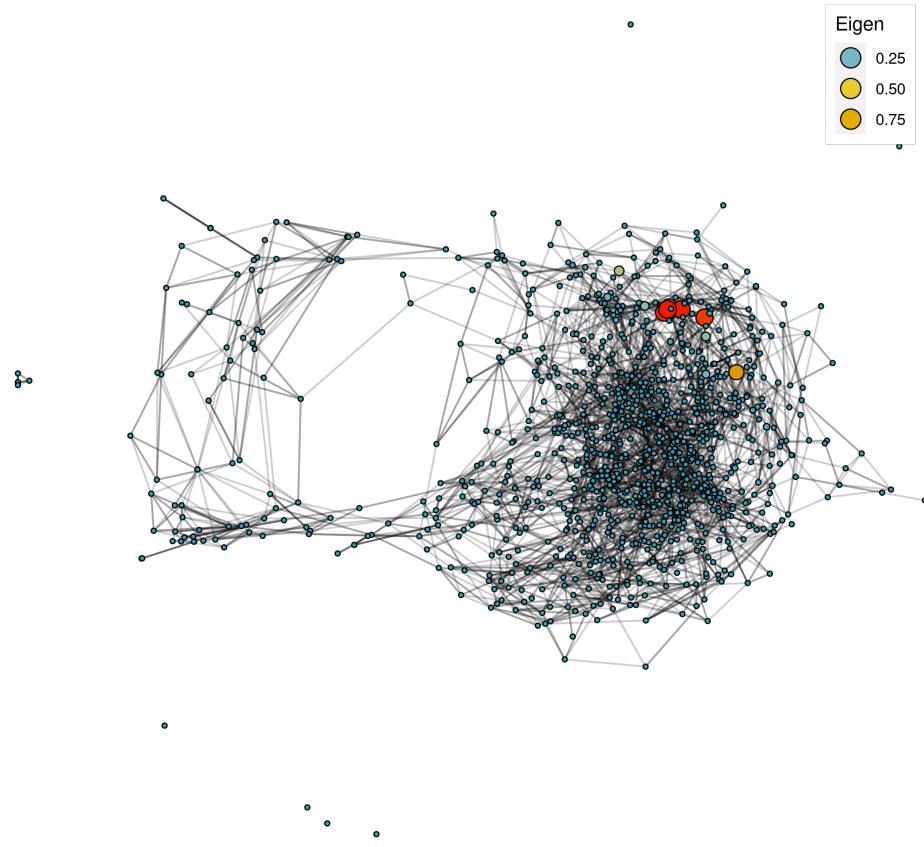


Figure 3.19: Overall network coloring nodes by their eigencentrality. Node size is also proportional to eigencentrality. A few nodes are popular people also connecting to other popular people.

by itself is not an exciting metric and needs to be compared with another density to get some useful information. For example, we can compare the friendship density of the high schools in Northern Norway, with the density in Southern Norway to see if there is a significant difference. We can do the same for nodes inside a network, such as comparing the friendship density of men against the friendship density of women.

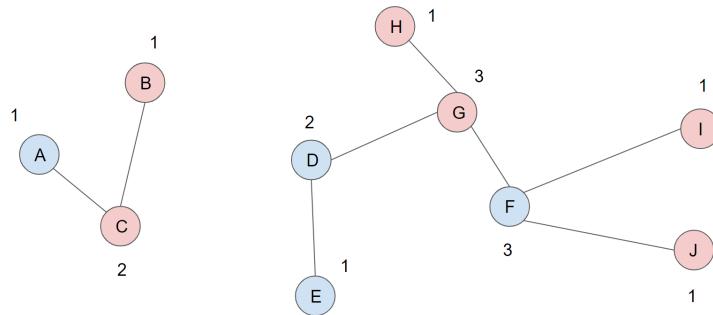


Figure 3.20: A network with two components. Nodes are labeled with the number of degree centrality in each. On the left component, we have an average degree of 1.3, while in the right component, we have an average degree of 1.71

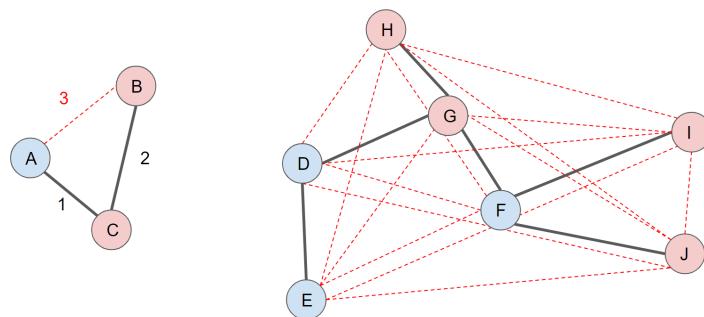


Figure 3.21: A network with two components. Black edges represent existing edges, and red thin, and dashed edges represent possible edges. On the left component, the 3 possible edges are shown and labeled. On the right component, no edge is labeled. The left component has a density of $2/3$ (66%), while the right component has a density of $6/49$ (12%). Notice that the right component, despite having higher average connectivity, also has a lower density.

Table 3.3: The average centralities measures for the whole population, each high school, and also men and women. Notice that in H5 we have a higher amount of disconnected triads, which causes outliers in the average value for closeness.

	Degree	Closeness	Between	Eigen	Bonacich
Population	5.34	0.0014	2362	0.0087	-0.0215
High school					
H1	5.4	0.0002	2696	0.0015	-0.0155
H2	5.06	0.0002	2197	0.0009	0.0084
H3	5.51	0.0002	2501	0.0015	-0.0423
H4	5.44	0.0002	2033	0.0011	-0.0154
H5	4.95	0.0148	1913	0.0000	0.0689
H6	4.23	0.0002	2258	0.0002	-0.2174
H7	5.76	0.0002	2526	0.0038	0.0294
H8	5.1	0.0002	2132	0.0627	-0.1456
Sex					
Men	5.31	0.0011	2405	0.0153	-0.0106
Women	5.37	0.0017	2317	0.0019	-0.0328

Homophily

Homophily is the core reason why social network studies work. Individuals tend to form strong bonds with people who are similar to them by factors such as marital status, race, economics, nationality, common interests, and many more [51–63]. This is also the biggest challenge when interpreting data, as we cannot be sure if individuals who are close to each other are influencing each other, or if they are simply sharing an environmental factor that influences everyone at the same time to no fault of the nature of their relationships.

Homophily is the ratio of, nodes that have an edge to another node that has the same property, against nodes that have an edge to a node of different property. Density only cares that there's an edge, while homophily cares about what types of nodes are being connected. Similar to density, homophily needs to be compared to another homophily number to gain some useful information.

First, we define edges that connect two nodes with the same attributes:

$$H_{any} \subseteq \{(x, y) | (x, y) \in V^2 \wedge x \neq y \wedge x_i = y_i\} \quad (3.3)$$

Second, we define edges that connect two nodes with the same given attribute:

$$H_{given} \subseteq \{(x, y) | (x, y) \in V^2 \wedge x \neq y \wedge x_i = y_i = A\} \quad (3.4)$$

Finally, we define edges that connect two nodes with either of them having a GIVEN attribute:

$$H_{either} \subseteq \{(x, y) | (x, y) \in V^2 \wedge x \neq y \wedge x_i = A \text{ or } y_i = A\} \quad (3.5)$$

For a given variable, there are two ways to calculate homophily:

$$Homophily_i = \frac{|H_{any}|}{|E|} \quad (3.6)$$

$$Homophily_{i=A} = \frac{|H_{given}|}{|H_{either}|} \quad (3.7)$$

In figure 3.3, we have 9 edges. Edges {B,C} and {H,G} connect red to red. Edges {A,D} and {D,E} connect blue to blue. Therefore, we have 4 edges connecting the same colors

out of 9 possible. $Homophily_{color} = 44.4\%$. Specifically for red, we have the same edges $\{B,C\}$ and $\{H,G\}$ connecting red to red, and edges $\{A,C\}$, $\{B,C\}$, $\{D,G\}$, $\{H,G\}$, $\{G,F\}$, $\{F,I\}$ and $\{F,J\}$ connecting at least one red. Then, $Homophily_{color=red} = 28.6\%$.

Is also possible to do a binomial test on the homophily to check if it is significant. This is simply measuring how many nodes of the same attributes we have, and how many are friends between them. If the number of edges is proportional to the number of nodes, that indicates that the number of relationships within the same group is what we would expect by random chance. But if we have too little, or too many relationships among them, it might indicate that the group is biased toward forming relationships among themselves or avoiding each other. Following the figure 3.3 example:

- How many edges are reds friending red? 2
- How many reds do we have? 6
- What is the probability of being red? 0.6

The calculated binomial test is:

$$\binom{6}{2} \cdot 0.6^2 \cdot (1 - 0.6)^{6-2} = 0.14$$

The probability $Homophily_{color=red} = 0.286$ is greater than we would expect by chance (0.14), however, is not greater by much. The p-value for a two-sided test is 0.23 indicating that we can't tell if the reds are biased in this graph.

Average path length

For fully connected networks, the Average Path Length (APL) is the average of all possible paths. If a network is not fully connected, then the average path length is infinite. In such cases is more interesting to find the APL per sub-network. APL is related to density. Forming 4 different interesting cases:

- Low APL and low density will have a close solution to a minimum spanning tree. This is a subset of edges that connects all the nodes without cycles and with the minimum possible edges or edge weight. This case is important because it shows the cheapest way to lay down roads connecting cities so all cities are connected, expending the least amount of money possible.

- High APL and low density mean all the nodes are connecting using very few edges. This is bad if we describe for example a network of roads, because it will have very long paths and at the same time all cars need to pass through the same common edges. This is good if we have an infectious disease because it means we can cut the network in half easily and avoid propagation.
- Low APL and high density. This case means that all the nodes are very well connected, and traveling from one to another is easy. As opposed to before, great for roads due to low traffic, but bad for diseases because is very difficult to slow down or stop the spread.
- High APL and high density. This means that you have many sub-networks that are very well connected, but the connection between sub-networks is done with a really low number of edges forming choking points.

Coverage

Coverage of a node refers to how many nodes you can reach from that particular node in equal or less a particular path distance (figure 3.22). This is also referred to as how many steps you need to reach any given pair of nodes if their distances are all equal.

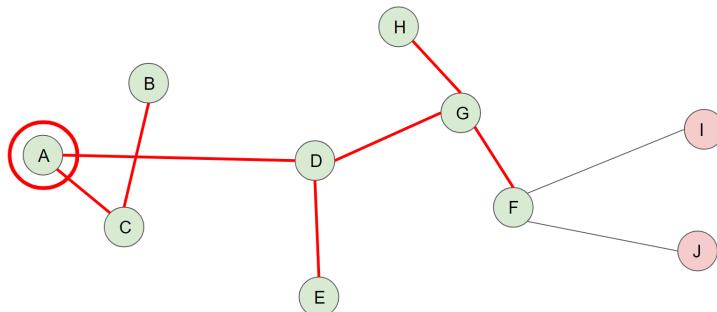


Figure 3.22: An example of coverage. For a given path of distance equal to 3, node A can reach all the green highlighted nodes using the red highlighted paths.

Reachability

Reachability refers to the average coverage per step in a network. Reachability is related to connectivity, as the network is more and more dense, is more easy to cover more nodes with fewer steps. In figure 3.23 we see an example. This is an important variable when we deal with infectious diseases, if a large area of the network is reachable very quickly, the disease will spread very quickly as well.

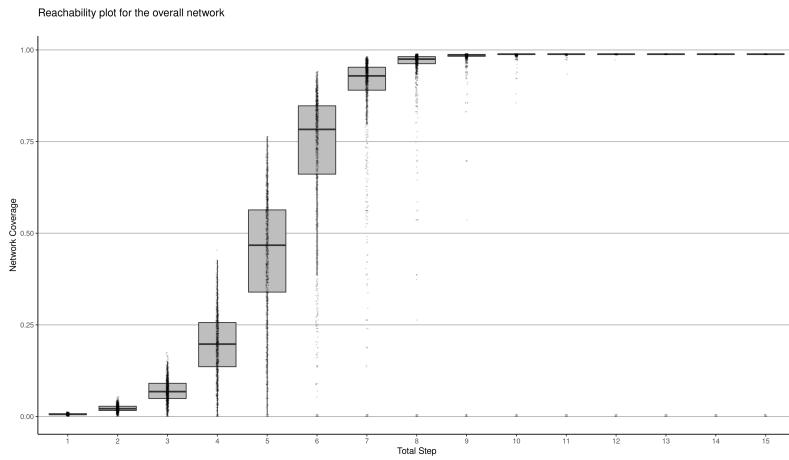


Figure 3.23: A reachability plot in the overall network. The x-axis represents the number of steps taken. The y-axis represents the proportion of network coverage. After 4 steps, we see that about 25% of the nodes manage to cover 25% of the network. In the next step, nearly 50% of the nodes, cover 50% of the network. In the next step, we see that the median surpasses the 75% coverage, so the node reach grows exponentially in this network. It never reaches 100% because some of the nodes are isolated, we see this better in the outliers that start showing from step 7 at the bottom of each box.

Random Graphs

In graph theory, random graphs refer to either a graph that has been generated randomly, or specifically regarding this section, describing graphs using a probability distribution [64]. The total number of possible graphs, with N nodes, where each node has A attributes, where each node can be connected to several nodes, is given by:

$$G = 2^{\binom{N}{2} + AN} \quad (3.8)$$

For a simple graph of only 100 nodes, and 2 possible attributes, we will already have 2^{5150} possible graphs. Working with random graphs is often computationally unattainable, however, there are several non-parametric simplified tests, such as bootstrapping [65] or jackknife [65] that can be performed on the graph.

Furthermore, the edge distribution probability is not necessarily random. For example, nodes N1 and N2 might have an almost certain connection of 99% probability of appearing in a random graph. The classical model of random graphs is the Erdős-Rényi model [66], which has two variants:

- **G(n, p):** In this model, a graph is constructed by connecting n nodes randomly. Each edge is included in the graph with probability p independent of every other edge. This model is most commonly used due to its simplicity given that edges are independent from each others. This might not be true in reality, for example, in a graph where a person is forced to choose exclusively between two or more communities.
- **G(n, M):** In this model, a graph is constructed by choosing exactly M edges from the possible $\binom{n}{2}$ pairs of n nodes.

These models can be used to describe and compare the properties of networks, such as connectivity, distribution of node degrees, the likelihood of forming clusters or communities, and so on.

ERGM

Similarly to random graphs, Exponential Random Graph Models (ERGM) [67] are a family of statistical models used to analyze network structure. ERGMs allow for dependencies. This means in ERGMs, the likelihood of an edge existing between two nodes can depend on other connections in the network, reflecting more realistic scenarios. Furthermore, ERGMs are particularly useful in scenarios where the network exhibits diads, triads, and so on, which are not adequately captured by basic random graph models.

The probability of observing a given network is modeled as an exponential function of a linear combination of network statistics, which are functions of the graph. These statistics are chosen based on hypotheses about the processes that might be driving network formation.

The probability of observing a graph is given by:

$$P(G) = \frac{1}{Z} \exp \left(\sum_k \theta_k g_k(G) \right) \quad (3.9)$$

where:

- θ_k are parameters to be estimated.
- $g_k(G)$ are network statistics, such as the number of edges, diads, triads, centralities, etc...).
- Z is a normalization constant that ensures the probabilities sum to 1.

A downside of ERGMs is that the formulation requires model expertise of the network, otherwise, the model might fall into what is known as "model degeneracy". This means the model places a high probability on a few networks that do not resemble the observed data. Therefore, careful selection of network statistics and constraints is necessary to avoid this issue. Furthermore, ERGMs are more complex to analyze and compute compared to the simplified tests of the random graph models.

Simulations

Finally, let's talk about how all these concepts can be put together in a practical case, which is how a disease can advance through the network. This topic has already been studied extensively; even a zombie apocalypse has been described and published. [68]

We can apply the same method to our network, to which we made a simulation engine capable of performing with the following infectious parameters:

- What is the probability of any given person being able to spread the disease to another connected person? (ie: do they stay in contact often? is the person using face masks?)
- What is the probability of a person receiving the disease? (ie: Does he wash his hands often? is he immunocompromised?)
- What is the probability of passing the disease and immunizing yourself once you are infected? How does this change over time? (ie: the Polio vaccine gives lifelong immunity, whereas chickenpox last for about 10 to 15 years)
- What is the probability of dying for each person? Does this increase over time once infected? (ie: comorbidity factors such as obesity and COVID-19)
- Is this disease recurrent? (ie: Epstein–Barr virus)

- Is this due to a biphasic virus? (ie: tick-borne encephalitis)
- Do we allow random jumps in the network? (ie: airborne diseases)
- How long until we develop a vaccine? How many can be vaccinated each unit of time?

For this example, we generated an unknown novel disease that has a 30% chance of transmission from person to person, as long as they are connected in the overall network. People can get immunized at any time and have a 10% chance of doing so, but they can also die at any time and have a 1% of so. We study six cases with 200 simulations each, one in which the disease follows its natural order, and the other five applying a vaccine to the population at step 10, with the restriction that we only have 30 vaccines per step. The vaccine is administered to people that we know that have never shown signs of infection, don't currently have the disease, didn't have the disease in the past, and are not dead. The vaccine is 100% effective and has no side effects. In the first vaccine case, we give the vaccine at random, the second case we give the vaccine first to people with a higher degree of centrality (most friends), third case higher closeness (shorter distance to rest of nodes), fourth case higher betweenness (connecting subnetworks), and fifth case higher eigencentrality (important people). The results for this simulation can be found in table 3.4.

Table 3.4: Comparison between the different simulation results. The first two cases can be seen in figure 3.24, and figure 3.25. Death rate and immunity rate are calculated at the final step of the simulation. Peak infection is calculated with respect to the maximum infection coverage in each simulation, and in which step did happen.

	No vaccine	Vaccines at step = 10 , 30 / step				
		Random	Degree	Closeness	Betweenness	Eigen
Total death (n)	66 ± 38	23 ± 19	18 ± 15	18 ± 17	16 ± 15	26 ± 19
Death rate (%)	6.30 ± 3.57	2.17 ± 1.77	1.71 ± 1.39	1.70 ± 1.55	1.53 ± 1.39	2.49 ± 1.80
Immunity rate (%)	62.15 ± 34.51	97.85 ± 1.73	98.35 ± 1.33	98.31 ± 1.51	98.51 ± 1.34	97.54 ± 1.74
Peak infection (%)	32.71 ± 18.29	12.23 ± 9.99	10.08 ± 8.16	9.10 ± 8.52	9.21 ± 8.33	13.98 ± 10.29
at step	23.20 ± 10.65	19.15 ± 7.67	17.47 ± 6.00	17.21 ± 8.70	15.86 ± 6.97	19.35 ± 9.05

In the first case (figure 3.24) if we are lucky, the disease will start in an isolated node, however, the overall immunity will be very low with the subsequent risk of the

disease restarting again in a more connected node. We can also see that both the peak infection and death rates are quite high, with about 66 students dying on average.

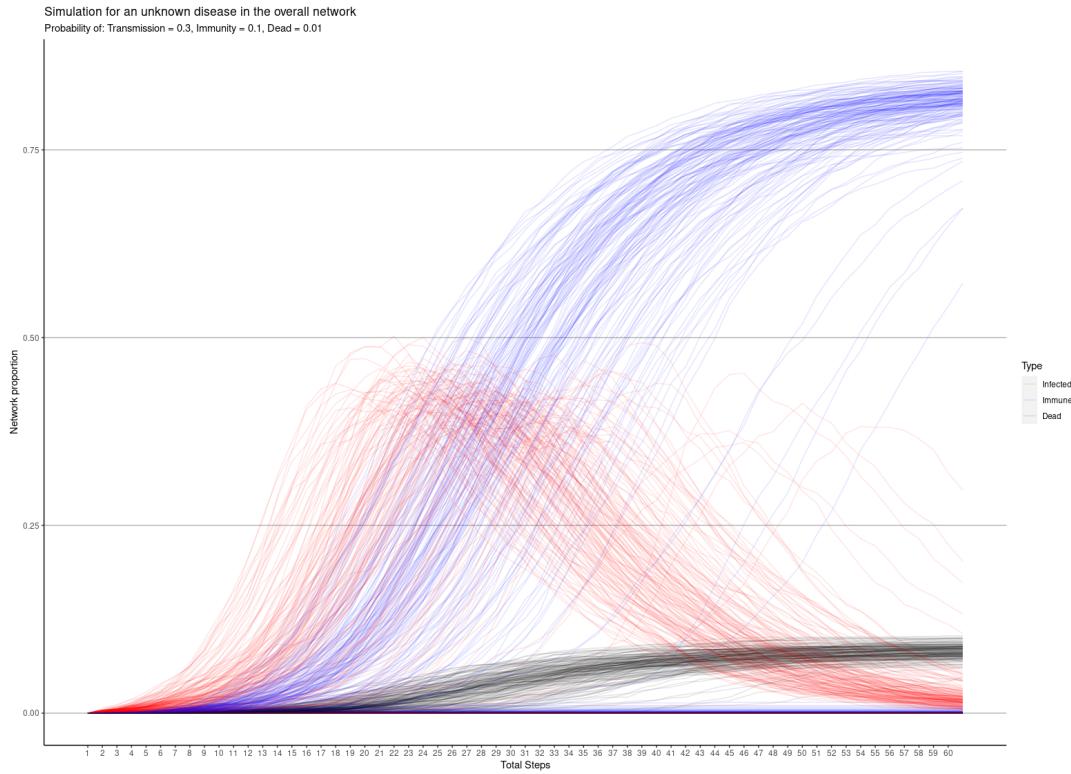


Figure 3.24: Simulation of the spread of an unknown disease in the network. In the x-axis, we have steps which is an arbitrary representation of time. In the y-axis the proportion of the network that each category has at any given time. Each red, blue, and red, line represents one of the 200 simulations. Red represents infected, Blue represents immunized, and Black represents dead.

For the second case (figure 3.25), we introduce a vaccine. As expected the death rate drops dramatically, with about 23 students dying on average and an almost entirely immunized population in all cases. The peak of infections is reduced by more than half and is easier to predict when the worst time is going to happen.

The rest of the cases produce similar figures with the slope of immunity being a bit more or less pronounced. The best results are when we apply the vaccine to people connecting subnetworks, in that way, we can isolate the disease very quickly. Notice that the number of deaths is lower when we immunize first people connecting sub-networks together ($n = 16$), rather than just giving it first to the most well-connected people ($n = 18$), which suggests that having prior knowledge of the network topology can help to save lives.

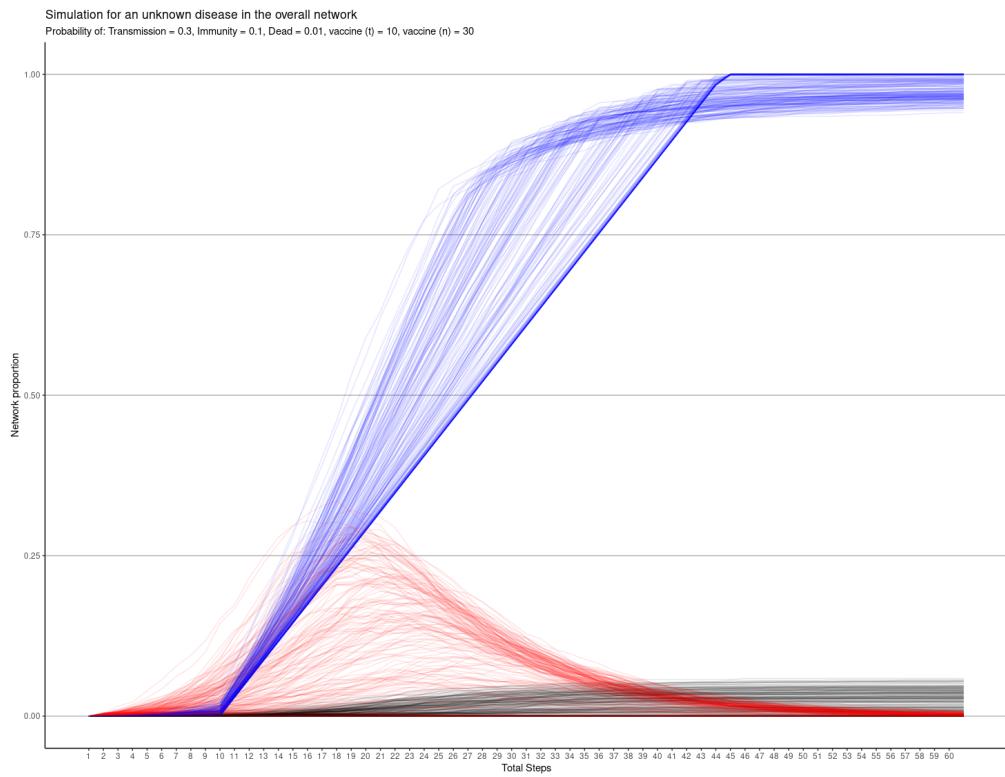


Figure 3.25: Simulation of the spread of an unknown disease in the network. In the x-axis, we have the steps in time. In the y-axis the proportion of the network that each category has at any given time. Each red, blue, and red, line represents one of the 200 simulations. Red represents infected, Blue represents immunized, and Black represents dead. A vaccine is introduced at $t = 10$ which suddenly spikes the immunization ratio and lowers dead rate

We also have the option of running this type of analysis against any layout of any network and making a comprehensive video simulation. This can be seen in figure 3.30.

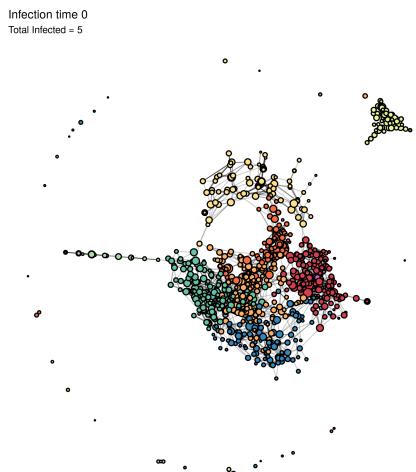
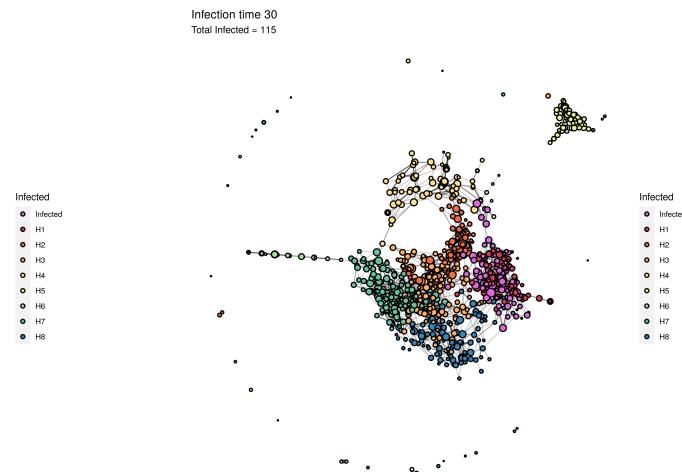
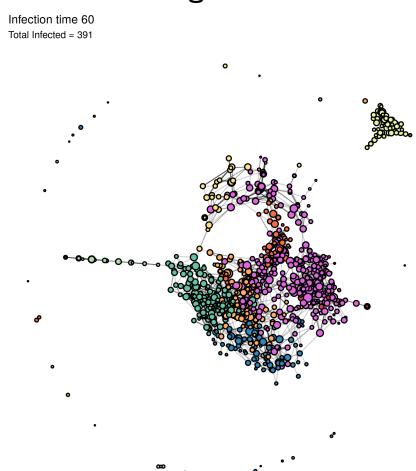
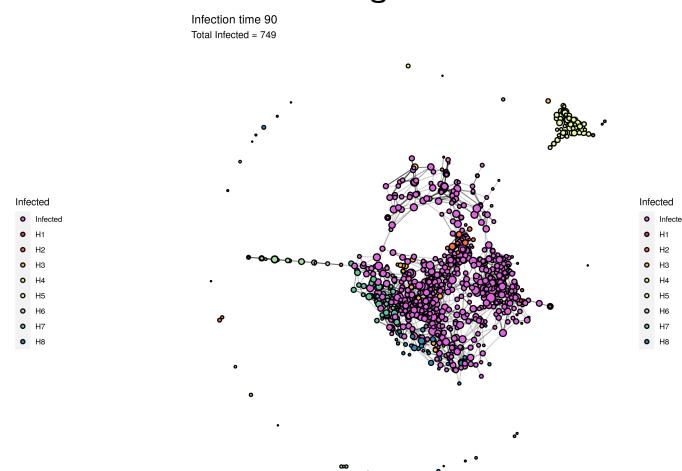
**Figure 3.26:** a**Figure 3.27:** b**Figure 3.28:** c**Figure 3.29:** d

Figure 3.30: A simulation of a disease advancing through the School Network using an MDS layout. Each sub-figure represents the disease advance after (a) 0 steps with 5 random initials infected (b) 30 steps (c) 60 steps (d) 90 steps. Infected people are highlighted in violet, and each school is highlighted in a different color. The disease starts in H1 and spreads to closer contacts. H5 and all isolated components are spared because is not connected and thus impossible to reach in this particular model. The complete video of the simulation can be found at <https://github.com/uit-hdl/mimisbrunnr/blob/main/results/network/simulation.mp4>

3.2 Machine learning

3.2.1 Introduction

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that consists of methods and algorithms that can learn from data and make decisions based on such data. The definition of what is AI, ML, and even statistics is blurry. For example, linear regression was introduced in the early 1800s [69] when no concept of AI existed, and regardless, it is considered a popular ML method. In this section, we offer an introduction to these topics for the reader unfamiliar with these concepts, as well as a more in-depth explanation of methods that are used in this thesis.

3.2.2 Basic concepts

Tasks

Tasks refer to the specific types of problems that a machine-learning algorithm aims to address. Some basic examples of these tasks include:

- **Classification:** This means predicting the category or class of a data instance. For example, in disease diagnosis, whether a patient is ill or not, and if so which type of disease.
- **Regression.** Similar to classification, regression tries to guess a value based on a data instance such as temperature in weather forecasting. However, regression is often referred to as continuous data rather than categorical data. Nevertheless, regression can be used to also classify categorical data, such as in the case of logistic regression.
- **Clustering:** Clustering is a type of problem that consist of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. For example, the characterization of patient clusters as similar patients needs similar treatment. The difference with classification is that classification knows the number of clusters beforehand, while clustering is not known and is up to the researcher to decide if the given clustering makes sense beyond the mathematical optimization provided by the algorithm.
- **Dimensionality Reduction:** Is the process of reducing a set of variables, which typically is too big to handle computationally, or has too many irrelevant variables,

to its most important variables minimizing the loss of the intrinsical relationship between these variables. Principal Component Analysis (PCA) and Partial Linear Regression (PLS) are classical statistical examples of this. While Autoencoders would be the classical ML method.

Types of Learnings

In the context of machine learning, 'learning' refers to the distinct methodologies employed by algorithms to understand when it is committing errors or getting the right answer while solving tasks.

- **Supervised Learning:** When the ML algorithm uses a predefined answer found in the data to evaluate the accuracy of the algorithm.
- **Unsupervised Learning:** Opposed to the previous one, the data does not contain the solution and is up to the algorithm to figure out by the patterns in the data.
- **Reinforcement Learning:** This is the third paradigm of the type of learning in which an agent tries to find the solution to a problem. Like unsupervised learning, reinforced learning does not need to know the solution for training, however, it requires an interpreter to tell whether the action taken by the agent was satisfactory or not. The agent will then try to maximize the amount of times that the actions are satisfactory.

Common terminology

Finally, some key concepts that are also needed to understand the methodology.

- **Label:** In supervised learning, when the data used to train the algorithm is already classified it is said to be labeled data.
- **Features:** These are the individual measurable properties of the problem being observed. Typically interchangeable with the term variables.
- **Loss function:** A loss function, also known as a cost function, quantifies the difference between the predicted values by the model and the actual target values. The loss function is used during the training phase to evaluate the performance of the model, and the objective is to minimize this loss during optimization, thereby improving the model's accuracy. Not all ML methods use a loss function for

optimization, but all of them use some sort of optimization function. For example, entropy is a method that can be used in Random Forests instead.

- **Gradient Descent:** In many ML processes, the parameters are updated through an optimization algorithm. Gradient descent is a popular algorithm to do so. The gradients used to update the weights are derived from the loss function.
- **Hyperparameters:** These are the different options that can be adjusted prior to training an algorithm and do not change while the algorithm is running.
- **Training and test sets:** The data used to train the algorithm is usually divided into two sets. One is the data used to train the algorithm (training set) and the other one is the data used to check if the algorithm has been trained correctly (test set). A key characteristic is that both sets need to have similar properties and a good representation of all the possible data (often known as being balanced) in order to train and test the algorithm effectively.
- **Validation set:** Unlike the test set, the validation set is used repeatedly during the model development process for tuning the hyperparameters, while the training set is done for model evaluation.
- **Cross validation:** Cross-validation, similarly to the train and test data set, is used for training and evaluation of the model. However, unlike a simple train/test split, cross-validation involves running a series of different training and testing splits to ensure that each observation from the original dataset has the chance of appearing in the training and test set.
- **Overfitting and Underfitting:** When an algorithm can't make a good generalization of the data it will be a bad predictor when used with real data. This is known as overfitting, such as for example having a single data point in the training set of a patient that has glucose above 200 mg/dL not being hypoglycemic, the model will decide that anyone close to that patient is not hyperglycemic either when in reality it would be more sensible to classify it as such. Underfitting is the opposite effect in which the model has a very poor predictive capability.

3.2.3 Algorithms

An algorithm in machine learning is a set of rules to perform a task such as classification, clustering, or regression. Each algorithm can be set up with different options, typically adjusting the hyperparameters of the algorithm.

The term algorithm is often interchanged with the term "model". However, algorithm refers to the highest mathematical abstraction definition and model should refer to a specific implementation of the algorithm.

- **Artificial Neural Networks (ANNs)** ANNs [70] simulate the behavior of neurons in the brain. They consist of layers of interconnected nodes or neurons. Each connection has a weight that is adjusted during the training process to minimize the loss of function. Typically, all features are represented, each in one input neuron. The simplest ANN consists in connecting all the input neurons to a single output neuron. This would be the same as trying to solve a linear regression equation, but letting the gradient descent algorithm figure out the coefficients instead of solving the derivative of the equation that optimize the error. Adding one or more hidden layers with non-linear activation functions allows ANNs to model complex, non-linear relationships that go beyond what linear regression can handle. Furthermore, it is not necessarily that we only have one output neuron either.

ANNs tend to be a poor model if the training data is not large enough. ANNs require tuning various hyperparameters such as the layer's architecture to achieve a good prediction rate and are typically very computationally expensive to train. Also, ANNs behave like a black box model and the variables importance is very hard to calculate. However, ANNs can reproduce better complex and non-linear relationships between input features.

- **Decision tree.** A decision tree [70] is a flowchart-like tree structure where each internal node represents a "test" on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label. Instead of a loss function, decision trees tend to use either entropy or gini impurity.

Entropy is used to calculate the information gained from a particular split. Information gain is the difference in entropy before and after the split. Decision tree algorithms aim to maximize information gain at each split, effectively reducing

uncertainty with each decision made. On the other hand, gini impurity indicates the likelihood of data being misclassified if it were given a random class label. Instead of maximizing, like entropy, gini needs to be minimized. Gini impurity is generally faster to compute than entropy, but entropy might be more sensitive to changes in the class probabilities of the nodes.

Decision trees, are much faster to train than ANNs and are very easy to understand the importance of each feature, but they are prone to overfitting and tend to have worse predictive value.

- **Random Forests (RF)** RF [70] simply creates multiple decision trees to make predictions and select the consensus of all the trees for categories and mean prediction (regression) for continuous data. RF randomly selects a subset of input features for each tree. It shares all the advantages and disadvantages of decision trees, but they offer a better prediction than a single decision tree.
- **K-Nearest Neighbor (KNN)** KNN [70] works by finding the closest data points in the training dataset—known as the nearest neighbors—to a new data point and predicts the output based on these neighbors. KNN is very easy to understand and implement. For any new data point, it simply calculates the distance from that point to all others, finds the nearest neighbors, and votes for the most popular output class.

As the number of features increases, the volume of the space increases exponentially. In practical terms, it means an exponential need for memory to store the entire dataset. It also needs to compute the distance of a new data point to all existing data points. Because of this, KNN tends to be used after a dimensionality reduction in the data if the dataset is very large. Finally, selecting the right number of neighbors can be tricky. Small numbers tend to overfit whereas large numbers tend to underfit.

- **Support-vector machines (SVM)** SVM [70] is a set of supervised learning methods used for classification, regression, and outliers detection. The basic model is defined by a separating hyperplane that best divides a dataset into two classes. SVMs are good with high non-linear dimensional data and have good generalization, however it shares the disadvantages of ANNs of computationally intensive and difficult to interpret.

- **Bayesian algorithms** are a set of techniques based on Bayes' Theorem, which describes the probability of an event based on prior knowledge of conditions that might be related to the event.

Bayesian methods share the same advantages of ANNs but with an extra disadvantage. Because the prior information is crucial, a very careful consideration and expertise in choosing priors and knowledge of the data distribution that is appropriate for the specific context of the problem is needed. As such, ANN tends to be a more attractive model.

3.2.4 Interpretability

An important concept is the interpretability aspect of the ML models [71]. It is crucial as ML is increasingly being used in critical areas such as healthcare, finance, and law where clear explanations are essential. In general, models offer easy interpretability in exchange for lower accuracy and the other way around. For example, a decision tree is very easy to follow and understand why it reaches the decision because the user just needs to follow a big IF/ELSE diagram. RF is also "only" following many of these diagrams and checking which is the consensus. But as discussed above, other methods such as ANN tend to have better prediction power. However, for ANN, we will have what is practically an untraceable gigantic weight matrix whose number depends on the backpropagation function iterated thousands of times.

In either case, ultimately what is important is to understand which features influenced the training of the model and in which range of values are relevant. For example, it's important to identify which variables are most important for predicting the risk of heart disease; variables such as age, blood pressure, and cholesterol levels may be more important than gender, BMI, or occupation.

A very good example of the importance of interpretability was the surprising bias in the model with high accuracy for COVID-19 diagnosis based on X-ray images [72], in which the model was simply finding letters outside the lung area because the X-ray machine in the hospital where most of the patients with COVID were tested used very prominent "R" characters to distinguish left and right side of the chest area. The model rightfully associated the "R" with COVID, leading to a very good prediction rate because the dataset was not properly preprocessed. However, a simple interpretability analysis of the model helps to debug simple mistakes such as this one. Here we take a look at the most common interpretability algorithms with emphasis on the ones used in this thesis:

Individual Conditional Expectation

Partial Dependence Plot (PDP) [73] is a visualization tool used in machine learning to analyze the effect of one or two features on the predicted outcome of a model, regardless of the values of other features. PDP can hide data patterns as it measure the global trend of a feature. In contrast, Individual Conditional Expectation (ICE) [74] each data instance individually, as a PDP is the average of the lines of an ICE plot. This method is model-agnostic and can be used in any ML model.

However, two important limitations are that ICE plots are only useful in displaying one feature at a time and that it assumes independence between features. Another limitation is that they are computationally expensive, however, other powerful methods share this limitation as well.

Local Surrogate

Surrogate models are interpretable models that are used to approximate the predictions of a more complex model. They work by training another ML model, usually something more simple like an RF or decision tree, to understand the original model; hence the name surrogate model.

Local interpretable model-agnostic explanations (LIME) [75] test individual data points against the original model. It tries to learn what happens when variations of the data are introduced. Once all the trials and variations are generated, this data is given to a simple surrogate model to check feature importance.

This method is also model-agnostic and computationally cheap. However, the disadvantages are that it only explain individual predictions without an overview of the whole model behavior and that changes in the variations introduced in the input data can result into different explanations of the features.

SHAP values

SHapley Additive exPlanations (SHAP) [76] calculates the exact contributions of each feature for individual predictions based on the Shapley values [77] to ensure that the explanations adhere to desirable mathematical properties. Shapley values are very computationally expensive to calculate as the time increases exponentially with the number of features to analyze. To overcome this, an approximation with Monte-Carlo sampling is used to approximate results when the number of features is high [78].

Compared to LIME, the Shapley value allows for more nuanced comparisons. Instead of just comparing a prediction to the average, you can compare it to the predictions for a specific subset of the data or even a single data point. This method can reveal more detailed insights into why a particular prediction was made, especially in comparison to more closely related instances or scenarios. Furthermore LIME assumes linear behavior locally, which is not necessarily true; while Shapley values do not.

However, beside the computational time, Shapley have two disadvantages with respect to LIME. First, Shapley uses all the features, while LIME can use a surrogate model such as Lasso or Ridge regression to dismiss features that are not useful. Second, Shapley doesn't give you a model to make predictions about the features, only their current importance score.

TreeSHAP, [79] is a variant of SHAP for tree-like models, like decision trees, or random forests. It means to increase computational performance, however, it offers a worse variable analysis performance [80, 81].

One final thought is that you can cluster data based on SHAP values with the built in method "Stacked SHAP explanations". Individual instances of the data offer individual SHAP values explaining why that particular instance has the model's predicted value. One just needs to cluster instances with similar explanations together.

Mean Decrease in Impurity

Mean Decrease in Impurity (MDI) [82] is a measure used in random forests to quantify the importance of each feature in predicting a target variable. MDI is calculated by summing the decreases in Gini impurity weighted by the number of samples that reach the node, for each time a feature is used to split the data in a tree, averaged over all trees in the forest.

MDI is a very straightforward interpretation of feature importance based on how much each feature contributes to reducing uncertainty in the model and is computed automatically during the training process of the RF. Furthermore, it is non-parametric as it does not assume any form of the relationship between features. However, if the number of trees in the RF is small, it can lead to variability in the importance scores. Each tree in a Random Forest is built on a different bootstrap sample of the data. A bootstrap sample is a randomly selected subset of the data (with replacement), which means each subset may have some data points repeated and others left out. This

sampling is great because introduces variability among the trees, but if it is not done enough (small number of trees), then MDI can be biased.

An alternative to MDI is Boruta [83]. Boruta is generally more robust in identifying important features, especially in cases where features are correlated or when the dataset includes irrelevant features; but at the cost of being computationally expensive.

3.3 Statistical Analysis

3.3.1 Introduction

This section gives an overview of the common statistical methods used throughout the thesis.

3.3.2 Contingency tables and Pearson X^2 test

Contingency tables are used to determine if the total number of samples in a combination of two categorical variables (i.e.: Sex and BMI), happens to be within the range of the values that we would expect by chance [84]. Several tests can be performed to determine if these combinations are suspicious or not. One in particular is the Pearson Chi-square test which uses a Chi-square distribution for a specific degree of freedom. The degree of freedom is related to the number of variables in the calculation and they represent the number of independent pieces of information available to estimate or calculate statistics. The Chi-square value is calculated by taking the sum of the squared differences between the observed and expected data, dividing by the expected data, and then multiplying that result by the number of variables. The resulting value is compared to the values in the Chi-square table to determine if the null hypothesis can be rejected or not.

If the p-value of the analysis is significant, it means that some combinations of variables are suspiciously high or low. Each combination can, later on, be tested using a simple binomial test to highlight values far away from the expected ones.

The Pearson Chi-square test requires that sample data should be collected in a way that the classification into one category does not depend on the classification into another category. There also need to be enough samples in each cell, generally defined as 5; if this condition is not met, Fisher's Exact Test or Yates' correction can be used instead.

Here we display a simple example:

Step 1: Table with the data

Suppose we want to test whether there is an association between gender and smoking. This is how a hypothetical data collection might look like:

Table 3.5: Table with example data for contingency analysis.

Sex/Smoking	Yes	No	Total
Men	30	10	40
Women	20	30	50
Total	50	40	90

If the smoking status were random, the men's numbers should be near 20 for each category; with a bit more towards yes because there are more smokers than no smokers. Same for women, which number should be near 25 for each category because there are more women than men. However, it seems that males tend to lean too much towards yes, and women towards no.

Step 2: Calculate Expected Frequencies

The expected frequency for each cell in a contingency table is calculated using the formula:

$$E_{ij} = \frac{(R_i \times C_j)}{N} \quad (3.10)$$

where E_{ij} is the expected frequency for cell i, j , R_i is the total for row i , C_j is the total for column j , and N is the grand total. For example, for "Men" and "Yes" the formula would be:

$$E_{\text{Men, Yes}} = \frac{(40 \times 50)}{90} \approx 22.22 \quad (3.11)$$

Step 3: Find the Chi-square Statistic

The Pearson Chi-square statistic is calculated using:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3.12)$$

where O_{ij} are the observed frequencies, and E_{ij} are the expected frequencies.

Calculating for each cell:

$$\chi^2 = \frac{(30 - 22.22)^2}{22.22} + \frac{(10 - 17.78)^2}{17.78} + \frac{(20 - 27.78)^2}{27.78} + \frac{(30 - 22.22)^2}{22.22} \quad (3.13)$$

$$\chi^2 \approx 2.72 + 3.39 + 2.19 + 2.72 \approx 11.02 \quad (3.14)$$

Step 4: Interpret the Result

We compare the calculated χ^2 statistic to the critical value from the χ^2 distribution table at the desired significance level (in our case, let's say 0.05) with degrees of freedom $df = (R - 1)(C - 1)$ where R and C are the number of rows and columns in the table. For this 2x2 table, $df = 1$.

If the calculated χ^2 is greater than the critical value (3.84 for $df = 1$ and $\alpha = 0.05$), reject the null hypothesis that there is no association between the variables. In this example, 11.02 is greater than 3.84. The p-value is 0.0009 which would suggest that there is indeed a bias between sexes and smoking in our data.

3.3.3 Logistic regression

Logistic regression is a type of statistical analysis used to predict the probability of an event occurring based on a set of variables [85]; in our case, we use many results to tell what the probability of a subject having an effect as the number of friends increases or decreases. It is named after the logistic function which takes an S-shaped curve that can take any input value and map it to a value between 0 and 1, representing the probability of an event occurring. It is primarily used when the dependent variable is categorical or binary, such as yes/no, true/false, or success/failure. The logistic function is defined as:

$$\text{logistic}(\eta) = 1/(1 + \exp(-\eta)) \quad (3.15)$$

For classification, it is better to have the output modeled as a probability, which of course ranges from 0 to 1. Therefore, we can convert a linear regression equation into a logistic function such as this:

$$P(y^{(i)} = 1) = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_p x_p^{(i)}))} \quad (3.16)$$

The logistic function transforms the weighted sum into a probability. Consequently, we must restructure the equation to ensure that only the linear term appears on the right side of the formula for proper interpretation.

$$\ln \left(\frac{P(y = 1)}{1 - P(y = 1)} \right) = \log \left(\frac{P(y = 1)}{P(y = 0)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (3.17)$$

This formula shows how the log odds compare to the linear model. The odds ratio tells us how much more likely it is for the dependent variable to occur compared to not occurring, given a change in the independent variable. Probability and odd ratios are not directly interchangeable because they have different scales and have a non-linear relationship. The main disadvantage of the logistic regression is the interpretation of the odd ratio weights since they are multiplicative and not additive. However, an approximation from odds to probability can be approximated by:

$$\text{Probability} = \text{odds}/(1 + \text{odds}) \quad (3.18)$$

Logistic regression assumes that independent variables should not be highly correlated with each other. It also sensitive to outliers, and it does not work well with small sample sizes.

3.3.4 Autocorrelation models

An autocorrelation outcome model is a type of statistical model used to describe certain types of dependent variables where the current value of the variable is regressed on its own previous values. This type of model is particularly useful in time series forecasting [86]

and in contexts where data points are not independent but instead exhibit some form of temporal or spatial dependency. In the context of network analysis, these models [87] strive to find how a network has changed until it finds an equilibrium point at the current time. Furthermore, it was used in previous works [88] which led to the proposal of this PhD project.

Let's define the model as:

$$Z = \alpha W Z + X \beta + \epsilon \quad (3.19)$$

And suppose we would like to predict the academic grade in a particular school. Each of the elements can be described as:

- **Z:** This is the dependent variable vector whose values are being modeled, for example, a series of observations over time or across different entities or locations. In our example, this would be a vector with the final grade of each student.
- **α :** This parameter measures the magnitude of the network effect, indicating how much the connections or interactions within the network influence the value of Z. Higher α suggests stronger influence. In our example, it would be how much students are able to influence each others.
- **W:** This is a matrix that represents the structure of the network or the weights of connections between different units or entities within the dataset. This can be for example, the Laplacian matrix described in section 3.1.4.
- **X:** This is a matrix of independent variables or covariates that also affect the dependent variable Z. For example, number of hours studied per week, alcohol consumption frequency, whether the student goes to class in person, online, or neither, etc...
- **β :** The vector of regression coefficients. Each element of β corresponds to the effect of a particular independent variable in X.
- **ϵ :** This represents the vector of stochastic errors, assumed to be independent of one another. These errors capture the random fluctuations in Z that are not explained by the network effects or the independent variables.

The errors ϵ may be interdependent. In such a case, ϵ itself is modeled as an autoregressive process. This is a common setup in spatial econometrics and network analysis, where the residuals (errors) from a regression may be correlated due to spatial or network-related dependencies. Such network autocorrelation can be modeled via the inclusion of a term $\bar{\epsilon}_W = W\epsilon$. The previous model can be rewritten as:

$$Z = X\beta + \epsilon, \epsilon = \rho W\epsilon + v \quad (3.20)$$

Let's break down the model and discuss how to estimate the unknown parameter (ρ):

- **ρ :** is the parameter that measures the strength of the network autocorrelation among the errors. It is similar to the previous α , but in this case, it quantifies how much the error for one observation is influenced by the errors of its neighbors or connected units.
- **v :** is a vector of independent random perturbations, typically assumed to be normally distributed with mean zero and some variance (σ^2).

The objective is to estimate the unknown parameter (ρ), which quantifies the degree of autocorrelation in the residuals due to the network structure. When the structure of the data is known, this is done by Maximum Likelihood Estimation (MLE) [89]. In network analysis, more often than not the structure of the network is so big that makes the likelihood methods too time-consuming to calculate. In such cases, a pseudolikelihood is used [90]. It provides a practical compromise, balancing computational feasibility with statistical rigor. If the model can be expressed in a way that some variables are uncorrelated with the errors but correlated with the predictors, you can use Generalized Method of Moments (GMM) [91].

Further refinement of the model can be introduced by measuring the W lagging over time [87]. This would be the natural extension of our work if data exists with multiple time points of the same individuals but different W per time point.

3.4 Inflammation

3.4.1 Introduction

Inflammation has been a well-known condition to mankind since the first time a caveman kicked a rock too hard, and everyone reading this text has experienced inflammation in their life. Text from ancient Egyptians used frankincense and myrrh to reduce pain and swelling [92]. Hippocrates was the first to describe using the word οἴδημα (/oídēma/, edema) and identify it as an important component of the healing process [93]. In the modern era, the discovery of anti-inflammatory drugs like aspirin and corticosteroids in the 20th century revolutionized the treatment of inflammatory conditions. [94]. In this thesis, we evaluated if the inflammation response is similar between individuals. Sharing an experiment of copying behaviors among friends can lead to similar immune responses, including inflammation outcomes. This is what motivated the writing of [Result II](#).

Inflammation has 4 main characteristics: heat, pain, redness, and swelling. The combination of these may lead to a 5th one which is loss of function in the affected area. The nature of each of these conditions will be revealed throughout this section. Inflammation is a natural process that the body uses to fight infections and eliminate the cause of inflammation, clear out the area of elements that shouldn't be there, and repair the tissue if damaged. Similar to fever, it has a bad name with the general population even for mild cases despite being a beneficial and necessary process for recovery. People tend to self-medicate in excess with antipyretics and anti-inflammatories which is counterproductive for both healing processes [95]. This process of acute inflammation is inherent to the natural healing process of an organism and overall, the natural pathways of inflammation do not pose a risk to the organism's life, and it is counterproductive to interfere with it, as such interference may ultimately impair the healing process in the long run. On the other hand, autoimmune reactions and chronic inflammations are the undesirable form of inflammation that may lead to lasting damage; often in the form of scar tissue.

At the end of this document (appendix C) there is a brief description of the 92 different inflammatory biomarkers upon which the social influence in the inflammation article is founded. However, comprehension of these biomarkers is a complex subject and requires an elaborate introduction to the intricacies of the inflammatory processes. Furthermore, the comprehensive background provided in this thesis is essential for the understanding of the complex interplay between mathematical models and biological

processes. Half of this section aims to aid in understanding the background of how these 92 biomarkers work. It also serves as a biological base of knowledge to further understand the immuno-evasion properties of *S. Aureus* or why spa-typing this bacteria is important, explaining the homeostatic properties of vitamin D, or the mechanism of action of several of the OTC medications. This foundational knowledge is indispensable for interpreting the data correctly and for the models to have practical relevance in predicting and managing health outcomes. Thus, the detailed introduction serves not only as an educational tool but also to bridge diverse disciplines. Each section has been meticulously cross-referenced, allowing readers, especially those without biological background, to trace the discussion back to fundamental concepts as needed, trying to ensure that if a reader encounters unfamiliar terms, has forgotten concepts, or is simply curious to understand a deeper level, they can easily refer to earlier sections for a thorough explanation.

3.4.2 Immunology basic concepts

This section serves as an introduction to basic and easy concepts related to immunity.

Reactive Oxygen Species

Reactive oxygen species (ROS) are a natural by-product of cellular metabolism and immune system responses which the influence of heavy metals, tobacco, radiation, or microplastics can overproduce. Excess amounts of ROS damage DNA, lipids, and proteins. This is bad for cells, but good if those cells are bacteria or viruses that we want to destroy.

Extracellular matrix

The extracellular matrix is the material that fills the spaces between cells in tissues and provides the structural support that allows tissues to maintain their shape and function properly. During inflammation, fibroblasts play an important role in the repair and remodeling of damaged tissues. They are activated by cytokines and growth factors that are released by immune cells. As a result, they fill the extracellular matrix with collagen, elastin, and fibronectin, which translate into tissue healing.

Antigens

Antigens (Ag) can be proteins, peptides, polysaccharides, lipids, or nucleic acids. Antigens can be parts of a normal functioning cell or part of a foreign pathogen or substance. They serve as an identification card of whatever cell is holding the Ag.

Major Histocompatibility Complex

The major histocompatibility complex (MHC) proteins are found on the surface of almost all cells with nuclei in the body. They catch antigens inside the cell and present them to the outside. There are three types of MHC, and overall the main difference is what type of leukocytes or complement protein they interact with.

Antigen presenting cells

And Antigen presenting cells (APCs) is a leukocyte that can present foreign pathogens or abnormal tumorous cells antigens to other cells of the immune system. The professional APCs are dendritic cells, macrophages, and B-cells. Any other cell in the body that has a nucleus also has an MHC, and can also be an APC, but these are called non-professional.

CD proteins

The cluster of differentiation (CD) is a protocol used to identify cell surface molecules. There are 371 known CD molecules in humans. Generally, they play a role in cell signaling. Two important CD proteins are:

- **CD4 protein** is a co-receptor for the T cell receptor (TCR) on T-helper cells, helping to enhance the binding and sensitivity of the TCR to antigen-MHC2 complexes. This is the primary receptor for the HIV, which is why this virus targets immune cells leading to immunodeficiency. Other cells such as monocytes and dendritic cells can also express CD4. Any cell that expresses CD4 on the surface is known as a CD4+ cell.
- **CD8 protein** is another TCR that binds to antigen-MHC1 complexes. It is mostly expressed on cytotoxic T cells but can be found on natural killer cells and dendritic cells. Any cell that expresses CD8 is known as a CD8+ cell.

Cytokines

Cytokines are a group of small signaling molecules that coordinate various processes in the human body, such as immune response activation, changing the states in the inflammation cycle, and the formation of new blood cells (hematopoiesis). They are produced by both immune and non-immune cells. The major groups of cytokines are:

- **Chemokines:** A family of chemotaxis cytokines secreted by cells to make leukocytes move in a particular direction, hence the name "kino" (movement). They are classified according to their behavior and protein structure. For example, a "C-C motif" just refers to the type of structure in which it has two adjacent cysteines. A "C-X-C motif" is the same, but in between cysteines there is another amino acid, hence the X. "Chemotaxis" means that it allows for movement, while "chemotactic" refers to how good or bad a particular protein does so.
- **Growth factors:** These cytokines are involved in regulating cell growth, proliferation, and differentiation. In particular, we will be interested in fibroblast growth factors (FGF), a family of cell signaling proteins produced by macrophages. Their function is to regulate normal cell development in animals.
- **Interferons:** Alert other non-immune cells about ready their defenses because the infection is happening and tell immune cells to come and kill infected cells. They are explained in section 3.4.3.
- **Interleukins:** The internet of leukocytes, they are explained in section 3.4.3.
- **Tumor necrosis factors:** Are both pro-inflammatory and anti-inflammatory signals needed in each stage of inflammation and will be explained in section 3.4.3

Types of immune responses

- **Humoral vs Cell-mediated.** A humoral immune response works against extracellular pathogens, while a cell-mediated response works against intracellular pathogens and abnormal cells. As a general rule, killing your own cells is a very bad idea that leads to many unwanted problems, and the least preferred type of immune reaction. Therefore, there are a lot of efforts promoting behaviors and diets that promote humoral response and can kill the pathogens before they can infect host cells, or prevent cancer from growing.

- **Innate vs Adaptive** Overall, innate immunity provides quick but weak protection against a wide range of pathogens. While adaptive immunity takes more time to activate but is more efficient against each specific type of pathogen and provides long-term protection. The function of cells named in this section will be introduced formally in section 3.4.2.
 - Innate immunity is present from birth (hence the name) and involves physical and chemical barriers that prevent pathogens from entering the body, such as skin, mucous membranes, and stomach acid. Innate immunity also includes cells such as natural killer cells, neutrophils, and macrophages that can recognize and attack foreign invaders, as well as the complement system, which can activate a series of proteins to help identify and destroy pathogens.
 - Adaptive immunity, on the other hand, is a specific, second line of defense that develops over time as the immune system gains experience and adapts to specific pathogens. This type of immunity involves the production of antibodies that lock to the pathogen's antigens. The first time a pathogen enters the body, neither the antibodies nor the cells manufacturing them exist, and it takes a long time, usually around 2 to 3 weeks, for the immune system to mount an effective response. Later on in life, once the pathogen enters the body again, such antibodies are ready to use and the response is shortened to about 2 to 3 days.

NF-κB

Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), is a collection of proteins that controls the transcription of DNA. In our context, it regulates both the innate and adaptive immune response [96]. NF-κB can be up or down-regulated by other mechanisms discussed later in this document, and by extension controlling immune reaction. It does it by promoting TNFA, IL1B, and IL18.

Homeostasis

The immune system must be balanced between slacking off and letting foreign agents invade and destroy the body, and being overactive and letting the immune system itself be the one that destroys the body. Proper functioning of the immune system requires a delicate balance between activation and regulation. This balance is known as immune system homeostasis. Overall, this work falls on the T cells, by maintaining the appropriate

levels of cytokines, which maintain the appropriate levels of everything else.

This is also important during inflammation episodes. There are parts of the immune system that are pro-inflammatory and anti-inflammatory. Both have their roles and both are needed for a successful resolution of an inflammation process, and the existence of neither of them is necessarily good or bad. Their presence is welcome within the time frame in which they are needed, and no more.

Leukocytes

Leukocytes, also known as white blood cells, are a type of immune cell that helps protect the body against infections and diseases. Neutrophils, Eosinophils, Basophils, and Mast cells, are a type of leukocytes that contain granules and are classified as granulocytes. Granules are essentially a bunch of sacks in the cytoplasm organ that release cytokines and will be discussed later with more detail in section 3.4.2. They are also known as polymorphonuclear leukocytes (PMN) as they have several nuclei. However, neutrophils are the most abundant of these types and they are also referred to as PMN despite sharing the group with other leukocytes.

Natural killer cells are known as large granular lymphocytes (LGL). They also have granules in their cytoplasms, but they are not considered to be granulocytes. They have a different root of pluripotent differentiation (figure 3.31), and as such their granules have different functions related to direct killing rather than helping the immune system.

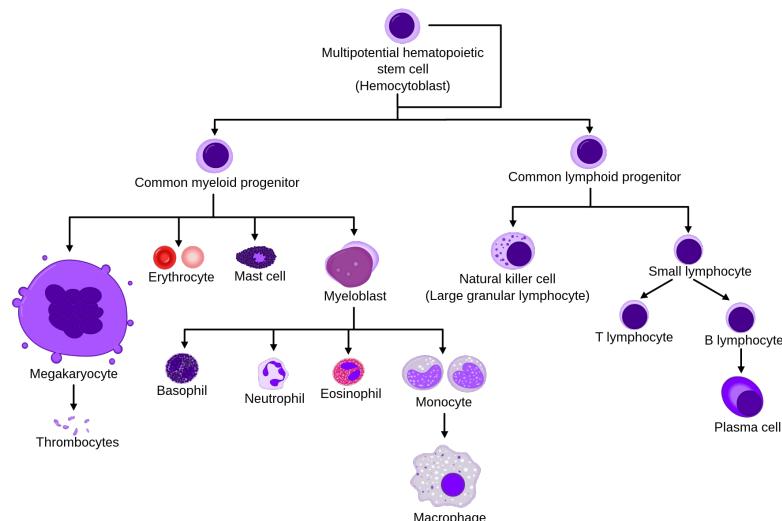


Figure 3.31: Overview of the formation of blood cellular components (hematopoiesis). All cellular blood components are derived from pluripotent stem cell roots. Image from [Wikimedia](#).

Monocytes develop from the same stem root, and while they occasionally display another type of granule (azurophilic granule), are not classified as granulocytes either. They also have only one nucleus but its shape is elongated and indented, which might be confused with having several nuclei.

- **Neutrophils.** These are the most abundant and are responsible for the phagocytosis of bacteria and other foreign substances. They have a similar role as macrophages, but neutrophils are the first to act and are not professional APCs. They are found in the blood from where they migrate to inflammation sites through a process known as chemotaxis. This involves the detection of a chemical signal (chemoattractant) released by cells at the site of infection. The neutrophil follows the gradient of the chemoattractant until it reaches the site of infection.
- **Eosinophils.** They play a role in the body's immune response to parasitic infections, allergies, and asthma. They contain granular structures which are bright red dyed with eosin, hence the name. High levels are correlated with eosinophilic asthma and hypereosinophilic syndrome.
- **Basophils.** Basophils are responsible for releasing histamine and other inflammatory molecules in response to allergens, parasites, and other types of foreign substances. Basophils make up less than 1% of all white blood cells in the body.
- **Mast cells.** Another type of leukocyte which are found throughout the body, particularly in tissues that are in contact with the external environment like the skin, lungs, and digestive tract. They are involved in the regulation of the immune system. When they are overactive, they can upregulate the immune system too much causing allergies, asthma, and autoimmune disorders.
- **Monocytes.** Monocytes are the largest type of white blood cell and play a role in the phagocytosis of foreign substances. Alongside neutrophils, they belong to the phagocytes group. Monocytes are further subclassified as:
 - **Macrophages** are large white blood cells that are primarily responsible for phagocytosis. Macrophages play a key role in presenting antigens to other cells of the immune system. They are mostly concentrated in the spleen, liver, and lymph nodes. When the proper signaling arrives, they migrate to blood vessels, and from there to the inflammation site. Macrophages are further divided into two types:

- * **M1** are activated by pro-inflammatory cytokines or lipopolysaccharides (LPS). They produce pro-inflammatory cytokines themselves and kill microbes by phagocytosis or ROS-like substances.
- * **M2** are activated by anti-inflammatory cytokines. These macrophages collaborate in tissue repair.
- **Dendritic cells** are specialized immune cells that are found primarily in tissues that are in contact with the external environment, such as the skin, lungs, and intestines. Dendritic cells are known for their ability to capture and process antigens, which they then present to other immune cells to activate an immune response. Dendritic cells are particularly important in the initiation of adaptive immune responses,
- **Natural killer cells.** Natural Killer (NK) cells identify and destroy human cells. Typically those cells would be infected or cancerous cells presenting an abnormal antigen in the MCH1. But they can also be healthy cells that lack the proper MHC1, such as in transplants from another body, cancer cells that typically display less MHC1, or simply that NK cells can't recognize MCH1 as self due to an autoimmune disorder. Unlike B and T cells, NK cells do not require prior activation which allows them to act very fast preventing further cancer growth or incubation time. They also produce cytokines that help regulate the immune response. On the downside, they do not have memory or clonal expansion capabilities.
- **Lymphocytes.** Lymphocytes are involved in the immune response, producing antibodies against foreign substances and attacking infected or cancerous cells. There are divided into further categories:
 - B-lymphocytes are involved in producing antibodies that will help other leukocytes to recognize and kill the pathogen, or prevent the pathogen from doing its function correctly. The basic sub-classification is:
 - * **B-cells** are activated by APC, and are also APC themselves. They can help regulate the activity of T-cells, but this is not their main function. They are responsible for producing antibodies. Antibodies are also known as Immunoglobulin (Ig), and are sub-classified into the following categories:

- **IgA:** Found in high concentration in saliva, tears, breast milk, and mucous membranes.
- **IgD:** These antibodies are found on the surface of B cells when they exist in the bone marrow and are co-expressed with IgM. But to this day their function is not fully understood.
- **IgE:** Bind to allergens such as pollen, animal dander, and food particles, triggering the release of histamine.
- **IgG:** Common in the blood and tissue fluids, they have a role in the adaptative immune response. They are also the only type of antibody that crosses the placenta from a mother to the fetus, providing temporary immunity.
- **IgM:** First to be produced in response to an infection and are effective against general bacteria and viruses.

When a B-cell changes producing one Ig to a different one is known as class switching. Later on, we will also talk about lipid mediator class switching, but despite the name, the two concepts are unrelated.

- * **Plasma cells** Are specialized B-cells that can only generate one type of antibody which is triggered by one particular antigen. They serve as a specialized and quick response in current infections and have a short time lifespan.
 - * **Memory B cells** are the same as plasma cells, but they have a long time lifespan and provide long-term immunity for years. When memory B cells finally die, you need a memory shot vaccine for their related disease. They can be activated by the pathogen directly, but is more common that they are awakened by memory T cells.
- T cells are involved in killing infected cells and regulating the immune system producing cytokines. They are called "T" because they mature in the thymus, but their further sub-classification and function range widely:

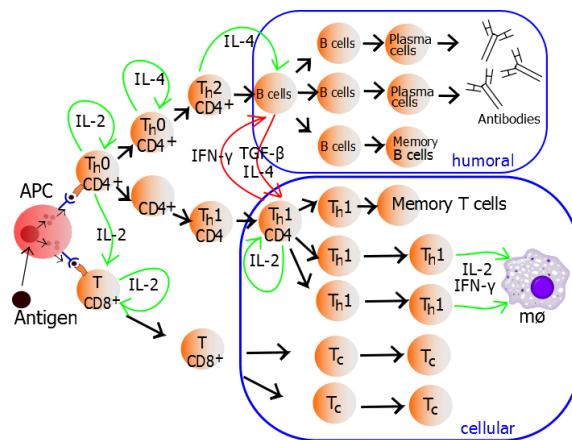


Figure 3.32: Overview of the T-cells activation and differentiation. Image from "[Medical Gallery of Mikael Häggström 2014](#)". [WikiJournal of Medicine](#)

- * **Helper T cell (Th):** binds to APC and stimulates the production of antibodies and cytokines. When T-cells are immature they are known as "Th" or "Th0". Then they get activated by APC turning into two main types, "Th1" or "Th2". Th1 cells lead to a cell-mediated response, and Th2 cells lead to a humoral immune response. "Th17" type is a third type that produces IL-17, a pro-inflammatory substance especially good at fighting extracellular pathogens and fungi. TH $\alpha\beta$ is the fourth type and provides host immunity against viruses by activating Natural Killer cells. Any Th cell that expresses the CD4 protein, is also known as CD4+ cells.
- * **Cytotoxic T cells:** These cells directly kill infected cells same as NK cells. The cells are known by many other names, including TC, cytotoxic T lymphocyte, CTL, T-killer cell, cytolytic T cell, CD8+ T-cell, and killer T cell, which may cause confusion with other T cell names. Cytotoxic T cells are part of the adaptive immune system and are highly specific in recognizing and attacking infected or abnormal cells that display a specific antigen on their MHC. In contrast, NK cells do not recognize specific antigens, but rather detect and target cells that have decreased expression of normal surface proteins, such as MHC1. Cytotoxic T cells can develop a long-lasting immunological memory after encountering an antigen, allowing for a more rapid and efficient response upon re-exposure.

- * **Regulatory T cells:** These cells help control the immune response by suppressing the activity of other T cells, preventing them from attacking healthy cells.
- * **Natural killer T cells:** Not to be confused with either Cytotoxic T cells (T-killer cells) or Natural Killer cells (NK cells). They can also be CD4+ and CD8+ cells. They recognize microbial lipids and enhanced humoral immunity. Upon activation, they produce large quantities of Interferon gamma (IFN- γ), IL-4, and granulocyte-macrophage colony-stimulating factor, IL-2, IL-13, IL-17, IL-21, and Tumor Necrosis Factor α (TNF α) among others. Lack of NKT shows autoimmune diseases, autoinflammatory diseases, and cancers.
- * **Memory T cells:** There are 5 types of memory T cells and 3 major ways of activation. They also share some functions with memory B cells. However, simplifying their function consists of recognizing previously encountered pathogens and activating the appropriate memory B cells. They have a relatively short lifespan and need to be manufactured constantly.

Granules

Granules are sacs contained in the cytoplasm and contain inflammatory mediators. The two main types of cells that have granules are lymphocytes of the granulocytes class and endothelial cells. In the case of the endothelial cells, granules are stored inside the Weibel-Palade body. Here we list the main inflammatory mediators released from granules during an inflammation process.

- **Von Willebrand factor.** von Willebrand factor (vWF) is a protein that plays a key role in the initial stages of clotting by binding to platelets and forming a plug at the site of injury. In an inflammatory process produced by an injury, the most important mechanism is to first stop the bleeding quickly so the body does not die from blood loss. Afterward, other mechanisms will promote vasodilation so the leukocytes can go there to do their job. Its deficiency might be caused by the most common hereditary blood-clotting disorder in humans von Willebrand disease (vWD), which is presented in the form of too much bleeding tendency, such as bruising, nosebleeds, heavy menstrual periods, and even severe internal bleeding in the worse cases.

- **Selectins.** Selectins are a family of cell adhesion molecules (CAMs). The work of selectin in epithelial cells is to slow down leukocyte rolling alongside the interior of blood vessels and bring it out of the blood vessel so it can go into the source of inflammation and deal with it. There are three subsets:
 - **P-selectin.** The release of histamine or thrombin in the inflammation sites makes endothelial cells raise P-selectin stored in the cytoplasm up to their cell wall. These are proteins in the shape of "hooks" that catch leucocytes or platelets near the inflammation site.
 - **E-selectin.** Release of IL-1 and TNF- α by macrophages in the inflammation site induces expression of E-selectin on endothelial cells bringing leukocytes inside. They work similarly to P-selectin, but E-selectin takes longer to be activated and is produced and released rather than stored in the cells.
 - **L-selectin** is expressed in leukocytes and they hook to the P/E-selectin expressed in the endothelial cells. L-selectin is also present on the surface of the human embryo and works the same way as with leukocytes facilitating its adhesion to the endometrium.
- **Bradykinin.** Bradykinin is a vasodilator, leading to increased blood flow into the affected area. It also increases vascular permeability, allowing inflammatory cells and proteins to move from the bloodstream into the affected tissues. Additionally, bradykinin activates pain receptors, leading to the sensation of pain associated with inflammation.
- **Histamine.** Mast cells and basophils can be activated by IgE and Complement component 3 A (C3a). This releases histamine which is an organic nitrogenous compound acting as a chemokine. When this happens, more P-selectin is expressed on the endothelial cells, and at the same time, endothelial cells open up so other leukocytes can migrate to the inflammation site. This process is illustrated in figure 3.33.

However, histamine is also released when allergens bind to mast-cell-bound IgE antibody sites. This is commonly known as an allergy and is the mechanism that can make people's life miserable in the form of bronchial smooth muscle contraction (cough), vasodilation (redness), itchy perception, urticaria, sneezing, hyper-secretion from glandular tissue (nose mucus), urinary bladder contractions

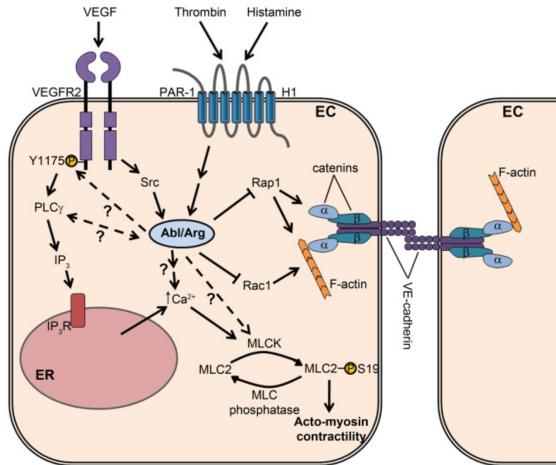


Figure 3.33: Overview of the histamine/thrombin cascade signaling which activates PS19 (P-selectin) to migrate up to the cell wall. This mechanism also activates the side chains VE-cadherin to open allowing for leukocytes to abandon the blood stream. Source: Wikimedia.

(urging need to urinate), and nasal congestion due to vascular engorgement. This is typically alleviated by the use of antihistamine medication. In the worst-case scenario, histamine can lead to an anaphylactic shock which causes death typically by collapsing the respiratory airways.

- **Cytokines.** A description of the different types of cytokines can be found in section 3.4.2.
- **Eicosanoids / Arachidonic acid** Eicosanoids are a family of lipids that act like hormones. They derive from arachidonic acid, and sometimes the two names are used interchangeably. Arachidonic acid is released during cell damage. Arachidonic acid is found in meat and eggs from animals that have free range and can exercise and move freely, as such, a lack of proper quality items in the diet can't initiate proper resolution in the inflammation cycle. They are suppressed by steroids and promoted by tissue injuries, thrombin, bradykinin, and epinephrine. An overview of the actions of eicosanoids can be seen in figure 3.34. Details regarding eicosanoids are discussed later in the sections 3.4.3 and 3.4.3.
- **Serotonin.** Serotonin is a well-known neurotransmitter better known for regulating mood, appetite, and sleep. But it also controls the activation and proliferation of T and B cells. Overall, the function of serotonin in immunity is complex and not fully understood yet, but some studies suggest that serotonin may have a protective role against infections [97].

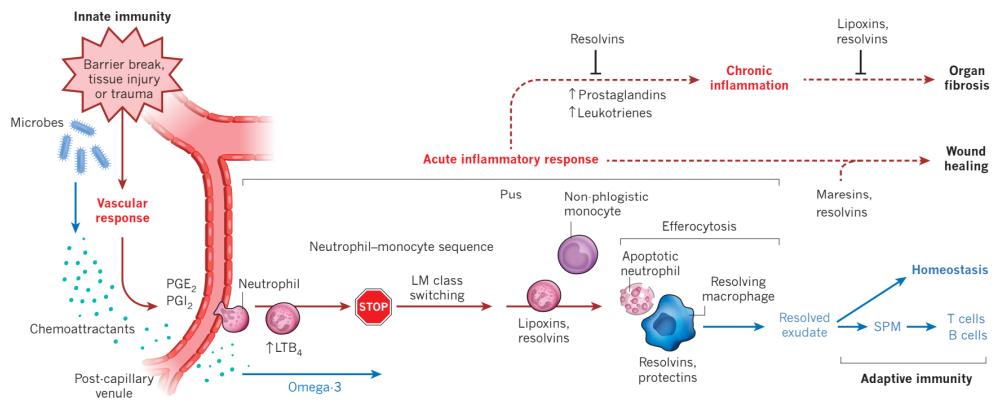


Figure 3.34: Overview of the roles of lipid mediators in vascular and leukocyte response, from the initiation of inflammation to the resolution of the event. Image from "Pro-resolving lipid mediators are leads for resolution physiology"

Complement system

The complement system is a collection 32 of proteins that can kill bacteria directly, or enhance other cells in the immune system via opsonization or chemotaxis production. The functioning of the complement system is a very beautiful topic that involves the complex synchronization of proteins working together like cogs in a machine. Unfortunately is beyond the scope of this document. Here we are just going to list a simplification of the reactions present in the 3 possible pathways, highlighting the proteins important in inflammation that are mentioned in this document.

- **Initiation.** When a couple of IgG, or a single IgM, are attached to a bacteria, is possible that by random chance a C1 protein gets attached to these antibodies. If this happens, the complement system becomes initiated.
- **Activation.** The C1 protein has two main parts, the C1q and C1rs. C1rs is attached to the Igs, while C1q can cleave other proteins. C1q cleaves C2 making C2a and C2b, and also cleaves C4 making C4a and C4b. C4b and C2a bind together making the C4bC2a complex, which cleaves C3 making C3a and C3b. C3b binds with C4bC2a making the "C4b2a3b Convertase". This attaches to the bacteria. When this finally occurs, then it is said that the complement system is activated.
- **Amplification.** C3b can also bind to another C3b and attach to the bacteria. This also can cleave C3 proteins, which translates into an exponential growth of the C3bC3b complex attached to the bacteria and C3a flooding the bloodstream.

C3a activates basophils into releasing histamine (section 3.4.2), which promotes the expression of P-selectin, which promotes the traveling of neutrophils into the inflammation site.

- **Termination.** C4b2a3b Convertase can also cleave C5 into C5a and C5b. C5a creates a chemotaxis gradient for neutrophils which kill the bacteria. Furthermore, neutrophils and other immune cells catch the bacteria by attaching to C3bC3b which is already attached to the bacteria wall. Here, C1q also stimulates the neutrophils to start the destruction of whatever they just grabbed. As C1q is supposed to only be attached to pathogens, this prevents them from destroying healthy cells accidentally caught by neutrophils.

The complement system can also terminate bacteria by itself forming the MAC complex, which consists of several other proteins that attach to C5b, which makes a hole in the bacteria wall of approximately 25 times the diameter of water molecules, allowing the insides of the bacteria to leak out.

3.4.3 Immunology advance concepts

Now that the basics of immunology have been explained, we can dive into more complex mechanisms that fully explain the inflammation cycle.

Type 1 vs Type 2 immune response

The Type 1 immune response is characterized by the activation of Th1 cells which secrete cytokines that stimulate the production of antibodies by B cells and activation of cytotoxic T cells. It is primarily effective against intracellular pathogens such as viruses and bacteria, and the development of cell-mediated immunity. Type 1 immunity response is mainly mediated by Th1, cytotoxic T-lymphocytes, and NK cells. Type 1 immunity implies the production of pro-inflammatory cytokines and IgG and IgM antibodies.

The Type 2 immune response involves the activation of Th2 cells which secrete cytokines that activate eosinophils, basophils, and mast cells, causing allergic inflammatory responses and increased production of antibodies by B cells. It is primarily effective against extracellular pathogens such as parasites, worms, and allergens and the development of humoral immunity. Type 2 immunity response is mainly mediated by Th2 cells, eosinophils, basophils, and mast cells. Type 2 immunity generally implies the production of anti-inflammatory cytokines but also involves both types such as in the case of IL-5 which promotes allergic reactions. It also mainly promotes IgA and IgE.

Nitric Oxide

NO is a free radical generated by phagocytes and is toxic to bacteria and some parasites, as well as to many human cells, leading to apoptosis. The damage to human cells is intentional and is believed to be used as a way of getting rid quickly of cells that are promoting pro-inflammation mechanisms. It is activated by IFN- γ and TGF α , and is suppressed by IL-4, IL-10, and TGF β .

JAK-STAT pathway

Janus kinase signal transducers and activators of the transcription proteins (JAK-STAT) pathway is related to endocrinology and its main function is to couple with a receptor that attaches to the growth hormone and brings it inside the cell. However, the receptor of JAK-STAT can also bind to several cytokines. In particular IL-2, IL-6, and interferons.

What is important to understand for the scope of this document is that the JAK portion function by phosphorylating each other, which phosphorylate the STAT part, which goes into the cell nucleus, which will promote gene transcription, which promotes protein translation.

Pro-inflammatory Eicosanoids

In section 3.4.2 we introduced how eicosanoids are included inside granules. Here we are going to expand on them and list the pro-inflammation ones.

- **Thromboxane.** They work in the first stage of tissue damage by acting as a pro-coagulation agent, a vasoconstrictor, a hypertensive agent, and facilitating platelet aggregation.
- **Prostaglandins.** prostaglandins (PG) are a complex family of eicosanoids which act both as pro-inflammation when the recruitment of the immune system is needed, and anti-inflammatory when the damage has been resolved. There are four different subclasses:
 - **PGD2:** Produced by mast cells and is primarily involved in allergic responses.
 - **PGE2:** It is responsible for the pain and fever associated with inflammation.
 - **PGF2 α :** Involved in the contraction of the smooth muscle in the uterus.
 - **PGI2:** Produced by the endothelial cells lining blood vessels and plays a

role in vasodilation. It shares PGH₂ precursors with Thromboxane which is a vasoconstrictor. You either have vasodilation or vasoconstriction, and both pathways are never active at the same time. Vasoconstriction reaction however occurs much more rapidly.

- **Leukotrienes.** Leukotrienes (LT) are both autocrine signaling (the cell signal itself) and paracrine signaling (the cell signal its neighbors). They serve to regulate immune responses bringing neutrophils. The LTC₄, LTD₄, LTE₄, and LTF₄ prolonged slow contraction of smooth muscle and have a major bronchoconstrictor role in asthma. Nonsteroidal anti-inflammatory drug (NSAIDs) block prostaglandins but promote leukotrienes, prostaglandins are initially pro-inflammatory but later on, switch to anti-inflammation, while leukotrienes are always pro-inflammation. This is the reason why NSAIDs are not always recommended for mild inflammations, and as long as the pain is bearable, it is better in the long run to avoid its use. Especially if side effects on the liver and stomach are taken into consideration.

Specialized pro-resolving mediators

Again, in section 3.4.2 we introduced how eicosanoids are included inside granules. Here we are going to talk about the switch from a pro-inflammation state to an anti-inflammatory effect which leads to the resolution of inflammation. These are known as specialized pro-resolving mediators (SPMs).

- **Lipoxin.** Lipoxins reverse the actions of the pro-inflammatory mediators and initiate tissue repair response [98]. Among many other things, they inhibit chemotaxis, transmigration, superoxide generation, NF-κB activation, generation of pro-inflammatory cytokines, suppress the production of IgM and IgG antibodies, reduce the perception of pain due to inflammation, induce the production of elements that neutralize oxidative stress and oxidant-induced tissue damage and block the actions of some leukotrienes. [99]
- **Resolvins.** Resolvins play an important role in resolving inflammation by promoting the clearance of cellular debris, bacteria, and other inflammatory mediators. They also inhibit neutrophil recruitment, decrease pro-inflammatory cytokine production, and promote tissue repair and regeneration [100]. The importance of resolvins lies in their ability to control the duration and intensity of inflammation, which is crucial for preventing the development of chronic inflammatory diseases.

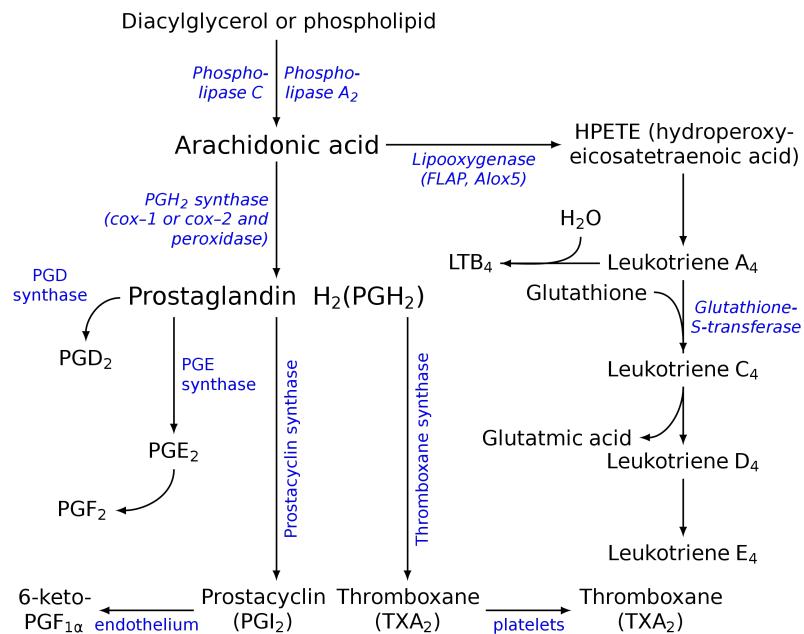


Figure 3.35: Overview of the prostaglandin pathways. Image from [Wikipedia](#)

They are formed from the metabolism of omega-3 polyunsaturated fatty acids, in particular, eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) (figure 3.36). Humans convert alpha-linolenic acid (ALA) to EPA very inefficiently, so is recommended to take food rich in EPA directly such as salmon, mackerel, herring, cod liver, some algae, and human milk. DHA can be converted from EPA but is recommended to also take DHA-rich food such as salmon, caviar, anchovies, mackerel, or herrings.

- **Protectins / Neuroprotectins.** Protectins reduce inflammation induced by oxidative stress and inhibit the pro-apoptotic signal. Can potentially protect respiratory cells from viral infections. Blocks formation of pro-inflammatory prostaglandins inhibits platelet-aggregating by thromboxane thus blocking the platelet aggregation responses such as those described for *S. Aureus* in section 3.5.3, and stimulate the efferocytosis [101, 102].
- **Maresins.** Its name derives from **MACrophage mediator in RESolving INflammation**. They are involved in resolving inflammation and allergic reactions, wound healing, apoptotic human neutrophils by human macrophages, reduced lung inflammation, suppressing the production of IL-5 and IL-13, and reducing the production of LTB4.

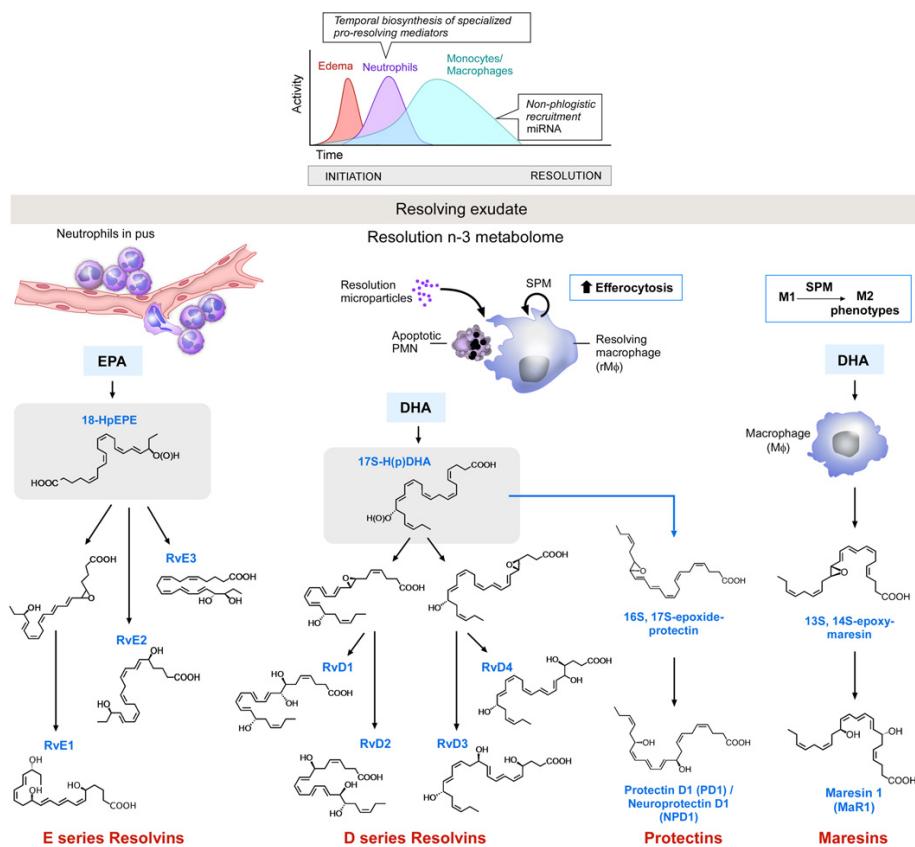


Figure 3.36: Overview of EPA and DHA conversion into resolvins, protectins, and maresins. Image from "Pro-resolving lipid mediators are leads for resolution physiology"

- **Eoxins.** Eoxins are proinflammatory eicosanoids first described in 2008 [103] which are suggested to contribute to the inflammation of airways during allergies and some cancers [104]. They still have an unknown function in human physiology or pathology. But their production is stimulated in eosinophils by pro-inflammatory mediators PD2, LTC4, and IL-5.

APR

Acute phase reactants (APR) are proteins that are produced by the liver and are high in plasma when there is a cause of inflammation or infection and correlate with disease activity. They are generated in response to IL-6. They go down when the cause of inflammation or infection has been resolved. However they have no diagnostic value; it is just an indicator that something wrong is going on, but there are too many diseases that cause APR to be high.

- **CRP.** C-reactive protein (CRP) is a type of APR. It is called "C" because it was first discovered reacting with the C-polysaccharide of the Capsule of *Streptococcus pneumoniae*. It attaches to bacteria or dying cells which promotes phagocytosis by macrophages. It also interacts with C1q and enhances its ability to bind to pathogens, and is also believed to interact with C3; although the mechanisms of interaction between CRP and the complement system are not fully understood. It increases very rapidly and falls very rapidly which means the half-life is barely 7 hours. Increased CRP is correlated with an increased risk of cardiovascular disease (CD) only if the patient history also correlates with CD.
 - **ESR.** The erythrocyte sedimentation rate (ESR) is another acute phase reactant that is used to measure inflammation. It measures how quickly red blood cells clump together and settle to the bottom of a test tube. Inflammation causes the blood to become thicker and stickier, which causes the red blood cells to settle faster. In opposition to CRP, ESR remains high after the resolution of the inflammation or infection and takes up to 7 days to regress to normal levels.

Interleukins

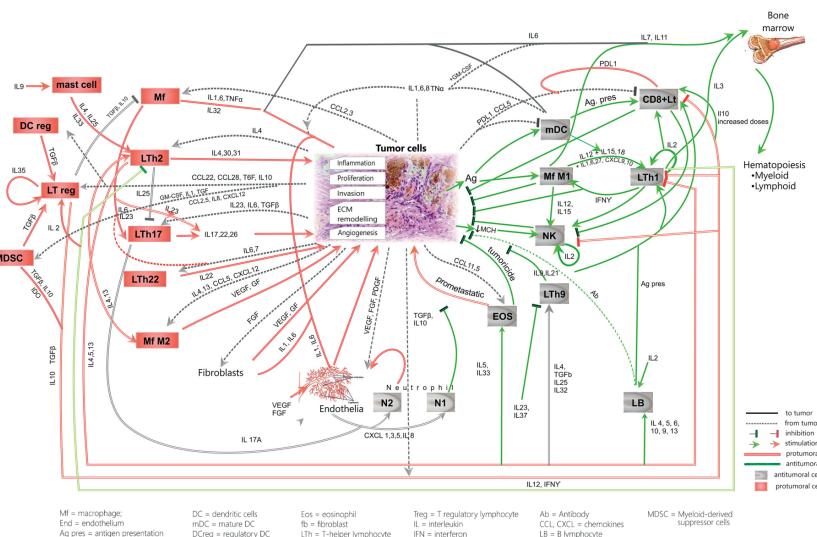


Figure 3.37: Overview of all interleukins affecting tumoral growth. Image reproduced from "PRO-AND ANTITUMOR ROLE OF THE INTERLEUKINS 1 TO 41" [8]

Interleukin is a type of cytokine used both by the immune system and for therapeutic purposes. There are more than 50 interleukins encoded in the human genome, with several subtypes and variants. While interleukins mostly have a beneficial effect, some of

them can also contribute to unwanted inflammation, autoimmune diseases, and cancer. They can stimulate or inhibit the proliferation, differentiation, activation, migration, and survival of white blood cells or other immune cells, as well as modulate the production of antibodies, chemokines, or other mediators.

The number of each interleukin is arbitrary and might lead to confusion. For example, Interleukin 1 family members includes IL-1 α , IL-1 β , IL-18, IL-33, IL-36 α , IL-36 β , IL-36 γ , IL-36ra, IL-37, and IL-38. It must be taken as a unique ID with no other meaning. Here we are going to explain the function of the ILs present in the Olink panel.

The difference between interleukins can be summarized by whether they are pro or anti-inflammatory, the type of immune cells with which they interact, the type of immune cells or other human cells that produce them, and their secondary effects on the body. Individually, interleukins are very easy to understand. But their complex interactions with each other, and with other parts of the immune system make them very difficult to visualize and keep track as can be exemplified in figure 3.37 and table 3.6.

- **IL-1.** Interleukin 1 (IL-1) is a cytokine that is involved in regulating the immune response and inflammation. It is produced by macrophages, monocytes, dendritic cells, and certain types of T cells. It is involved in both innate and adaptive immune responses and plays a key role in the body's defense against pathogens.

Activates and promotes the production of several proteins involved in the acute phase of inflammation. In particular highlights that activates PGE2 which leads to fever, and it increases the concentration of TNF and IL-1 in the brain which may break the blood-brain barrier. Increases APR. It also triggers the production of chemokines, activates T-cells, and promotes B-cell maturation and proliferation. It also promotes IL-2 and fibroblast growth factors.

There are two main forms of IL-1: IL-1 alpha and IL-1 beta. While they are both involved in the same functions, they are produced by different cells and are regulated differently:

- **IL-1 alpha** is produced and stored by epithelial cells, endothelial cells, and fibroblasts. IL-1 alpha is stored in the cytoplasm of cells, which means that it can be released rapidly in response to stress or injury.
- **IL-1 beta** is produced when needed and is not stored. Is secreted by activated monocytes and macrophages. IL-1 beta is produced as an inactive precursor.

Both of them need to be activated, which is a role primarily done by the caspase family of proteins. Caspase-8 is one of the proteins which we use in our analysis and is explained in section C.2.6.

Steroids can block IL-1, which is of special interest in the treatment of autoimmune diseases.

- **IL-2.** Interleukin 2 (IL-2) is produced mainly by activated T cells and NK cells. IL-2 functions as a growth factor for Th, cytotoxic T cells, and regulatory T cells, promoting their proliferation and activation. It also activates NK cells and cytotoxic T cells, increasing their cytotoxicity (ability to kill infected or cancerous cells).

The interleukin-2 receptor (IL-2R) is the receptor complex that binds and responds to IL-2. It is expressed on the surface of T cells, NK cells, B cells, and dendritic cells. IL-2R mediates signaling through various intracellular pathways, including the JAK-STAT pathway. The signaling regulates gene expression, cell proliferation, survival, and differentiation in response to IL-2. IL-2 can also bind to the IL-12 receptors which will be discussed later, but in essence, further promotes Th1 and NK cells.

Changes in IL-2R expression and signaling can impact immune cell function and contribute to immune dysfunction and disease.

- **IL-4.** This induces differentiation from Th0 to Th2 and the production of IL-4 Th2 cells, which creates a positive feedback loop. IL-4 is also produced by mast cells, eosinophils, and basophils.

Secondary functions include decreasing the production of Th1 because Th0s are converted to Th2s instead. M1 macrophages are also decreased and promoted into M2 instead. IL-4 is usually coupled with secretion of IL-10 and TGF- β which further reduce inflammation and M2 production. It also decreases the production of IFN γ , by decreasing Th1 indirectly decreases IL-12 also.

Its functions are similar to those of IL-13. Overproduction of IL-4 is related to allergies.

- **IL-5.** It is primarily produced by Th2 cells, mast cells, and eosinophils. It activates, stimulates, differentiates, and recruits eosinophils. It also makes B cells focus

producing IgA. It has been associated with allergies, in particular with allergic rhinitis and asthma.

It can also suppress the production of pro-inflammatory cytokines, TNF- α , and IL-1 β , reducing the activation of inflammatory cells.

- **IL-6.** Is a cytokine that plays a variety of roles in the body. It acts as a pro-inflammatory cytokine and as an anti-inflammatory myokine; this is a type of signaling that occurs in the skeletal muscle cells which has an effect on nearby cells as well as the endocrine system. In particular, it inhibits the effects of TNF- α and IL-1.

It is produced by macrophages in response to Pathogen Associated Molecular Patterns (PAMPs). IL-6 initiates PGE2 production which leads to fever. Furthermore, in muscle and fatty tissue increase energy consumption leading to increased body temperature. It also influences the liver increasing APR, and is believed to be the reason why obese individuals have higher CRP. This is closely related to long-term inflammation due to obesity.

IL-6 is responsible for stimulating acute phase protein synthesis, as well as the production of neutrophils in the bone marrow. It supports the growth of B cells and is antagonistic to regulatory T cells.

IL-6 stimulates the inflammatory response in multiple auto-immune processes, such as MS, Neuromyelitis optica spectrum disorders (NMOSD), diabetes, lupus, atherosclerosis, rheumatoid arthritis, and many more.

As a myokine, it suppresses inflammation caused by stress in bones and muscles during exercise and promotes bone re-absorption. Myokines function is still poorly understood but is believed to have a beneficial impact as a response to PA [105] which is related to IL-10 too. Furthermore, IL-6 also regulates glucose metabolism and energy homeostasis.

- **IL-7.** It is secreted by stem cells, keratinocytes, dendritic cells, hepatocytes, and epithelial cells. It stimulates the differentiation of stem cells into the lymphoid path (NK cells, B cells, and T cells) as opposed to IL-3 which stimulates the opposite branch. However elevated levels of IL-7 promote acute lymphoblastic leukemia and T-cell lymphoma. Furthermore, it plays a critical role in the survival of T cells and their development in the thymus.

- **IL-8 / CXCL8.** It is also known as neutrophil chemotactic factor. Its name has officially changed to CXCL8. IL-8 main mission is to attract neutrophils into infection sites, but it also works on other granulocytes. Is produced by any cells with toll-like receptors that are involved in the innate immune response; but mainly macrophages, epithelial cells, muscle cells in the airway passage, and endothelial cells.

IL-8 is readily stored in Weibel-Palade bodies of endothelial cells and has a quick response time. It increases intracellular calcium ions, histamine release, ROS, and the expression of CAMs (section 3.4.2).

- **IL-10.** IL-10 and its receptor play a very influential role in the anti-inflammation process and overall it should be considered as such. However, it can also promote pro-inflammatory effects by stimulating B cell activity. Its activity is related also to IL-19, IL-20, and IL-24.

It is produced primarily by monocytes, but also secondarily by Th2, mast cells, regulatory T cells, and activated T cells and B cells.

It has 4 receptors in total which can down-regulate the JAK-STAT activation and block NF- κ B activity. They are present in Th1, macrophages, and B cells, and their function is predominantly to stop the pro-inflammation activity. In Th and some macrophage cells it down-regulates the production of pro-inflammatory cytokines TNF α , IL-1 β , IL-2, IL-3, IL-12, and others not discussed in this document. In cells with TLRs can block the production of IFN γ . On CD4+ cells can suppress their antigen expression which was discussed in section 3.4.2. On bacteria can directly inhibit LPS.

Another important aspect of IL-10 is that promotes the COX activation shown in figure 3.35, suppressing LT production and enhancing PG which in turn promotes the SPMs discussed in section 3.4.3.

Finally, also comments that PA increases the levels of IL-10 by about 30x [106] and seems to be linked with IL-6, suggesting that PA promotes an anti-inflammation environment in the body.

- **IL-12.** IL-12 is produced by dendritic cells, macrophages, neutrophils, and B-cells. Its receptors can be found in T cells and NK cells.

Its main function is to promote as much killing as possible. It increases Th1 activity which increases macrophages and cytotoxic T cells. It also increases NK activity. IL-2 can also bind to the IL-12 receptors adding to this effect. It also stimulates the production of IFN- γ , and TNF- α , and overrides the activity of IL-4.

- **IL-13.** Produced by Th2, CD4, natural killer T cells, mast cells, basophils, and eosinophils.

It is an interleukin very similar to IL-4. IL-13 can bind to both IL-4 receptors and IL-13 receptors. However, IL-13 lacks the same ability to differentiate Th0 from Th2. Instead, it specializes more in the airway hyperresponsiveness and mucus production proper of allergies.

- **IL-15.** Produced by monocytes, macrophages, dendritic cells keratinocytes (skin), fibroblasts (extracellular matrix), myocytes (muscles), and nerve cells.

It is an interleukin very similar to IL-2. But unlike the IL-4 / IL-13 pair, in this case, IL-15 can only bind with its own receptor which is expressed in a wider range of cells; lymphocytes and NK cells. IL-2 activates cells, while IL-15 promotes the survival and activation of the same cells, plus those extra with IL-15 receptors.

- **IL-17.** It is produced by many types of both immune and non-immune cells, but primarily by Th17 cells in order to initiate the adaptive immune response against bacteria or fungi.

The IL-17 receptor expression has been found in T cells, B cells, monocytes, neutrophils, epithelial cells, fibroblasts, and endothelial cells. Interleukin 17 has 5 known types of receptors (IL17RA, IL17RB, IL17RC, IL17RD and IL17RE). The main difference between them is which type of cell expresses each. IL17RA is the most widely expressed and can also bind with other ILs of the IL17 family. The mechanisms of each of the receptors are very complex which include the activation of pathways not even mentioned in this document, and as such, they are not going to be described further. However, it will be mentioned that they are the target of several drugs for IL-17 inhibitor and monoclonal antibodies techniques which aim for the treatment of several autoimmune diseases.

- **IL-18.** It is produced by many types of cells in the body, and its receptor is also found in a wide variety of cells. IL-18 is involved in the activation of NK cells, T cells, and B cells. The binding of IL-18 to IL-18 receptor activates NF- κ B

which promotes the production of pro-inflammatory cytokines in both immune and non-immune cells.

- **IL-20.** IL-20 is mainly produced by monocytes, granulocytes, and dendritic cells in response to IL-1 β , IL-17, IL-22, TNF, and LPS. The main targets of IL-20 are keratinocytes, endothelial cells, and adipocytes. It plays a role in several biological processes:
 - Regulation of skin homeostasis: IL-20 is produced by skin cells and helps to regulate the normal function of keratinocytes and fibroblasts. It can also modulate the expression of various genes involved in skin development and differentiation.
 - Wound healing and tissue repair: IL-20 has been shown to stimulate the proliferation and migration of epithelial cells, which play a critical role in wound healing and tissue repair. It can also enhance the production of collagen and other extracellular matrix proteins, which are essential for tissue regeneration.
 - Cancer development: Because it promotes tissue growth, especially in the skin, it can also promote the development of tumor cells that derive from the skin. However, it is also known to reduce tissue damage in chronic inflammation which suppresses cancer. So the mechanism of IL-20 with respect to cancer is ambiguous and needs further study.
 - Inflammatory responses: IL-20 can stimulate the production of TNF α , IL-1I, L-6 and IL-8, leading to an pro-inflammatory response. It is also a chemotaxin itself. Can modulate the function of dendritic cells, T cells, and B cells. It can also regulate the production of antibodies
- **IL-22.** IL-22 is produced by the immune cells Th1, Th17, NKT cells, neutrophils, and macrophages, which are already in the inflammation site. It targets several non-immune cell types. It promotes cell growth and healing of tissue, while at the same time promoting the synthesis of several proteins with germicidal properties.
- **IL-24.** Released mainly by activated monocytes, macrophages, and Th2 cells through TLR. Target cells are those in the skin, lungs, and reproductive tissues. It suppresses cell proliferation during wound healing, and particularly it has been found to have an important role in destroying cancerous cells. It promotes TNF- α , IFN γ , and IL-1, which in turn promote cell apoptosis.

Table 3.6: Overview of all interleukins explained in this document

ID	Type	Chemotactic	Producers	Targets	Main
IL-1	Pro	Yes	Dendritic cells, Macrophages, Monocytes, T-cells	Pro-inflammatory proteins, PGE2	Fever, pro-apoptosis, and pro-inflammation
IL-2	Pro	No	NK cells, T-cells	NK cells, T-cells	Increases immune activity
IL-4	Anti	No	Basophils, Eosinophils, Mast cells, Th2	M1, Th0, Th2	Make Th2 and anti-inflammatory
IL-5	Both	Yes	Eosinophils, Mast cells, Th2	Eosinophils	Allergic rhinitis and asthma
IL-6	Both	Yes	Macrophages	B-cells, Neutrophils, T-cells	Fever and Autoimmune
IL-7	Neither	No	Stem cells / Dendritic cells / Keratinocytes, Hepatocytes, Epithelial cells	T-cells	T-cell regulator
IL-8	Pro	Yes	Macrophages / Endothelial cells, Epithelial cells, Smooth muscle cells	Neutrophils	CAM activator
IL-10	Anti	No	Monocytes	Macrophages, B cells, Th1	Anti-inflammatory
IL-12	Pro	No	B-cells, Dendritic cells, Macrophages, Neutrophils	NK cells, T-cells	Kill
IL-13	Anti	No	Basophils, Eosinophils, Mast cells, NK/T cells, Th2	Th0, Th2 and M1	Anti-inflammatory and Allergies
IL-15	Pro	No	Dendritic cells, Macrophages, Monocytes / Fibroblasts, Keratinocytes, Myocyte, Nerve cells	NK cells, Innate lymphoid cells	Increases immune activity and survival
IL-17	Pro	No	Th17	B-cells, Monocytes, Neutrophils, T-cells / Epithelial cells, Endothelial cells, Fibroblasts	Increases pro-inflammatory cytokines
IL-18	Pro	No	Several	Several	Type 1 immunity
IL-20	Pro	Yes	Dendritic cells, Granulocytes, Monocytes	Adipocytes, Endothelial cells, Keratinocytes	Healing
IL-22	Neither	No	Macrophages, Neutrophils, NK/T cells, Th1, Th17	Several	Healing and germicidal
IL-24	Pro	No	Macrophages, Monocytes, Th2	Several	Anti-healing and Anti-cancer
IL-33	Anti	No	Several	Mast cells, Th2	Promotes Th2 cytokines

- **IL-33.** It is expressed by a wide variety of cells. It can target IL1 family receptors mainly expressed in Th2 cells and mast cells.

IL-33 main function is to promote Th2 cytokines production. This has anti-inflammatory effects such as those described in IL-4, but also pro-inflammatory effects due to allergic reactions as described for IL-13. But overall, the same as with IL-4 and IL-13, is an anti-inflammatory cytokine that leads to resolution. IL-33 is known to play an important role in the development of allergies and asthma. It activates Th2 cells and eosinophils, which are the key mediators of allergic inflammation. IL-33 also contributes to airway remodeling, mucus production, and bronchoconstriction.

IL-33 has both pro-tumor and anti-tumor effects depending on the type and stage of cancer. In some cancers, IL-33 promotes tumor growth and metastasis by stimulating angiogenesis and cell proliferation. However, in other cancers, IL-33 promotes anti-tumor immunity by activating natural killer cells and CD8+ T cells.

Interferon

Interferon is the way cells have to communicate to other cells that a virus is about to kill them, and they need to be ready and prepared so they don't die next. If a virus manages to sneak its RNA/DNA into a cell, it usually means that the cell is unsalvageable and needs to be eliminated. The cell itself can recognize this. The cell will initiate the Interferon regulatory factor (IRF) by two means. First, it can recognize viral parts using its own Patter Recognition Receptors (PRRs). The second way is using cGAS-STING cytosolic DNA sensing pathway; this tool detects DNA in the cytosol of the cell. DNA should never be outside the nucleus, so chances are that DNA in the cytosol is viral DNA and needs to be eliminated. Either way, this activation leads to further activation of NF- κ B in activated B cells.

- **Alpha and Beta interferons.** IRF will stimulate the cell DNA to express Interferon alpha (IFN- α) and Interferon beta (IFN- β) which will be released from the cell and go into nearby similar friendly cells and NK cells.

On similar cells IFN- α and IFN- β will stimulate the differentiation of protein kinase R. This protein clip viral RNA/DNA, so when the virus that was lurking around injects its genetic material into the cell, protein kinase R will be ready to cut it and make it useless. On the other hand, it also alerts nearby NK cells to come

by and start checking the MHC1s. The infected cell will usually fail to provide a valid MHC1 and the NK cell will eliminate it.

Finally, IFN- α and IFN- β also lead to the activation of the JAK-STAT pathway (section 3.4.3), which leads to the activation of downstream Interferon-stimulated genes (ISGs), which leads to the production of more interferons creating a local positive-feedback loop which alert many nearby cells at once.

- **Gamma interferon.** IRF will also stimulate the expression of IFN- γ . IFN γ is produced mostly by NK, NKT, CD4 Th1, and CD8 cytotoxic T lymphocytes. Its function is to activate nearby macrophages, signaling them to start proliferating and to express even more MHC1/2 on the surface so they can communicate with other leukocytes better. Macrophages can also release IFN- α and IFN- β . It can also activate dendritic cells and B-cells, and induce production of TNF- α and IL-6.

An important property of IFN γ is that can increase the production of CXCL10 which is an anti-angiogenic chemokine that will be discussed later. In essence, it means that IFN γ aids in suppressing the creation of new blood vessels. This suppresses healing, but also suppresses tumor grow and is an important

Toll like receptors

Toll-like receptors (TLRs) are a group of membrane receptors that are expressed on APC that enhance inflammatory response. They bind to PAMPs, and depending on which type of TLR was binding to which type of PAMP, they will signal the production of certain cytokines or chemokines. There are 13 known TLRs, 10 of which are present in humans. Three main gene regions are activated by TLRs:

- activator protein-1 (AP-1) which regulates transcriptions of several cytokines and growth factors. Furthermore, it controls many cell processes including differentiation, proliferation, and apoptosis.
- IRF (discussed in section 3.4.3)
- NF- κ B (discussed in section 3.4.2)

TLR activation can also lead to excessive inflammation and tissue damage, thus dysregulation of TLR signaling has been implicated in various inflammatory disorders and autoimmune diseases.

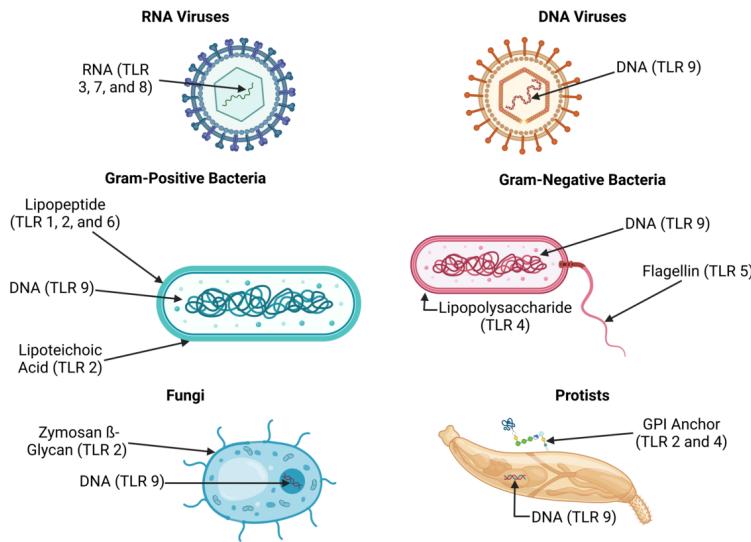


Figure 3.38: Overview of factors that activate human TLR. Reproduced from <https://commons.wikimedia.org>

Tumor necrosis factors superfamily

The tumor necrosis factor (TNF) superfamily is a collection of 19 proteins. They are mostly expressed in immune cells and they regulate immune response, inflammation, proliferation, differentiation, apoptosis, and embryogenesis. They can also leave the cell and act as cytokines. Here we discuss the 3 major ones, while in the Olink panel section, we will see the details of other TNFs and TNF receptors.

- **LT α / TNF β / TNFSF1** Lymphotoxin- α (LT α) mainly acts as a lymphoid tissue organizer by promoting the development and maintenance of secondary lymphoid organs such as lymph nodes and spleen. LT α can also signal through the Lymphotoxin- β receptor (LT β R) to stimulate the production of pro-inflammatory cytokines and chemokines by stromal or epithelial cells. This protein was formally known as Tumor Necrosis Factor β , and is referred to as such in the Olink panel.
- **LT β / TNFC / TNFSF3** Lymphotoxin- β (LT β) is involved in the regulation of immune cell trafficking and inflammatory responses. LT β can bind to three distinct receptors: LT β R, TNF receptor 2 (TNFR2), and herpes virus entry mediator (HVEM) (also known as TNFRSF14). The binding of LT β to LT β R promotes the development of lymphoid tissues and contributes to the initiation of adaptive immune responses. The binding of LT β to TNFR2 or HVEM can activate NF- κ B signaling and induce the expression of pro-inflammatory genes. This protein was formally known as Tumor Necrosis Factor C.

- **TNF / TNF α / TNFSF2** The name for this protein might be confusing. Tumor Necrosis Factor is protein 2 of the many Tumor Necrosis Factor superfamily proteins (TNFSF2). Now it is called simply Tumor Necrosis Factor. It was formerly known as tumor necrosis factor-alpha (TNF α) and is referred to as such in the Olink panel and throughout this thesis to avoid misunderstandings with the TNF superfamily.

It shares many biological functions and signaling pathways with LT α and LT β . TNF α is produced by a variety of immune and non-immune cells in response to infections, tissue damage, or other inflammatory stimuli, but in particular, it is chemotaxis that attracts neutrophils and promotes leukocytosis. It will increase APR, and it causes fever by activating PGE2.

TNF α can bind to two receptors: TNFR1 and TNFR2, which are expressed in almost all cell types. The binding of TNF α to its receptors triggers a cascade of signaling events that lead to the activation of inflammatory pathways, cell proliferation, and cell death.

It also acts as an adipokine. TNF α promotes insulin resistance. This is a condition in which higher levels of insulin are present in the bloodstream, but they are unable to move efficiently glucose into the cells for energy. As such is associated with obesity-induced type 2 diabetes.

3.4.4 Inflammation cycle

We have finally reached the point in which we can understand everything that happens during an inflammation event. An inflammation will start with a specific stimulus, and from there on all the mechanisms that lead to inflammation and the resolution of inflammation will take place at the same time. However some will act faster than others; first, the clotting of the blood happens very quickly, then the inflammation and destruction of the pathogen happen, and finally, the mechanisms for resolution will mount up and decrease the inflammation.

Stimuli

In order to start an acute inflammation process is necessary to provide the body with a stimulus, for example, pathogens, toxins, radiation, or physical trauma [107]. The stimuli can be classified as:

- External:
 - Non-microbial
 - * Allergens are a type of antigens that stimulate the secretion of IgE to fight a particular allergen instead of a normal parasitic infection. This includes plants, metals, insect stings, penicillin, many types of food, vaccines, animals, and many more.
 - * Irritants can be any object that enters the body and causes irritation, such as splinters and dirt. It can also be substances such as acids, alkalis, and solvents.
 - * Toxic compounds like chemicals such as recreational drugs or snake venoms, physical, such as coal dust or asbestos, or ionizing radiation.
 - Microbial
 - * Virulence factors, such as everything discussed in section 3.5 regarding *S. Aureus* colonization, immuno-evasion, and immunosuppression.
 - * PAMPs. These are small molecule patterns present in pathogens but not the host, which help the immune system target the pathogens. Common PAMPs are found in:
 - Viruses: double-stranded RNA (dsRNA)
 - Gram-Negative bacteria: LPS and Peptidoglycans described in section 3.5.2.
 - Gram-Positive bacteria: Lipoteichoic acid and bacterial lipoproteins (sBLP) described in section 3.5.2.
- Internal: Internal Damage Associated Molecular Patterns (DAMPs) are host cell molecules that are released when there is tissue damage, or the cell just dies.

These factors are not mutually exclusive. For example, a bacteria can have PAMPs, and start destroying human cells which then will release DAMPs. Irritants do not contain PAMPs, but in contact with the skin, they can form a complete antigen and cause inflammation, like with for example poison ivy.

Clotting

When an injury occurs, even before dealing with any infection, it is necessary to stop the bleeding so a person does not die from blood loss. The first step in blood clotting is the formation of a platelet plug, which aggregates at the site of injury [108] and becomes activated, causing them to stick together and form a temporary seal over the site of the injury [109]. As platelets aggregate, they release vWF (section 3.4.2) so they can stick together, and thromboxane (section 3.4.3) to reduce blood flow to the site of injury. This forms a quick temporal fix that prevents further bacteria from entering the body and provides the basic structure for tissue repair.

Once the platelet plug has formed, a more stable blood clot is formed later on using proteins called clotting factors. These clotting factors work together to convert prothrombin, into thrombin, which in turn converts fibrinogen into fibrin. Fibrin forms a net that reinforces the platelet plug, creating the final stable clot. Fibrin also happens to help *S. Aureus* infiltration as described in section 3.5.2.

Acute phase

Once the clot is in place, then it is possible to activate vasodilation and bring in the immune system so it can deal with the infection. Leucocytes have external PRRs in the cell surface which are activated in contact with PAMPs or DAMPs, which trigger the immune and inflammation response [110]. These PRRs are non-specific, meaning that the leucocyte does not know what causes the stimuli, just that something bad is happening and an immune response needs to happen to whatever it is. There is also no cell memory associated with the PRRs.

Additionally, some non-immune cells such as epithelial cells, endothelial cells, and some stromal cells can also express certain types of PRRs that can recognize viral components such as TLRs or RIG-I-like receptors (RLRs) which are explained in section 3.4.3.

When PAMPs interact with a granulocyte as described in section 3.4.2 , they release histamines, bradykins, and eicosanoids that cause vasodilation in blood vessels and also open a gap in the endothelium releasing blood into the area close to where the DAMP activity is happening, which causes swelling. It also causes muscle relaxation which causes vasodilation (localized hyperemia). You get more blood in the area which turns the area red. It also helps endothelium cells to express more selectins which attract neutrophils to the inflammation site via extravasation. Then neutrophils phagocytose

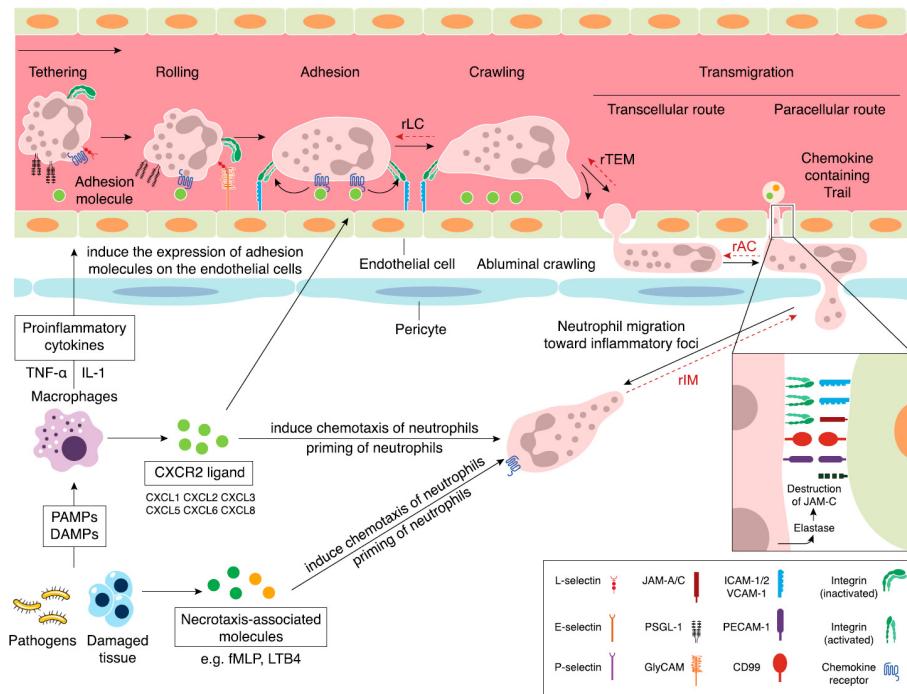


Figure 3.39: Overview of cell migration from a blood vessel to an inflammation site. Image reproduced with permissions from "Deep insight into neutrophil trafficking in various organs" [9].

the bacteria. Then dendritic cells present pieces of the bacteria to T-lymphocytes which activates the adaptative immune system if necessary.

Switching and resolution

Specialized pro-resolving mediators (SPMs) are produced to inhibit pro-inflammatory mediators and regulate neutrophils. This mechanism is done by the Lipid-mediator-class switching illustrated in figure 3.34 which is described in sections 3.4.3 and 3.4.3.

Tissue repair

Macrophages eat dying or dead cells which provide room for new cells [111]. They also secrete growth factors that promote the angiogenesis of temporal capillary vessels. Fibroblasts synthesize collagen in the area of interest. Mild damage in a tissue gets repaired to a normal state, while in severe damage the tissue is replaced by a non-functional fibrous scar.

Chronic inflammation

If everything goes well, the infection has been neutralized, and the inflammation has subsided. However, sometimes errors can occur. Typically this involved not being able to clear the site of inflammation from the cause. For example DAMPs or any other bacterial debris [112]. In worse cases, external bodies such as undetected wood splints, or metal allocated inside the muscle that can't be removed surgically.

Dysregulation of cytokine signaling can result in excessive production of pro-inflammatory cytokines such as TNF α [113], IL-1 [114], or Interleukin 6 (IL-6) [115]. Otherwise, failure or delay of activation of SPMs mechanisms.

Finally, autoimmune diseases such as celiac disease, rheumatoid arthritis, lupus, and many others lead to chronic inflammation.

3.5 *Staphylococcus aureus*

3.5.1 Introduction

The word "Staphilus" derives from the Greek "σταφυλόκοκκος", composed of "staphylé" meaning bundle, and "coccus" meaning grape. This refers to their bundle of grapes-like arrangement. "aureus" comes from Latin origin meaning golden, which is the golden-orange characteristic color of this bacteria as it is rich in carotenoid pigments. *S. aureus* was discovered in 1880 by Alexander Ogston who noticed a formation of bacteria in pus during a procedure he was performing [116]. Wounds caused by *S. aureus* infections were fatal for most patients until the 1940s when it was discovered that benzylpenicillin could cure such infections. But unfortunately, shortly after [117], *S. aureus* evolved into a penicillin resistance strain which became widespread by the end of the century. This led to the development of methicillin [118] in the 1960s, which was a better option to treat infections caused by the bacterium. However, again, the bacteria evolved to be antibiotic resistant, called *Methicillin-resistant Staphylococcus aureus* (MRSA). Once this strain was characteristic of the hospital, but today it is widespread (ranging from 2% to 80%) in both human and cattle population [119], and thus it is important to understand the *S. aureus* social spread which motivated the writing of Paper A.

Nearly 30% of humans are carriers of *S. aureus* [120] which are usually present in the skin and the upper respiratory tract. Under normal circumstances the bacteria is harmless, but it can cause a wide range of diseases, ranging from minor skin diseases such

as pimples or follicles the most common ones, to life-threatening ones such as pneumonia, endocarditis, and sepsis. *S. aureus* is in the top five most common intrahospital infections and is the most common cause of wound infection after surgery, causing around 500.000 hospital infections in the US alone [121], of which 10% end up in death-related to such infections from the non-antibiotics resistance strain alone [122].

In 2019, there were nearly 14 million infection-related deaths globally, of which *S. aureus* is the world's leading bacterial pathogen in the number of deaths with more than 1 million [123], and second in years of life lost [123].

3.5.2 Microbiology basic concepts

Prokaryote and Eucaryote

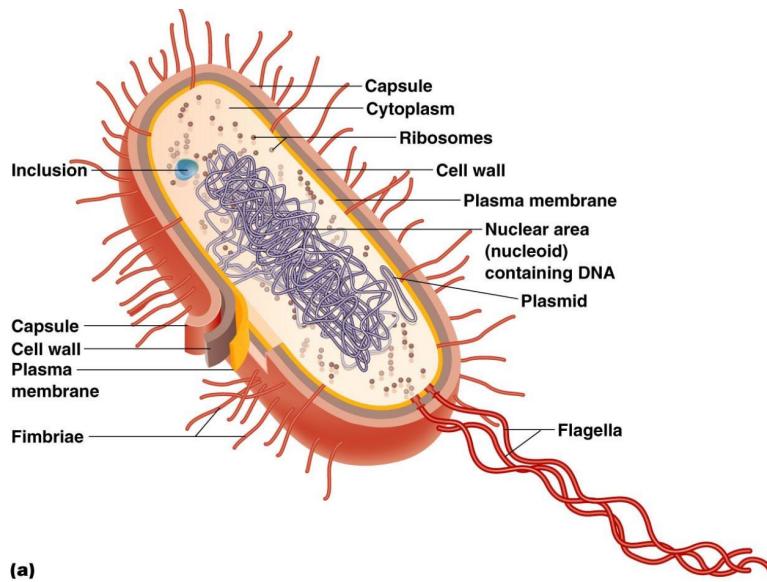
Prokaryotes are unicellular organisms that lack a nucleus encasing the genetic material and membrane-bound organelles. Their genetic material is a circular DNA molecule located in a region called the nucleoid. They are generally smaller in size and have simple structures compared to eukaryotes. However, they have both a cell wall and a cell membrane. This cell wall is made of peptidoglycan. Prokaryotes are unable to perform cellular respiration and don't have any cytoskeleton. *S. Aureus* belongs to this category.

Eukaryotes, on the other hand, are more complex organisms that have a nucleus enclosing genetic material and various membrane-bound organelles, such as mitochondria, endoplasmic reticulum, Golgi apparatus, and lysosomes. Their genetic material consists of several linear DNA strands contained within the nucleus. They can be unicellular or multicellular organisms. They have a cell membrane, but not necessarily a cell wall (humans do not). Eukaryotes have a cell wall made up of cellulose or chitin.

Gram-positive and Gram-negative

Gram staining is a technique developed by Hans Christian Gram in 1884 which is used to differentiate between bacteria (prokaryotes cells). The first type is called Gram-positive, which has a simple structure of only one cell wall made of peptidoglycan and teichoic acid, followed by a cell membrane made of a lipid bilayer, lipoteichoic acid, proteins, and carbohydrates. The second type is called Gram-negative and it has an outer cell membrane, followed by a much thinner cell wall, followed by the inner cell membrane.

The space between membranes is denominated the intermembrane space or periplasm. Both, the cell wall plus cell membrane in Gram-positive bacteria, and the outer membrane, cell wall, and inner membrane in Gram-negative bacteria are called the cell envelope. Outside the cell envelope, we might find the final layer, the glycocalyx, a shell made of polysaccharides, proteins, and lipids, which come from the surface of a bacterium, plus extracellular debris that adheres to the bacterium. Alternatively, the bacterium might have an "S-layer" (surface layer), a monomolecular layer composed of only one or two identical proteins or glycoproteins.



Copyright © 2007 Pearson Education, Inc., publishing as Benjamin Cummings.

Figure 3.40: Overview of a rod-shaped Gram-positive cell with a capsule. Reproduced with permissions from Pearson Education.

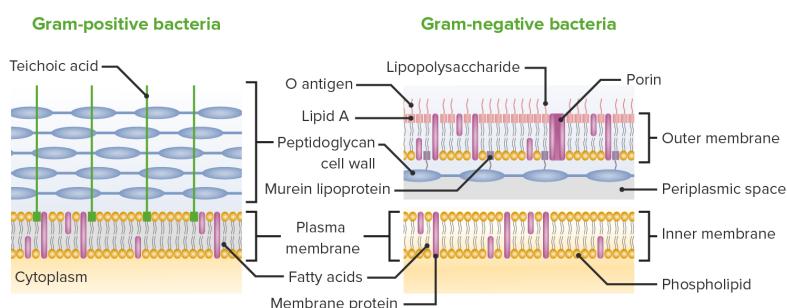


Figure 3.41: Differences between Gram-positive and Gram-negative cell wall and cell membranes. Reproduced with permission from <https://www.lecturio.com/>

I - Glycocalyx

The glycocalyx is a gelatinous layer just outside the cell envelope. An irregular gel-like glycocalyx that varies in shape and density is called a slime layer, whereas a distinct more

rigid one is called a capsule. In both cases, it serves as a protective function, preventing desiccation, detergents, heat, antibiotics, and phagocytosis. And also in both cases, it contributes to the bacterial structural integrity.

In addition to protecting the cell from environmental stresses, the glycocalyx particularly the slime layer plays a role in attachment to substrates and host tissue. This is particularly inconvenient for the host, as the slime layer can help them to stick to tissues or implants and evade the immune system. This is critical for the formation of biofilm, which will be discussed later.

II - Outer membrane

In Gram-negative only. The outer membrane contains Porins, which are beta barrel proteins (holes) that open a channel so specific types of molecules can be transported inside or outside of the cell. Porins also exist in Gram-positive mycobacteria.

It also contains LPS. These are large molecules that are toxic to humans but are not released from the bacteria towards the host, so they are denominated endotoxin (within-toxin) and are often a synonym of LPS. They are characterized as being very antigenic (polysaccharide O-antigen) and pyrogenic (lipid A) by activating IL-1 (fever and inflammation) and TNF α (recruiting white blood cells and endothelial activation). Detail about these cytokines are explained in section 3.4.2.

III - Intermembrane space

In Gram-negative only. The cell wall is contained within the intermembrane space. So everything that is proper of the cell wall is also in here. Within this space, molecules can accumulate, in particular β -lactamase, which is very significant in antibiotic resistance.

IV - Cell wall

Gram-positive bacteria are composed of many peptidoglycan layers and teichoic acid which is specific to each bacteria species and allows them to bind to fibronectin which will become relevant later on. A peptidoglycan is simply a chain made of peptides and carbohydrates that gives structural support to the cell and gives protection against osmotic damage. In Gram-negative bacteria, it only has a very thin cell wall composed of fewer layers of peptidoglycans.

V - Inner membrane

Gram-positive bacteria are the only ones that have lipoteichoic acid. It is a regulator of autolytic wall enzymes (prevent the cell from killing itself) and it has specific antigenic properties.

Both Gram-positive and Gram-negative have a phospholipid bilayer, carbohydrates, and proteins. In particular, enzymes are responsible for cell wall synthesis and penicillin-binding protein (PBP). When penicillin binds to PBP, the cell cannot make cell walls, and it dies either from collapsing or osmotic damage. Human cells are unaffected by penicillin because eukaryotes cells regulate osmotic damage directly in the cell membrane, and because their structure is supported by their internal cytoskeleton. Bacteria's inner membrane can also protect from osmotic damage, but it relies more on the cell wall for that function.

Morphology and arrangement

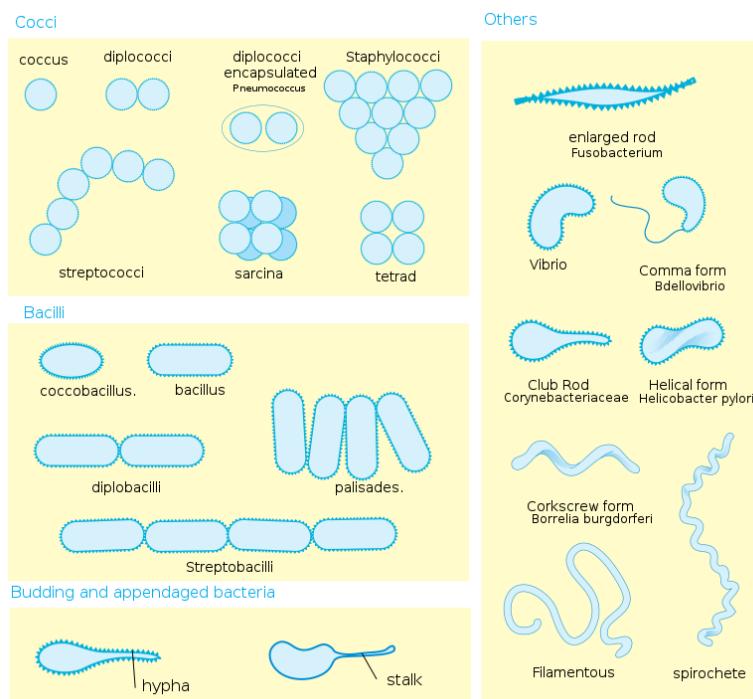


Figure 3.42: Different types of bacteria morphologies. Source: Wikimedia.

Both Gram types can be either cocci (round) or rod-shaped. Rods are also referred to as bacilli, but these are not to be confused with the genus *Bacilli*. Bacteria are italicized in text, as such *Staphylococcus Aureus* is italicized in this thesis. A better example is

Haemophilus influenzae a bacterium, which shall not be confused with influenza, not italicized, hence a virus. *Staphylococcus* refers to the genus, while *aureus* refers to the species. As the name suggests, *Staphylococcus Aureus* is a cocci bacterium arranged in clusters. Other less common morphologies are included in figure 3.42.

Gram-positive characterization

I - Spore formation

Rods can be spore-forming or non-spore-forming. Both are sub-categorized into aerobic, and anaerobic. Furthermore, both are divided into Motile and Immotile. They are beyond the scope of this thesis.

Gram-negative bacteria never make spores. However some Gram-positive bacteria can make endospores under some conditions, typically unfavorable environments. This effectively means that they transform from a vegetative state (capable of reproduction) into a spore state, which is a dormant state in which the bacteria are encased and protected by multiple layers of Calcium-Dipicolinic acid. This is the reason why some bacteria are heat resistant, and boiling things, such as your food, can kill some bacteria but not all of them. And why we use autoclaves to sterilize instruments at 121°C at 15psi for nearly 60 minutes.

II - Catalase

Catalase is an enzyme that catalyzes the conversion of H₂O₂ into H₂O and O₂. Coccis can be catalase positive forming cocci bundles such as *S. Aureus*, or catalase negative forming cocci rows such as *Streptococci*. A quick test with no microscope, stains, or culture, for bacterial infection is to drop some H₂O₂ on it, and if it bubbles (O₂) it means you have a *Staph* group, and is immune to oxidative burst by neutrophils.

Catalase positive can be divided further using a coagulase test. If the test is positive it means that the sample is a *Staph aureus*. If the test is negative you flood the sample with an antibiotic called novobiocin. If the bacteria die, it means that you have a *Staph epidermidis*, but if it is resistant it means that you have *Staph saprophyticus*

Catalase negative are organized in chains, such as the *Streptococcus*, and are classified according to their hemolytic activity. This means they can break down or not red blood cells. Here I will give you some clinically relevant examples.

- **Partial hemolytic (α)** Do an optochin resistance test (another antibiotic), if it is sensitive then is *S. Pneumoniae*, if it is resistance you have viridans group streptococci (vgs).
- **Total hemolytic (β)** Do another antibiotic resistance test with bacitracin. If it is sensitive you have a group A streptococcus (GAS) such as *Streptococcus pyogenes*. If it is resistance, you have a group B streptococcus (GBS) such as *S. agalactiae*.
- **No hemolytic (γ)** Try to grow it in a NaCl solution. If it grows, you have *Enterococci*. If it kills the bacteria, you have *S. Bovis*.

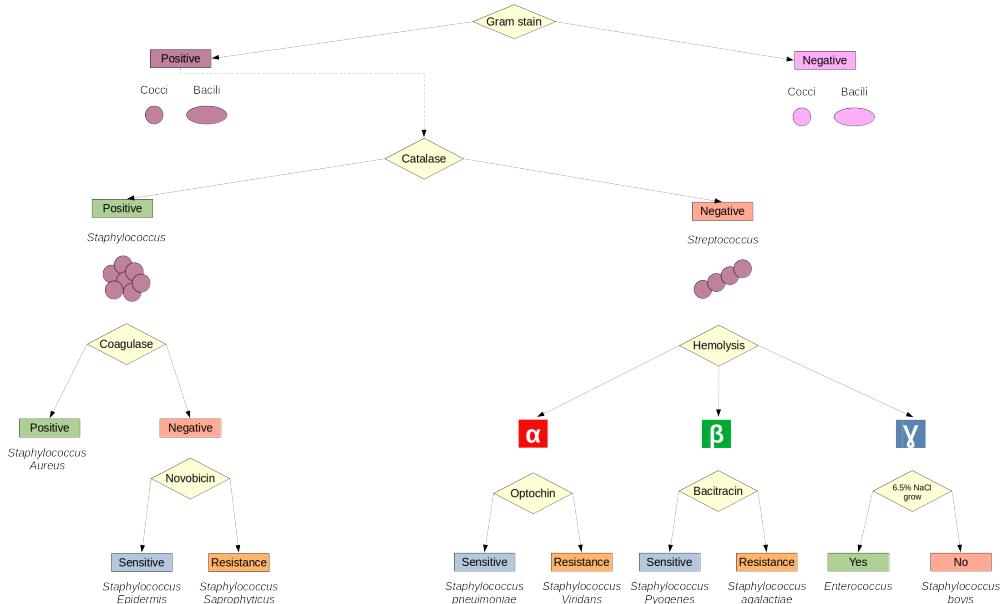


Figure 3.43: Overview of *S. Aureus* identification and other bacteria. Self-made figure.

III - Fibrin

In all cases, all coagulase-positive bacteria lack endospores. "Endo" means with-in, meaning that even though they cannot make their own spores, they can use external tools to make spores. This is the case of Fibrinogen which is found in the human body, and *Staph* uses it to convert it into Fibrin, utilizing coagulase to do so. Effectively, this is employed to hide from the immune system. As a result, coagulase-positive bacteria are highly localized but vary in infection size (folliculitis, abscesses, furuncles, and carbuncles). While coagulase-negative bacteria tend to be widespread (sepsis, cellulitis, necrotizing fasciitis)

Localized bacteria have the advantage that they can be a threat with an incision and drainage as opposed to a widespread infection. The disadvantage is that it exerts much more pressure in the area, especially inside the skull when dealing with brain abscesses.

3.5.3 *S. aureus* main characteristics

The mechanisms of the spread of *S. aureus*, as well as the severity of its associated illnesses, are quite diverse.

Bacterial properties

S. Aureus is a prokaryote Gram-positive bacteria capable of growing in both aerobic and anaerobic, and a variety of acidic or based places, although it prefers aerobic and neutral acidic environments such as the skin.

Structural components

I - Glycocalix

The *S. Aureus* can have both a capsule and a slime layer depending on the strain. The *S. Aureus* capsule inhibits phagocytosis as with many other bacteria capsules. Finally, the capsule is prone to contain adhesin proteins which help *S. Aureus* adhere to the epithelium of the mucosa, such as skin, nasopharynx, oropharynx, gastrointestinal tract, and in neonates umbilical stump and peri-anal area. The *S. Aureus* slime layer is prone to forming biofilms. [124]

A biofilm is a community of microorganisms, including bacteria, that stick together on a surface. The biofilm is composed of many layers of cells and extracellular polymeric substances (EPS). The EPS is a sticky layer mostly formed by polysaccharides secreted by the bacteria that helps them stick to anything, particularly room surfaces, human tissues, or medical equipment. The EPS also provides a barrier against harmful agents and promotes nutrient exchange and communication between adjacent cells.

Biofilms are often associated with infections. They are very difficult to detect or treat. The immune system is often unable to break through the protective exopolysaccharides layer. Antibiotics stay in the blood for a short time and are also unable to penetrate the biofilm, and in the meantime, bacteria hide inside. Biofilms are also the ideal breeding ground for new strains given their protected environment.

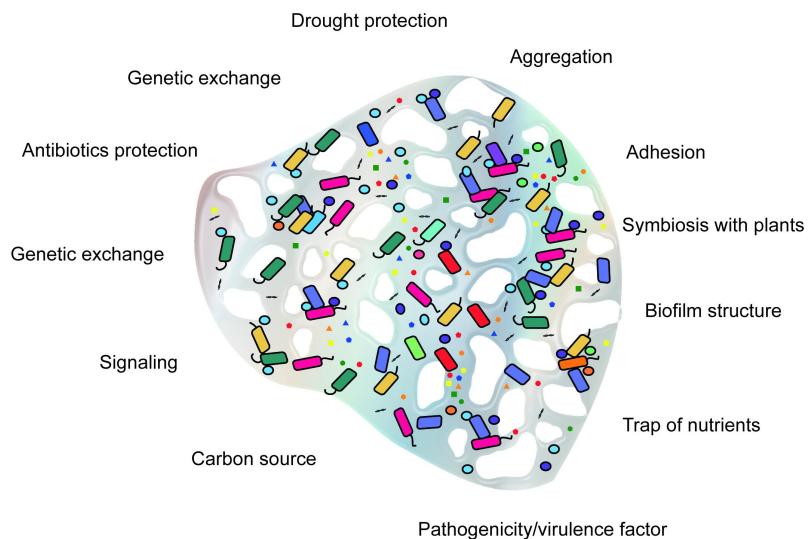


Figure 3.44: Conceptual framework of the functions of microbial extracellular polymeric substances (EPS) in soil. Reproduced from <https://doi.org/10.3389/fmicb.2018.01636>

II - Cell Wall

The peptidoglycan in the cell wall in the *S. aureus* provides surface adhesion proteins that have a very special function. They adhere to the peptidoglycan cell wall (bacteria) and at the same time to fibronectin, fibrinogen, collagen, or elastin (human tissue). Technically, it helps the immune system recognize the bacteria which is why they are called Microbial Surface Component Recognizing Adhesive Matrix Molecules (MSCRAMM).

- **Staphylococcal protein A (Spa)** arrest antibodies by binding to the constant part of the Immunoglobulin G (IgG), preventing antibody-mediated immune clearance of the *S. Aureus* [125]. Furthermore, this forms an antigen-antibody complex, which activates the classical complement pathway of the complement immune system, wasting it, and leading to hypocomplementemia. [126].
- **Other surface proteins** *S. Aureus* also have Staphylococcal surface protein C (SasC) [127] Staphylococcal surface protein G (SasG) [128] and Staphylococcal surface protein X (SasX) [129]. All of them promote adhesion to desquamated epithelial cells and contribute to biofilm formation.
- Fibronectin binding protein A (FnBPA) connect the bacteria with fibronectin and fibrinogen (human parts), which is critical for *S. aureus* virulence. It enables the bacterium to colonize and infect host tissues. In addition, FnBPA contributes to the formation of biofilms. Several small-molecule inhibitors and monoclonal antibodies against FnBPA have been developed and tested in pre-clinical studies

approaches for the treatment and prevention of *S. Aureus* infections associated with medical implants.

Enzymes

I - Coagulase

In vertebrates, thrombin converts fibrinogen into fibrin, which later on will form a blood clot aimed to stop hemorrhages. Fibrin also reduces thrombin activity and promotes angiogenesis. All of these are used by *S. Aureus* to its advantage, defining its coagulase abilities.

Clumping factor proteins A (ClfA) and Clumping factor proteins B (ClfB) transform fibrinogen into fibrin. They both bind to fibrinogen and promote clotting, forming a shell around the *S. Aureus*, which is the way *S. Aureus*, a coagulase-positive bacterium that lacks endospores as any other, can make something similar to endospores which helps it hide from the immune system on highly localized surfaces [130]. This also serves as an adhesion protein that binds the peptidoglycan cell wall to other surfaces and tissues.

Another feature of *S. Aureus* is its ability to secrete staphylocoagulase and von Willebrand factor-binding protein (vWbp), binding to human thrombin and forming staphylothrombin. Which further promotes fibrinogen to fibrin conversion, forming even more clots.

Interactions of *S. Aureus* and platelets play an important role in the pathogenesis of intravascular infections such as infective endocarditis (IE). A typical feature of *S. aureus* is the ability to generate thrombin activity through the secretion of two prothrombin-activating molecules, staphylocoagulase and vWbp, which bind to human prothrombin to form the enzymatically active staphylothrombin complex. The role of staphylothrombin in the interaction between *S. aureus* and platelets have not yet been studied. [131]

II - Hyaluronidase

Hyaluronidases are a family of enzymes that catalyze the degradation of hyaluronic acid. *S. aureus* produce hyaluronidase to obtain carbon from hyaluronan which is abundant in epithelial tissue. Is speculated that *S. aureus* uses Hyaluronidases to destroy the polysaccharide in cells, making it easier to infiltrate in the host.

III - Staphylococcal Fibrinolysin

It is a powerful fibrinolytic enzyme that helps the bacterium to dissolve blood clots to spread throughout the host. Staphylokinase plays a key role in the virulence of *S. Aureus* by allowing the bacterium to invade and spread through tissues, in particular the formation of abscesses. It also breaks IgG and Complement component 3 B (C3b), inhibiting phagocytosis, and contributing to hypocomplementemia.

IV - Lipase

S. Aureus produces lipase to digest lipids into fatty acids and glycerol, which later are turned into ATP. Lipids are also a major component of skin (and other tissues), which helps to colonize human skin more effectively. Lipids also happen to be hydrophobics. As such, when added to the biofilm, lipids help to protect the bacteria against antimicrobial agents.

V - Nuclease

A nuclease enzyme is used to break down DNA components and is a common beneficial enzyme required for the central dogma of molecular biology. In the hands of *S. Aureus* however, it breaks down the host DNA into its individual components which the bacterium uses as a source for ATP and other common cellular requirements. This is especially important when the bacteria infiltrate parts of the body where resources are scarce.

Neutrophils can commit suicide as a last method of attacking pathogens by projecting their DNA and using it as a net to catch the invader. This is known as neutrophil extracellular traps (NETs). Nuclease can break down this trap, allowing it to continue infecting the host. Furthermore, this can also damage healthy cells' DNA prompting apoptosis. All combine to contribute to tissue damage and inflammation.

VI - Collagen adhesin

Collagen adhesin (CNA) is a protein that facilitate adhesion to collagen-rich tissue [132] [133]. Especially important in the biofilm formation for ocular keratitis and septic arthritis.

VII - Iron-regulated surface protein

Iron-regulated surface protein A (IsdA) and Iron-regulated surface protein B (IsdB) facilitate the adhesion to desquamated epithelial cells and promote iron acquisition, typically obtained from the effect of hemolysin. It helps with nasal colonization. [127] [128]

Toxins

I - Hemolysin

This leads to the destruction of the red blood membrane, releasing their hemoglobin, which is used as a source of iron, which is essential for bacterial growth. Secondarily, it can aid immuno evasion because the destruction of red blood cells will lead to inflammation. This means that the immune system will divert resources to soften the inflammation instead of fighting the bacteria. Inflammation can further increase blood permeability allowing the *S. aureus* to transverse blood vessels invading different tissues.

II - Exfoliative Toxins

In between skin cells, keratinocytes, there is a protein called desmoglein-1 which attaches them together packing the skin tight. Exfoliative toxins will damage this connection so they cannot stay linked together, leading to patches of the skin falling off. This is the reason why Nikolski's sign symptom manifests in some *S. Aureus* related diseases.

III - Enterotoxins

This targets the enterocytes in the Gastrointestinal Track (GIT), creating pores on them leading to the destruction of the cell membrane. The sodium inside those cells and other electrolytes leak out, making it more difficult for the rest of the GIT to absorb nutrients, which causes diarrhea. This also damages the epithelium cells of the GIT causing gastroenteritis.

IV - Panton–Valentine leukocidin (PVL)

Panton–Valentine leukocidin (PVL) creates pores in leukocytes which lead to ions imbalance making the leukocytes die. This leads to inflammation, particularly in the lungs.

V - TSST-1

Toxic Shock Syndrome Type-1 (TSST-1) is a toxin which can be released by *S. Aureus*. This toxin acts as a superantigen. Antigen Presenting Cells have major histocompatibility complex 2 (MHC2) that interact with other cells like T-cells via their CD4 protein. TSST-1 acts as a bridge between the two and hyperstimulates their response leading to a cytokine storm of IL-1, IL-2, TNF α , and IFN- γ . All of these proteins lead to an inflammatory reaction that acts on the skin creating a rash [134]. They also increase capillary permeability causing vasodilation of the blood vessels leading to both hypotension and hypovolemic shock. Finally, they also increase prostaglandins in the hypothalamus, which leads to fever.

3.5.4 *Staphylococcus aureus* diseases

S. aureus is usually harmless and will just colonize the skin and nasopharyngeal tract of the host. However, it is also an opportunistic bacterium that easily attaches to and infects many tissues and can develop much more serious complications there [135]. These diseases range widely and can be caused by either the bacteria or the toxins produced by them.

What causes the disease?

- **Bacteria.** This is when a direct bacterial invasion. To diagnose it you need to look for IgG related to the bacteria or to perform bacterial cultures. When there are too many, it leads to bacteremia. Generally, if you want to treat it, you do it with antibiotics.
- **Toxins.** This is when the bacteria produce toxins that affect the host and produce symptoms. To diagnose it you need to look for IgG related to the toxin, not the bacteria. When there are too many toxins, it leads to toxemia. They do not respond at all to antibiotics, and samples will show no growth in bacterial agar plates.

First, we are going to list the diseases caused by the bacteria, and then, the diseases caused by the toxins.

Skin lesions and infections

They are highly localized due to the coagulase factors. The common ones are superficial abscesses and folliculitis (hair infection). If this becomes severe and deep, they will turn into furuncles (multiple and deeper folliculitis). If it becomes worse, they will turn into carbuncles.

Carbuncle is a deeper folliculitis with multiple sinus tracks, usually on the back and neck. If it goes deeper it passes the fat layer of the skin and reaches, muscles, bones, or blood vessels. Reaching the blood vessels means reaching the blood, which goes everywhere, thus spreading the bacteria in multiple organ tissues. It can also generate septic embolisms in which pus is dislodged from the original site transverse the blood and might land in an organ. Septic embolisms can also be generated in the newly infected organs reaching other random organs via blood vessels.

S. Aureus is the most common pathogen associated with wound infections and the most isolated bacteria from chronic wound infections [136] such as diabetic foot ulcers. Wound infections are also a common source of septic embolism, which might be caused due trauma (dirty injury), or surgical equipment, which might not be completely clean due to the biofilm properties of *S. Aureus* discussed in section 3.5.3. Or even not clean at all such as dirty needles reutilization in drug users.

Cellulitis is also a common complication of *S. Aureus*, however, due to the coagulase mechanisms it tends to stay focalized rather than spread, so is more likely that cellulitis is caused by strep bacteria.

Depending on the severity, abscesses are generally treated with incision and drainage. If it is a severe case, systemic antibiotics. If it is so severe that it reaches the bloodstream then intravenous antibiotics.

Catheter associated infections

When surgical equipment, prosthetics, needles, or catheters are inserted into a patient, they need to transverse the skin layer which is rich in *S. Aureus*. Even if the equipment is completely clean, this may lead to *S. Aureus* being attached to the equipment, forming a biofilm around it, and staying in the bloodstream until the device is removed. At the same time, the biofilm can release bacteria leading to bacteremia [137] in the blood and spreading everywhere else.

Pyomyositis

Pyomyositis is a bacterial infection of the skeletal muscles which, typically results in an abscess. In the case of *S. Aureus*, secondary pyomyositis which is localized is more frequent. Internal imaging such as CT scans or MRI is required to assess the lesion extension.

Endocarditis

Types of bacterial endocarditis are:

- **Acute.** High-virulence organisms are invading and destroying an intact and healthy heart valve. The *S. Aureus* acts in this group.
- **Subacute.** Weak bacteria are attacking a weak valve. Examples of these are the Haemophilus, Aggregatibacter Cardiobacterium, Eikenella, and Kingella organisms group (HACEK).

It should present with a fever, due to the *S. Aureus* MSCRAMM and cloths, and with a new murmur due to the patient having a previously healthy valve. Since the bacteria is already in the blood, it can be diagnosed using a blood culture.

Lung abscesses

Similar to endocarditis, *S. Aureus* goes into the blood and lands in the lungs. Usually causing an excessive amount of liquid (pus) to be allocated in the pleural space in the lung (pleural empyema). It causes productive cough, hemoptysis, or even necrotizing pneumonia. Besides the blood culture, it can also be diagnosed by sputum culture.

Brain abscesses

Similar to endocarditis and lung abscesses, *S. Aureus* goes into the blood and lands in the brain, causing meningitis or brain abscesses. Previously we discussed that *S. Aureus* is heavily localized, it doesn't spread but in exchange can increase pressure in the infected area. This is not a problem in places where it has space to grow, such as the skin. But the skull is 100% pack with no room for anything else which leads to the increase of intracranial pressure (ICP). This leads to any sort of focal neurological symptoms, as well as classical ICP such as morning frontal headache, projectile vomiting, and blurry vision. When doing a lumbar puncture, it might present an increase in exit pressure.

Osteomyelitis and septic arthritis

Once again, *S. Aureus* goes into the blood, which goes into the bone, bone marrow, and joints. However, it can also be caused by direct trauma from something dirty that penetrates the skin and reaches the bone, such as a knife or a bullet.

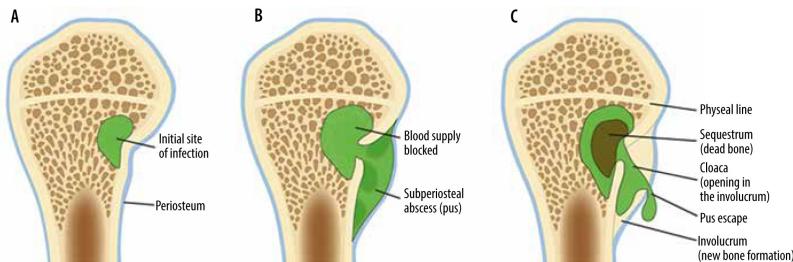


Figure 3.45: An artist's drawing of osteomyelitis progression, subacute to chronic. A) Initial site of osteomyelitis involving the medial aspect of the long bone metaphysis/intraosseous abscess (green). B) The abscess extends through the cortex into the subperiosteal space forming a subperiosteal abscess (shades of green). C) Chronic osteomyelitis with a detached central necrotic bone fragment/sequestrum (brown) within the intraosseous abscess (green) with peripheral new bone formation/involucrum and a cortical and periosteal opening in the involucrum/cloaca allowing the pus to escape. Reproduced with permission from <https://doi.org/10.5114/pjr.2022.113825>

Is likely to be caused by the metaphysis of the bone due to the stagnated blood circulation in that area, and due to the cartilage blocking the path to the epiphysis. It might lead to septic arthritis (inflammation of the joints), fracture, growth disruption, and vertebral osteomyelitis which can destroy the vertebral cord and develop neurological symptoms. Infections in bones are difficult to treat, which might lead to chronic osteomyelitis, developing into further complications such as sclerosis or the need for amputations.

Staphylococcal Scalded Skin Syndrome

Staphylococcal scalded skin syndrome (SSSS) is a skin infection. This infection mainly affects patients with weak immune systems, such as babies, children, old individuals, or HIV patients. SSSS is characterized by blistering of the skin similar to a sunburn-like rash, and the skin may feel like it is scalded. It leads to peeling and shedding of the top layer of skin just with mild pressure to the touch (Nikolsky's sign).

It doesn't present any inflammation, no cytolysis, no leukocytosis, and no bacterial growth in cultures because this is not caused by the *S. Aureus*, it is caused by the exfoliative toxins A and B. It heals on its own in about 7 to 10 days and requires only palliative care. However, immunocompromised patients have a mortality rate of 60%.

Ocular infections

Keratitis is a condition in which the cornea becomes inflamed. This can be caused by introducing *S. Aureus* in the eye, via injury, or dirty contact lenses. Similarly, conjunctivitis can occur. If the infection penetrates deeper into the eye it can turn into endophthalmitis. [133]

Bullous Impetigo

Is a yellowish crust sore on the skin, normally in areas with skin folds. Is a localized form of SSSS and is only caused by a specific strain. Is treated with systemic antibiotics such as oral cephalexin. Nonbullous impetigo can be caused by staph or strep bacteria and it can be less severe. In such cases, topical antibiotics can be used.

Food poisoning

Clients of a dirty restaurant who experienced vomiting or diarrhea were likely affected by an enterotoxin (section 3.5.3). *S. Aureus* have Staphylococcal enterotoxin A (SEA) and Staphylococcal enterotoxin B (SEB). They are heat resistant and remain even after cooking the food, meaning that you killed the bacteria, but not the toxin. Is likely that the *S. Aureus* ended up in the food due to the chef touching his nose with his fingers, regardless of whether he wore sanitary gloves or not. If food is left uncooked at room temperature it will give *S. Aureus* a better chance to reproduce and create even more toxins. *S. Aureus* also survives in salt.

SEA causes gastroenteritis, it presents no fever and it leads to nausea, vomiting, and watery diarrhea. It usually appears after 5 hours of incubation and resolves itself in less than 24 hours. Treatment includes normal saline for severe cases of dehydration and NSAIDs for severe pain.

SEB cause gastroenteritis and enterocolitis. SEB is a superantigen, which causes the immune system to release a large number of cytokines, which means fever and inflammation.

Toxic Shock Syndrome

The mechanism described for TSST-1 in section 3.5.3, leads to Toxic Shock Syndrome. It usually happens when external objects are left inside the body for too long and do not have an antibiotic coating, such as surgical cures or tampons.

Necrotizing Pneumonia

The mechanism described for PVL in section 3.5.3 might kill lung cells which leads to this disease.

3.5.5 Antibiotics resistance strains

Methicillin-sensitive Staphylococcus aureus (MSSA) is vulnerable to various common Methicillin-family antibiotics such as oxacillin, cloxacillin, dicloxacillin, and nafcillin. As the name suggests, it is also vulnerable to methicillin, which back in time was great because it was a β -lactamase resistant penicillin. However, methicillin also kills the kidneys via interstitial nephritis, which is why methicillin is not used anymore and similar alternatives of the same antibiotics family are used.

As time passed a new strain developed resistance to this family of antibiotics, called MRSA. This is due to the bacterium's brand new *mecA* gene that encodes and produces a PBP named PBP2a. This is present in the bacteria cell membrane instead of the original PBP, however, because the structure is different, penicillin (or Methicillin-like antibiotics) can't bind to it anymore. The solution was simple, use Vancomycin to kill the MRSA, especially the hospital-acquired MRSA.

As time passed a new strain came, *Vancomycin-resistant Staphylococcus aureus* (VRSA), which has a brand new *vanA* gene that alters the peptidoglycan cell wall structure, so the Vancomycin can't prevent cell wall formation anymore. VRSA is treated with Linezolid, daptomycin, telavancin, or ceftaroline. Linezolid is really good but expensive.

In 2001 there was the first reported case of *Linezolid resistance in Staphylococcus aureus* (LZRSA) [138]. Then 2006 in China [139], 2008 in the US [140], 2006 to 2008 in Japan [141], 2008 to 2009 in Spain [142], and 2010 in Italy [143]. As the reader might have imagined, we already have LZR *Staphylococcus Aureus*. The incidence of LZR staphylococci is still low, but this may change any day due to antibiotic abuse. Limiting hospital stances and antibiotic abuse is critical to limit new antibiotic-resistant strains. LZRSA is currently treated with daptomycin and tedizolid, however, these two antibiotics have severe adverse effects.

3.5.6 Vaccines

Several vaccines against *S. Aureus* have been developed since 1999 [124], however, due to the bacteria's immunoevasive properties, this task has been proven challenging.

Furthermore, the safety of these vaccines is still being studied [144]. Nevertheless, the *Staphylococcus aureus* 4-Antigen Vaccine vaccine has shown immune Responses up to 36 months [145].

3.5.7 Epidemiology

Prevalence

In the classic literature, there are three main groups suggested to define a person infected with *S. Aureus*, persistent carriers [120, 146, 147], intermittent carriers, and non-carriers. In 2004 "culture rules" were defined to identify each group [148]. Two cultures are taken a week apart. A person with both positive cultures is considered a persistent carrier, only one positive culture is defined as intermittent, and a no positive culture is non-carrier. This method shows a 0.99 negative predictive value [149], which is particularly important in confirming that those identified as non-carriers are very likely to be true non-carriers which maximizes the avoidance of individual misdiagnosed negative thus preventing infection. An alternative method was suggested in 2009 [150] suggesting that non-carrier and intermittent carriers are very similar in their nasal elimination kinetics and should be grouped together.

Persistent nasal carriages are strongly associated with the main diseases described in section 3.5.4 [151]. *S. Aureus* primary reservoir in humans is the nose, in particular, the squamous epithelium of the anterior nares [147]. In the general population [152], the carriage rate in the respiratory tract is approximately 27% in the nose, 10% in the neck, and 15% in the pharynx. The skin shows a carriage of 27% in hands, 20% in forearms, 22% in the perineum, 15% abdomen, and chest, 10% in ankles, 8% at the axilla, and 5% in the vagina. The carriage rate has decreased significantly from 50% during 1930s ($r = -0.55$, p -value <0.001) [152]. *S. aureus* nasal carriage seems to be more predominant among males [153, 154], white ethnicity [153], and younger patients [155, 156]. *S. aureus* nasal carriage is also increased in patients with obesity [154], diabetes mellitus [157], patients undergoing hemodialysis [158], and liver disease [159].

While we are vigilant about the transmission of *Staphylococcus aureus* in everyday life settings such as school, work, or at home (community-acquired), the greater concern lies with hospital-acquired (nosocomial) infections, where the risks are significantly higher. MRSA clones have led the nosocomial infections in the USA and other parts of the world [160], and are the second cause of nosocomial bloodstream infections [152]. *S. aureus* was described as "*the most rampant pathogen causing ventilator-associated*

"pneumonia" [161]. And to make things worse, MRSA is no longer only a nosocomial pathogen [162] and a worrisome trend is the community-acquired increase in the last years. Approximately 150,000 MRSA infections occur annually in the EU and EEA, resulting in over 7,000 deaths [163]. The European Centre for Disease Control emphasized the importance of the spectrum of activity of prescribed antibiotics [163]. Prescribing broad-spectrum antibiotics when not necessary can lead to the development of resistance as these antibiotics affect a wide range of bacteria, including beneficial bacteria, potentially allowing resistant bacteria to thrive and spread.

In the context of Norway [164], only 0.8% MRSA were detected in human clinical isolates during 2021 in blood cultures and cerebrospinal fluids. A total of 20 MRSA strains were detected. Most of the cases were reported as superficial wound infections or abscesses. No Linezolid resistant isolates were found. The number of *S. aureus* skin isolated has been stable from 2017, ranging from 10.1% to 11.1%, while the MRSA total cases have decreased since 2017 from roughly 50 per 100.000 people a year, to 32 in 2021. In cattle and also during 2021, no MRSA was detected from nasal swabs in horses, or in pig herds [164]. A special isolated is MRSA CC398 spa-type t011, which is prevalent in animals and a possible case of zoonosis. This was found in cattle in previous years; 3 in 2014 and 5 in 2019.

In the specific context of Tromsø, the most common spa types were t012 (8.4%), t084 (7.6%), and t065 (4.9%). The three large clonal complexes, CC012, CC065, and CC084 comprised 62.4% of the studied population ($n = 4026$). T012 is associated with nasal carriage and TSST-1 (section 3.5.3) [165], T084 is associated with not having the gene for breaking fibrinolysis (section 3.5.3) and seems to be mainly spread in meat products and poultry workers [166], while T065 is associated with community-acquired MRSA (section 3.5.5), which is particularly concerning because it represents the spread of antibiotic resistance into the general community [167] instead of healthcare settings.

Main Routes of Transmission

- **Direct Contact.** *S. Aureus* can be transmitted through direct skin-to-skin contact due to the skin lesions described in section 3.5.4. This is common in environments where people are in close contact, such as hospitals [168] or schools [169]. There is also an increased risk of transmission by people with direct contact with livestock [170] and pets [171].

- **Indirect Contact.** *S. Aureus* can also spread through contact with objects or surfaces that have been contaminated. *S. aureus* is able to be transmitted from dry surface biofilm via different types of gloves [172], as well as in counters [168], beds [168], or vitals monitors [168]. This is also possible at homes [171], and furthermore, doctors and their households of healthcare professionals have been shown to act as reservoirs for nosocomial MRSA [173].
- **Foodborne Transmission.** Improperly handled or stored food can become contaminated with *S. Aureus* as described in section 3.5.4. *S. Aureus* has been found in nearly 30% of food products analyzed [174], leading in cereals (46%), with the lower (27%) in dairy, eggs, and vegetables.
- **Droplet Transmission.** Although less common, *S. Aureus* can be spread through droplets in the air when an infected person coughs or sneezes [168], which later on, transmit into any of the objects described above.

3.6 Vitamin D

3.6.1 Introduction

Vitamin D is crucial in any population, but in the Arctic, it has a special interest due to the lack of sun exposure which limits the availability of this vitamin. If friends share similar vitamin D levels, then it is possible to promote behaviors that enhance vitamin D absorption which motivates the writing of Paper B.

The human body requires what is called essential nutrients, which are compounds that the body cannot synthesize by itself, or is unable to synthesize in large enough quantities. These essential nutrients consist of macro-nutrients, vitamins, minerals, choline, and water [175]. Vitamins are organic molecules that the body needs in small quantities for correct metabolism. They are presented in two groups, water-soluble vitamins, and fat-soluble vitamins. Water soluble means they dissolve in water, while fat soluble means that they dissolve in fat. Vitamin D, along with vitamins A, E, and K, belongs to the fat-soluble group, meaning that they are absorbed through the intestinal tract with the help of lipids. Luckily, vitamin D is already present in fat-rich foods such as salmon. Humans need a constant intake of water-soluble vitamins as the body is unable to store them, except for B9 (folate) and B12 (cobalamin) which are not stored either but can last for weeks to years respectively.

There are two primary functions for vitamin D, to absorb calcium and phosphate from the gut into the blood; and to inhibit Parathyroid hormone (PTH) production. There are plenty of secondary functions such as immune homeostasis, which showed a worrisome trend during the COVID-19 epidemic in which patients deficient in vitamin D levels have the highest risk of developing acute respiratory distress syndrome prompting ICU admission or death, as well as being more susceptible to contracting SARS-CoV-2 [176, 177].

3.6.2 Sources and metabolism

There are two methods by which your body obtains Vitamin D, sun exposure (pre-vitamin D3) and food intake or dietary supplements (vitamin D2 and D3) [178]. Regardless of intake method, vitamin D needs to undergo two chemical transformations known as hydroxylation to be activated and do its functions. The first one occurs in the liver and transforms vitamin D into 25-hydroxyvitamin D (25(OH)D), using 25-hydroxylase. This first form of vitamin D is the variable that we measure in the blood serum later on. The second hydroxylation happens in the kidneys and forms 1,25-dihydroxyvitamin D (1,25(OH)2D), using 1- α -hydroxylase, which is the final and physiologically active form of vitamin D. This form is known as calcitriol. If calcitriol becomes excessive, then it is converted to 24,25-dihydroxycholecalciferol, which is less active. This prevents hypervitaminosis D, and it is why this condition is very rare [179] unless a person takes an overwhelming amount of vitamin D supplements [180].

Both vitamins D2 (ergocalciferol) and D3 (cholecalciferol) raise 25(OH)D levels. The metabolism and actions of vitamins D2 and D3 are identical, and they only differ in the chemical morphology present in their side-chain structure. Evidence suggest that vitamin D3 increases 25(OH)D levels greater and longer than vitamin D2 [181–184]. Dietary supplements of 25(OH)D3 are three to five times as potent as vitamin D3 supplements [185, 186]. In any case, all form of dietary vitamin D is absorbed in the small intestine via passive diffusion using intestinal membrane carrier proteins [187]. As stated before, the presence of fat helps the diffusion and absorption of vitamin D, although some vitamin D can still be absorbed without fat.

3.6.3 VDR

Lastly, calcitriol needs to enter the cells; this is where the Vitamin D receptor (VDR) comes into play. This protein is mostly found in the cells of the small intestine, immune system, kidneys, and bones. Calcitriol then binds to VDR, which forms a protein complex

that enters the nucleus of cells and binds to DNA, up or down-regulates the expression of hundreds of genes [188, 189]. This includes calcium absorption (promotes calbindin), bone formation, cell growth [190], and immune function. VDR regulates the production of cytokines in the immune system, generally acting as an anti-inflammatory agent, facilitating humoral response, and leading to homeostasis. In figure 3.46 we see an overview of these functions.

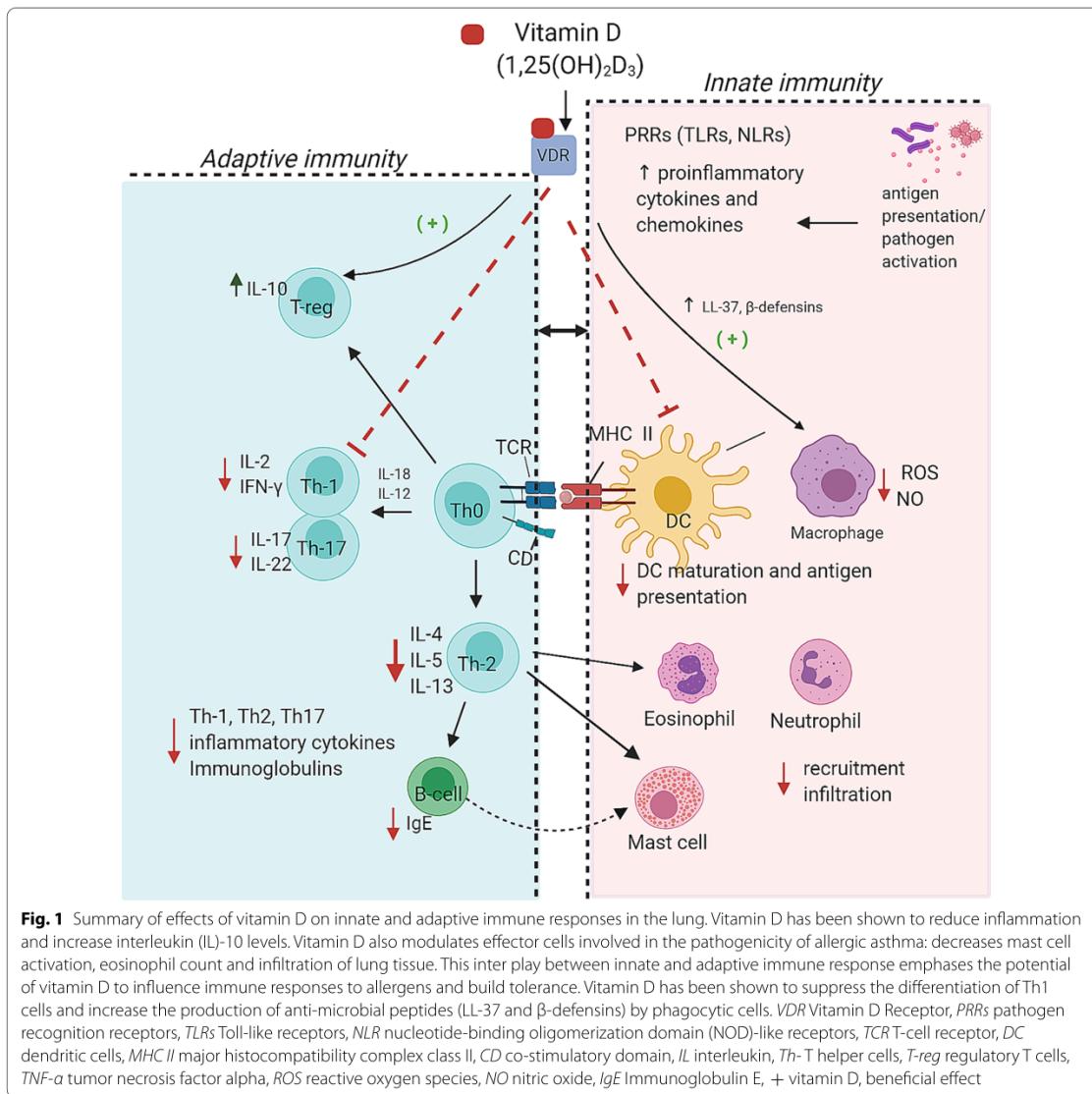


Figure 3.46: Vitamin D effects on the immune system. Figure reproduced under CC4.0 license from "Recent advances in vitamin D implications in chronic respiratory diseases" [10].

3.6.4 Calcium

Calcium ions (Ca^{2+}) are an indispensable neurotransmitter and play a critical role in muscle contraction. Without proper calcium levels in your blood, your heart will just stop working or work incorrectly. On top of that, calcium promotes healthy bone mineralization and bone growth. Finally, calcium promotes sperm hyper-activation, causing them to increase their mobility by "shaking" more violently, increasing their chances of reaching the oocyte, thus a better chance of fertilization [191–193]. Calcium is the most abundant mineral in blood, and the 5th most abundant element in the body after Hydrogen and Oxygen (water), Carbon (every cell wall), and Nitrogen (the backbone of amino acids, RNA, and DNA).

When calcitriol is present in the blood, it stimulates the epithelial cells in the intestine to increase the production of calbindin-D proteins. This increases the absorption of calcium from the brush border to the basolateral membrane, where calcium finally enters the bloodstream. While vitamin D is in charge of absorbing calcium, it is not in charge of maintaining healthy levels of calcium in the blood serum. This is the task of PHT. If calcium levels are too low, then PTH:

- Binds to osteoblast, which increases Receptor activator of nuclear factor kappa-B ligand (RANKL), which transforms pre-osteoclasts into osteoclasts. Osteoclast literally breaks down your bones apart in order to maintain calcium ions as they should be. The self-destruction of bones might sound brutal but the body prioritizes having a working heart before having a working skeleton. Details of this process are explained in section C.2.85.
- Makes a person urinate less calcium by increasing renal re-absorption in the kidneys. More precisely in the distal convoluted tubule. Ultimately, this calcium is reabsorbed into the blood, but in exchange, the person will urinate more phosphate.
- Because the person is running low on calcium, it will increase the production of $1,25(\text{OH})_2\text{D}$ in the kidney, by increasing the production of $1-\alpha$ -hydroxylase.
- Because breaking down bones and having too much calcium in the blood is also not good, PTH also inhibits the production of PTH, making a negative feedback loop that prevents flooding the blood serum with calcium everywhere.

Having too little calcium in blood serum levels could be caused by:

- Not eating enough calcium.
- The parathyroid gland is not working (hypoparathyroidism), so no PTH to take calcium from the bones into the blood.
- The person is taking too little vitamin D. Even if a person drinks an entire cow's worth of milk, without activation of the calbindin proteins, calcium cannot be absorbed by the gut at significant levels.
- The kidneys are not working. If kidneys do not make calcitriol, then the body won't have the ultimate form of vitamin D regardless of how much it takes from the Sun or food.
- The liver is not working. As to a similar reason to the previous point, if the first hydroxylation didn't happen, so can't the second one.
- Everything is fine, but the calcium is binding to something else. Due to hyperphosphatemia (eating too much bad phosphate, generally in fast food and soft drinks), alkalosis (blood is too acidic, higher albumin, which binds to calcium), or too much fatty acids in the blood.

On the opposite side, if the body has too much vitamin D and too much calcium, then it is time to save the calcium and prevent vitamin D toxicity. This is what Calcitonin is for, which is secreted in the thyroid glands. Calcitonin simply decreases calcium levels by inhibiting osteoclasts, so they don't break down the skeleton anymore. Having too much calcium in the blood is not that dangerous compared to having too little, thus calcitonin has less homeostasis power than PTH and calcitriol.

Having a bit extra calcium is not that bad, and the body really needs to have a lot of calcium ($>3\text{mmol/L}$) to start showing symptoms, which include short QT intervals (cardiac arrhythmia, too many Ca^{2+} ions in the heart make current flow to not work as intended due shorter absolute refractory), muscle weakness (Calcium pump not working properly plus an accumulation of lactic acid), kidney stones (they are trying to get rid of the extra calcium constantly, thus leading to calcium deposit on its way out via urine), or depression. This may be due:

- The person is taking too much vitamin D and eating too much calcium at the same time. Cut down vitamin D and the extra calcium will just abandon the body with the rest of the waste.

- The parathyroid gland is working too much (Hyperparathyroidism). PTH is stealing calcium from the bones way more than it should.
- Bad tumors, as they secrete PTH peptides (PTHrP) to mimic PTH, so they can get more calcium and grow better.
- The kidneys are working too much, increasing the calcium reabsorption constantly. Normally due to the use of diuretic substances.

3.6.5 Phosphate

After calcium, the second most common mineral in blood serum is phosphate, present in the inorganic form of phosphorus. This is crucial for the structure of DNA and RNA. Plays a central role in the transportation of energy in the form of ATP. All cellular membranes are composed of phospholipids. Calcium phosphate salts stiff bones and is a central component for healthy teeth development.

Everything discussed before regarding calcium absorption is also applied to phosphate. On top of that, under normal circumstances, low levels of phosphate in blood may also be caused by:

- Not eating enough phosphate-rich food
- The blood is too acidic (acidosis). This might be due to diabetes-related ketoacidosis, alcohol abuse, or severe respiratory alkalosis.
- The phosphate is binding to something else, such as magnesium or sodium, which happens during electrolyte disorders. This is especially common in patients with chronic kidney diseases as they take phosphate binders.

Both calcium and phosphate prevent hypocalcemic tetany, a condition similar to tetanus, in which the involuntary contraction of muscles leads to spasms and cramps.

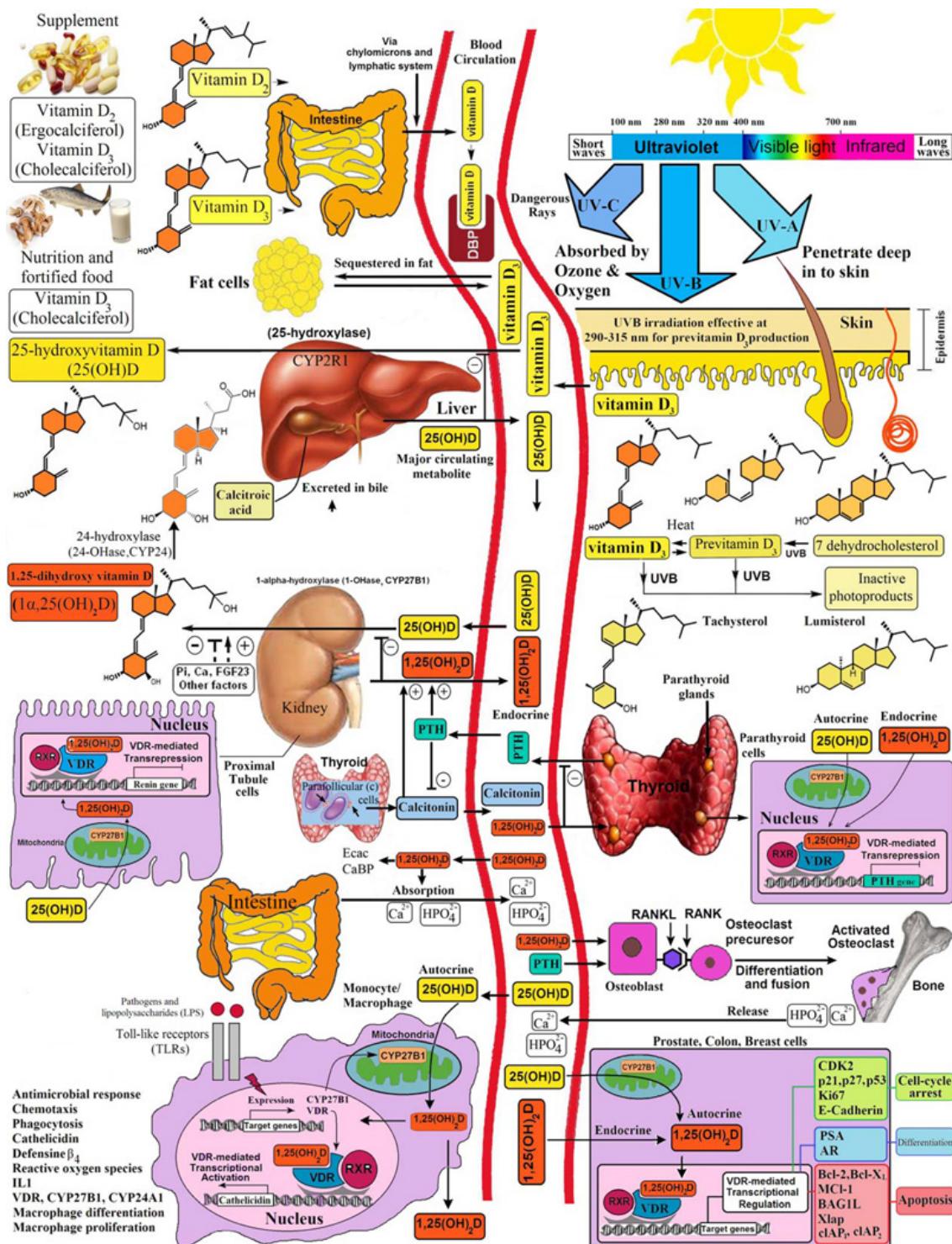


Figure 3.47: Overview of vitamin D, calcium, and phosphate interactions. Reproduced with permission from: [Vitamin D for Health: A Global Perspective](#)

3.6.6 Deficiency and toxicity

Vitamin D screening has become more common in primary health care. Previously we stated that sunlight is the method by which most people acquire vitamin D, however, that does not mean that people take enough vitamin D either via Sun or dietary sources. In table 3.7 we can see the prevalence of vitamin D deficiency around the world.

Table 3.7: Vitamin D deficiency ($25(\text{OH})\text{D} < 50\text{nmol/L}$) prevalence around the world.

Region	Prevalence (%)
Europe [194]	40
– Greece [195]	62
– Germany [195]	44
– Netherlands [195]	29
– Ireland [195]	27
– UK [195]	57
– Denmark [195]	37
– Finland (foreigners) [195]	64
– Finland (nationals) [195]	7
– Iceland [195]	34
– Norway [195]	28
— Tromsø [195]	19
USA [194]	24
Canada [194]	37
Asia [196]	58
Australia [195]	23
Africa [197]	34

Different studies use different population backgrounds, testing methods, and times of the year of sampling. This table has been standardized as well as possible for easy comparison following these simplification rules:

- Vitamin D deficiency is defined as $25(\text{OH})\text{D} < 50\text{nmol/L}$.
- Standardized $25(\text{OH})\text{D}$ levels based on Vitamin D Standardization Program (VDSP) calibration before other tests, if possible.
- Summer levels over winter levels (and the other way around in the southern hemisphere)
- Adult population ($y>18$)

Ironically, places with high solar irradiance show the highest vitamin D deficiency

among their population. One possible explanation for this could be that it is common for people to go to work at 08:00 and stay there until 16:00 while driving back and forth to their destination. Because office windows and car windshields block UVB, the person might have the feeling that he has been exposed to the Sun all day, but in reality, no UVB has been absorbed. For the rest of the day, the person only has a few hours more of the Sun at its lowest elevation over the horizon. For most cities (where the majority of the population resides), this also means that tall buildings are blocking the Sun, leaving you only a sea of shadows to navigate. Dietary changes over the last years favoring junk food culture didn't help either.

Vitamin D deficiency can be due to any insufficient intake of food or Sun as described so far. But even if you take enough vitamin D, there might be other factors that prevent its absorption. Due to vitamin D being fat soluble, anything that hinders the gut's ability to absorb fat such as liver disease, cystic fibrosis, celiac disease, Crohn's disease, and ulcerative colitis will lower your vitamin D absorption [198]. Some of these diseases can also prevent not only from absorbing vitamin D but even from eating good sources of vitamin D in the first place.

Any medication or drug that interferes with normal liver function, in particular medicines to treat seizures (Phenytoin) causes vitamin D deficiency by interfering with the transformation to 25-hydroxyvitamin D in the liver. [199–201].

Obese people don't have a lower capacity to synthesize vitamin D via skin and sun, but the adipose tissue will sequester more of vitamin D. More weight also means more cells that demand more calcium so more need for vitamin D. Furthermore, due to obesity bone stress is higher (more calcium) and muscle strain is also higher (calcium pumps overworking). Overall, evidence suggests that obese people require higher vitamin D [202, 203]. Finally, obese people can have bariatric surgery, thus removing part of the digestive tract in charge of absorbing fats and thus decreasing vitamin D absorption [204, 205].

Hypervitaminosis D is almost exclusively occurring due to high intake of dietary supplements [206–208]. Activation of vitamin D₃ in the skin gives rise to various non-vitamin D forms that limit the formation of vitamin D₃. This is why toxicity due to sun exposure is extremely rare (other conditions can still occur such as first and second-degree burns, skin aging, cancer, blindness, etc...). Tanning saloons have been linked to causing hypervitaminosis D due to UVB radiation. [209–211]

The upper limit of vitamin D toxicity can change depending on your maximum intake of calcium. As vitamin D increases the calcium absorption in the gastrointestinal tract, if you take a lot of calcium, this would lead of course to an increase in absorption, developing symptoms of hypercalcemia [206]. Hypercalcemia due to hypervitaminosis D works the same as regular hypercalcemia and can develop symptoms of nausea, vomiting, muscle weakness, neuropsychiatric disturbances, pain, loss of appetite, dehydration, polyuria, excessive thirst, and kidney stones [212–215]. Continuous hypercalcemia can develop for the worse and cause renal failure, and calcification of soft tissue (coronary vessels or heart valves, which by themselves will develop arrhythmias and death).

Other types of harmful vitamin or mineral level events might occur with medication interaction. This includes any medication that reduces fat absorption due to vitamin D being fat soluble, hence not being able to circulate in blood [216–219]. Any medication that modifies cholesterol because pre-vitamin D₃ is derived from cholesterol via sun exposure [211, 219–221]. Any medication that reduces calcium absorption or excretion can modify vitamin D metabolism; this is quite common with both steroids and anti-inflammatories and diuretics (less calcium leaving your body means more calcium inside your body leading to hypercalcemia or hyperparathyroidism) [219, 222–227].

With all that said, different endocrine societies, from different countries, have different RDAs and different upper limit levels, which also change with time as protocols get updated [228]. In table 3.8 we can see the recommended intake for a healthy individual, sex-independent, pregnancy-independent, and lactation status-independent, which tries to summarize all of these different opinions.

Table 3.8: Vitamin D daily recommendation.

Age	RDAs	Upper limits
0-6 months	10 mcg (400 IU)	25 mcg (1000 IU)
7-12 months	10 mcg (400 IU)	38 mcg (1500 IU)
1-3 years	15 mcg (600 IU)	63 mcg (2500 IU)
4-8 years	15 mcg (600 IU)	75 mcg (3000 IU)
9-70 years	15 mcg (600 IU)	100 mcg (4000 IU)
70+ years	20 mcg (800 IU)	100 mcg (4000 IU)

3.6.7 Epidemiology

Vitamin D has been related to several diseases, such as bone-related diseases, cancer, cardiovascular diseases, depression, multiple sclerosis, diabetes, and obesity. It's also linked to the modulation of inflammatory processes and testosterone production. However, except for bone-related diseases where there is somewhat strong consensus, there is contradictory evidence between vitamin D levels and the rest of the health issues. This might be due 25(OH)D tests not being properly standardized and poor experiment design [195, 229, 230].

Bone health

A vitamin D deficiency will lead to a lower amount of calcium and phosphate in the blood. As a result, the main problem with vitamin D deficiency is conditions such as abnormal bone growth, rickets, brittle bone diseases, osteomalacia, or osteoporosis [231].

Rickets is a rare disease in developed countries that increased significantly from 10% to 70% in places in Africa, the Middle East, and Asia. [232] And has a higher prevalence in a recent year than in the past, up until the introduction of vitamin D and Calcium supplements in infant food [232–234]; a possible explanation for this is the introduction of the immigrant population in first world countries who have different vitamin D absorption rates at an even higher latitude and differences in dietary preferences. Most forms of rickets are due to vitamin D deficiency in children, while severe rickets causes developmental delay, hypocalcemic seizures, tetanic spasms, cardiomyopathy, and dental abnormalities [233, 234]. Almost all patients with rickets had been partially or exclusively breastfed [233, 235]. Osteomalacia is the same as rickets, but we use rickets for children and osteomalacia for adults. There are different definitions and diagnostic criteria to define osteomalacia, but prevalence is about 3.7% with a higher rate in women [236].

Bones are constantly being destroyed and built again. However, as people grow older, the ratio in which they are destroyed is higher than being built again. Over time, bone density plummets, and osteoporosis develop. Osteoporosis is characterized by the deterioration of bone tissue leading to bone fractures. About 15% of people older than 50 and 70% older than 80 suffer osteoporosis worldwide [237]. Osteoporosis is caused by the lack of calcium intake, while osteomalacia and rickets are caused by the lack of vitamin D intake; however, a lack of vitamin D intake contributes to the severity of osteoporosis [191].

For the general older adult population, some studies show that vitamin D and calcium supplementation increase slightly bone density, display reduce fracture rates. [191, 238–240], while other studies show no difference [240–247] .

For the United States, bone density, mass, and fracture risk are correlated with 25OHD in white individuals and Mexicans but not for black individuals [248–250].

Muscle weakness

Vitamin D assists in the development of muscle fibers. Proper support of the bone structure is needed for optimal bone health, so indirectly, vitamin D also helps bone development not only via calcium absorption but also due to muscle growth. Inadequate levels of vitamin D can lead to myopathy (muscle weakness). Sadly, experiments with vitamin D supplements also include calcium supplements. As such, being able to differentiate between the advantages of calcium and the advantages of vitamin D independently is challenging.

Studies are also not standardized in the amount of nutrients or the time in which they are administered; this is important as nutrient absorption varies with the time of the day and with the food that you are eating. For example, calcium inhibits iron absorption (do not eat red meat with dairy products), while vitamin C enhances iron absorption [251, 251, 252] (for example, having orange juice shortly before or after red meat consumption is recommended). Another classic example is that proteins are better absorbed with carbohydrates as insulin enhances amino acid transport into muscle cells and prevents amino acid conversion into energy. Or, as discussed before, vitamins can be water or fat-soluble, and each group requires a different food setup to maximize absorption; all of these variables are rarely taken into account during the intake of supplements in studies.

For the general population, studies have shown either inconsistencies [253] on the effects of vitamin D supplementation with respect the muscle strength and muscle decline, or no correlation [254, 255].

Cancer

Some studies suggest that vitamin D inhibits cancer formation (carcinogenesis) and slows tumor growth. They also suggest that it has an anti-inflammatory effect, can modulate the immune system, triggers cells to self-destruct (this is good, proapoptotic),

and destroys or interferes with the fine network of blood vessels needed by tumors to grow and metastasize (antiangiogenic) [191, 256].

Observational studies show a correlation between low 25OHD and cancer mortality [240, 257, 258]. Some clinical trials support the hypotheses of these observational studies [259–261]. And particular clinical trial support that vitamin D supplementation in a general population delays cancer appearance by 5.3 years in median [262]. Others suggest no correlation effect [240, 263]. In both cases, there is confusing wording on the effect of vitamin D or calcium supplementation, suggesting that individuals with low 25OHD have an increased risk of cancer mortality that is fixed by supplementation, while individuals with normal 25(OH)D levels have no increased risk and thus the supplementation is useless because no association is linked between supplementation and cancer mortality. This makes confusing the effect of supplementation and mortality because it is not the supplementation itself that reduces the cancer mortality, but having appropriate 25(OH)D levels. However, we have already established that low levels are cured by the use of supplementation provided that you don't suffer from any of the co-morbidities that prevent 25(OH)D absorption.

For particular cancer types:

- **Breast cancer:** contradictory evidence, from inverse levels 25(OH)D and mortality to the opposite, passing through no correlation. [264–270].
- **Lung cancer:** No association between circulating concentrations of vitamin D and risk of lung cancer [271].
- **Pancreatic cancer:** No correlation and positive levels of 25(OH)D associated with higher cancer risk [272–274].
- **Colorectal cancer:** Inverse levels 25(OH)D and mortality especially in women, no correlation, and positive correlation between vitamin D supplements and calcium supplements with the development of polyps [264, 275–277].
- **Prostate cancer:** Contradictory evidence between 25(OH)D levels (all) and cancer risk, mortality, and length [278–286].

Cardiovascular diseases

In the context of Cardiovascular diseases (CVD), Vitamin D regulates the Renin-Angiotensin-Aldosterone system (RAAS) [287], this is a hormone system that regulates blood pressure, systematic vascular resistance (how much your blood pressure needs to be in order to flow), electrolyte balance and fluid. Also regulates vascular cell growth, fibrotic pathways (scarring), and inflammatory pathways. Indirectly, vitamin D enhances the absorption of calcium and potassium, which are both elemental pieces in the cardiac action potential for both pacemaker cells and contractile cells, especially the latter as the absolute refractory in contractile cells is essential to prevent tetanus (shorter absolute refractory equals fast heart rate).

Vitamin D deficiency has been shown to be associated with vascular dysfunction, arterial stiffening left ventricular hypertrophy, and hyperlipidemia (too much fat in the blood) [288]. As such, observational studies have linked vitamin D with lower CVD risk.

Observational studies have found a positive association between high 25(OH)D and lower CVD events (myocardial infarction, ischemic heart disease, heart failure, and stroke) and mortality [289]. There is also a study that shows an association between high and low 25(OH)D levels and CVD [290]. And other studies found a correlation between low 25(OH)D and high CVD [291, 292].

However, clinical trials contradict everything in the previous paragraph [240, 262, 293], with some studies suggesting that it protects against cardiac failure, but not against myocardial infarction or stroke [294].

Supplements have also been shown to reduce total cholesterol and Low-Density Lipoprotein (LDL), but not High-Density Lipoprotein (HDL) [295], has a contradictory effect on blood pressure for normal weight patients [296, 297], and when taken with calcium increase blood pressure in overweight and obese patients while having low 25(OH)D also increase blood pressure [296, 298].

Depression and other mental illnesses

Vitamin D receptors are present in several areas of the brain, and it is believed to be involved with depression, dementia, schizophrenic-like disorders, hypoxic brain injury, and other mental illnesses; as well as neuronal development and a decrease of microglial inflammatory function. [299–303].

But once again, clinical trials found that the administration of vitamin D supplements were not linked with the reduction of depressive symptoms [304–308]. Worth mentioning that none of these studies tested the combination of vitamin D supplements, plus antidepressants, in individuals with low 25OHD.

Multiple sclerosis

Multiple sclerosis (MS) is a chronic disease of autoimmunity or oligodendroglialopathy origin [309] where an inflammation of the cover of nerve cells in the brain and spinal cord disrupts the ability of the nervous system to transmit signals properly. As a result, it can present almost any neurological symptoms. MS occurs less frequently near the equator and more frequently near the poles [310]. This led researchers to investigate whether sun exposure has anything to do with MS, which is of course related to vitamin D absorption [240].

Studies have shown the presence of low 25(OH)D before and after MS begin [310–313]. Others show that normal 25(OH)D levels reduce the risk of contracting MS, and decrease the time in between relapses once it has started [314]. Clinical trials once more, contradict these findings [310, 315].

Diabetes and glucose homeostasis

The human body needs to maintain a certain level of sugar in blood constantly, too much sugar causes nerve damage, and too little sugar causes death; this process is known as glucose homeostasis. When too much sugar is present, the body releases a hormone known as insulin which transforms sugar into glycogen and is stored in the liver and muscles (about 100 gr and 300 gr respectively). When too little sugar is present, the body releases glucagon, a hormone that does the opposite to insulin and transforms glycogen back into sugar.

If the liver stores too much glycogen, then it will start transforming sugar into fat instead (lipogenesis). The opposite pathway, as in converting fat back into sugar or energy (lipolysis), doesn't start until two conditions are met. First, the liver needs to consume all the stored glycogen (takes about a day), and second, the alanine-lactate acid accumulating in the muscles has also been converted into pyruvate and into glucose (gluconeogenesis, another day). This last stage also generates too much ammonia which is toxic. So far, two days have passed and the body has not started to burn any significant fat. This is why short-term starvation and fasting to burn fat is useless, and on top of

that not healthy for the body. After two days the body starts burning fat for about a week, and after that, it will start breaking down proteins from your muscles and organs until death. The other more healthy way to burn fat is to have high epinephrine and low insulin in the blood, which is achieved during exercise; the highest the intensity the more epinephrine is released, but mid to low-level exercise can also promote fat burning as well, although takes more time.

With normal levels of insulin, and some of the previously described pathways did not happen as they should, it might be due to insulin resistance, where the body needs more insulin than normal for all of this to work properly; normally caused by obesity and not exercising, sometimes leading to the need of injecting external insulin due type 2 diabetes. Vitamin D stimulates the secretion of insulin via VDR on the pancreatic beta cells, which reduces insulin resistance, and it is also involved in the insulin signaling pathways [316–318]. Some studies have shown a relationship between vitamin D and the risk of diabetes [191], and others have shown an inverse correlation between vitamin D and blood sugar [319].

Some clinical trials tell us that all of this seems to be incorrect and there is no relationship. Including vitamin D not improving insulin sensitivity in overweight and obese patients [320], supplements not having an effect on glucose homeostasis or insulin resistance [240, 321], no prevention of going from pre-diabetes to diabetes [322], or overweight or obese with normal 25(OH)D levels but true for those with lower 25(OH)D levels [316, 323], and no benefits on individuals that already have diabetes [318].

Obesity

As vitamin D is fat-soluble, more fat deposits in the body will sequester more vitamin D. Observational studies suggest that weight has an inverse correlation with 25(OH)D levels [202, 203, 324, 325], and even reduce weight gain in postmenopausal women [326] but vitamin D supplementation doesn't help with weight loss [327].

Testosterone and male reproductive organs

It has been observed that vitamin D supplements can increase testosterone in males [328], and secondarily increase fertility due to calcium absorption. [191, 193, 329]. It also has been observed that, while vitamin D has an essential role in the function of testicles and the prostate [330], supplements it do not increase testosterone in healthy males [330, 331], or in males with advance heart failure [332].

3.7 Over-the-counter medicine

3.7.1 Introduction

Over-the-counter (OTC) medicine [333] refers to a type of medication that is sold without the need for a prescription and can be acquired even outside of pharmaceutical outlets such as supermarkets. These types of medicine have a low dosage of the active agent and are used to alleviate common symptoms such as pain, fever, cough, allergies, and digestive issues.

The OTC medicine market is substantial. Worldwide, there was an estimated spend of 3.68 billion GBP (adjusted) between 2013 to 2019 on codeine-based products alone [334]. In the US in 2015, there was a spend of 32 billion USD on OTCs with an average of nearly 350 USD per household per year [335]. Worth mentioning that there are specific marketing strategies to make OTC appealing to children under 6, with all products indicating in different typography the type flavoring, having cases such as 5% of the cough syrups being chocolate flavored [335]. OTC are fairly safe when used in moderation and responsibly, but the accessibility and appeal of these medications also open the door to a variety of potential misuse; for example, from 2010 to 2013, 43% of the emergency visits in children under 6 related to medicine exposure were due to OTC overdose [335]. In recent years, the effectiveness of OTCs for the general public has been questioned [336] due to the increase in their misuse. About 16% of the population misuse these medicines, about 2% abuse them, and up to 7% recognize addiction [333,337–339]. In more extreme cases, other in theory non-OTC medicines abuse, such as antibiotics, becoming de-facto OTC has led Greece to have one of the most antimicrobial resistance rates in Europe [340].

Common reasons for its use are the false belief that if taken more the symptoms will go away faster, or to the contrary, for self-harming purposes [333]. In Norway [341], teenagers report using these drugs mostly for physical pain, to alleviate stress and fatigue, due to familiar conflicts, and to appear socially successful; while the common user is a person with constant pain, binge drinking, having school problems, bad sleeping habits, lower life ambitions and high spare time. About 55% of OTC medicines are sold in non-pharmaceutical outlets [341]. A study in the Chinese market [342] showed that despite the increase in marketing budget expended by pharmaceutical, the inclusion, and diffusion of new OTCs is very slow and that the mouth-to-mouth recommendation of peers seems to be much more effective.

3.7.2 Principles and direct adverse effects

Analgesics

Analgesics are a class of drugs commonly known as painkillers. They reduce or block the perception of pain signals in the brain. NSAIDs usually overlap their effect with analgesic medicine but will be discussed in the antiinflammatories section.

The most common active component of analgesics is paracetamol (acetaminophen) [343]. The majority of acetaminophen is metabolized safely by conjugation with glucuronic acid (glucuronidation) and sulfate (sulfation) in the liver. These reactions produce non-toxic, water-soluble metabolites that are easily excreted by the kidneys.

About 2% of it transforms into N-acetyl-p-benzoquinone imine (NAPQI), which remains in the liver for a longer time before it also moves into the kidneys. NAPQI is toxic. When taken in a low amount, NAPQI is quickly detoxified in the liver by conjugation with glutathione, a powerful antioxidant found in the liver. The resulting non-toxic compounds are then excreted via the kidneys. Problems arise when analgesics are taken in higher doses or at a higher frequency, the liver may become overwhelmed and unable to detoxify all of the NAPQI. When glutathione is depleted, NAPQI accumulates in the liver, binds to cellular proteins, and causes oxidative stress which leads to hepatocyte (liver cell) death and can result in acute liver failure.

In contrast, non-OTC analgesics are usually opioids such as morphine. They work by binding to specific receptors in the brain and spinal cord called opioid receptors [344]. Activating these receptors blocks the transmission of the most severe pain signals. However, they are at risk of addiction and their use is severely restricted. Analgesics are not used as recreational drugs directly but are used in combination with other substances such as alcohol or cannabinoids [345].

Antihistamines

Antihistamines are a class of medications that are commonly used to treat allergic rhinitis and allergic-like symptoms. They work by blocking histamine, hence their name. Histamine is released when allergens bind to mast-cell-bound Immunoglobulin E (IgE) antibody sites [346]. This is commonly known as an allergy and is the mechanism of bronchial smooth muscle contraction, urinary bladder contractions, vasodilation, visceral hypersensitivity, itch perception, urticaria, sneezing, hyper-secretion from glandular tissue, and nasal congestion due to vascular engorgement.

Different antihistamine medications can cause both vasoconstriction and vasodilation. In general, vasodilation may be more effective at relieving symptoms such as congestion and mucus production, while vasoconstriction may be more effective at reducing inflammation and swelling.

Histamine binds to histamine receptors, of which there are four types, but the two most relevant to antihistamine medications are H1 and H2 receptor blockers. First-generation H1 antihistamines can block the neurotransmitter acetylcholine. Acetylcholine, among its many functions, plays a role in the constriction of blood vessels. When acetylcholine's action is blocked, this can lead to a relative decrease in vascular tone (the tension of blood vessel walls), which may result in vasodilation. Second-generation H1 antihistamines have a much lower affinity for cholinergic receptors and therefore have fewer anticholinergic (or sedative) effects. H2 Antihistamines are commonly used to treat excess stomach acid. H2 receptors are found in the stomach lining and are responsible for stimulating acid secretion. However, H2 receptors are also present in the blood vessels and contribute to vasodilation. By blocking H2 receptors, H2 antihistamines can reduce acid secretion and also decrease vasodilation, thus increasing vasoconstriction.

Both cases can be detrimental. Vasoconstriction can increase blood pressure in the heart and reduce blood flow in the kidneys. For patients with decreased kidney functionality or hypertension, this would be dangerous. Vasodilation increases blood flow, which is generally beneficial for normal kidney function. However, it is detrimental in cases when the patient is incapable of filtering waste, and excess fluid from the blood would lead to a buildup of toxins and fluid in the body. Lastly, worth mentioning that the first generation of antihistamines developed during the 1930s have a detrimental flaw and they tend to cross the brain-blood barrier [347]. The blood-brain barrier is a specialized system of blood vessels that helps to protect the brain from harmful substances and toxins.

The effects of drug abuse related to antihistamines have a huge variety, but as a general rule, they are used as vasodilators, which promote a calming and sedating effect. They can enhance the effect of other substances such as making opioids more hallucinogenic, especially the first generation [345].

Antiinflammatories

NSAIDs drugs such as ibuprofen work by inhibiting the production of prostaglandins. Prostaglandins cause blood vessels to dilate, which can increase blood flow to the affected

area and cause redness and swelling. They also sensitize nerve endings to pain, which can cause pain and discomfort. NSAIDs work by inhibiting the activity of an enzyme called cyclooxygenase (COX), (subdivided into COX1 and COX2) which is responsible for the production of PG [348]. These are discussed in section 3.4.3; about how the inflammation process needs to pass from a pro-inflammatory state into an anti-inflammatory state.

NSAIDs can be divided broadly into three categories, COX1 inhibitors, COX2 inhibitors, and both COX1 and COX2 inhibitors. Prostaglandins PGE2 and PGI2 participate in the synthesis of protective mucus and gastric flow, which is why a normal side effect of COX1 inhibitors is gastrointestinal bleeding and ulcers [348]. COX1 also participates in the production of thromboxane which promotes platelet aggregation, which is why antiinflammatories such as aspirins can cause an increase in bleeding. On the other hand, prostaglandins PGI2 and PGH2 are vasodilators and share COX precursors with thromboxane which is also a vasoconstrictor. Simply put, there must be an equilibrium between COX1 (vasoconstrictor) and COX2 (vasodilator). NSAIDs that inhibit COX2 increase blood pressure which can lead to heart infarction [348]. These side effects can be mitigated with the usage of prostaglandin analog medicines.

Due to their vasoconstriction nature, these medicines can be used as stimulants, which can also induce psychotic symptoms, paranoia, and visual hallucinations. [345]

Cough syrups

Cough syrups are used to stop unwanted coughing which may cause discomfort, or even physical injuries, in the upper respiratory tract. Their mechanism of action is not fully understood but their effects are accomplished by reducing the signaling between the laryngeal nerves and vagus nerve [349].

The three main antitussive components are codeine, Dextromethorphan (DXM), and benzonatate [350]. Codeine breaks down into codeine-6-glucuronide and morphine, which stimulates the μ -opioid receptors [344] causing euphoria, constipation, and also cough suppression. DXM is the principal component in OTC medicines, and it has similar side effects of codeine but in the form of hallucinogenics, and in particular as a dissociative. Benzonatate is a non-narcotic drug that works as a local anesthetic in the whole respiratory tract.

Due to their action on the peripheral nervous system, the main side effect of these drugs is reduced consciousness in the form of drowsiness. In higher dosages, it can

lead to hallucinations, paranoia, perceptual distortions, delusional beliefs, ataxia, and out-of-body experiences [345].

Laxatives

The main component of these drugs is loperamide. Loperamide works by binding to peripheral μ -opioid receptors in the gastrointestinal tract [351]. It inhibits the release of acetylcholine and other neurotransmitters that stimulate the contractions of the intestinal wall. This leads to a reduction in peristalsis, or the wave-like contractions of the smooth muscle in the intestinal wall which slows down the passage of stool and reduces the frequency and urgency of bowel movements. This also allows for the gastrointestinal tract to have more time absorbing fluids, increasing the hardness of the fecal matter.

Similar to antitussive medications, antidiarrheals relax the peripheral nervous system. Thus, these drugs are used also as opioids, to alleviate symptoms of opioid withdrawal, or as a psychoactive. [345]

3.7.3 Adverse drug interactions

Anticoagulant

A classical example of adverse drug interaction is Warfarin. Warfarin is an anticoagulant medication that is used to prevent blood clots from forming or from growing larger. NSAIDs and aspirin also have an anticoagulant effect by inhibiting platelet aggregation, although through different mechanisms than warfarin.

Taking warfarin with these medications is discouraged because their additive anti-coagulant effect can lead to a significantly increased risk of bleeding because both the platelet function and the clotting factor production are impaired. This includes gastrointestinal bleeding which can be caused by NSAID irritation in the gut, and exacerbated by warfarin.

Antidepressants

Monoamine oxidase inhibitors (MAOIs) are a class of antidepressant drugs that work by inhibiting the activity of one or both monoamine oxidase enzymes (MAO-A and MAO-B) [352]. These enzymes are responsible for breaking down neurotransmitters such as serotonin in the brain. By inhibiting these enzymes, MAOIs increase the levels of these neurotransmitters, which can help improve mood and reduce symptoms of depression.

Many over-the-counter cold and cough medications contain sympathomimetic amines, such as pseudoephedrine, phenylephrine, or ephedrine. These substances act as decongestants by constricting blood vessels in the nasal passages. When taken with MAOIs, the breakdown of these sympathomimetic amines is inhibited, leading to their increased levels and prolonged action. The danger from this is an increase, or longer, vasoconstriction spike.

Furthermore, dextromethorphan, a cough syrup, can increase serotonin levels. When combined with MAOIs, the risk of accumulation of serotonin in the brain is increased; which may lead to confusion, agitation, muscle twitching, or sweating; an effect often sought if someone tries to use misused cough syrup as a recreational drug, but which can derive into seizures, irregular heartbeat, and unconsciousness.

Blood pressure regulators

As seen in the introduction, NSAIDs of the COX2 can either nullify or enhance other blood pressure medications leading to heart infartation.

Other non-OTC supplements

Saint John's Worth is a popular herbal remedy sold in parapharmacies that has a moderate effect as an antidepressant [353]. However, what at first glance is just an innocent tea drink, interacts severely with many medications by modifying the proper xenobiotic pathways in the liver (mainly CYP3A4 and CYP2C9 enzymes) [354].

This can lead to death, by suppressing anticoagulants, such as warfarin mentioned above, anti-HIV medication, and Digoxin which is a medication used to threaten mainly atrial fibrillation in the heart. But can also have other detrimental side effects such as making other medications stop working, such as hormonal contraceptives which lead to pregnancy, Xanax which leads to anxiety and panic attacks, or several antidepressants which may lead to increased depressive disorders.

3.7.4 Other ethical considerations

Legal framework

There are plenty of different regulations when it comes to OTCs [355]. Within the EU alone, 16 of the 30 European countries allow selling OTCs in non-pharmacy outlets; in 13 countries, the dosage is restricted and the medicine is secure in lockers, and

in others purchasing them online without restriction is allowed. Greece or Lithuania have total freedom of sales, whereas Hungary or Poland only allows pharmacies, and the pharmacies must be owned by, at least 51%, the pharmacist and not by a big corporation. Pharmacies in Denmark, Luxembourg, Slovenia, and Finland are owned by the state. In Iceland and Norway, there was a deregulation of the ownership to promote competition, but now there is a de-facto oligopoly where two and three pharmacy groups control about 90% of the market. In Japan, pharmaceutical access is supposed to control the abuse of individuals and ensure that purchases are done in moderation but this regulation is rarely followed by pharmacies [356]; and in many other places, while high dosages such as Ibuprophen 1000mg is restricted, nothing keep a person from taking 5 times a dosage of 200mg instead.

Right to Health

In the previous paragraph, we can see that there is not a unique approach to these medicines. On the one hand, people should be able to purchase cheap medicines whether they need to, but this leads to misuse or addiction. On the other hand, harsh regulation would prevent people's access to medicine or strain the medical system with an unnecessary burden. If the market is liberated, manufacturers tend to market their product as seen in the introduction, and rarely do they put effort into educating the public on the possible long-term detrimental side effects [356].

Education

One can think that education is the key to preventing the public from misuse of OTC. In a questionnaire done to 133 2nd year medical students of a tertiary care hospitals: "*Most of them did not knew about the contraindications of the OTC drugs and only a few knew about the potential drug-food interactions*" [357].

The OTC-SOCIOMED initiative (<http://www.otcsociomed.uoc.gr/joomla/index.php/key-findings>) concluded [358] that interventions in GPs prevent overprescribing OTCs and that sale points should be limited to pharmacies; which was further recommended to Greece in particular because, as seen above, they allowed a total liberalization. In contrast, the Congress in the US passed a reform act [359] that: "*allow drug companies a more efficient pathway to marketing OTC products, making OTC drug development more attractive*", and to prevent misuse pharmaceutical companies should create a database to monitor proper use, but at the same time the act recognize that these companies "*...more likely to accede to industry goals such as rushed reviews and*

lower standards for approval when it is financially linked to industry. Such organizations cite to reports of increased pressure to meet deadlines and anecdotal evidence of drugs being approved without sufficient scientific support."

Nurses in school settings suggest proper nurse training, alternative methods such as cold / heat for inflammation and pain, healthy habits, and more importantly, not using OTCs so children stop complaining and can be returned to class quickly. [360]

No literature was found regarding the effectiveness of packaging regulation like what is done in tobacco products, such as limiting colors, font size, displaying size effect, and the like.

Chapter 4: Methodology

4.1 Fit Futures

The Fit Futures (FF) study [361] is a cohort with repeated health surveys among high school students in the Norwegian municipality of Tromsø and the neighboring municipality of Balsfjord. All first-year high school students in Tromsø and Balsfjord were invited. This dataset is the main dataset used across all results of the thesis and is divided into FF1 and FF2. FF1 is used in Paper A, Paper B, Result I, Result II, Result III and Result IV; while a subset of FF2 is used in Result III.

FF1 was conducted from September 2010 to May 2011 for 8 months. FF1 included students from eight schools consecutively. A total of 1117 youths were invited 93% attended, 508 girls (48.9%), and 530 boys. The age ranges from 15 to 28 years old, with 822 (79.2%) being 16 years old or younger, and 52 (5%) older than 18 years old. Older students with special educational needs or mental disabilities have the right to study at the high school level in Norway. In figure 4.1 we can see the geographical distribution of the schools and in table 4.1 information specific to each school.

Fit Futures 2 (FF2) is a follow-up survey conducted from November 2012 to June 2013. FF2 invited all participants in FF1 and all new students from the third year at the eight high schools. Altogether 870 high school students were recruited in the FF2 study, and 78% of these attended both surveys. FF2 included the same variables present in FF1, but we did not have access to the complete FF2 data, only the FF2 anthropometrical data. In FF2, 694 students (66.9% of the FF1 total) completed the anthropomorphic measurements, 378 girls (54.5% of total participants), and 316 boys. Some students in the vocational training program did not get permission from work to attend the FF2 follow-up measures, which contributed to lowering the total student count in the follow-up study.

In FF1 and FF2, the participants had a one-day visit to The Clinical Research Unit

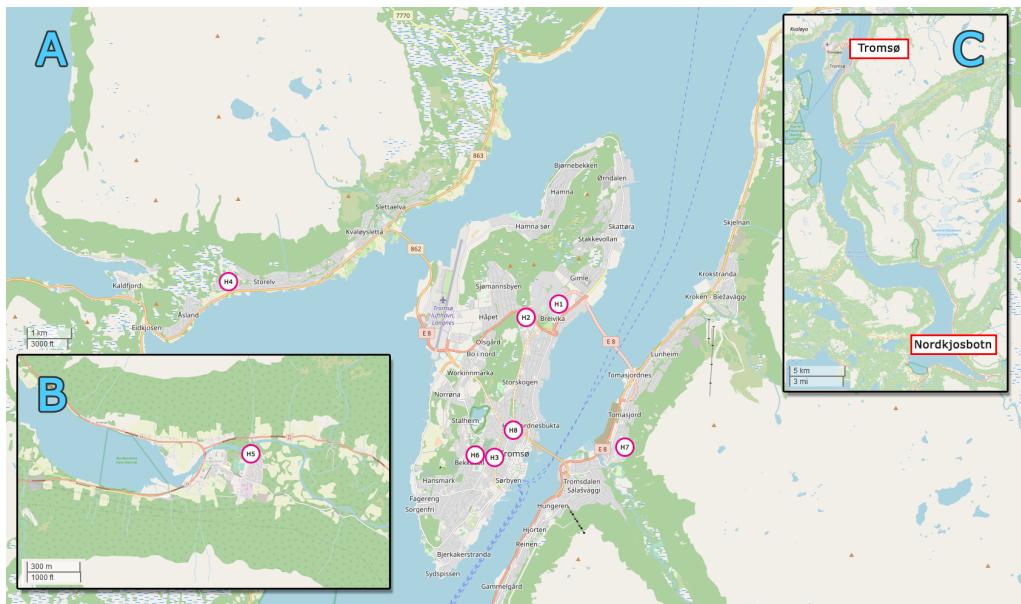


Figure 4.1: Geographical location of all eight high schools included in the Fit Futures study. "A" refers to the area around Tromsøya, and "B" is the area in the town of Nordkjosbotn (Balsfjord). "C" shows the distance between "A" and "B".

at the University Hospital of North Norway (UNN), which included clinical examinations, microbiological samples, blood samples, a web-based general questionnaire (available in the appendix, chapter 7), and an interview [362]. All procedures were performed by trained research study nurses.

FF11 and FF12 refer to short follow-ups performed during the FF1 period. In these sub-surveys, not all the data was gathered again; only a sub-sample such as the swabbing of the *S. aureus*.

4.1.1 Social network assessment

The social network was constructed based on the following questions in the interview. These were written and answered in Norwegian, here we provide the English translation: "*Which students have you had the most contact with the last week? Name up to 5 students at your own school or other schools in Tromsø and Balsfjord.*". Reciprocity in the nomination was not mandatory. For each of the nominations, five "yes/no" questions assessed the type of contact they had with their nominations: "*Do you have physical contact?*", "*Are you together at school?*", "*Are you together at sports?*", "*Are you together at home?*", "*Are you together at other places?*". This resulted in five social networks: Physical Network, School Network, Sports Network, Home Network, and Other Network. Adding all the relationships together formed a sixth network that

was called the Overall Network. Illustrations of all networks are presented in figure 4.2.

Not 100% of the students participated in the study, and some of the relationship information between them is lost (table 4.2). The final analysis shows that some of the lost 134 IDs, were very popular, with up to 9 friends nominated.

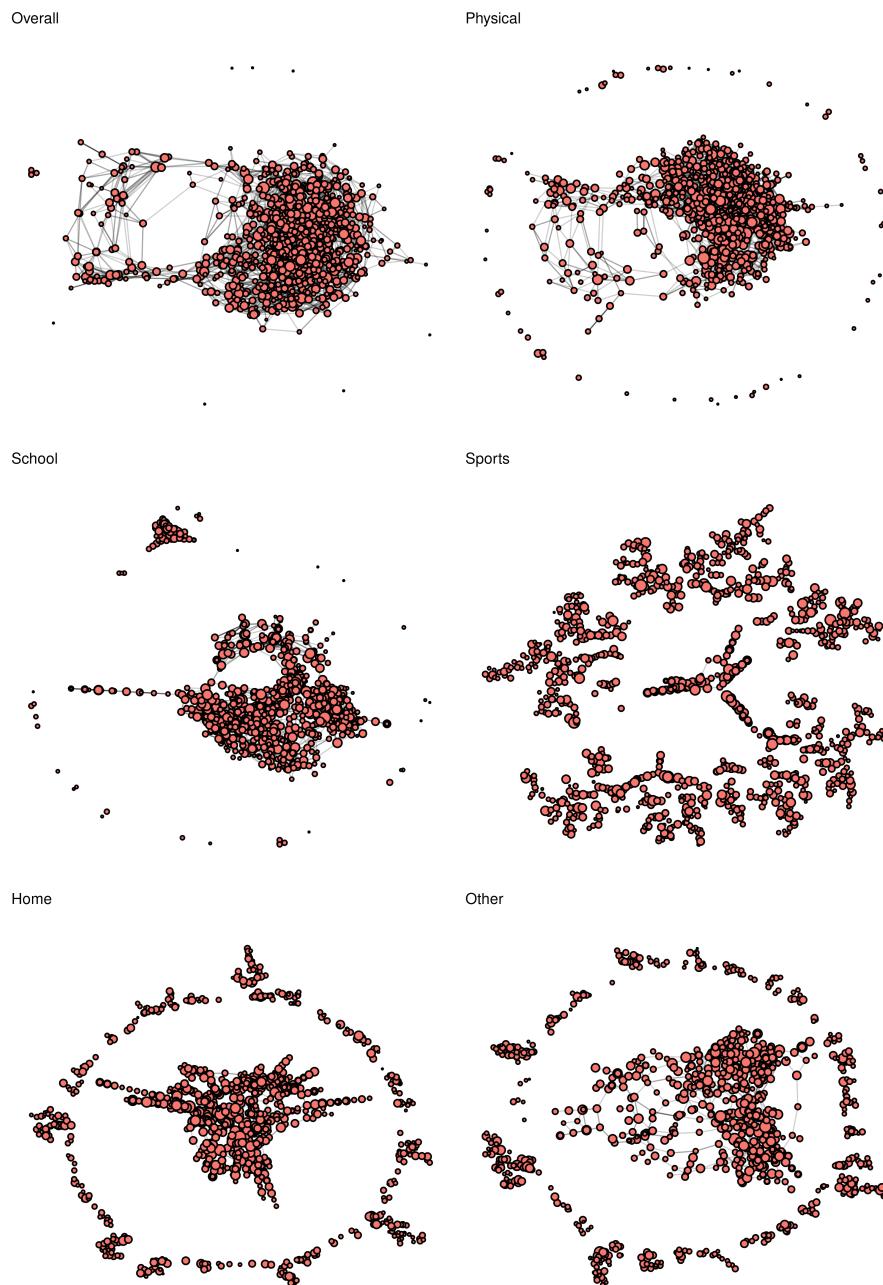


Figure 4.2: All networks in FF1 following a MDS layout. Each student is represented by one node in the network. Each relationship is represented by an undirected edge, i.e., a line, in the network.

Table 4.1: Information with the school names, study program, the total number of students in FF1, astronomical season, and whether a significant amount of students' blood samples were taken during the polar night.

ID	Name	Studies program	FF1 Students	Extraction	Polar night
H1	Breivika videregående skole	Vocational	207	Autumn	No
H2	Breivang videregående skole	Vocational and General	142	Autumn	Yes
H3	Kongsbakken videregående skole	Vocational and General	168	Winter	No
H4	Kvaløya videregående skole	Vocational and General	98	Spring	No
H5	Nordkjosbotn videregående skole	Vocational and General	85	Spring	No
H6	Norges Toppidrettsgymnas Tromsø	Sports	26	Spring	No
H7	Tromsdalen videregående skole	Sports and General	192	Winter	Yes
H8	Tromsø maritime skole	Vocational	120	Winter	Yes

To evaluate if the friends mentioned were representative of the participant's social network, the following question was asked: "*To what degree does this table of friends give an overview of your social network? Please indicate on a scale from 0 (small degree) to 10 (high degree).*" Nominated friends that did not participate in FF1 were excluded from the analysis (n=134). In figure 4.3 we can see a histogram with all the answers.

Table 4.2: Summary of lost connections at data cleaning.

Concept	Total	Relative
Total IDs:	1177	100 %
- Total Deleted IDs:	139	11.81 %
- Total Remaining IDs:	1038	88.19 %
Total Edges:	4125	100 %
- Total Deleted Edges:	473	11.47 %
- Total Remaining Edges:	3652	88.53 %

4.1.2 Host risk factors

All questions related to sex, use of recreational drugs, dietary habits, chronic diseases, medication usage, sport frequency, sedentism, and so on, are self-reported using a web-based questionnaire.

4.1.3 Hormonal contraceptive

Information on current hormonal contraceptive use was obtained from the interview. Hormonal contraceptives (HC) were categorized into combination contraceptives and progestin-only contraceptives. The combination contraceptives were further divided into groups according to high and low ethinylestradiol daily dosage. High dosage was defined as HC containing $\geq 30 \mu\text{g}$ ethinylestradiol. Low dosage was defined as contraceptives

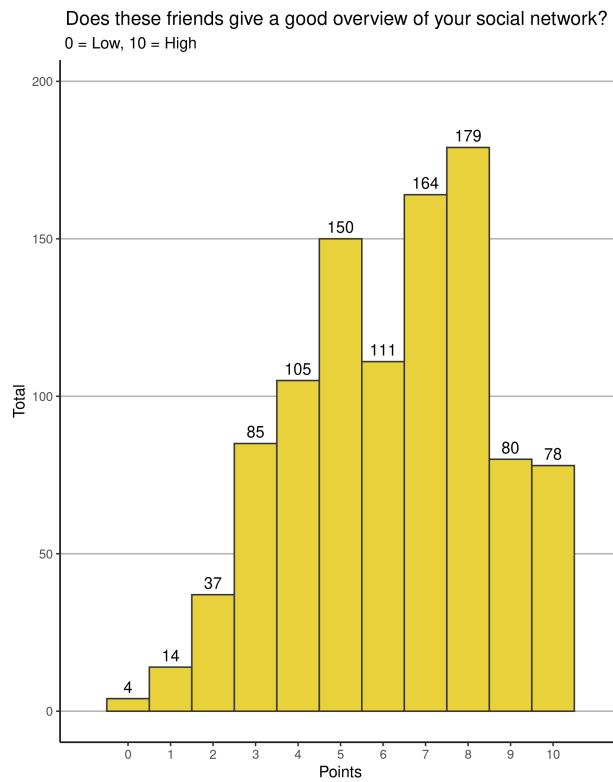


Figure 4.3: Histogram with all the answers to the question: “*To what degree does this table of friends give an overview of your social network? Please indicate on a scale from 0 (small degree) to 10 (high degree).*”

containing $\leq 30 \mu\text{g}$ ethinylestradiol. The classification for each brand can be seen in table 4.3.

Table 4.3: Types of hormonal contraceptives classification and their respective brands.

Hormonal type	Contraceptive brand
Non-hormonal	Condoms
Progestin only	Cerazette Nexplanon Depo-provera Implanon
Low Estradiol	Mercilon Yasminelle Loette 28 Nuvaring
High Estradiol	Marvelon Yasmin Microgynon Oralcon Diane Synfase Evra Zyrona
Unknown	Any other brand/type

4.1.4 *S. aureus* assessment

A first set of nasal and throat swab samples was taken at the research center, and a second set of samples was taken at school after a mean interval of 17 days. All 1038 students were sampled on both occasions, the first batch contained 1028 valid samples, and the second batch 988. A NaCl (0.9%)-moistened sterile rayon-tipped swab rotated three times with gentle pressure was used to sample both vestibule nasi (nose sample), and an additional swab was used to sample both tonsillar regions (throat sample). The swabs were immediately placed in a transport medium (Amies Copan, Brescia, Italy) and stored at 4°C for a maximum of 3 days. All samples were analyzed at the Department of Microbiology and Infection Control, UNN, both by direct culture [363] and enrichment broth (Bacto Staphylococcus medium broth, (Difco Laboratories, Sparks, MD, USA - [364])), using blood agar for growth control (Oxoid, UK) and chromID-plates (SAID) for *S. aureus* detection (bioMérieux, Marcy l'Etoile, France). A summary of these methods can be found in the supplementary materials. The growth of any bacterial colonies on agar plates was registered as a valid culture. The most dominating *S. aureus* colony type was frozen at -70°C in glycerol-containing liquid media after confirmation by Staphaurex plus agglutination test (bioMérieux, Marcy l'Etoile, France).

From these results, *S. aureus* nasal or throat persistent carriage was defined as having two *S. aureus* positive cultures for each niche respectively. [365, 366] Two definitions of *S. aureus* persistent carriage was used in the analysis; one based on direct culture, and one based on enrichment broth. All results for every possible combination between the first or second sample, nasal or throat, direct culture or enrichment broth, and so on, can be found in figure 4.4.

4.1.5 SPA typing

SPA-typing is a technique used to identify the *S. aureus* strain. The gene encoding Spa is highly variable among strains of *S. aureus*, making it a good target for distinguishing different types of bacteria [367]. The technique involves targeting the Spa located in the cell wall are described in the supplementary materials.

For the analysis of *S. aureus* genotype, only data from throat isolates were available ($n = 746$). All *S. aureus* isolates from throat samples were subjected to spa-typing. The frozen cultures were inoculated on blood agar (Oxoid) and incubated overnight at 37°C. Two or three colonies were transferred to 200 µl sterile H₂O and vortexed. The isolates were later spa-typed [368].

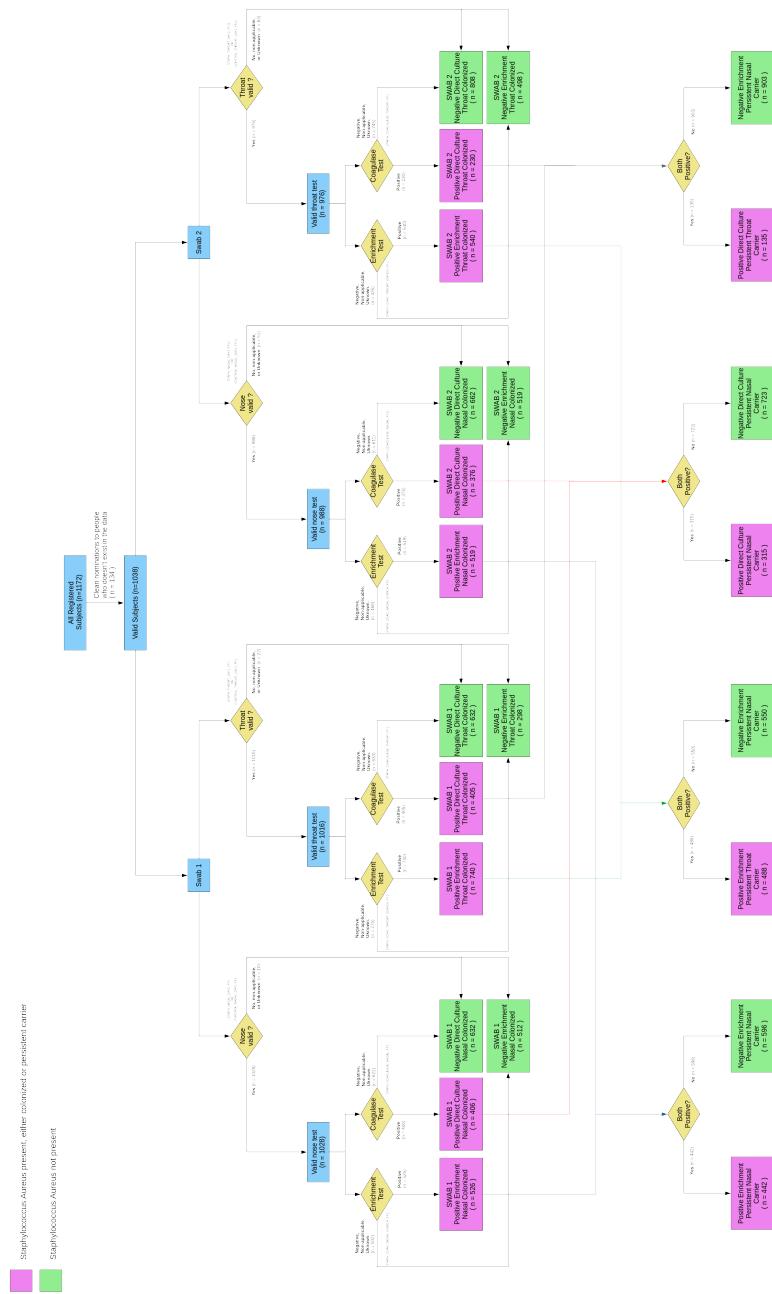


Figure 4.4: All the *S. aureus* combinations. A PDF version of this image can be found at the GitHub repository: <https://github.com/rafanozal/PhDThesis/blob/main/Images/carrierDefinition.pdf>

4.1.6 Anthropometry assessment

All anthropometric measurements were measured on an electronic scale with participants wearing light clothing and no footwear. BMI is calculated as weight (kg) divided by the squared height (m^2) with no correction for sex or age. FF1 has a total of 1034 valid samples, and FF2 has a total of 694.

The BMI comes as a real number and we categorize it following the World Health Organization (WHO) definition [369]. "Underweight" if the BMI is lower than 18.5, "Healthy" is above 18.5 and below 25, "Overweight" is between 25 and 30, and "Obese" is greater than 35. The new column is a categorical variable with this information for each person.

The WHO also provides BMI-for-age growth charts that take into consideration an individual's age and sex to determine their BMI percentile. A BMI percentile below the 5th percentile is considered underweight, while a BMI percentile between the 85th and 94th percentiles is considered overweight, and a percentile above the 95th percentile is considered obese. This definition is NOT used since being a relative respects an average rather than a constant value (i.e.: a student being the less obese of a group of people does not make the student not obese); instead, a constant reference point is used as described in the previous paragraph for better comparison across time.

4.1.7 Vitamin D assessment

Blood samples were collected by nurses at the UNN, centrifugated, and plasma serum was frozen at -70°C in the Biobank at the UiT: The Arctic university of Norway (UiT). All samples ($n = 890$) were sent to the Hormone Laboratory, Haukeland University Hospital, Bergen, Norway; and analyzed by high-pressure liquid chromatography-mass spectroscopy (LC-MS/MS). A sample from all blood vials was reanalyzed at University College Cork, Cork, Ireland, by LC-MS/MS again as a part of the VDSP [370], and standardization was applied to the rest of the samples [371]. 25(OH)D was used as a marker for vitamin D levels. This combines both sources of provitamin D + UVB, and D2+D3 from diet. It has a longer half-life span in blood than other available metabolites. Both 25(OHD)D₂ and 25(OHD)D₃ were measured at the same time.

4.1.8 OLINK Target 96 Inflammation

Serum levels of 92 proteins were analyzed at the Clinical Biomarkers Facility, SciLifeLab, (Uppsala, Sweden), using the Target 96 Inflammation panel from Olink Holding AB (Uppsala, Sweden) [372]. A brief description of the biological role of each marker, alongside the OLINK significance levels for each value, and Bioprot database entry, can be found in the appendix (section C. A total of 936 samples were analyzed this way.

4.2 Data cleaning

This section will present a summary of the important parts of the methodology followed in the data-cleaning process as well as the shortcomings encountered due to the experiment design or the original data-registering process. In the Appendix (chapter 7), some samples from the 25 tables with the original variables' names and a description of what each variable represents are provided, but we do not provide examples from tables that could potentially be used to identify students.

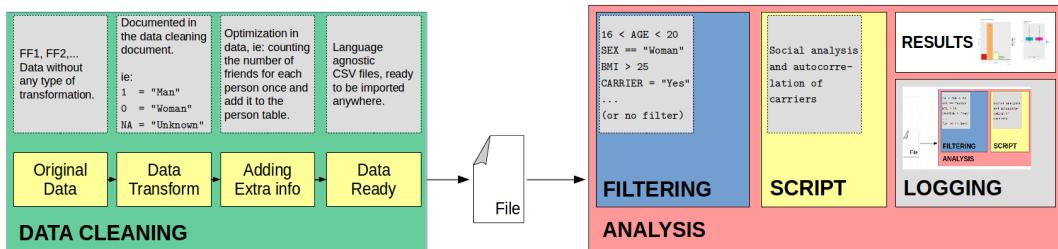


Figure 4.5: Overview of the data cleaning process. To the left, each step that transforms the raw data into a more practical version. This process is done only once. To the right, the subsequent analyses are performed as many times as needed.

Using our data will save time as it has been thoroughly cleaned, numerical data which is meant to be categorical has been given meaningful strings, diseases with just a verbal description with no code for International Statistical Classification of Diseases and Related Health Problems (ICD-10) has been assigned manually, and most importantly the data has been transformed into tables following proper Boyce-Codd normalization [373].

4.2.1 Naming

We changed their name to something more human-friendly and CSV and latex-compatible. This means using upper and lower cases appropriately, "ID" instead of "pers_key_ff1".

Samples are labeled as S1 and S2 instead of the given FF11 or FF12 sub-time period. Direct culture and enrichment broth are simplified to "direct" (from STAPH in any case which is confusing) and "enrich". Units are included where possible in the variable name ("FE_FF1" to "Fe_(μ mol/L)"). All FF id references are deleted as they are divided into Table_FF1, Table_FF2, and so on. Adding variable name continuity to quickly discern between binary and categorical variables ("FAT_FISH_FF1" to "FatFishFrequency") so the user already knows what to expect in each column. Deleting redundancy in names ("FRIEND2_CONTACT_SCHOOL_FF1" to "Friend2School").

4.2.2 Dates

For all tables where a date appears, dates are standardized from several data formats (Posix, "mm/dd/YY", "dd/mm/YY", "YYYY-MM-DD") to "YYYY-MM-DD" format. Some of the dates could be transformed to have more precision down to "HH:mm:ss" resolution, but this information is irrelevant to us anyway so it is discarded. SPSS is the programming language that is used to encode the original data, and it uses the start of the Gregorian Calendar in 1582 as the default date for POSIX origin. Because of this chosen date reference, there might be some days' error in coercion in these dates because we do not know which timezone, or time origin, we take as a reference. We delete all timezones from UTC to simply a date, which we assume is GMT +1/+2 depending on the time of the year. Since the time difference is one to two hours, and we have a day resolution level, this loss of detail does not matter. It is advisable to use another non-SPSS method in the future as this time error coercion might render biological data which is time-critical unusable.

4.2.3 General tables

S.aureus

These tables contain the *S.aureus* information. Note that none of the FF11 variables are described in the metadata files. The information was retrieved from Fit Future experiment designers.

Swabbing information

All lab comments that are just registered as "OK" by lab technicians for all samples are discarded to avoid data redundancy. Also, all status and event variables are redundant information and are later discarded. All leading and trailing white spaces are deleted.

None of these variables are described in the metadata and information.

What remains, for each unique nasal and throat swab, is whether the swabbing was performed successfully, with irregularities and the given reason, the swab was repeated, or the swab was not performed at all. We also have the freezer ID of where to find each sample as well as the freezing date.

Blood serum and Blood Technical information

Several variables indicate if some value is above or below the healthy limit which are discarded. The reference on whether some value is healthy or not is marked during the analysis according to the given references for each value. Within the blood serum scope, there are also a bunch of columns named EVENT0 to EVENT9, and EVENTA to EVENTL, which are empty, no description is given in the metadata or any other source, and have no information at all. As such all of them are deleted. All LCMSMS (Mass spectrometry measures) have the same values as their normal measurements counterparts, so those are also skipped.

4.2.4 Relational Normalization

The original data does not have any type of relational meaning between columns. To add future benefits, we need to fix this issue due to privacy security [374], data logical consistency [374], and computational time efficiency. This section describes all tables related to relational data. The medicine, contraceptive, and disease tables are read from the original dataset and then transformed later into a structure that follows a proper relational database property.

The data should be organized in Boyce–Codd normal form [373]. Boyce–Codd normalization is important because it ensures that there are no data dependencies between columns, which translates into guaranteeing the integrity and consistency of data which is critical for accurate analysis. This is also crucial in datasets with large amounts of data, but this is not the case. Summarizing the normalization process, all variables and data in those tables need to be transformed so they follow these fundamental principles of normalization:

- **0NF** No information is lost and no information is duplicated.
- **1NF** No columns, or multicolumns (3NF), which contain sets of values.

- **2NF** Single column primary key (person ID) which is also used as superkey (4NF).
- **3NF** Eliminating the transitive functional dependencies.
- **BCNF** Always satisfies lossless join condition.

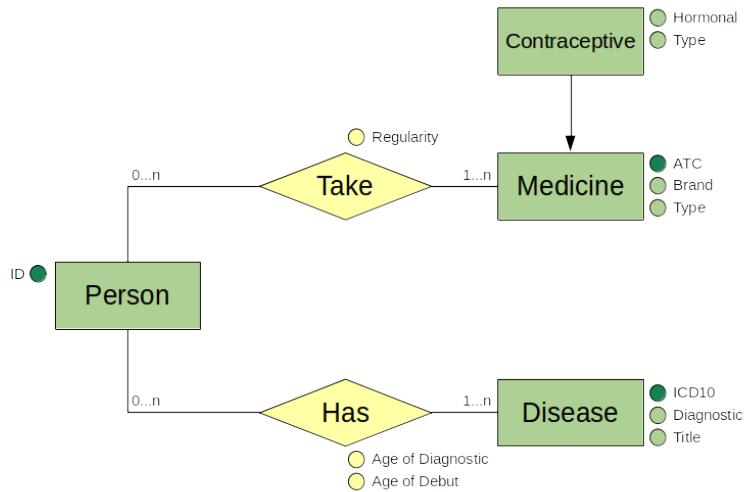


Figure 4.6: BCNF transformation from the original data. The database schema for "Person" can be obtained upon request.

4.2.5 IDs transformation

Each individual has a personal key described in the "pers_key_ff1" variable. The original key looks like this: "12345678" and is simply an 8-digit unique key for each of the 1038 individuals in that file. To avoid visual cluttering and math optimization, we substitute the original IDs with an integer number that goes from 1 to 1038, assigned randomly to each individual.

We have two special IDs. An ID equal to 0 means that a person has a friend that is not in our ID table, for example, it could be a student in another school that is not in Tromsø, or Balsfjord; in short, people who were not part of this study. An ID equal to -1 means no friend. This ID numbering keeps consistency later on when we do the filtering. This way, all the variables for each of the 5 friends have an integer, and is easier and faster to do math, indexing, and filtering. ID swapping log is registered in case identifying the original person is necessary but kept confidential.

4.2.6 Mapping

In this subsection, we discuss the relevant issues that arise during the mapping and transformation of the original variables. From the subsequent 29 tables generated during the mapping, we only provide a few relevant examples in this thesis.

S. aureus

Changes for Nasal and Throat variables are identical. This is also true for the variables for sample 1 and the variables for sample 2. In total, we have Nasal Sample 1, Nasal Sample 2, Throat Sample 1, and Throat Sample 2. Notice that all SPA-type variables remain the same.

School and education

Three columns represent "HighSchool", "Class", and "Programme". All of those have a numerical ID to represent the high school, a numerical ID to represent each class, and so on. While those can work ok the way they are, I choose to change them by adding a letter to each column to categorize it properly; these are not numerical values, these are categories. So high-school "1" becomes "H1", high-school "2" becomes "H2", class "10" becomes "C10", program "23" becomes "P23", and so for all values.

Sociology

Two variables expressed if "You live with 1 to 2 Siblings (yes/no/NA)", and "You live with 3 or more siblings (yes/no/NA)". Ideally, the questions should be: "How many siblings do you have? (number)", "How many siblings live with you? (number)". But we do not have that. However, we can convert those two variables into a single categorical variable that is expressed as "How many siblings live with you? (NA / Zero / One or Two / Three or more)". Regarding working status, the original data is also divided into too many variables that should not be, as they are sometimes mutually exclusive. The first one is "Is your mother studying?", "Is your mother a housewife?", and "Is your mother disabled?". All three of those are okay to have them separated as any combination of those (yes/no), plus working status, is possible. The rest of the possible answers refer to "Full time", "Part-time", "Unemployed", "Pensioned", "Deceased", "Don't Know", and "Other". All of those are grouped into the same variable within the same variable. Regarding ethnicity, the original question is not clear, and people have not answered ethnicity directly; instead, they have a combination of their country of origin, country of residence, country of parents, and their own race or cultural background. So here is

the attempt to clear all that data into something useful. The final variable is a string with as many ethnicities as needed ("Norwegian", "Norwegian-Sami", "Belgium-Spain-France"). People who said they are Norwegian, Sami, or Kven, have a combined ethnicity if answered more than one of those, and also combined with whatever they write in the "Other" column where applicable.

We must express our concern that sharing the precise text from the "Other" category may lead to the identification of individuals. Researchers seeking access to the FF data can, however, apply for access to obtain the exact methodology.

Finally, we have these variables which are just a straightforward mapping "*Who do you live with? / What is your parents' educational background?*" In this case, while would be a fringe case, is possible to live with multiple combinations of these at the same time. For example, parents can be divorced while the student lives with both of them half of the time, each having a stepmother/stepfather, while the grandfather also lives inside one of the houses. So all of these variables are kept independent.

Puberty and Sleeping

These tables and serve as an example in which categorical data is overdone due to the string field limitations of SPSS/STATA data.

The variable for "At what age did your pubic hair start growing?" is, originally, saved as a categorical variable, that gets the values "1", "2", "3", "4", "5", "6", "7", that means from 9 to 15 years accordingly. This is changed from a categorical to a numerical value, however, the data is of course censored to the left and to the right since we do not have any other option. Ideally, we should have a real number instead of the category. For the rest of the men variables we have standardized all answers and eliminated references to the variable in each of the options, so instead of "Facial hair has not yet started growing", or "Voice has not yet started changing", both are mapped to "Haven't Started".

In table 4.4 we see that modeling a time of the day, into 18 different categories is not practical. A better solution would be, at the questionnaire level, to ask "At what time do you go to sleep?" so we can have a numerical value instead. This particular mapping is later transformed into "How many minutes since noon passes until sleeping time", which is a more sensible data format.

Table 4.4: Original values for the sleeping habits.

Variable	Original	Transformed
SleepingPills	1	Not used
	2	Less frequently than every week
	3	Every week, but not daily
	4	Daily
	NA	Didn't Answered
BedTimeHourCat	1	18:00 or earlier
	2	18:30
	3	19:00
	4	19:30
	5	20:00
	6	20:30
	7	21:00
	8	21:30
	9	22:00
	10	22:30
	11	23:00
	12	23:30
	13	00:00
	14	00:30
	15	01:00
	16	01:30
	17	02:00 or later
	NA	Didn't Answered

4.2.7 Diseases

All diseases are transformed into the proper relational table described earlier in section 4.2.4. There are plenty of transformations that need to be curated manually.

Common Diseases I

The questionnaire keeps track of diseases in three different ways. First, there are 7 diseases that the original data track explicitly, but do not register any ICD10 code. Those are "Diabetes" (unspecified which type), "Itchy Skin", "Hand Eczema", "Rhinitis", "Asthma", "Atopic Eczema" and "Psoriasis". The ICD10 codes can be seen in table 4.5.

Table 4.5: Table with the common 7 chronic diseases asked in a subsection of the questionnaire.

Questionary	Medical	ICD10	Comment
Diabetes	Other specified diabetes mellitus	E13	
Itchy Skin	Pruritus, unspecified	L29.9	
Hand Eczema	Dyshidrosis	L30.1	
Rhinitis	Allergic rhinitis, unspecified	J30.9	
Asthma	Asthma	J45.9	
Atopic Eczema	Atopic dermatitis, unspecified	L20.9	Skin and subcutaneous tissue
Psoriasis	Psoriasis, unspecified	L40.9	

Common Diseases II

The second way to track diseases is that during the interview, the student can tell up to 5 chronic diseases, and the ICD10 code is also registered by the person performing the interview. Again, in order to safeguard privacy, the initial response will not be displayed throughout this thesis. Instead, solely the conclusive compilation of ICD10 codes and their corresponding medical terms will be provided in table 4.6.

Table 4.6: Table with the up to 5 self-reported chronic diseases asked during the interview.

Medical	ICD10
Allergic rhinitis, unspecified	J30.9
ADHD	F90.9
Celiac disease	K90.0
Eczema	L30.9
Food allergy	T78.4
Migraine, unspecified	G43.909
Lactose intolerance	E73.9
Depression	F32.9
Anemia, unspecified	D64.9
Insomnia	F51.9
Diabetes Type 1	E10
Anxiety	F41.9
Tension headache (TTH)	G44.2
Gastritis	K29
Arthritis	M13.4
Hypothyroidism	E03
Asthma	J45.9
Eating Disorder	F50.9

It can be seen that some that were in the previous table are also repeated here. Redundant diseases are deleted. If the disease has more information (i.e.: "Diabetes

type 1" instead of "Diabetes"), we keep the most specific record. There are also some cases in which the person registered "Rhinitis" but assigned different "J30.X" codes to them. In this case, we changed the name "Rhinitis" to the ICD10 referred name. The disease "Migraine" is registered but no ICD10 was given, so we assigned the "G43.909" code.

Other Diseases

Finally, we need to look for all diseases that are written in the "Other" column. There is no further information here besides what the patient describes in one line of text, and no ICD10 code is given. People reporting two diseases at once are registered as two independent diseases. If the description here is more specific than in previous tables, then the information is updated. We have a total of 1133 instances of an individual linked to a disease, from which more than 10% came from cleaning this part of the data. We also have a total of 98 unique diseases, of which more than 75% come from this section alone.

4.2.8 Medicines

All medicine information, including contraceptives, is transformed into a proper normalization table as shown before. Is not described for all cases, but if possible, we include regularity information, meaning how often the person takes this medication. Notice that in the medicine table, we later also include the contraceptives that are hormonal in the list of medicines that this woman is taking.

There is one section of the interview where it is asked about regular medication. In the first one, detailed information is asked of the patient (i.e.: what is the name of the medication, the ATC code, and so on). There is also a section in the questionnaire where we have redundant information and it is only asked in a "yes/no" format (i.e.: "Have you taken sleeping pills in the last 4 weeks?"). In the second case, no regularity, no brand, and no ATC code are provided. We also have no information regarding whether this was a one-time-only event for this medication, if it was taken without a prescription, or any other extra information. Because of this, the second part is ignored since we cannot add a generic drug, or when specifically when it was consumed in the last 4 weeks. This affects the question regarding painkillers, sleeping pills, antidepressants, ADHD medication, and tranquilizers. People who take any of these regularly, have already filled this information in the first part of the questionnaire.

4.3 Statistical and Machine Learning

4.3.1 Software

Analysis was performed using R version 3.6.3, R Studio 1.3.1093, and Python version 3.9. Notable packages used were "ggplot2", "igraph", "ggraph", "ergm", "tensorflow", "sklearn", and "SHAP".

4.3.2 Simulations

To determine whether there is bias within the relationships in a network we use bootstrapping as described in section 3.1.7, simulating 1000 networks to avoid making any assumptions about the underlying population. We do this by using a similar approach as common non-parametric tests [375]. Our approach consists of counting how many relationships connect two nodes with the same attributes in our network (i.e., *S. aureus* carrier with *S. aureus* carrier) and comparing this number with the same number given by the simulations. In figure 4.7 we can see an example of a real network, and 3 simulations with an arbitrary node distribution, counting also the same-to-same relationships for each example.

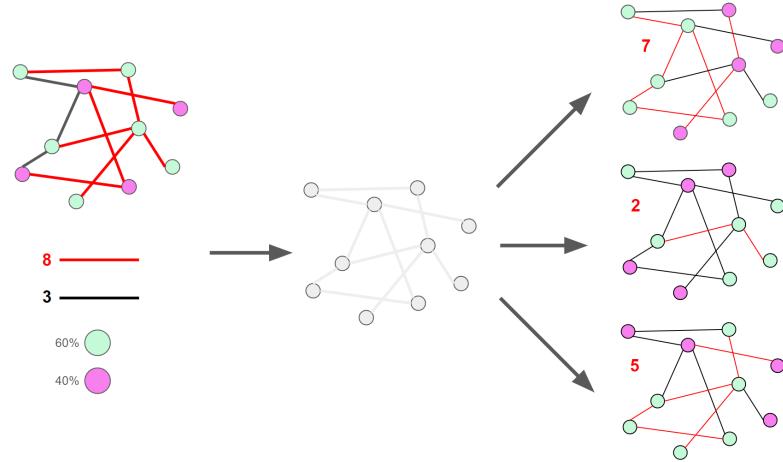


Figure 4.7: In this figure we see the overall process of simulating new networks. On the left, we have a representation of a real network that has 60% green nodes and 40% magenta nodes as attributes. The edges connect nodes, and 8 edges connect nodes with the same attributes (red) while 3 edges connect nodes with different attributes (black). On the center, we strip the nodes of the attributes but we keep the topology of the network, keeping nodes and edges as they originally were. On the right, we have 3 simulations (top, center, bottom) that use the center topology, but distribute the attributes randomly as described on the original network on the left. The simulated network will also have a total of edges connecting nodes sharing their attributes, in this case, the top has 7, the center has 2 and the bottom has 5; the average of the simulated network connecting nodes with the same attributes would be 5.7. This example has only 3 simulations, which is very limited, but seems to indicate that the real network (left) has some sort of bias connecting attributes as the 8 original same-to-same relationships are bigger than the 5.7 average of the simulations using the same attributes' distribution.

Unlike in figure 4.7, we obtain a bigger sample of simulations, which gives us a distribution of 1000 values from which we can extract a mean and a standard deviation. We then perform simple hypothesis testing, like a t-test, using the real amount of same-to-same relationships against a normal distribution given by the simulated mean and standard deviation. Let's describe the whole process as follows:

1. Original Graph Description. Let $G = (V, E)$ be a graph where:
 - V is the set of nodes, with $|V| = N$.
 - Each node $v \in V$ has an attribute $A(v)$ which can take one value from $\{A_1, A_2, \dots, A_X\}$.

Define the distribution of attributes as:

$$p_i = \frac{|\{v \in V : A(v) = A_i\}|}{N} \quad (4.1)$$

where p_i is the proportion of nodes that have attribute A_i .

Define Z as the number of "same-to-same" relationships:

$$Z = |\{(v, w) \in E : A(v) = A(w)\}| \quad (4.2)$$

2. Attribute Removal and Simulation

- (a) Remove all attributes from the nodes, resulting in a graph $G' = (V, E)$ with no attributes.
- (b) Simulate 1000 new graphs $\{G'_1, G'_2, \dots, G'_{1000}\}$ where each graph keeps the topology of G' but assigns attributes randomly to each node according to the distribution $\{p_1, p_2, \dots, p_X\}$.

3. Simulation Analysis

For each graph G'_k , calculate the number of "same-to-same" relationships:

$$Z_k = |\{(v, w) \in E : A_k(v) = A_k(w)\}| \quad (4.3)$$

where $A_k(v)$ is the attribute of node v in the simulation k .

Collect these values to form a distribution $\{Z_1, Z_2, \dots, Z_{1000}\}$ with mean μ_Z and variance σ_Z^2 :

$$\mu_Z = \frac{1}{1000} \sum_{k=1}^{1000} Z_k \quad (4.4)$$

$$\sigma_Z^2 = \frac{1}{999} \sum_{k=1}^{1000} (Z_k - \mu_Z)^2 \quad (4.5)$$

4. Statistical Testing

Perform a two-sided t-test to compare the observed number of "same-to-same" relationships Z from the original graph with the simulated distribution:

$$t = \frac{Z - \mu_Z}{\sigma_Z / \sqrt{1000}} \quad (4.6)$$

The p-value is calculated from the t-distribution with 999 degrees of freedom to determine if Z is significantly different from μ_Z . If the p-value is small (typically less than 0.05), it suggests that the original graph has a bias towards "same-to-same" relationships compared to what would be expected by chance given the attribute distribution.

We expand on this concept by instead of using the distribution of attributes in the general population of a node (i.e., *S. aureus* carrier prevalence being 30%), using the distribution of one specific category (i.e., *S. aureus* carrier prevalence in women being 20%) and repeating the same process for each category present in each attribute of interest (ie: women and men in sex, from underweight to obese in BMI, and so for). This gives us a new mean which then can be compared with the previous simulated distribution. In this way, we can check how much each of the categories deviates with respect to each other, and we can identify which category has a higher or lower risk for the outcome variable; if any. Let described as follows:

1. Original Graph Description with Attributes A and B . Let $G = (V, E)$ be a graph where:

- V is the set of nodes, with $|V| = N$.
- Each node $v \in V$ has a primary attribute $A(v)$ which can take one value from $\{A_1, A_2, \dots, A_X\}$ and a secondary attribute $B(v)$ which can take one value from $\{B_1, B_2, \dots, B_Y\}$.

Define the distribution of attributes as:

$$p_i = \frac{|\{v \in V : A(v) = A_i\}|}{N} \quad (4.7)$$

where p_i is the proportion of nodes that have attribute A_i .

Define Z as the number of "same-to-same" relationships, this works for both A and B:

$$Z = |\{(v, w) \in E : A(v) = A(w)\}| \quad (4.8)$$

The distribution of B depending on the value of A :

$$p_{i|j} = P(A(v) = A_i | B(v) = B_j)$$

where $p_{i|j}$ is the conditional probability of node v having attribute A_i given that it has attribute B_j .

2. Attribute A Simulation

- (a) Remove all attributes from the nodes, resulting in a graph $G' = (V, E)$ with no attributes.
- (b) Simulate 1000 new graphs $\{G'_1, G'_2, \dots, G'_{1000}\}$ where each graph keeps the topology of G' but assigns attributes randomly (ignoring B) to each node according to the distribution $\{p_1, p_2, \dots, p_X\}$.

3. Attribute B Simulation

- (a) Remove all attributes from the nodes, resulting in a graph $G' = (V, E)$ with no attributes.
- (b) Similarly, simulate 1000 new graphs for each category of B $\{G'_{B1}, G'_{B2}, \dots, G'_{BY}\}$,

where each graph assigns attributes A based on the conditional distributions $p_{i|j}$ corresponding to each B_j .

4. Simulation Analysis

For each graph G'_{Ak} , calculate the number of "same-to-same" relationships:

$$Z_{Ak} = |\{(v, w) \in E : A_k(v) = A_k(w)\}| \quad (4.9)$$

Collect these values to form a distribution with mean μ_{ZA} and variance σ_{ZA}^2 .

For each category B_j and each graph G'_{Bjk} , calculate the number of "same-to-same" relationships:

$$Z_{Bjk} = |\{(v, w) \in E : A_{Bjk}(v) = A_{Bjk}(w)\}| \quad (4.10)$$

Collect these values for each B_j to form distributions with means μ_{ZBj} and variances σ_{ZBj}^2 .

5. Statistical Testing

Perform a two-sided t-test to compare the observed number of "same-to-same" relationships μ_{ZA}, σ_{ZA}^2 from the simulated distribution with respect to the "same-to-same" relationships μ_{ZBj} from the simulated distribution with conditional probabilities.

$$t_{Bj} = \frac{\mu_{ZBj} - \mu_{ZA}}{\sigma_{ZA}^2 / \sqrt{1000}} \quad (4.11)$$

This comparison helps identify if any specific B attribute significantly influences the "same-to-same" relationship bias in the original graph.

Ideally, this technique should be done not by simulating similar networks, but by

simulating every possible network and comparing those in which bias happens to those in which bias does not happen. However, it is impossible to find every possible network within reasonable computational time as described in section 3.1.7. So we need to reduce the number of possible networks based on some assumptions [64]. This allegedly gives the model properties that make it similar enough to all the possible networks. In our case, we use the same frequency tables with a network with the same topology as constriction. We also assume that the virulence of *S. aureus* would cluster carriers with carriers and vice-versa. As BMI homophily is high and the Chi-square table also suggests so, we also assume that this happens with subjects with similar BMI. Finally, we also assume that friends share similar environments and activities and this would be reflected in their vitamin D levels.

We use this approach in [Paper A](#), [Paper B](#), [Result I](#), and [Result IV](#).

4.3.3 Friendship ratio

For [Result II](#), we devise this method as an alternative to simulations using categorical data. Biomarkers levels are a continuous variable and as such we cannot use the simulation approach unless we categorize them into something similar to "low level", "medium level", and "high level", losing some information in the process. Instead, we compare numerical levels between friends and non-friends biomarkers one by one. We do this by finding the ratio of, the average square difference between each person's biomarker level and friend's biomarker levels, and the average square difference between each person's biomarker levels and non-friends biomarker levels. Let's define it as follows:

1. Graph Definition. Let $G = (V, E)$ be a graph where:

- V is the set of nodes, $|V| = N$.
- Each node $v \in V$ has a real number value $R(v)$ which can take one value from $\{R_1, R_2, \dots, R_X\}$.

2. Define the distance d between any two nodes v_i and v_j as the square of the difference between their values:

$$d(v_i, v_j) = (R(v_i) - R(v_j))^2 \quad (4.12)$$

3. Calculation of Average Distances. For each node v in the graph:

- Let $F(v)$ denote the set of friends (nodes directly connected to v) and $NF(v)$ denote the set of not-friends (nodes not directly connected to v).
- Calculate the average distance to friends and not-friends:

$$\text{Avg}_F(v) = \frac{1}{|F(v)|} \sum_{u \in F(v)} d(v, u) \quad (4.13)$$

$$\text{Avg}_{NF}(v) = \frac{1}{|NF(v)|} \sum_{u \in NF(v)} d(v, u) \quad (4.14)$$

4. Ratio of Averages. Calculate the overall average of the averages for friends and not-friends across all nodes:

$$\text{Avg}_F = \frac{1}{N} \sum_{v \in V} \text{Avg}_F(v) \quad (4.15)$$

$$\text{Avg}_{NF} = \frac{1}{N} \sum_{v \in V} \text{Avg}_{NF}(v) \quad (4.16)$$

Then, calculate the ratio of these overall averages:

$$\text{Ratio} = \frac{\text{Avg}_{NF}}{\text{Avg}_F} \quad (4.17)$$

5. Interpretation of the Ratio

- If $\text{Ratio} \gg 1$, not-friends tend to be further away than friends, indicating that clusters of friends are more similar to each other.
- If $\text{Ratio} \ll 1$, friends tend to be further away than not-friends, indicating

that not-friends are more similar to each other.

- If $\text{Ratio} \approx 1$, friends and not-friends seem similar.

We however do not have a sensible threshold cut-off for when a value is significantly different from 1. We arbitrarily suggest that values greater than 1.1 or smaller than 0.9 are the significant ones.

4.3.4 Statistical methods

On Paper A, we tested whether the observed statistics are in the range of the ERGM simulated graph as described in section 3.1.7. On Paper A, Paper B, we used homophily as described in section 3.1.7; and to some extend, we reference this in the rest of the results as well to justify strong connections at high school level. On Paper A, Paper B, Result I and Result IV we used X^2 test as described in section 3.3.2. On Result III, we used ANN and RF as described in 3.2. In particular, SHAP and MDI methods are explained in 3.2.4 and 3.2.4. On Paper A, we used an autocorrelation model explained in section 3.3.4. Finally, logistic regression was used in Paper A, Paper B, Result I as described in section 3.3.3.

4.4 Ethical considerations

The Regional Committee for Research Ethics approved the Fit Futures study (REK North application ID 16773) and the analysis as part of the Tromsø Staph and Skin study - Fit Futures (REK North application ID 23432).

Chapter 5: Summary of main papers

5.1 Paper A

Social network analysis of *Staphylococcus aureus* carriage in a general youth population.

We explored the prevalence of *S. aureus* and risk factors in the FF1 population. Carriage prevalence was 30.4% for direct culture and 42.6% for enrichment broth. Both direct culture and enrichment broth showed a significant difference between males and females; with males having 36.4% and 48.1% prevalence, and females having 24% and 36.8% prevalence respectively. No other host factor was significant.

Students who attend the same high school tend to share the same Spa-type between them, which indicates that they share the same source of infection. The simulations indicated that school transmission is significant in the school network if the direct culture is used, and significant in the overall, physical, and school network if the enrichment culture is used. Simulation regarding Spa-type similarity indicated that transmission is relevant in all networks.

Males showed to have less connectivity than females, however, they have a higher prevalence. Autocorrelation regression also indicates that transmission happens in the network, but only the direct culture indicates that sex is a relevant factor. Our simulations indicate that sex and Physical Activity (PA) are relevant in both cases, which seems to indicate that women are at more risk of person-to-person transmission due to their higher connectivity. Autocorrelation also shows that Body mass index (BMI) and PA were relevant in both cases and the study program and alcohol for enrichment only. Our simulations indicate BMI and Alcohol for the direct culture only.

Finally, we estimated that a random student has an average increased risk of transmission of 3.5% with logistic regression, and an increased risk of 5% with auto-correlation, for each additional friend who is *S. aureus* carrier.

5.2 Paper B

"Friends are the sunshine of life" Social influence on vitamin D in a general youth arctic population.

Vitamin D is of special interest in the Arctic region; here we explored if friends tend to have similar levels. While vitamin D is not contagious from person to person, similar levels would indicate that friends share the same environment, activities, diets, or habits that promote or hinder vitamin D absorption.

First, we presented all possible factors that affect 25(OH)D levels in the blood. Diseases and medications were not relevant for this population. Then we checked which variables had a bias for each high school. This is because Ultraviolet B (UVB) influence overwhelmingly affects vitamin D absorption, and each high school had a different date for blood extraction across the year, differing traveling to sunny regions due to school calendar holidays, and different solarium habits. Without stratification, several levels are significant, but once the high schools are investigated one by one, only sex in H8, PA in H3, and holiday traveling in H1 and H3 were significant variables.

We also found out that women influence other women into going to the solarium, however, men do not influence other men. Currently, teenagers are banned from entering solariums in Norway due to their increased risk of skin lesions and cancer.

Among non-solarium goers, using logistic regression, we estimated that people with friends who have normal vitamin D levels ($>50 \text{ nmol/l}$) have a 7.25% chance of having normal vitamin D levels themselves for each additional normal vitamin D friend. We also checked this for high schools, and the influence was also significant among 5 of the 8 high schools.

Finally, we found contradictory results in the vitamin D levels concerning diet and vitamin D supplements. That is, people, eating fatty fish which is high in D3, or taking supplements that are extremely high in D3, do not show elevated levels of 25OHD in their blood. This is impossible. In the discussion part, we mentioned how different memory-based dietary assessment methods (M-BMs) can be used to improve the validity of dietary data, as well as Metabolic Equivalent of Task (MET) for PA.

5.3 Paper C

An introduction to network analysis for studies of medication use.

We presented how Network Analysis (NA) can help study medication usage regarding co-medications and drug interaction in the Norwegian population.

NA is underutilized in Drug Prescription Networks (DPN). As such we provided examples of how this type of analysis can help to analyze the relationships between prescriptions, health professionals, and patients. We accompanied this with a comprehensive tutorial on how to apply these methods using R and Stata syntax.

To accomplish this, we presented networks using the Norwegian Prescription Database (NorPD) in the elderly population, and another network of severe drug-drug interactions (DDIs) using the Norwegian Electronic Prescription Support System (FEST). In the results, we presented several statistics with their explanation for the reader to understand this type of analysis.

5.4 Result I

Social network influences on obesity in a general youth population.

We studied how friendship and social contact influence obesity. We saw that students tend to cluster together based on their BMI. Simulations indicate that students being friends with other students of the same BMI does not seem to be at random.

BMI increased almost $1\text{kg}/\text{m}^2$ on average from FF1 to FF2. We also show that students who belong to the "Healthy" group in FF1 have fewer chances of belonging to the same group in FF2 as their number of friends with $\text{BMI} > 25$ increases.

5.5 Result II

Social network influences on inflammatory response in a general youth population.

We found several results that suggest friends share similar inflammatory profiles among them.

A student's average biomarker level tends to be similar to his / her friend's average levels as well once sex and high school are accounted for. Similarities increase if we compare samples taken early in the academic year with samples taken later on. Furthermore, we introduced a distance metric which also suggests that friends have similar levels to non-friends.

We also tried to compare inflammation induced via social contact with inflammation resulting from obesity complications. Some biomarkers that correlated with anthropometric variables are not present in the social influence results; suggesting that the inflammation effect has a mix of both.

5.6 Result III

Measuring social influence with random forest regression and artificial neural networks.

We ran two machine learning models to predict FF2 BMI based on FF1 variables sex, BMI, smoking, snuff, alcohol, and sport frequency habits; as well as the number of underweight, healthy, overweight, and obese friends. We ran these models in 6 different subsets in which students increased, decreased, or stayed in the same BMI group from FF1 to FF2.

We used MDI and SHAP to measure the most important variables according to the ML models. After the initial BMI, in most cases, the total number of friends in each BMI group was evaluated as more important than the non-social variables.

5.7 Result IV

Frequency consumption of medication and social network influence in a general youth population.

We studied the possibility that students can influence each other's usage of over-the-counter medicines. These are often misused with unwanted side effects, or abused due to their potential as recreational drugs.

We saw a huge disproportion of reported diseases with respect to the reported medicines that are relevant for these diseases. In particular a spike in the use of anti-inflammatories and painkillers. Consumption by sex indicates that women tend to consume more of these medicines than men.

Hormonal contraceptives and painkillers seem to be associated with high schools. Anti-inflammatories and painkillers seem to be associated with sex. Simulations indicate that women who are friends share the same hormonal contraceptive brand.

Chapter 6: Discussion

6.1 Papers and results

6.1.1 Paper A

S. aureus transmission has been the subject of study on several occasions. Whether it is within the same household [171, 376], same hospital and ICUs [377, 378], or worldwide MRSA [379]. Also, social transmission of the beneficial or detrimental pathogen has been studied for HIV [380], and microbiota sharing among family and their pets [381].

To our knowledge, this is the first study to analyze social transmission in a general youth population. In 2023 another paper studied the spread of *S. aureus* in schools in England reaching similar results and conclusions to us [169].

We hypothesize that school is an important factor, not because the school itself is relevant, but because students spend more time together. However seasonal immunology is also a factor to be considered [382], and so far we have not been able to analyze this effect with confidence with only a one-time data point. For example, temperature, humidity, and vitamin D levels seem to drive the seasonal immunity for influenza [383]. It would be interesting to determine if a similar effect is relevant for *S. aureus*, especially considering that it is an opportunistic bacterium, adjusts a time series model accordingly, so the risk of transmission can also be adjusted depending on the time of the year.

Our results in **Result III** also show inflammation processes correlated as the academic year progresses and with schools individually. It is reasonable to think that infection risk increases in function over time, and so does the immune reaction that follows.

Defining what is a carrier of *S. aureus* is a difficult task in itself that took a considerable level of debate to resolve because we used two different growth methods (direct and enriched broth). An alternative method to define a carrier or infected individual would be by using fuzzy logic. Fuzzy logic is a branch of mathematical logic that deals

with reasoning and decision-making in situations that involve uncertainty and imprecision. Classical logic operates on binary values (true/false), whether fuzzy logic allows for degrees of truth, in which logical propositions can be partially true or partially false. This approach has been used to improve chronic disease classification and decision-support systems [384, 385], and it would be interesting to see if it can perform well in graph models designed to predict disease spread in cases such as this in which a person can be a carrier of a bacterium, in different body tissues, at different enrichment levels of the sample, while also being asymptomatic.

Sequencing the spa gene of both MSSA and MRSA assigns spa types to detect and control outbreaks. However, Spa-typing focuses on a single DNA location (locus), the spa gene. This can limit its discriminatory power compared to methods that analyze multiple genetic loci or the whole genome [386]. Consequently, spa-typing may not differentiate between strains that are closely related but differ in attributes like virulence or antibiotic resistance. Globally, spa types t008 and t002 are the most distributed, with variations in prevalence across continents influenced by migration and local epidemiological factors [387]. Within Europe, spa types t032, t008, and t002. Within Scandinavia, t084 in Norway, t002 in Sweden, and t067 in Finland were reported. T084 is consistent with our study as it is top number 2 in our population, tied with t024 at 8% prevalence. T002 (top 3 global, top 1 Sweden) was top number 7 accounting for the 3% of our spa-typed samples. We did not report on students with less than 5 isolated (less than 1% spa-typed) due to privacy concerns. A multidrug-resistant Bengal Bay clone ST772-MRSA-V was found in Norway during the period of FF1 and FF2 [388], but this strain was not found in our population. Worth mentioning that, at the time, almost the entire population of the study identified as Norwegian or Sami, and no isolated prevalent in other countries was expected to dominate or even show in our population.

6.1.2 Paper B

There are plenty of studies comparing ethnicity with vitamin D [389–395]. There is also at least one study comparing socioeconomic factors with vitamin D deficiency [396]. Although these variables tend to have a strong social component, to our knowledge no previous study has tried to study the social aspect of vitamin D. Other previous studies have measured the vitamin D prevalence in Tromsø [397–399], but again not the social aspect of it.

This paper is of particular interest in our Arctic population given how poorly vitamin D is absorbed via UVB radiation in this area. We can counter bad habits such as the described negative effect of solarium and recommend better activities such as PA so it has a spillover effect on the network. There are also plenty of external social influences regarding PA alone [400] which are not within the reach of this data and would be interesting to see and compare at the same time. Another nice follow-up would be how these recommendations would be effective in an adult and elderly population as their PA decreases while their traveling increases but as their capacity for vitamin D metabolism decreases with age as well [401].

One interesting approach from a public intervention point of view is that teenagers with no European background are especially vulnerable to vitamin D deficiency [194–197, 232, 232–234]. In Paper B we discussed how immigrant populations tend to form strong community bonds. So public interventions targeted to this population to increase vitamin D levels would be quite effective.

Here we also discuss how future projects should gather data. There are plenty of improvements that can achieve better results with M-BMs and dietary data, and using MET for PA. Within the vitamin D topic, there are plenty of contradictory results [402–404] due to poor standardization or poor experiment design [195, 229, 230]. For example, in our own data, we are not taking into consideration any interactions between foods [251, 251, 252]. It is paramount that we do not add confusion and noise to the literature; and that we start with proper data.

As previously commented, the immune system seems to be influenced as time passes by. Vitamin D promotes a homoeostatic effect on the immune system. If vitamin D is depleted over time, such as is the case in Tromsø where students come back from sunbathing in August and local UVB radiation is not high enough to fill them again, it would mean that unwanted type 2 immune reactions [405] are going to increase.

6.1.3 Paper C

Previous works have used network analysis to detect fraudulent opioid prescriptions [406] as well as being at risk of opioid abuse [407]. Another study from 2021 showed the capabilities of data mining using prescription networks in Italy [408]. Overall using SNA particularly to study prescriptions and find patterns has proven to be a useful tool for researchers and health professionals.

This paper, similarly to **Result III**, has been composed to demonstrate potential applications of a specific methodology. The results of DDIs are explored in other articles.

In our study, we saw that SNA can be useful in visualizing complex interactions and measuring clusters and central nodes with different metrics. Clusters in particular could be connected to pharmacological data to see the importance of pharmacokinetic interactions.

The bigger downside is to be able to draw the network in a meaningful way; not because this network is difficult to draw, but because plotting it is a general downside of SNA as we discussed during the background in section 3.1.6. Another downside to this network is that it is a bipartite network divided into prescriptions, patients, and medics. Bipartite networks are generally more complex to analyze. This makes the tool great for visualization and exploring but weak for hypothesis testing.

6.1.4 Result I

Here we saw that social dynamics behave in similar tendencies as in other teenage populations [409–413] where it is also shown that “*avoidance of overweight friends is the primary determinant of friendship patterns related to BMI.*” and “*a significantly greater proportion of obese/overweight versus non-overweight youth reported difficulty in making friends*”.

A questionnaire on dietary habits, including vitamin supplementation, was given to the participants. However, our analysis of the dietary response, as shown in **Paper B**, shows that the self-reported diet habits are not a reliable answer, as the estimated nutritional intake from questionnaire data (selected food items with average frequencies of intake) does not correlate with the nutritional data retrieved from blood samples. As such, no nutritional data was included in this study so there is no assessment of how one person’s diet influences another person’s eating habits as well.

Another limitation of this study, which we try to partially overcome with **Result III**, is that we do not address how to move students from unhealthy BMI groups to healthy BMI groups. The second limitation is how to increase the connectivity from Healthy BMI to other groups as it would seem that groups tend to be self-biased and close to people from other groups.

6.1.5 Result II

There are plenty of papers that have explored the effects of inflammation and isolation [27], but to our knowledge, this is the first time that it has been studied how non-isolation influences other people's inflammation. This is important with IL-6 and CRP in particular as they seem to be the leading factor in inflammation affecting loneliness. There also have been several studies linking biomarker levels with obesity. To name a few, ADA has been linked in mouse models with lower obesity and insulin resistance [414]. Axin-1 is correlated with glucose uptake in skeletal muscle [415]. BNGF has been linked with BMI and obesity regulation in a Scandinavian population [416]. Obesity-driven chemokine has been studied for CCL2, CCL13, CCL18-19, CCL23, CCL26, CXCL1, CXCL3 and CXCL14 [417]. Patients in the obesity group had higher IL-1beta, IL-1RA, IL-2, IL-4, IL-5, IL-6, IL-8, IL-9, IL-10, IL-15, IL-17A, MCP-1/CCL2, MIP-1alpha/CCL3, MIP-1beta/CCL4, G-CSF, GM-CSF, FGF, IFN-gamma, and TNF-alpha than control group [418].

Here we see a positive correlation between IL-6, waist, hip, weight, and BMI in both men and women, plus a correlation between IL-6 within some high schools, and IL-6 being quite similar in between women's friends. IL-6 acts as a pro-inflammatory cytokine and as an anti-inflammatory myokine, it suppresses inflammation caused by stress in bones and muscles during exercise and promotes bone re-absorption. Myokines function is still poorly understood but is believed to have a beneficial impact as a response to PA [105], which can be one of the reasons why women seem to have similar levels given that they do more PA than men in this population.

Another interesting result is that we can also see IL-10 and IL-13 correlating in women's friendships, and IL-10, IL-13, and IL-33 correlating with high school and sex. These are anti-inflammatory cytokines which would lead to believe social influence may alter the anti-inflammation profile of a student. Several chemokines related to immune hemostasis are also found across relevant results in both sexes, such as CCL 23, CCL 25, and CCL 28. In women's cluster of friends again, Brain-derived Neurotrophic factor (BDNF) is also associated with anti-inflammatory, chemotactic proteins that aid with diapedesis and extravasation of monocytes, and Macrophage colony-stimulating factor 1 (CSF1) it regulates osteoclast proliferation and differentiation and the regulation of bone resorption which might be interesting considering the sex differences of vitamin D levels seen in [Paper B](#).

A methodology shortcoming of this study is, as discussed previously in [Paper A](#), seasonal immunity effects are not taken into consideration. The study itself could be just a signal that we are measuring seasonality as each high school has different blood extraction dates and we are just seeing proteins reflecting immunity determined by time rather than high school influence.

For CRP, we found a very strong association for H6 women ($R^2 = 0.88$, $p-v < 0.001$), and a weak for H6 men ($R^2 = 0.22$, $p-v = 0.11$). Others APR would be interesting to study, in particular ESR given that it takes way longer than CRP to return to normal levels after an inflammation process.

6.1.6 Result III

In previous studies [409, 419] two strategies have been theorized regarding weight loss for overweight individuals. One that increases the total connectivity in general, and another one that increases the total healthy connectivity in particular. Obese students have low average connectivity, and they tend to be friends among themselves rather than with the general population. These results seem to indicate that the second strategy is better.

Initial BMI is the most important variable by far in all models. Individuals seeking weight changes should make necessary adjustments to their goals and refrain from relying on methods that promise quick results.

In all models, after FF1 BMI, MDI evaluated healthy social contact influence in the final BMI more than any other non-social host factors including sport frequency. In general, SHAP evaluates some of the total friends' variables very highly alongside sex and sports frequency. General high connectivity with healthy friends seems to be a good contributor to lower BMI or to prevent further BMI increase. General high connectivity with overweight friends seems to be a bad contributor to increasing BMI or preventing further BMI losses. Only group D seems to have a negative correlation, but the effect is small (-0.15 BMI from 0 to 3 overweight friends).

These results follow the previous analysis done in [Result I](#) where we see that FF2 BMI is proportional with connectivity with high BMI individuals. We also observed that students have a bias toward choosing friends of the same BMI category. How to break this bias is beyond the scope of this study, but social relationships seem to be as important as doing sports in aiding overweight and obese individuals to lose weight.

It would be interesting to figure out why Healthy students stopped being healthy and advanced to an Overweight or worse state. But model E seems hard to interpret by our model. Further data regarding if these people change their habits with time (i.e.: Sport was “hard” in FF1 but then “none” in FF2 two years later) would be useful.

Sports are evaluated as having high importance by MDI and SHAP to stay Healthy but seem to be of not much relevance when you increase or decrease your weight. Sports are generally also evaluated as a strong variable indicator. Sports increase the metabolism and total energy consumption, but if the energy intake is still greater than the energy spent in sports, then weight loss is not possible. In this dataset, we do not have a healthy diet adherence evaluation, as it could be for example the 14-Item Mediterranean Diet Adherence Screener (MEDAS). Further analysis that includes food questionnaires is needed to evaluate how diet is influenced by social contacts, and how important it is with respect the social influences.

In these results, we used some naive approaches to building machine learning models and used a limited set of explainability techniques. Our main goal was to show that this is useful and the results are important, rather than fine-tuning each model accordingly for each dataset; which in particular is self-evident with dataset E.

6.1.7 Result IV

Other works have presented the over-the-counter usage in Norway [341], and have also been speculated, such as "*Parents' symptom experience seems to influence their children's medicine use over and above medicine use indicated by symptoms. Two potential explanations are suggested: a socialization pathway and/or a pathway through adverse living conditions.*" [420], which links medicine abuse to the social network influence. Once again, to our knowledge, no other article has analyzed the social aspect.

In Paper C we discussed how other works detected substance abuse using SNA. Here something similar is achieved in specific high schools showing a disproportionate amount of medicines with respect to diseases, and how social influence affects this.

Painkiller usage is biased within high schools. This should not happen under normal circumstances, as both the diseases related to pain, and the use of over-the-counter drugs should be spread across high schools randomly. Pain-related diseases are indeed spread fairly balanced across high schools, but not the medicine part. The use of anti-inflammatory bias in women can be explained by the use of medication during menstrual

cramps. We can see that the use of Naproxen (brand name Naprocyn) is 100% females as this is an NSAID mainly used to treat such pains.

The main concern of these results is whether students are using painkillers as recreational drugs, the same as seen in other populations [421]. While these results should raise some attention, are not direct proof of anything. A proper follow-up questionnaire should be addressed and asked directly "*Have you ever used over-the-counter medicine/self-medication with non-medical purposes? If so, in which period of time and what frequency?*".

Previous work studied the relationship between pain threshold and friends' influence and inflammation response [422]. The bias that we saw in painkiller consumption at the high school level might be partially explained due to some high schools having lower pain resistance in general. However, we do not have access to the pain threshold data to investigate this approach further.

This data is also from the 2010s, being not a well-up-to-date reflection of current teenage drug activities. For example, in recent years "U-47700" has been popularized as a Non-Fentanyl Synthetic Opioids [423] and it could be that current teenagers are ditching painkillers for this drug, or any other alternative.

6.2 Network modelization

In our work, we decided to use an undirected graph as the main method for network analysis. This is due to three main reasons. First, we wanted to compare our work with previous work done at this university regarding pain [88] and expand further on it; this previous work also uses undirected graphs. Second, we wanted to compare this work with other literature, for example in **Paper A** we discuss reference "*Moldovan et al., 2019*" [424] where it study transmission using a multimodal network connecting people with rooms also undirected. We discussed the possibility of modeling the network also as people connecting with classrooms, schools, or other entities in which the direction between entities and people is not well defined. The third reason was the simplification of the methodology due to the abstract definition of friendship. For example, the question "*Do you have physical contact?*" is quite open to interpretation and can be positive from either a simple handshake to intimate contact. The main question "*Which students have you had the most contact with the last week?*" is also open to interpretation and the student may consider that contact with another student for just 10 minutes is not

contact, while for other 10 minutes is technically contact and as such counts as defined by the question. Due to all of this, we decided to take the option of using the undirected approach.

Another concern is the modelization using self-reported data. We justify the usage because the students also self-report that the network is quite representative (shown in figure 4.3). The alternative to no self-reporting data is to keep some sort of geographical tracking for the students, similar to the GPS tracking that is done in epidemiological studies with cattle. Such a direct invasion of privacy is very unlikely to pass the ethical committee of any experiment design. However, recently after the COVID epidemic, there has been some effort into contact tracking using cellphones and Bluetooth IDs, microphones, GPS location, cameras, Wi-Fi connection history, or Media access. Both centralized and decentralized approaches are vulnerable to the disclosure of location [425], not even going into the weaker encryption standards that cellphone users (our students in this hypothetical case) which are prevalent in the general population. This makes this approach even worse than strictly secured GPS tracking. In recent years other approaches have been proposed [426–431] both centralized and decentralized that improve privacy tracking and external attacks.

6.3 Direct influence vs common environment

Two influences can be detected using social network analysis. The first one is the influence that is shared directly by contact, such as in the case of Paper A, where bacteria jump from person to person. The other way is by sharing the same environment as with Result II; the inflammation process cannot jump from person to person but they both can share an environment in which allergens, and irritants, or toxic compounds are present, which can lead to common inflammation reactions. Realistically, in most cases, we will have a bit of both cases all the time. In Result II we can also have a shared microbe going around the schools. In Paper B students share an environment with the same amount of UVB, and even though we were not able to measure it, it is likely that they share a similar diet and PA.

A significant limitation of SNA is not being able to tell if direct influence or common environment or both are happening and needs the support of other classical statistical methods or direct measurements, to be able to tell which are the reasons why people have common levels of something. For example, in Result II we see that people belonging to the same school share inflammation processes as the school year progresses.

However, a significant advantage, is that social analysis can be done very quickly. It is unrealistic to expect constant monitoring of allergens, diets, pathogens, and so on in every environment in the population. But we can monitor and detect trends in people's health immediately. This might lead to us being able to tell that friends in a particular environment (school) are getting unhealthy levels or whatever concept we are interested in. If we were to only measure the general population instead of grouping by friends we could lose significant patterns as we show in **Result II**, in which analyses stratifying by sex only shows no significant results; even though the effects are already there. This could lead to being too late to do something of value to stop the health hazard in time.

6.4 Optimization of resources in public interventions

In the introduction (section 1.2) examples of how SNA optimized public health interventions and resources with little cost were presented. This is something that we would like to validate further with our results.

For example, in **Paper B** we see that women influence other women into going to the solarium, but men do not influence other men. Solariums are a horrible idea that leads to a significant increase in skin cancer lesions with insubstantial benefits [432]. As such we would like to convince the population to stop going to the solarium. Based on our results, an optimal approach would be to target women in an ad campaign rather than the general population. We see something similar in **Paper A** in which men are more common carriers of *S. aureus*, but women are more at risk due to social contact. In **Result III** we associated an increased risk of obesity due to social contact with overweight friends, and a lower risk with healthy-weight friends, so a better approach would be to increase network connectivity between these two groups. **Result IV** shows which high schools tend to have biased use towards painkillers, so we can concentrate efforts on drug prevention campaigns there.

A second important concept is the spillover effect [433–436] which is discussed in **Paper B**. In a network, especially one with a hierarchical topology, it is possible to target the top hierarchical behavior which would make the hanging nodes copy this behavior propagating the health effects throughout the network, instead of targeting all nodes at the same time which takes time and efforts which we might not be able to afford. This is similar to any marketing campaign in which internet influencers (top nodes) are paid to promote a product among the followers (hanging nodes).

6.5 Challenges in privacy

We encountered some data points that potentially allow for the identification of individual students. This exemplifies why the use of the Boyce - Codd normal form (BCNF), described in methodology (section 4.2.4), is important, as it mathematically guarantees that access to information contained in tables that a person shouldn't have access are kept in those tables and not outside of it.

These events are reported back to the head of the appropriate department and corrected. In a future manuscript, we will list all the "lessons learned" from the data cleaning process and help to develop better protocols for future epidemiology studies.

6.6 Challenges in reproducibility

In science, you have not discovered something until you discover it and somebody else reproduces it. All our code is open source (Affero GPL3.0 [437]), however, due to regulations in Norwegian law [438] and privacy ethics [439], the data is only available upon request. For example, a legal limitation is that subjects under 16 years old must sign special consent which includes limitation to the data access. This can limit the ability of other researchers to replicate and validate the results of our studies, which can lead to a lack of confidence in the findings. This also hinders the ability to build upon our methods. Any push to allow for more flexible data access can also hinder the trust of the public who gave us the data in the first place, hindering in this case any future project.

This is a very serious limitation that we need to overcome, and luckily there is one particular example that can be used to inspire future projects. There's however another project, the Tromsø Study [440], which has a more flexible approach. For example, all the metadata is [open and publicly available](#), including some basic descriptive statistics for many of the variables.

Despite the similarities between the two cases, data remains accessible only upon request. In the future, it would be beneficial to adopt a more open approach to data collection while ensuring the public's trust is maintained. In 2017, a systematic review found only one study discussing how data could be more open: "*this systematic review of the literature has uncovered a lack of evidence-based incentives for researchers to share data, which is ironic in an evidence-based world*" [441] Another similar study of

2020 [442], highlights the paradox of how open data is widely supported by researchers, publishers, governmental institutions, and is even a necessary condition for asking for funding; and yet the sharing of the data is heavily restricted due to publication pressure and fear for competition.

6.7 Challenges in framework developing

We have successfully developed a framework that automatizes most of the tedious scripting aspects. However, we found that R was quite limited in comparison with other programming languages [443]. As such, we are putting our effort into developing this framework into other languages in the future which would be less complicated than trying to keep updating and maintaining our current packages.

When it comes to performance, large-scaling computing is not an option in R, especially since R does not have built-in support for multithreading. While techniques such as the use of RCPP library [444] can integrate C++ into R, is just an unnecessary middleman. On top of this, R and the de-facto Integrated Development Environment (IDE) RStudio, have limited memory management capabilities, which can lead to memory leaks and slow down the performance of R scripts even further.

While performance is a deal breaker in the long run, this does not affect our data too much since we barely have about 1000x1000 tabular matrices which is not much. However, what pushes us to abandon further support for the future is the lack of proper object-oriented programming and lack of standardization. R does not have strong support for object-oriented programming. R aimed to retain the core functionalities and syntax of S while adding modern features and extending its capabilities. However, S was designed as a language for data analysis and a graphical representation [445], primarily used for statistical modeling and visualization which back in the 1970s did not have in mind the very strong paradigms of object-oriented programming that C++ would properly develop years later during the 1980s. As a result, R was developed in the 1990s, trying to use the 1970s syntax, but somewhat trying to appeal to modern functionality. Even though R does not have to deal with pointers, and memory management like in C++, R has a steeper learning curve than Python or C++ in the long run due to its lack of core personality. Making Python a much better option for beginners a C++ a better option for software professionals. In both cases, all the statistical-related functionality that R specializes in is also available in both of these languages.

From our own experience, R sometimes might appear difficult to work for people who know other programming languages like computer science professionals, but loved by people who are used to working with SPSS or Stata. We tried to satisfy this last group, but now the limitations have become clear, and we should be pushing scripting and programming to be done in Python or C/C++ as in both cases, in the long run, will produce better scripts and final products with much less hassle.

6.8 Future projects

Several small-molecule inhibitors and monoclonal antibodies against FnBPA [446, 447] are being tested in intrahospital environments [448]. All the social influence techniques we have discussed and presented so far, especially those related to *S. aureus* hospital-acquired infections [378, 449], can be applied to study nosocomial infections related to *S. aureus* to show the effectiveness of preventing transmission using these inhibitors.

SPMs is highly associated with a fish-rich diet. Friends tend to have similar diets, so the similar good inflammatory process might be due to sharing similar pro-inflammatory or anti-inflammatory diets. As discussed in Paper B, we need to refine the dietary data in FF or use another dataset, before we try this approach.

A forthcoming manuscript uses lessons learned in this thesis regarding FF for future epidemiological studies. This is of high interest because future epidemiological studies need to optimize the organization of the data better to minimize the work described in section 4.2, which is currently repeated over and over by different researchers. Other practices while analyzing the data need to be revised as well.

In particular, Boyce-Codd Normal Form is an advanced form of database normalization designed to reduce redundancy and eliminate undesirable characteristics like insertion, update, and deletion anomalies in relational databases. It minimizes information redundancy and helps in reducing the space and maintaining the consistency of the database. It makes it easy to detect anomalies in the data, and thanks to this effort in normalization, failures in the privacy protocol have been reported and fixed. There is also the importance of performance, which can be maximized only by using operations that can be performed more quickly due to the reduced number of joins and simpler query structures. The trade-off is that data can no longer be saved in a big blob of tabular data as currently is maintained and passed to researchers. This makes data selection a bit more complicated; if a researcher has permission to analyze diseases, now a committee

needs to select, not only the proper tables but also the proper columns in each.

We had very limited access to the FF2 dataset, which did not include the social network. We also did not get any data from FF3. There are plenty of studies that can be done using longitudinal data between FF1, FF2, and FF3. For example, on the topic of how is health affected at a later age by a person's social network as the year progresses?; how does *S. aureus* colonization evolve as the social network changes? , does inflammation markers change when we get new friends? does over-the-counter misuse get better or worse over time?

Chapter 7: Conclusions

This document showed the multiple applications that SNA have in our population within a broad range of topics. In Paper A we saw how *S. aureus* transmission is more associated with women's social interactions and how schools have distinctive strains measured by their Spa-typing. Result II also showed the importance of school interaction concerning immune response and inflammation as the academic year progresses. Result I and Result III showed the importance of friends in the quest for a healthy BMI. Paper B shows solarium influence in women and underlying influences related to social contacts about vitamin D levels. Result IV also hints at how hormonal contraceptive usage is shared among communities and some concerning usage of OTC medicine shared in schools.

We showed the application of non-parametric simulations and machine learning methods in studying social influences. We developed an easy-to-use framework in R and showed easy-to-use examples in Paper C, but due to several drawbacks, mainly maintenance of big libraries, is going to be ditched in favor of the Python and C++ implementations in the future.

We show several metrics to quantify these influences. However further work and discussion with health professionals are needed to determine the best way to impact policy-making based on these results at a lower cost.

In summary, this thesis provides evidence that suggests social influence is currently present in many topics, and accounting for this factor is as important, if not more, than accounting for other classical variables such as smoking or PA. We have also sadly experienced recently the necessity to account for social interactions during the SARS-CoV-2 pandemic, which also demonstrated a good example of why this type of analysis can be critical.

Bibliography

- [1] F. Jónsson, *De Gamle Eddadigte*, F. Jónsson, Ed. København, Dankmark: G.E.C. GADS Forlag, 1932. [Online]. Available: <http://www.septentrionalia.net/etexts/eddadigte.pdf>
- [2] G. JÓNSSON. (1943) Fáfnismál. [Online]. Available: <http://www.heimskringla.no/wiki/F%C3%A1fnism%C3%A1l>
- [3] ——. (1943) Fáfnismál. [Online]. Available: <http://www.heimskringla.no/wiki/V%C3%B3lusp%C3%A1>
- [4] J.-J. Rousseau., *Du contrat social; ou, Principes du droit politique*, M.-M. Rey, Ed. Amsterdam, Netherlands: Marc-Michel Rey, 1762. [Online]. Available: <http://metalibri.wikidot.com/title:du-contrat-social>
- [5] L. McGovern, “The relative contribution of multiple determinants to health,” Health Affairs, Tech. Rep., aug 2014. [Online]. Available: <https://doi.org/10.1377/hpb20140821.404487>
- [6] J. D. Wright, *International Encyclopedia of the Social and Behavioral Sciences*, J. D. Wright, Ed. University of Central Florida, Orlando, FL, USA: Elsevier Ltd, 2015. [Online]. Available: <https://www.sciencedirect.com/referencework/9780080970875/international-encyclopedia-of-the-social-and-behavioral-sciences>
- [7] (1613) Königsberg by joachim bering. https://upload.wikimedia.org/wikipedia/commons/1/15/Koenigsberg%2C_Map_by_Bering_1613.jpg.
- [8] V. C. Ovidiu Farc, *PRO-AND ANTITUMOR ROLE OF THE INTERLEUKINS 1 TO 41*. Institutul Cantacuzino, 2019.
- [9] Y.-M. Hyun and C.-W. Hong, “Deep insight into neutrophil trafficking in various organs,” *Journal of Leukocyte Biology*, vol. 102, no. 3, pp. 617–629, Jun. 2017.

- [Online]. Available: <https://doi.org/10.1189/jlb.1ru1216-521r>
- [10] M. Gaudet, M. Plesa, A. Mogas, N. Jalaleddine, Q. Hamid, and S. A. Healy, "Recent advances in vitamin d implications in chronic respiratory diseases," *Respiratory Research*, vol. 23, no. 1, Sep. 2022. [Online]. Available: <https://doi.org/10.1186/s12931-022-02147-x>
- [11] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, vol. 357, no. 4, pp. 370–379, Jul. 2007. [Online]. Available: <https://doi.org/10.1056/nejmsa066082>
- [12] J. G. Trogdon, J. Nonnemaker, and J. Pais, "Peer effects in adolescent overweight," *Journal of Health Economics*, vol. 27, no. 5, pp. 1388–1399, Sep. 2008. [Online]. Available: <https://doi.org/10.1016/j.jhealeco.2008.05.003>
- [13] N. A. Christakis and J. H. Fowler, "The collective dynamics of smoking in a large social network," *New England Journal of Medicine*, vol. 358, no. 21, pp. 2249–2258, May 2008. [Online]. Available: <https://doi.org/10.1056/nejmsa0706154>
- [14] K. A. Aschbrenner, C. Bobak, E. J. Schneider, J. A. Naslund, M. F. Brunette, and A. J. O'Malley, "Egocentric social networks and smoking among adults with serious mental illness," *Translational Behavioral Medicine*, vol. 8, no. 4, pp. 531–539, Apr. 2018. [Online]. Available: <https://doi.org/10.1093/tbm/ibx014>
- [15] J. N. Rosenquist, "The spread of alcohol consumption behavior in a large social network," *Annals of Internal Medicine*, vol. 152, no. 7, p. 426, Apr. 2010. [Online]. Available: <https://doi.org/10.7326/0003-4819-152-7-201004060-00007>
- [16] M. M. Ali, A. Amialchuk, and S. Nikaj, "Alcohol consumption and social network ties among adolescents: Evidence from add health," *Addictive Behaviors*, vol. 39, no. 5, pp. 918–922, May 2014. [Online]. Available: <https://doi.org/10.1016/j.addbeh.2013.11.030>
- [17] S. C. Mednick, N. A. Christakis, and J. H. Fowler, "The spread of sleep loss influences drug use in adolescent social networks," *PLoS ONE*, vol. 5, no. 3, p. e9775, Mar. 2010. [Online]. Available: <https://doi.org/10.1371/journal.pone.0009775>
- [18] N. A. Christakis and J. H. Fowler, "Social network sensors for early detection of

- contagious outbreaks," *PLoS ONE*, vol. 5, no. 9, p. e12948, Sep. 2010. [Online]. Available: <https://doi.org/10.1371/journal.pone.0012948>
- [19] S. R. Friedman, A. Neaigus, B. Jose, R. Curtis, M. Goldstein, G. Ildefonso, R. B. Rothenberg, and D. C. D. Jarlais, "Sociometric risk networks and risk for HIV infection." *American Journal of Public Health*, vol. 87, no. 8, pp. 1289–1296, Aug. 1997. [Online]. Available: <https://doi.org/10.2105/ajph.87.8.1289>
- [20] R. Fernández-Peña, J. L. Molina, and O. Valero, "Satisfaction with social support received from social relationships in cases of chronic pain: The influence of personal network characteristics in terms of structure, composition and functional content," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, p. 2706, Apr. 2020. [Online]. Available: <https://doi.org/10.3390/ijerph17082706>
- [21] S. Lin, L. Faust, P. Robles-Granda, T. Kajdanowicz, and N. V. Chawla, "Social network structure is predictive of health and wellness," *PLOS ONE*, vol. 14, no. 6, p. e0217264, Jun. 2019. [Online]. Available: <https://doi.org/10.1371/journal.pone.0217264>
- [22] S. Tasselli, "Social networks of professionals in health care organizations," *Medical Care Research and Review*, vol. 71, no. 6, pp. 619–660, Nov. 2014. [Online]. Available: <https://doi.org/10.1177/1077558714557079>
- [23] K. Sabot, D. Wickremasinghe, K. Blanchet, B. Avan, and J. Schellenberg, "Use of social network analysis methods to study professional advice and performance among healthcare providers: a systematic review," *Systematic Reviews*, vol. 6, no. 1, Oct. 2017. [Online]. Available: <https://doi.org/10.1186/s13643-017-0597-1>
- [24] R. F. Hunter, K. de la Haye, J. M. Murray, J. Badham, T. W. Valente, M. Clarke, and F. Kee, "Social network interventions for health behaviours and outcomes: A systematic review and meta-analysis," *PLOS Medicine*, vol. 16, no. 9, p. e1002890, Sep. 2019. [Online]. Available: <https://doi.org/10.1371/journal.pmed.1002890>
- [25] S. B. Gesell, S. L. Barkin, and T. W. Valente, "Social network diagnostics: a tool for monitoring group interventions," *Implementation Science*, vol. 8, no. 1, Oct. 2013. [Online]. Available: <https://doi.org/10.1186/1748-5908-8-116>

- [26] T. R. Moore, M. C. Pachucki, L. Calancie, A. R. Korn, E. Hennessy, and C. D. Economos, "Coalition-committees as network interventions: Baseline network composition in context of childhood obesity prevention interventions," *Systems*, vol. 9, no. 3, 2021. [Online]. Available: <https://www.mdpi.com/2079-8954/9/3/66>
- [27] N. R. Smith, P. N. Zivich, and L. Frerichs, "Social influences on obesity: Current knowledge, emerging methods, and directions for future research and practice," *Current Nutrition Reports*, vol. 9, no. 1, pp. 31–41, Jan. 2020. [Online]. Available: <https://doi.org/10.1007/s13668-020-00302-8>
- [28] J. McGlashan, K. de la Haye, P. Wang, and S. Allender, "Collaboration in complex systems: Multilevel network analysis for community-based obesity prevention interventions," *Scientific Reports*, vol. 9, no. 1, Aug. 2019. [Online]. Available: <https://doi.org/10.1038/s41598-019-47759-4>
- [29] R. M. Pinto, "USING SOCIAL NETWORK INTERVENTIONS TO IMPROVE MENTALLY ILL CLIENTS' WELL-BEING," *Clinical Social Work Journal*, vol. 34, no. 1, pp. 83–100, Dec. 2005. [Online]. Available: <https://doi.org/10.1007/s10615-005-0005-5>
- [30] R. J. Rosychuk, I. Colman, and B. H. Rowe, "Comparison of cluster detection using patients and events of medically treated self-inflicted injuries in alberta, canada," *Health Services and Outcomes Research Methodology*, vol. 9, no. 2, pp. 100–116, May 2009. [Online]. Available: <https://doi.org/10.1007/s10742-009-0048-0>
- [31] C. A. Latkin and A. R. Knowlton, "Social network assessments and interventions for health behavior change: A critical review," *Behavioral Medicine*, vol. 41, no. 3, pp. 90–97, Jul. 2015. [Online]. Available: <https://doi.org/10.1080/08964289.2015.1034645>
- [32] R. S. Sadashivam, S. L. Cutrona, T. M. Luger, E. Volz, R. Kinney, S. R. Rao, J. J. Allison, and T. K. Houston, "Share2quit: Online social network peer marketing of tobacco cessation systems," *Nicotine and Tobacco Research*, p. nt187, Jul. 2016. [Online]. Available: <https://doi.org/10.1093/ntr/ntw187>
- [33] L. Danon, T. A. House, J. M. Read, and M. J. Keeling, "Social encounter networks: collective properties and disease transmission," *Journal of The Royal*

- Society Interface*, vol. 9, no. 76, pp. 2826–2833, Jun. 2012. [Online]. Available: <https://doi.org/10.1098/rsif.2012.0357>
- [34] K. Wang, K. Brown, S.-Y. Shen, and J. Tucker, “Social network-based interventions to promote condom use: A systematic review,” *AIDS and Behavior*, vol. 15, no. 7, pp. 1298–1308, Aug. 2011. [Online]. Available: <https://doi.org/10.1007/s10461-011-0020-1>
- [35] A. Llupià, J. Puig, G. Mena, J. M. Bayas, and A. Trilla, “The social network around influenza vaccination in health care workers: a cross-sectional study,” *Implementation Science*, vol. 11, no. 1, Nov. 2016. [Online]. Available: <https://doi.org/10.1186/s13012-016-0522-3>
- [36] B. Ljubic, D. Gligorijevic, J. Gligorijevic, M. Pavlovski, and Z. Obradovic, “Social network analysis for better understanding of influenza,” *Journal of Biomedical Informatics*, vol. 93, p. 103161, May 2019. [Online]. Available: <https://doi.org/10.1016/j.jbi.2019.103161>
- [37] N. Marquetoux, M. A. Stevenson, P. Wilson, A. Ridler, and C. Heuer, “Using social network analysis to inform disease control interventions,” *Preventive Veterinary Medicine*, vol. 126, pp. 94–104, Apr. 2016. [Online]. Available: <https://doi.org/10.1016/j.prevetmed.2016.01.022>
- [38] A. Ortiz-Pelaez, D. Pfeiffer, R. Soares-Magalhães, and F. Guitian, “Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (FMD) epidemic in the UK,” *Preventive Veterinary Medicine*, vol. 76, no. 1-2, pp. 40–55, Sep. 2006. [Online]. Available: <https://doi.org/10.1016/j.prevetmed.2006.04.007>
- [39] C. Corcoran and J. M. Clark, “Adaptive network modeling of social distancing interventions,” *Journal of Theoretical Biology*, vol. 546, p. 111151, Aug. 2022. [Online]. Available: <https://doi.org/10.1016/j.jtbi.2022.111151>
- [40] D. Centola, “Considering network interventions,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 52, pp. 32 833–32 835, Dec. 2020. [Online]. Available: <https://doi.org/10.1073/pnas.2022584118>
- [41] R. J. Trudeau, *Introduction to graph theory*, 2nd ed., ser. Dover Books on Mathematics. Mineola, NY: Dover Publications, Feb. 1994.

- [42] K. Faust, "A puzzle concerning triads in social networks: Graph constraints and the triad census," *Social Networks*, vol. 32, no. 3, pp. 221–233, Jul. 2010. [Online]. Available: <https://doi.org/10.1016/j.socnet.2010.03.004>
- [43] ——, "7. very local structure in social networks," *Sociological Methodology*, vol. 37, no. 1, pp. 209–256, Aug. 2007. [Online]. Available: <https://doi.org/10.1111/j.1467-9531.2007.00179.x>
- [44] T. M. J. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and Experience*, vol. 21, no. 11, pp. 1129–1164, Nov. 1991. [Online]. Available: <https://doi.org/10.1002/spe.4380211102>
- [45] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," *Information Processing Letters*, vol. 31, no. 1, pp. 7–15, Apr. 1989. [Online]. Available: [https://doi.org/10.1016/0020-0190\(89\)90102-6](https://doi.org/10.1016/0020-0190(89)90102-6)
- [46] S. B. Martin, W. M. Brown, R. Klavans, and K. W. Boyack, "Drl: Distributed recursive (graph) layout." *Journal of Graph Algorithms and Applications*, 5 2008. [Online]. Available: <https://www.osti.gov/biblio/1145621>
- [47] T. N. a. Gábor Csárdi [aut], "The graphopt layout algorithm," Apr 2023. [Online]. Available: https://rdrr.io/cran/igraph/man/layout_with_graphopt.html
- [48] A. Buja, D. F. Swayne, M. L. Littman, N. Dean, H. Hofmann, and L. Chen, "Data visualization with multidimensional scaling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 444–472, Jun. 2008. [Online]. Available: <https://doi.org/10.1198/106186008x318440>
- [49] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, Jan. 1978. [Online]. Available: [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- [50] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *The Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, Jan. 1972. [Online]. Available: <https://doi.org/10.1080/0022250x.1972.9989806>
- [51] J. Stehlé, F. Charbonnier, T. Picard, C. Cattuto, and A. Barrat, "Gender homophily from spatial behavior in a primary school: A sociometric study," *Social Networks*, vol. 35, no. 4, pp. 604–613, Oct. 2013. [Online]. Available:

- <https://doi.org/10.1016/j.socnet.2013.08.003>
- [52] J. Moody, "Race, school integration, and friendship segregation in america," *American Journal of Sociology*, vol. 107, no. 3, pp. 679–716, Nov. 2001. [Online]. Available: <https://doi.org/10.1086/338954>
- [53] Z. Qian and D. T. Lichter, "Social boundaries and marital assimilation: Interpreting trends in racial and ethnic intermarriage," *American Sociological Review*, vol. 72, no. 1, pp. 68–94, Feb. 2007. [Online]. Available: <https://doi.org/10.1177/000312240707200104>
- [54] J. E. Cheadle and P. Schwadel, "The 'friendship dynamics of religion,' or the 'religious dynamics of friendship'? a social network analysis of adolescents who attend small schools," *Social Science Research*, vol. 41, no. 5, pp. 1198–1212, Sep. 2012. [Online]. Available: <https://doi.org/10.1016/j.ssresearch.2012.03.014>
- [55] J. M. McPherson and L. Smith-Lovin, "Homophily in voluntary organizations: Status distance and the composition of face-to-face groups," *American Sociological Review*, vol. 52, no. 3, p. 370, Jun. 1987. [Online]. Available: <https://doi.org/10.2307/2095356>
- [56] P. P. Sergio Curranini, Matthew O. Jackson, "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, vol. 77, no. 4, pp. 1003–1045, 2009. [Online]. Available: <https://doi.org/10.3982/ecta7528>
- [57] G. Kossinets and D. J. Watts, "Origins of homophily in an evolving social network," *American Journal of Sociology*, vol. 115, no. 2, pp. 405–450, Sep. 2009. [Online]. Available: <https://doi.org/10.1086/599247>
- [58] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, Aug. 2001. [Online]. Available: <https://doi.org/10.1146/annurev.soc.27.1.415>
- [59] J. A. Smith, M. McPherson, and L. Smith-Lovin, "Social distance in the united states," *American Sociological Review*, vol. 79, no. 3, pp. 432–456, Apr. 2014. [Online]. Available: <https://doi.org/10.1177/0003122414531776>
- [60] F. Karimi, M. Génois, C. Wagner, P. Singer, and M. Strohmaier, "Homophily influences ranking of minorities in social networks," *Scientific Reports*, vol. 8, no. 1,

- Jul. 2018. [Online]. Available: <https://doi.org/10.1038/s41598-018-29405-7>
- [61] E. Lee, F. Karimi, C. Wagner, H.-H. Jo, M. Strohmaier, and M. Galesic, "Homophily and minority-group size explain perception biases in social networks," *Nature Human Behaviour*, vol. 3, no. 10, pp. 1078–1087, Aug. 2019. [Online]. Available: <https://doi.org/10.1038/s41562-019-0677-4>
- [62] C. Avin, H. Daltrophe, B. Keller, Z. Lotker, C. Mathieu, D. Peleg, and Y.-A. Pignolet, "Mixed preferential attachment model: Homophily and minorities in social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 555, p. 124723, Oct. 2020. [Online]. Available: <https://doi.org/10.1016/j.physa.2020.124723>
- [63] A. Asikainen, G. Iñiguez, J. Ureña-Carrión, K. Kaski, and M. Kivelä, "Cumulative effects of triadic closure and homophily in social networks," *Science Advances*, vol. 6, no. 19, May 2020. [Online]. Available: <https://doi.org/10.1126/sciadv.aax7310>
- [64] D. Bevan, "Introduction to random graphs by alan frieze and michał karoński, pp. 464, £49.99 (hard), ISBN 978-1-107-11850-8, cambridge university press (2016)." *The Mathematical Gazette*, vol. 101, no. 551, pp. 376–377, Jun. 2017. [Online]. Available: <https://doi.org/10.1017/mag.2017.110>
- [65] T. A. Snijders and S. P. Borgatti, "Non-parametric standard errors and tests for network statistics," 1999. [Online]. Available: http://www.analytictech.com/borgatti/papers/snijders_borgatti-density_significance.pdf
- [66] P. Erdős and A. Rényi, "On random graphs. i." *Publicationes Mathematicae Debrecen*, vol. 6, no. 3–4, p. 290–297, Jul. 2022. [Online]. Available: <http://dx.doi.org/10.5486/PMD.1959.6.3-4.12>
- [67] *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge University Press, Nov. 2012. [Online]. Available: <http://dx.doi.org/10.1017/CBO9780511894701>
- [68] P. A. Munz, I. Hudea, J. P. Imad, and R. J. Smith, "When zombies attack!: Mathematical modelling of an outbreak of zombie infection," 2009.
- [69] S. M. Stigler, *The history of statistics*. London, England: Harvard University

- Press, Jul. 1986.
- [70] S. S. Al-Zaiti, A. A. Alghwiri, X. Hu, G. Clermont, A. Peace, P. Macfarlane, and R. Bond, "A clinician's guide to understanding and critically appraising machine learning studies: a checklist for ruling out bias using standard tools in machine learning (ROBUST-ML)," *European Heart Journal - Digital Health*, vol. 3, no. 2, pp. 125–140, Apr. 2022. [Online]. Available: <https://doi.org/10.1093/ehjdh/ztac016>
- [71] A. Somani, A. Horsch, and D. K. Prasad, *Interpretability in Deep Learning*. Springer International Publishing, 2023. [Online]. Available: <https://doi.org/10.1007/978-3-031-20639-9>
- [72] O. D. T. Catala, I. S. Igual, F. J. Perez-Benito, D. M. Escriva, V. O. Castello, R. Llobet, and J.-C. Perez-Cortes, "Bias analysis on public x-ray image datasets of pneumonia and covid-19 patients," *IEEE Access*, vol. 9, p. 42370–42383, 2021. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2021.3065456>
- [73] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics*, vol. 29, no. 5, Oct. 2001. [Online]. Available: <http://dx.doi.org/10.1214/aos/1013203451>
- [74] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, p. 44–65, Jan. 2015. [Online]. Available: <http://dx.doi.org/10.1080/10618600.2014.907095>
- [75] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," 2016. [Online]. Available: <https://arxiv.org/abs/1602.04938>
- [76] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
- [77] L. S. Shapley, *A value for n-person games*. Cambridge University Press, Oct. 1988, p. 31–40. [Online]. Available: <http://dx.doi.org/10.1017/CBO9780511528446>.

003

- [78] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowledge and Information Systems*, vol. 41, no. 3, p. 647–665, Aug. 2013. [Online]. Available: <http://dx.doi.org/10.1007/s10115-013-0679-x>
- [79] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," 2018.
- [80] M. Sundararajan and A. Najmi, "The many shapley values for model explanation," 2019.
- [81] D. Janzing, L. Minorics, and P. Blöbaum, "Feature relevance quantification in explainable ai: A causal problem," 2019.
- [82] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: <https://doi.org/10.1023/a:1010933404324>
- [83] M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," *Journal of Statistical Software*, vol. 36, no. 11, 2010. [Online]. Available: <http://dx.doi.org/10.18637/jss.v036.i11>
- [84] M. Bonamente, "Contingency tables and diagnostic tests," in *Statistics and Analysis of Scientific Data*. Springer Nature Singapore, 2022, pp. 181–211. [Online]. Available: https://doi.org/10.1007/978-981-19-0365-6_10
- [85] V. Bewick, L. Cheek, and J. Ball, "Medical statistics, statistics review 14: Logistic regression," *Critical Care*, vol. 9, no. 1, p. 112, 2005. [Online]. Available: <https://doi.org/10.1186/cc3045>
- [86] L. Anselin, *Spatial Econometrics: Methods and Models*, ser. Studies in Operational Regional Science. Springer Netherlands, 1988. [Online]. Available: <https://books.google.no/books?id=3dPIXClv4YYC>
- [87] A. J. O'Malley and P. V. Marsden, "The analysis of social networks," *Health Services and Outcomes Research Methodology*, vol. 8, no. 4, pp. 222–269, Oct. 2008. [Online]. Available: <https://doi.org/10.1007/s10742-008-0041-z>
- [88] S. Engebretsen, A. Frigessi, K. Engø-Monsen, A.-S. Furberg, A. Stubhaug, B. F.

- de Blasio, and C. S. Nielsen, "The peer effect on pain tolerance," *Scandinavian Journal of Pain*, vol. 18, no. 3, p. 467–477, May 2018. [Online]. Available: <http://dx.doi.org/10.1515/sjpain-2018-0060>
- [89] R. J. Rossi, *Mathematical statistics*. Nashville, TN: John Wiley & Sons, Dec. 2018.
- [90] J. Besag, "Statistical analysis of non-lattice data," *The Statistician*, vol. 24, no. 3, p. 179, Sep. 1975. [Online]. Available: <http://dx.doi.org/10.2307/2987782>
- [91] F. Hayashi, *Econometrics*. Princeton, NJ: Princeton University Press, Oct. 2000.
- [92] O. Kadioglu, S. Jacob, S. Bohnert, J. Naß, M. E. Saeed, H. Khalid, I. Merfort, E. Thines, T. Pommerening, and T. Efferth, "Evaluating ancient egyptian prescriptions today: Anti-inflammatory activity of ziziphus spina-christi," *Phytomedicine*, vol. 23, no. 3, pp. 293–306, Mar. 2016. [Online]. Available: <https://doi.org/10.1016/j.phymed.2016.01.004>
- [93] A. Touwaide and N. G. de Santo, "Edema in the corpus hippocraticum," *American Journal of Nephrology*, vol. 19, no. 2, pp. 155–158, 1999. [Online]. Available: <https://doi.org/10.1159/000013442>
- [94] D. N. Granger and E. Senchenkova, "Inflammation and the microcirculation," *Colloquium Series on Integrated Systems Physiology: From Molecule to Function*, vol. 2, no. 1, pp. 1–87, Jul. 2010. [Online]. Available: <https://doi.org/10.4199/c00013ed1v01y201006isp008>
- [95] C. N. Serhan, "Pro-resolving lipid mediators are leads for resolution physiology," *Nature*, vol. 510, no. 7503, pp. 92–101, Jun. 2014. [Online]. Available: <https://doi.org/10.1038/nature13479>
- [96] E. M. Smith, M. Gregg, F. Hashemi, L. Schott, and T. K. Hughes, "Corticotropin releasing factor (CRF) activation of NF-kb-directed transcription in leukocytes," *Cellular and Molecular Neurobiology*, vol. 26, no. 4–6, pp. 1019–1034, Apr. 2006. [Online]. Available: <https://doi.org/10.1007/s10571-006-9040-1>
- [97] A. Roumier, C. Béchade, and L. Maroteaux, "Serotonin and the immune system," in *Serotonin*. Elsevier, 2019, pp. 181–196. [Online]. Available: <https://doi.org/10.1016/b978-0-12-800050-2.00010-3>

- [98] M. C. Basil and B. D. Levy, "Specialized pro-resolving mediators: endogenous regulators of infection and inflammation," *Nature Reviews Immunology*, vol. 16, no. 1, pp. 51–67, Dec. 2015. [Online]. Available: <https://doi.org/10.1038/nri.2015.4>
- [99] N. Sharma-walia and J. Chandrasekharan, "Lipoxins: nature's way to resolve inflammation," *Journal of Inflammation Research*, p. 181, Sep. 2015. [Online]. Available: <https://doi.org/10.2147/jir.s90380>
- [100] K. Moro, M. Nagahashi, R. Ramanathan, K. Takabe, and T. Wakai, "Resolvins and omega three polyunsaturated fatty acids: Clinical implications in inflammatory diseases and cancer," *World Journal of Clinical Cases*, vol. 4, no. 7, p. 155, 2016. [Online]. Available: <https://doi.org/10.12998/wjcc.v4.i7.155>
- [101] M. Lagarde, E. Véricel, M. Liu, P. Chen, and M. Guichardant, "Structure-function relationships of non-cyclic dioxygenase products from polyunsaturated fatty acids: Poxytrins as a class of bioactive derivatives," *Biochimie*, vol. 107, pp. 91–94, Dec. 2014. [Online]. Available: <https://doi.org/10.1016/j.biochi.2014.09.008>
- [102] C. N. Serhan, J. Dalli, R. A. Colas, J. W. Winkler, and N. Chiang, "Protectins and maresins: New pro-resolving families of mediators in acute inflammation and resolution bioactive metabolome," *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1851, no. 4, pp. 397–413, Apr. 2015. [Online]. Available: <https://doi.org/10.1016/j.bbalip.2014.08.006>
- [103] S. Feltenmark, N. Gautam, Å. Brunnström, W. Griffiths, L. Backman, C. Edenius, L. Lindbom, M. Björkholm, and H.-E. Claesson, "Eoxins are proinflammatory arachidonic acid metabolites produced via the 15-lipoxygenase-1 pathway in human eosinophils and mast cells," *Proceedings of the National Academy of Sciences*, vol. 105, no. 2, pp. 680–685, Jan. 2008. [Online]. Available: <https://doi.org/10.1073/pnas.0710127105>
- [104] H.-E. Claesson, "On the biosynthesis and biological role of eoxins and 15-lipoxygenase-1 in airway inflammation and hodgkin lymphoma," *Prostaglandins and Other Lipid Mediators*, vol. 89, no. 3-4, pp. 120–125, Sep. 2009. [Online]. Available: <https://doi.org/10.1016/j.prostaglandins.2008.12.003>
- [105] K. Ostrowski, P. Schjerling, and B. K. Pedersen, "Physical activity and plasma interleukin-6 in humans - effect of intensity of exercise," *European Journal of*

- Applied Physiology*, vol. 83, no. 6, pp. 512–515, Dec. 2000. [Online]. Available: <https://doi.org/10.1007/s004210000312>
- [106] K. Ostrowski, T. Rohde, S. Asp, P. Schjerling, and B. K. Pedersen, “Pro- and anti-inflammatory cytokine balance in strenuous exercise in humans,” *The Journal of Physiology*, vol. 515, no. 1, pp. 287–291, Feb. 1999. [Online]. Available: <https://doi.org/10.1111/j.1469-7793.1999.287ad.x>
- [107] A. L. Kiss, “Inflammation in focus: The beginning and the end,” *Pathology and Oncology Research*, vol. 27, Jan. 2022. [Online]. Available: <https://doi.org/10.3389/pore.2021.1610136>
- [108] R. Li and J. Emsley, “The organizing principle of the platelet glycoprotein Ib–IX–v complex,” *Journal of Thrombosis and Haemostasis*, vol. 11, no. 4, pp. 605–614, Apr. 2013. [Online]. Available: <https://doi.org/10.1111/jth.12144>
- [109] N. C. for Biotechnology Information, “Genetics and Genomics in Medicine,” <https://www.ncbi.nlm.nih.gov/books/NBK7014/>, [Online; accessed June 22, 2023].
- [110] D. Li and M. Wu, “Pattern recognition receptors in health and diseases,” *Signal Transduction and Targeted Therapy*, vol. 6, no. 1, Aug. 2021. [Online]. Available: <https://doi.org/10.1038/s41392-021-00687-0>
- [111] J. Y. Cho, *Recent Advancements in Microbial Diversity*, E. Inc, Ed. Elsevier, 2022. [Online]. Available: <https://doi.org/10.1016/c2019-0-04189-1>
- [112] M. Miyasaka and K. Takatsu, Eds., *Chronic Inflammation*. Springer Japan, 2016. [Online]. Available: <https://doi.org/10.1007/978-4-431-56068-5>
- [113] Y. Li, C. Klein, and D. Kotlarz, “Dysregulation of cell death in human chronic inflammation,” *Cold Spring Harbor Perspectives in Biology*, vol. 12, no. 7, p. a037036, Dec. 2019. [Online]. Available: <https://doi.org/10.1101/cshperspect.a037036>
- [114] A. M. Nouri, G. S. Panayi, and S. M. Goodman, “Cytokines and the chronic inflammation of rheumatic disease. i. the presence of interleukin-1 in synovial fluids,” *Clin. Exp. Immunol.*, vol. 55, no. 2, pp. 295–302, Feb. 1984.
- [115] M. Aliyu, F. T. Zohora, A. U. Anka, K. Ali, S. Maleknia, M. Saffarioun, and G. Azizi, “Interleukin-6 cytokine: An overview of the immune

- regulation, immune dysregulation, and therapeutic approach," *International Immunopharmacology*, vol. 111, p. 109130, Oct. 2022. [Online]. Available: <https://doi.org/10.1016/j.intimp.2022.109130>
- [116] A. Lyell, "Alexander ogston, micrococci, and joseph lister," *Journal of the American Academy of Dermatology*, vol. 20, no. 2, pp. 302–310, Feb. 1989. [Online]. Available: [https://doi.org/10.1016/s0190-9622\(89\)70035-9](https://doi.org/10.1016/s0190-9622(89)70035-9)
- [117] W. M. M. Kirby, "Extraction of a highly potent penicillin inactivator from penicillin resistant staphylococci," *Science*, vol. 99, no. 2579, pp. 452–453, Jun. 1944. [Online]. Available: <https://doi.org/10.1126/science.99.2579.452>
- [118] A. F. N. Author's Last Name, *Intellectual Property Rights and the Life Science Industries. Past, Present, and Future. 2nd Edition.* 5 Toh Tuck Link, Singapore 596224: World Scientific Publishing, 2009.
- [119] M. Shoaib, A. I. Aqib, I. Muzammil, N. Majeed, Z. A. Bhutta, M. F. e Alam Kulyar, M. Fatima, C.-N. F. Zaheer, A. Muneer, M. Murtaza, M. Kashif, F. Shafqat, and W. Pu, "MRSA compendium of epidemiology, transmission, pathophysiology, treatment, and prevention within one health framework," *Frontiers in Microbiology*, vol. 13, Jan. 2023. [Online]. Available: <https://doi.org/10.3389/fmicb.2022.1067284>
- [120] J. Kluytmans, A. van Belkum, and H. Verbrugh, "Nasal carriage of staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks," *Clinical Microbiology Reviews*, vol. 10, no. 3, pp. 505–520, Jul. 1997. [Online]. Available: <https://doi.org/10.1128/cmr.10.3.505>
- [121] E. Klein, D. L. Smith, and R. Laxminarayan, "Hospitalizations and deaths caused by methicillin-resistant staphylococcus aureus, united states, 1999–2005," *Emerging Infectious Diseases*, vol. 13, no. 12, pp. 1840–1846, Dec. 2007. [Online]. Available: <https://doi.org/10.3201/eid1312.070629>
- [122] T. Aratani, H. Tsukamoto, T. Higashi, T. Kodawara, R. Yano, Y. Hida, H. Iwasaki, and N. Goto, "Association of methicillin resistance with mortality of hospital-acquired staphylococcus aureus bacteremia," *Journal of International Medical Research*, vol. 49, no. 11, p. 030006052110588, Nov. 2021. [Online]. Available: <https://doi.org/10.1177/03000605211058872>

- [123] K. S. Ikuta, L. R. Swetschinski, G. Robles Aguilar, F. Sharara, T. Mestrovic, A. P. Gray, N. Davis Weaver, E. E. Wool, C. Han, A. Gershberg Hayoon, A. Aali, S. M. Abate, M. Abbasi-Kangevari, Z. Abbasi-Kangevari, S. Abd-Elsalam, G. Abebe, A. Abedi, A. P. Abhari, H. Abidi, R. G. Aboagye, A. Absalan, H. Abubaker Ali, J. M. Acuna, T. D. Adane, I. Y. Addo, O. A. Adegbeye, M. Adnan, Q. E. S. Adnani, M. S. Afzal, S. Afzal, Z. B. Aghdam, B. O. Ahinkorah, A. Ahmad, A. R. Ahmad, R. Ahmad, S. Ahmad, S. Ahmad, S. Ahmadi, A. Ahmed, H. Ahmed, J. Q. Ahmed, T. Ahmed Rashid, M. Ajami, B. Aji, M. Akbarzadeh-Khiavi, C. J. Akunna, H. Al Hamad, F. Alahdab, Z. Al-Aly, M. A. Aldeyab, A. V. Aleman, F. A. N. Alhalaiqa, R. K. Alhassan, B. A. Ali, L. Ali, S. S. Ali, Y. Alimohamadi, V. Alipour, A. Alizadeh, S. M. Aljunid, K. Allel, S. Almustanyir, E. K. Ameyaw, A. M. L. Amit, N. Anandavelane, R. Ancceanu, C. L. Andrei, T. Andrei, D. Anggraini, A. Ansar, A. E. Anyasodor, J. Arabloo, A. Y. Aravkin, D. Areda, T. Aripov, A. A. Artamonov, J. Arulappan, R. T. Aruleba, M. Asaduzzaman, T. Ashraf, S. S. Athari, D. Atlaw, S. Attia, M. Ausloos, T. Awoke, B. P. Ayala Quintanilla, T. M. Ayana, S. Azadnajafabad, A. Azari Jafari, D. B. B, M. Badar, A. D. Badiye, N. Baghcheghi, S. Bagherieh, A. A. Baig, I. Banerjee, A. Barac, M. Bardhan, F. Barone-Adesi, H. J. Barqawi, A. Barrow, P. Baskaran, S. Basu, A.-M. M. Batiha, N. Bedi, M. A. Belete, U. I. Belgaumi, R. G. Bender, B. Bhandari, D. Bhandari, P. Bhardwaj, S. Bhaskar, K. Bhattacharyya, S. Bhattarai, S. Bitaraf, D. Buonsenso, Z. A. Butt, F. L. Caetano dos Santos, J. Cai, D. Calina, P. Camargos, L. A. Cámera, R. Cárdenas, M. Cevik, J. Chadwick, J. Charan, A. Chaurasia, P. R. Ching, S. G. Choudhari, E. K. Chowdhury, F. R. Chowdhury, D.-T. Chu, I. S. Chukwu, O. Dadras, F. T. Dagnaw, X. Dai, S. Das, A. Dastiridou, S. A. Debela, F. W. Demisse, S. Demissie, D. Dereje, M. Derese, H. D. Desai, F. N. Dessalegn, S. A. A. Dessalegni, B. Desye, K. Dhaduk, M. Dhimal, S. Dhingra, N. Diao, D. Diaz, S. Djalalinia, M. Dodangeh, D. Dongarwar, B. T. Dora, F. Dorostkar, H. L. Dsouza, E. Dubljanin, S. J. Dunachie, O. C. Durojaiye, H. A. Edinur, H. B. Ejigu, M. Ekholuenetale, T. C. Ekundayo, H. El-Abid, M. Elhadi, M. A. Elmonem, A. Emami, L. Engelbert Bain, D. B. Enyew, R. Erkembayar, B. Eshrati, F. Etaee, A. F. Fagbamigbe, S. Falahi, A. Fallahzadeh, E. J. A. Faraon, A. Fatehizadeh, G. Fekadu, J. C. Fernandes, A. Ferrari, G. Fetensa, I. Filip, F. Fischer, M. Foroutan, P. A. Gaal, M. A. Gadanya, A. M. Gaidhane, B. Ganesan, M. Gebrehiwot, R. Ghanbari, M. Ghasemi Nour, A. Ghashghaei, A. Gholamrezanezhad, A. Gholizadeh, M. Golechha, P. Goleij, D. Golinelli,

- A. Goodridge, D. A. Gunawardane, Y. Guo, R. D. Gupta, S. Gupta, V. B. Gupta, V. K. Gupta, A. Guta, P. Habibzadeh, A. Haddadi Avval, R. Halwani, A. Hanif, M. A. Hannan, H. Harapan, S. Hassan, H. Hassankhani, K. Hayat, B. Heibati, G. Heidari, M. Heidari, R. Heidari-Soureshjani, C. Herteliu, D. Z. Heyi, K. Hezam, P. Hoogar, N. Horita, M. M. Hossain, M. Hosseinzadeh, M. Hostiuc, S. Hostiuc, S. Hoveidamanesh, J. Huang, S. Hussain, N. R. Hussein, S. E. Ibitoye, O. S. Ilesanmi, I. M. Ilic, M. D. Ilic, M. T. Imam, M. Immurana, L. R. Inbaraj, A. Iradukunda, N. E. Ismail, C. C. D. Iwu, C. J. Iwu, L. M. J, M. Jakovljevic, E. Jamshidi, T. Javaheri, F. Javanmardi, J. Javidnia, S. K. Jayapal, U. Jayarajah, R. Jebai, R. P. Jha, T. Joo, N. Joseph, F. Joukar, J. J. Jozwiak, S. E. O. Kacimi, V. Kadashetti, L. R. Kalankesh, R. Kalhor, V. K. Kamal, H. Kandel, N. Kapoor, S. Karkhah, B. G. Kassa, N. J. Kassebaum, P. D. Katoto, M. Keykhaei, H. Khajuria, A. Khan, I. A. Khan, M. Khan, M. N. Khan, M. A. Khan, M. M. Khatatbeh, M. M. Khater, H. R. Khayat Kashani, J. Khubchandani, H. Kim, M. S. Kim, R. W. Kimokoti, N. Kissoon, S. Kochhar, F. Kompani, S. Kosen, P. A. Koul, S. L. Koulmane Laxminarayana, F. Krapp Lopez, K. Krishnan, V. Krishnamoorthy, V. Kulkarni, N. Kumar, O. P. Kurmi, A. Kuttikkattu, H. H. Kyu, D. K. Lal, J. Lám, I. Landires, S. Lasrado, S.-w. Lee, J. Lenzi, S. Lewycka, S. Li, S. S. Lim, W. Liu, R. Lodha, M. J. Loftus, A. Lohiya, L. Lorenzovici, M. Lotfi, A. Mahmoodpoor, M. A. Mahmoud, R. Mahmoudi, A. Majeed, J. Majidpoor, A. Makki, G. A. Mamo, Y. Manla, M. Martorell, C. N. Matei, B. McManigal, E. Mehrabi Nasab, R. Mehrotra, A. Melese, O. Mendoza-Cano, R. G. Menezes, A.-F. A. Menthis, G. Micha, I. M. Michalek, A. C. Micheletti Gomide Nogueira de Sá, N. Milevska Kostova, S. A. Mir, M. Mirghafourvand, S. Mirmoeeni, E. M. Mirrakhimov, M. Mirza-Aghazadeh-Attari, A. S. Misganaw, A. Misganaw, S. Misra, E. Mohammadi, M. Mohammadi, A. Mohammadian-Hafshejani, S. Mohammed, S. Mohan, M. Mohseni, A. H. Mokdad, S. Momtazmanesh, L. Monasta, C. E. Moore, M. Moradi, M. Moradi Sarabi, S. D. Morrison, M. Motaghinejad, H. Mousavi Isfahani, A. Mousavi Khaneghah, S. A. Mousavi-Aghdas, S. Mubarik, F. Mulita, G. B. B. Mulu, S. B. Munro, S. Muthupandian, T. S. Nair, A. A. Naqvi, H. Narang, Z. S. Natto, M. Naveed, B. P. Nayak, S. Naz, I. Negoi, S. A. Nejadghaderi, S. Neupane Kandel, C. H. Ngwa, R. K. Niazi, A. T. Nogueira de Sá, N. Noroozi, H. Nouraei, A. Nowroozi, V. Nuñez-Samudio, J. J. Nutor, C. I. Nzoputam, O. J. Nzoputam, B. Oancea, R. M. Obaidur, V. A. Ojha, A. P. Okekunle, O. C. Okonji, A. T. Olagunju, B. O. Olusanya,

- A. Omar Bali, E. Omer, N. Otstavnov, B. Oumer, M. P A, J. R. Padubidri, K. Pakshir, T. Palicz, A. Pana, S. Pardhan, J. L. Paredes, U. Parekh, E.-C. Park, S. Park, A. Pathak, R. Paudel, U. Paudel, S. Pawar, H. Pazoki Toroudi, M. Peng, U. Pensato, V. C. F. Pepito, M. Pereira, M. F. P. Peres, N. Perico, I.-R. Petcu, Z. Z. Piracha, I. Podder, N. Pokhrel, R. Poluru, M. J. Postma, N. Pourtaheri, A. Prashant, I. Qattea, M. Rabiee, N. Rabiee, A. Radfar, S. Raeghi, S. Rafiei, P. R. Raghav, L. Rahbarnia, V. Rahimi-Movaghar, M. Rahman, M. A. Rahman, A. M. Rahmani, V. Rahmanian, P. Ram, M. M. A. N. Ranjha, S. J. Rao, M.-M. Rashidi, A. Rasul, Z. A. Ratan, S. Rawaf, R. Rawassizadeh, M. S. Razeghinia, E. M. M. Redwan, M. T. Regasa, G. Remuzzi, M. A. Reta, N. Rezaei, A. Rezapour, A. Riad, R. K. Ripon, K. E. Rudd, B. Saddik, S. Sadeghian, U. Saeed, M. Safaei, A. Safary, S. Z. Safi, M. Sahebazzamani, A. Sahebkar, H. Sahoo, S. Salahi, S. Salahi, H. Salari, S. Salehi, H. Samadi Kafil, A. M. Samy, N. Sanadgol, S. Sankararaman, F. Sanmarchi, B. Sathian, M. Sawhney, G. K. Saya, S. Senthilkumaran, A. Seylani, P. A. Shah, M. A. Shaikh, E. Shaker, M. Z. Shakhmardanov, M. M. Sharew, A. Sharifi-Razavi, P. Sharma, R. A. Sheikhi, A. Sheikhy, P. H. Shetty, M. Shigematsu, J. I. Shin, H. Shirzad-Aski, K. M. Shivakumar, P. Shobeiri, S. A. Shorofi, S. Shrestha, M. M. Sibhat, N. B. Sidemo, M. K. Sikder, L. M. L. R. Silva, J. A. Singh, P. Singh, S. Singh, M. S. Siraj, S. S. Siwal, V. Y. Skryabin, A. A. Skryabina, B. Socea, D. D. Solomon, Y. Song, C. T. Sreeramareddy, M. Suleman, R. Suliankatchi Abdulkader, S. Sultana, M. Szócska, S.-A. Tabatabaeizadeh, M. Tabish, M. Taheri, E. Taki, K.-K. Tan, S. Tandukar, N. Y. Tat, V. Y. Tat, B. N. Tefera, Y. M. Tefera, G. Temesgen, M.-H. Temsah, S. Tharwat, A. Thiagarajan, I. I. Tleyjeh, C. E. Troeger, K. K. Umapathi, E. Upadhyay, S. Valadan Tahbaz, P. R. Valdez, J. Van den Eynde, H. R. van Doorn, S. Vaziri, G.-I. Verras, H. Viswanathan, B. Vo, A. Waris, G. T. Wassie, N. D. Wickramasinghe, S. Yaghoubi, G. A. T. Y. Yahya, S. H. Yahyazadeh Jabbari, A. Yigit, V. Yiğit, D. K. Yon, N. Yonemoto, M. Zahir, B. A. Zaman, S. B. Zaman, M. Zangiabadian, I. Zare, M. S. Zastrozhan, Z.-J. Zhang, P. Zheng, C. Zhong, M. Zoladl, A. Zumla, S. I. Hay, C. Dolecek, B. Sartorius, C. J. L. Murray, and M. Naghavi, "Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the global burden of disease study 2019," *The Lancet*, vol. 400, no. 10369, p. 2221–2248, Dec. 2022. [Online]. Available: [http://dx.doi.org/10.1016/S0140-6736\(22\)02185-7](http://dx.doi.org/10.1016/S0140-6736(22)02185-7)
- [124] R. Parastan, M. Kargar, K. Solhjoo, and F. Kafilzadeh, "Staphylococcus aureus biofilms: Structures, antibiotic resistance, inhibition, and vaccines,"

- Gene Reports*, vol. 20, p. 100739, Sep. 2020. [Online]. Available: <https://doi.org/10.1016/j.genrep.2020.100739>
- [125] E. E. Radke, S. M. Brown, A. J. Pelzek, Y. Fulmer, D. N. Hernandez, V. J. Torres, I. P. Thomsen, W. K. Chiang, A. O. Miller, B. Shopsin, and G. J. Silverman, "Hierarchy of human IgG recognition within the staphylococcus aureus immunome," *Scientific Reports*, vol. 8, no. 1, Sep. 2018. [Online]. Available: <https://doi.org/10.1038/s41598-018-31424-3>
- [126] Y. M. Ali, N. J. Lynch, P. Khatri, I. E. Bamigbola, A. C. Y. Chan, M. Yabuki, G. A. Demopoulos, J. L. Heaney, S. Pai, H. Baxendale, and W. J. Schwaeble, "Secondary complement deficiency impairs anti-microbial immunity to klebsiella pneumoniae and staphylococcus aureus during severe acute COVID-19," *Frontiers in Immunology*, vol. 13, Apr. 2022. [Online]. Available: <https://doi.org/10.3389/fimmu.2022.841759>
- [127] T. J. Foster, J. A. Geoghegan, V. K. Ganesh, and M. Höök, "Adhesion, invasion and evasion: the many functions of the surface proteins of staphylococcus aureus," *Nature Reviews Microbiology*, vol. 12, no. 1, pp. 49–62, Dec. 2013. [Online]. Available: <https://doi.org/10.1038/nrmicro3161>
- [128] A. G. Cheng, H. K. Kim, M. L. Burts, T. Krausz, O. Schneewind, and D. M. Missiakas, "Genetic requirements for staphylococcus aureus abscess formation and persistence in host tissues," *The FASEB Journal*, vol. 23, no. 10, pp. 3393–3404, Jun. 2009. [Online]. Available: <https://doi.org/10.1096/fj.09-135467>
- [129] E. Josefsson, O. Hartford, L. O'Brien, J. M. Patti, and T. Foster, "Protection against experimental staphylococcus aureus arthritis by vaccination with clumping factor a, a novel virulence determinant," *The Journal of Infectious Diseases*, vol. 184, no. 12, pp. 1572–1580, Dec. 2001. [Online]. Available: <https://doi.org/10.1086/324430>
- [130] K. A. Lacey, M. E. Mulcahy, A. M. Towell, J. A. Geoghegan, and R. M. McLoughlin, "Clumping factor b is an important virulence factor during staphylococcus aureus skin infection and a promising vaccine target," *PLOS Pathogens*, vol. 15, no. 4, p. e1007713, Apr. 2019. [Online]. Available: <https://doi.org/10.1371/journal.ppat.1007713>
- [131] T. Vanassche, A. Kauskot, J. Verhaegen, W. Peetermans, J. van Ryn,

- O. Schneewind, M. Hoylaerts, and P. Verhamme, "Fibrin formation by staphylothrombin facilitates *staphylococcus aureus*-induced platelet aggregation," *Thrombosis and Haemostasis*, vol. 107, no. 06, pp. 1107–1121, 2012. [Online]. Available: <https://doi.org/10.1160/th11-12-0891>
- [132] J. M. Patti, B. L. Allen, M. J. McGavin, and M. Höök, "MSCRAMM-MEDIATED ADHERENCE OF MICROORGANISMS TO HOST TISSUES," *Annual Review of Microbiology*, vol. 48, no. 1, pp. 585–617, Oct. 1994. [Online]. Available: <https://doi.org/10.1146/annurev.mi.48.100194.003101>
- [133] K. MURUGAN, M. USHA, P. MALATHI, A. S. AL-SOHAIBANI, and M. CHANDRASEKARAN, "Biofilm forming multi drug resistant *staphylococcus* spp. among patients with conjunctivitis," *Polish Journal of Microbiology*, vol. 59, no. 4, pp. 233–239, 2010. [Online]. Available: <https://doi.org/10.33073/pjm-2010-036>
- [134] P. M. Schlievert, R. J. Roller, S. H. Kilgore, M. Villarreal, A. J. Klingelhutz, and D. Y. M. Leung, "Staphylococcal TSST-1 association with eczema herpeticum in humans," *mSphere*, vol. 6, no. 4, Aug. 2021. [Online]. Available: <https://doi.org/10.1128/msphere.00608-21>
- [135] K. A. Berry, M. T. A. Verhoef, A. C. Leonard, and G. Cox, "Staphylococcus aureus adhesion to the host," *Annals of the New York Academy of Sciences*, vol. 1515, no. 1, pp. 75–96, Jun. 2022. [Online]. Available: <https://doi.org/10.1111/nyas.14807>
- [136] M. Bhattacharya, D. J. Wozniak, P. Stoodley, and L. Hall-Stoodley, "Prevention and treatment of *staphylococcus aureus* biofilms," *Expert Review of Anti-infective Therapy*, vol. 13, no. 12, pp. 1499–1516, Nov. 2015. [Online]. Available: <https://doi.org/10.1586/14787210.2015.1100533>
- [137] L. E. López-Cortés, J. Gálvez-Acebal, and J. Rodríguez-Baño, "Therapy of *staphylococcus aureus* bacteremia: Evidences and challenges," *Enfermedades Infecciosas y Microbiología Clínica*, vol. 38, no. 10, pp. 489–497, Dec. 2020. [Online]. Available: <https://doi.org/10.1016/j.eimc.2020.01.018>
- [138] S. Tsiodras, H. S. Gold, G. Sakoulas, G. M. Eliopoulos, C. Wennersten, L. Venkataraman, R. C. Moellering, and M. J. Ferraro, "Linezolid resistance in a clinical isolate of *staphylococcus aureus*," *The Lancet*,

- vol. 358, no. 9277, pp. 207–208, Jul. 2001. [Online]. Available: [https://doi.org/10.1016/s0140-6736\(01\)05410-1](https://doi.org/10.1016/s0140-6736(01)05410-1)
- [139] R. N. Jones, T. R. Fritsche, H. S. Sader, and J. E. Ross, “LEADER surveillance program results for 2006: an activity and spectrum analysis of linezolid using clinical isolates from the united states (50 medical centers),” *Diagnostic Microbiology and Infectious Disease*, vol. 59, no. 3, pp. 309–317, Nov. 2007. [Online]. Available: <https://doi.org/10.1016/j.diagmicrobio.2007.06.004>
- [140] R. E. Mendes, L. M. Deshpande, M. Castanheira, J. DiPersio, M. A. Saubolle, and R. N. Jones, “First report of mediated resistance to linezolid in human staphylococcal clinical isolates recovered in the united states,” *Antimicrobial Agents and Chemotherapy*, vol. 52, no. 6, pp. 2244–2246, Jun. 2008. [Online]. Available: <https://doi.org/10.1128/aac.00231-08>
- [141] Y. Ikeda-Dantsuji, H. Hanaki, F. Sakai, C. Yanagisawa, T. Nakae, F. Sakai, K. Tomono, Y. Takesue, J. Honda, Y. Nonomiya, A. Suwabe, O. Nagura, K. Yanagihara, H. Mikamo, K. Fukuchi, M. Kaku, S. Kohno, K. Yoshida, and Y. Niki, “Linezolid-resistant staphylococcus aureus isolated from 2006 through 2008 at six hospitals in japan,” *Journal of Infection and Chemotherapy*, vol. 17, no. 1, pp. 45–51, 2011. [Online]. Available: <https://doi.org/10.1007/s10156-010-0085-1>
- [142] C. Seral, Y. Sáenz, S. Algarate, E. Duran, P. Luque, C. Torres, and F. J. Castillo, “Nosocomial outbreak of methicillin- and linezolid-resistant staphylococcus epidermidis associated with catheter-related infections in intensive care unit patients,” *International Journal of Medical Microbiology*, vol. 301, no. 4, pp. 354–358, Apr. 2011. [Online]. Available: <https://doi.org/10.1016/j.ijmm.2010.11.001>
- [143] R. E. Mendes, L. M. Deshpande, D. J. Farrell, T. Spanu, G. Fadda, and R. N. Jones, “Assessment of linezolid resistance mechanisms among staphylococcus epidermidis causing bacteraemia in rome, italy,” *Journal of Antimicrobial Chemotherapy*, vol. 65, no. 11, pp. 2329–2335, Sep. 2010. [Online]. Available: <https://doi.org/10.1093/jac/dkq331>
- [144] X. Xu, H. Zhu, and H. Lv, “Safety of staphylococcus aureus four-antigen and three-antigen vaccines in healthy adults: A meta-analysis of randomized controlled trials,” *Human Vaccines and Immunotherapeutics*, vol. 14, no. 2, pp. 314–321,

- Dec. 2017. [Online]. Available: <https://doi.org/10.1080/21645515.2017.1395540>
- [145] C. B. Creech, R. W. Frenck, A. Fiquet, R. Feldman, M. K. Kankam, S. Pathirana, J. Baber, D. Radley, D. Cooper, J. Eiden, W. C. Gruber, K. U. Jansen, A. S. Anderson, and A. Gurtman, "Persistence of immune responses through 36 months in healthy adults after vaccination with a novel staphylococcus aureus 4-antigen vaccine (SA4ag)," *Open Forum Infectious Diseases*, vol. 7, no. 1, Dec. 2019. [Online]. Available: <https://doi.org/10.1093/ofid/ofz532>
- [146] R. Williams, "Healthy carriage of staphylococcus aureus: its prevalence and importance," *Bacteriological reviews*, vol. 27, no. 1, pp. 56–71, 1963.
- [147] M. E. Mulcahy and R. M. McLoughlin, "Host–bacterial crosstalk determines staphylococcus aureus nasal colonization," *Trends in Microbiology*, vol. 24, no. 11, p. 872–886, Nov. 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.tim.2016.06.012>
- [148] J. L. Nouwen, A. Ott, M. F. Kluytmans-Vandenbergh, H. A. Boelens, A. Hofman, A. Van Belkum, and H. A. Verbrugh, "Predicting the staphylococcus aureus nasal carrier state: derivation and validation of a "culture rule"," *Clinical Infectious Diseases*, vol. 39, no. 6, pp. 806–811, 2004.
- [149] J. U. E. Sollid, A. S. Furberg, A. M. Hanssen, and M. Johannessen, "Staphylococcus aureus: determinants of human carriage," *Infect. Genet. Evol.*, vol. 21, pp. 531–541, Jan. 2014.
- [150] A. van Belkum, N. J. Verkaik, C. P. De Vogel, H. A. Boelens, J. Verveer, J. L. Nouwen, H. A. Verbrugh, and H. F. Wertheim, "Reclassification of staphylococcus aureus nasal carriage types," *The Journal of infectious diseases*, vol. 199, no. 12, pp. 1820–1826, 2009.
- [151] A. E. Simor, "Staphylococcal decolonisation: an effective strategy for prevention of infection?" *The Lancet Infectious Diseases*, vol. 11, no. 12, p. 952–962, Dec. 2011. [Online]. Available: [http://dx.doi.org/10.1016/s1473-3099\(11\)70281-x](http://dx.doi.org/10.1016/s1473-3099(11)70281-x)
- [152] H. F. L. Wertheim, D. C. Melles, M. C. Vos, W. van Leeuwen, A. van Belkum, H. A. Verbrugh, and J. L. Nouwen, "The role of nasal carriage in staphylococcus aureus infections," *Lancet Infect. Dis.*, vol. 5, no. 12, pp. 751–762, Dec. 2005.

- [153] A. M. Cole, S. Tahk, A. Oren, D. Yoshioka, Y.-H. Kim, A. Park, and T. Ganz, "Determinants of staphylococcus aureus nasal carriage," *Clinical Diagnostic Laboratory Immunology*, vol. 8, no. 6, p. 1064–1069, Nov. 2001. [Online]. Available: <http://dx.doi.org/10.1128/cdli.8.6.1064-1069.2001>
- [154] L. A. Herwaldt, J. J. Cullen, P. French, J. Hu, M. A. Pfaller, R. P. Wenzel, and T. M. Perl, "Preoperative risk factors for nasal carriage of staphylococcus aureus," *Infection Control and Hospital Epidemiology*, vol. 25, no. 6, p. 481–484, Jun. 2004. [Online]. Available: <http://dx.doi.org/10.1086/502426>
- [155] S. J. Peacock, A. Justice, D. Griffiths, G. D. I. de Silva, M. N. Kantzanou, D. Crook, K. Sleeman, and N. P. J. Day, "Determinants of acquisition and carriage of staphylococcus aureus in infancy," *Journal of Clinical Microbiology*, vol. 41, no. 12, p. 5718–5725, Dec. 2003. [Online]. Available: <http://dx.doi.org/10.1128/jcm.41.12.5718-5725.2003>
- [156] R. Parnaby, G. O'Dwyer, H. Monsey, and M. Shafi, "Carriage of staphylococcus aureus in the elderly," *Journal of Hospital Infection*, vol. 33, no. 3, p. 201–206, Jul. 1996. [Online]. Available: [http://dx.doi.org/10.1016/S0195-6701\(96\)90003-1](http://dx.doi.org/10.1016/S0195-6701(96)90003-1)
- [157] B. A. Lipsky, R. E. Pecoraro, M. S. Chen, and T. D. Koepsell, "Factors affecting staphylococcal colonization among niddm outpatients," *Diabetes Care*, vol. 10, no. 4, p. 483–486, Jul. 1987. [Online]. Available: <http://dx.doi.org/10.2337/diacare.10.4.483>
- [158] V. L. Yu, A. Goetz, M. Wagener, P. B. Smith, J. D. Rihs, J. Hanchett, and J. J. Zuravleff, "Staphylococcus aureus nasal carriage and infection in patients on hemodialysis," *New England Journal of Medicine*, vol. 315, no. 2, p. 91–96, Jul. 1986. [Online]. Available: <http://dx.doi.org/10.1056/NEJM198607103150204>
- [159] C. Chapoutot, G.-P. Pageaux, P.-F. Perrigault, Z. Joomaye, P. Perney, H. Jean-Pierre, O. Jonquet, P. Blanc, and D. Larrey, "Staphylococcus aureus nasal carriage in 104 cirrhotic and control patients a prospective study," *Journal of Hepatology*, vol. 30, no. 2, p. 249–253, Feb. 1999. [Online]. Available: [http://dx.doi.org/10.1016/s0168-8278\(99\)80070-1](http://dx.doi.org/10.1016/s0168-8278(99)80070-1)
- [160] H. Carleton, B. Diep, E. Charlebois, G. Sensabaugh, and F. Perdreau-Remington, "Community-adapted methicillin-resistant staphylococcus aureus(mrsa): Population dynamics of an expanding community reservoir of mrsa," *The Journal*

- of Infectious Diseases*, vol. 190, no. 10, p. 1730–1738, Nov. 2004. [Online]. Available: <http://dx.doi.org/10.1086/425019>
- [161] E. E. Bernardy, R. A. Petit, V. Raghuram, A. M. Alexander, T. D. Read, and J. B. Goldberg, “Genotypic and phenotypic diversity of *staphylococcus aureus* isolates from cystic fibrosis patient lung infections and their interactions with *pseudomonas aeruginosa*,” *mBio*, vol. 11, no. 3, Jun. 2020. [Online]. Available: <http://dx.doi.org/10.1128/mBio.00735-20>
- [162] M. D. King, B. J. Humphrey, Y. F. Wang, E. V. Kourbatova, S. M. Ray, and H. M. Blumberg, “Emergence of community-acquired methicillin-resistant *staphylococcus aureus* usa 300 clone as the predominant cause of skin and soft-tissue infections,” *Annals of Internal Medicine*, vol. 144, no. 5, p. 309, Mar. 2006. [Online]. Available: <http://dx.doi.org/10.7326/0003-4819-144-5-200603070-00005>
- [163] M. A. Borg and L. Camilleri, “What is driving the epidemiology of methicillin-resistant *staphylococcus aureus* infections in europe?” *Microbial Drug Resistance*, vol. 27, no. 7, p. 889–894, Jul. 2021. [Online]. Available: <http://dx.doi.org/10.1089/mdr.2020.0259>
- [164] S. authors, *NORM/NORM-VET 2021. Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway*, G. S. Simonsen, H. S. Blix, K. Grave, and A. M. Urdahl, Eds. FHI, 2022. [Online]. Available: <https://www.fhi.no/en/publ/2022/norm-og-norm-vet-usage-of-antimicrobial-agents-and-occurrence-of-antimicrob/>
- [165] B. Strommenger, C. Kettilitz, T. Weniger, D. Harmsen, A. W. Friedrich, and W. Witte, “Assignment of *staphylococcus* isolates to groups by spa typing, Smal macrorestriction analysis, and multilocus sequence typing,” *J. Clin. Microbiol.*, vol. 44, no. 7, pp. 2533–2540, Jul. 2006.
- [166] L. Piechowicz and K. Garbacz, “Poultry-like pa+ biotype of *staphylococcus aureus* CC346/084 clone in human population,” *Curr. Microbiol.*, vol. 73, no. 1, pp. 124–131, Jul. 2016.
- [167] R. Cohen, S. Paikin, T. Finn, F. Babushkin, E. Anuka, M. Baum, and A. Rokney, “Molecular epidemiology of methicillin-resistant *staphylococcus aureus* clinical iso-

- lates during 7.5 years in one regional hospital in israel," *J. Environ. Public Health*, vol. 2021, p. 6643108, Mar. 2021.
- [168] C. E. Adams and S. J. Dancer, "Dynamic transmission of *staphylococcus aureus* in the intensive care unit," *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, p. 2109, Mar. 2020. [Online]. Available: <http://dx.doi.org/10.3390/ijerph17062109>
- [169] A. J. van Tonder, F. McCullagh, H. McKeand, S. Thaw, K. Bellis, C. Raisen, L. Lay, D. Aggarwal, M. Holmes, J. Parkhill, E. M. Harrison, A. Kucharski, and A. Conlan, "Colonization and transmission of *staphylococcus aureus* in schools: a citizen science project," *Microbial Genomics*, vol. 9, no. 4, Apr. 2023. [Online]. Available: <https://doi.org/10.1099/mgen.0.000993>
- [170] I. N. Abdullahi, C. Lozano, L. Ruiz-Ripa, R. Fernández-Fernández, M. Zarazaga, and C. Torres, "Ecology and genetic lineages of nasal *staphylococcus aureus* and mrsa carriage in healthy persons with or without animal-related occupational risks of colonization: A review of global reports," *Pathogens*, vol. 10, no. 8, p. 1000, Aug. 2021. [Online]. Available: <http://dx.doi.org/10.3390/pathogens10081000>
- [171] F. Zhu, H. Zhuang, S. Ji, E. Xu, L. Di, Z. Wang, S. Jiang, H. Wang, L. Sun, P. Shen, Y. Yu, and Y. Chen, "Household transmission of community-associated methicillin-resistant *staphylococcus aureus*," *Frontiers in Public Health*, vol. 9, May 2021. [Online]. Available: <https://doi.org/10.3389/fpubh.2021.658638>
- [172] S. Tahir, D. Chowdhury, M. Legge, H. Hu, G. Whiteley, T. Glasbey, A. K. Deva, and K. Vickery, "Transmission of *staphylococcus aureus* from dry surface biofilm (DSB) via different types of gloves," *Infect. Control Hosp. Epidemiol.*, vol. 40, no. 1, pp. 60–64, Jan. 2019.
- [173] L. Sun, H. Zhuang, L. Di, X. Ling, Y. Yin, Z. Wang, M. Chen, S. Jiang, Y. Chen, F. Zhu, H. Wang, S. Ji, L. Sun, D. Wu, Y. Yu, and Y. Chen, "Transmission and microevolution of methicillin-resistant *staphylococcus aureus* st88 strain among patients, healthcare workers, and household contacts at a trauma and orthopedic ward," *Frontiers in Public Health*, vol. 10, Jan. 2023. [Online]. Available: <http://dx.doi.org/10.3389/fpubh.2022.1053785>
- [174] T. Liang, Z. Liang, S. Wu, Y. Ding, Q. Wu, and B. Gu, "Global prevalence of *staphylococcus aureus* in food products and its relationship with the occurrence

- and development of diabetes mellitus," *Medicine Advances*, vol. 1, no. 1, p. 53–78, Mar. 2023. [Online]. Available: <http://dx.doi.org/10.1002/med4.6>
- [175] A. L. Morris and S. S. Mohiuddin, "Biochemistry, nutrients," Jan. 2023.
- [176] I. Chiodini, D. Gatti, D. Soranna, D. Merlotti, C. Mingiano, A. Fassio, G. Adami, A. Falchetti, C. Eller-Vainicher, M. Rossini, L. Persani, A. Zambon, and L. Gennari, "Vitamin d status and sars-cov-2 infection and covid-19 clinical outcomes," *Frontiers in Public Health*, vol. 9, Dec. 2021. [Online]. Available: <http://dx.doi.org/10.3389/fpubh.2021.736665>
- [177] A. A. Gomaa, Y. A. Abdel-Wadood, R. H. Thabet, and G. A. Gomaa, "Pharmacological evaluation of vitamin d in covid-19 and long covid-19: recent studies confirm clinical validation and highlight metformin to improve vdr sensitivity and efficacy," *Inflammopharmacology*, vol. 32, no. 1, p. 249–271, Nov. 2023. [Online]. Available: <http://dx.doi.org/10.1007/s10787-023-01383-x>
- [178] D. D. Bikle, "Vitamin d: Production, metabolism and mechanisms of action," 2000.
- [179] M. Holick, "Environmental factors that influence the cutaneous production of vitamin d," *The American Journal of Clinical Nutrition*, vol. 61, no. 3, pp. 638S–645S, Mar. 1995. [Online]. Available: <https://doi.org/10.1093/ajcn/61.3.638s>
- [180] E. Marcinowska-Suchowierska, M. Kupisz-Urbańska, J. Łukaszkiewicz, P. Płudowski, and G. Jones, "Vitamin d toxicity—a clinical perspective," *Frontiers in Endocrinology*, vol. 9, Sep. 2018. [Online]. Available: <https://doi.org/10.3389/fendo.2018.00550>
- [181] L. Tripkovic, H. Lambert, K. Hart, C. P. Smith, G. Bucca, S. Penson, G. Chope, E. Hyppönen, J. Berry, R. Vieth, and S. Lanham-New, "Comparison of vitamin d₂ and vitamin d₃ supplementation in raising serum 25-hydroxyvitamin d status: a systematic review and meta-analysis," *The American Journal of Clinical Nutrition*, vol. 95, no. 6, pp. 1357–1364, May 2012. [Online]. Available: <https://doi.org/10.3945/ajcn.111.031070>
- [182] U. Lehmann, F. Hirche, G. I. Stangl, K. Hinz, S. Westphal, and J. Dierkes, "Bioavailability of vitamin d₂ and d₃ in healthy volunteers, a randomized placebo-controlled trial," *The Journal of Clinical Endocrinology and*

- Metabolism*, vol. 98, no. 11, pp. 4339–4345, Nov. 2013. [Online]. Available: <https://doi.org/10.1210/jc.2012-4287>
- [183] V. F. Logan, A. R. Gray, M. C. Peddie, M. J. Harper, and L. A. Houghton, “Long-term vitamin d₃ supplementation is more effective than vitamin d₂ in maintaining serum 25-hydroxyvitamin d status over the winter months,” *British Journal of Nutrition*, vol. 109, no. 6, pp. 1082–1088, Jul. 2012. [Online]. Available: <https://doi.org/10.1017/s0007114512002851>
- [184] L. Tripkovic, L. R. Wilson, K. Hart, S. Johnsen, S. de Lusignan, C. P. Smith, G. Bucca, S. Penson, G. Chope, R. Elliott, E. Hyppönen, J. L. Berry, and S. A. Lanham-New, “Daily supplementation with 15 ug vitamin d₂ compared with vitamin d₃ to increase wintertime 25-hydroxyvitamin d status in healthy south asian and white european women: a 12-wk randomized, placebo-controlled food-fortification trial,” *The American Journal of Clinical Nutrition*, vol. 106, no. 2, pp. 481–490, Jul. 2017. [Online]. Available: <https://doi.org/10.3945/ajcn.116.138693>
- [185] L. A. Graeff-Armas, I. Bendik, I. Kunz, R. Schoop, S. Hull, and M. Beck, “Supplemental 25-hydroxycholecalciferol is more effective than cholecalciferol in raising serum 25-hydroxyvitamin d concentrations in older adults,” *The Journal of Nutrition*, vol. 150, no. 1, pp. 73–81, Sep. 2019. [Online]. Available: <https://doi.org/10.1093/jn/nxz209>
- [186] J. M. Quesada-Gomez and R. Bouillon, “Is calcifediol better than cholecalciferol for vitamin d supplementation?” *Osteoporosis International*, vol. 29, no. 8, pp. 1697–1711, Apr. 2018. [Online]. Available: <https://doi.org/10.1007/s00198-018-4520-y>
- [187] M. C. Silva and T. W. Furlanetto, “Intestinal absorption of vitamin d: a systematic review,” *Nutrition Reviews*, vol. 76, no. 1, pp. 60–76, Aug. 2017. [Online]. Available: <https://doi.org/10.1093/nutrit/nux034>
- [188] S. Nagpal, S. Na, and R. Rathnachalam, “Noncalcemic actions of vitamin d receptor ligands,” *Endocrine Reviews*, vol. 26, no. 5, pp. 662–687, Mar. 2005. [Online]. Available: <https://doi.org/10.1210/er.2004-0002>
- [189] A. S. Dusso, A. J. Brown, and E. Slatopolsky, “Vitamin d,” *American Journal of Physiology-Renal Physiology*, vol. 289, no. 1, pp. F8–F28, Jul. 2005. [Online]. Available: <https://doi.org/10.1152/ajprenal.00336.2004>

- [190] S. Samuel and M. D. Sitrin, "Vitamin d's role in cell proliferation and differentiation," *Nutrition Reviews*, vol. 66, pp. S116–S124, Sep. 2008. [Online]. Available: <https://doi.org/10.1111/j.1753-4887.2008.00094.x>
- [191] Institute of Medicine, Food and Nutrition Board, and Committee to Review Dietary Reference Intakes for Vitamin D and Calcium, *Dietary reference intakes for calcium and vitamin D*, H. B. Del Valle, A. L. Yaktine, C. L. Taylor, and A. C. Ross, Eds. Washington, D.C., DC: National Academies Press, Mar. 2011.
- [192] F. Limongi, M. Noale, A. Marseglia, A. Gesmundo, M. Mele, S. Banni, G. Crepaldi, and S. Maggi, "Impact of cheese rich in conjugated linoleic acid on low density lipoproteins cholesterol: Dietary intervention in older people (CLADIS study)," *J. Food Nutr. Res.*, vol. 6, no. 1, pp. 1–7, Dec. 2017.
- [193] A. Ross, B. Caballero, R. Cousins, K. Tucker, and T. Ziegler, *Modern nutrition in health and disease: Eleventh edition*. Wolters Kluwer Health Adis (ESP), Dec. 2012.
- [194] K. Amrein, M. Scherkl, M. Hoffmann, S. Neuwersch-Sommeregger, M. Köstenberger, A. T. Berisha, G. Martucci, S. Pilz, and O. Malle, "Vitamin d deficiency 2.0: an update on the current status worldwide," *European Journal of Clinical Nutrition*, vol. 74, no. 11, pp. 1498–1513, Jan. 2020. [Online]. Available: <https://doi.org/10.1038/s41430-020-0558-y>
- [195] K. D. Cashman, K. G. Dowling, Z. Škrabáková, M. Gonzalez-Gross, J. Valtueña, S. D. Henauw, L. Moreno, C. T. Damsgaard, K. F. Michaelsen, C. Mølgaard, R. Jorde, G. Grimnes, G. Moschonis, C. Mavrogianni, Y. Manios, M. Thamm, G. B. Mensink, M. Rabenberg, M. A. Busch, L. Cox, S. Meadows, G. Goldberg, A. Prentice, J. M. Dekker, G. Nijpels, S. Pilz, K. M. Swart, N. M. van Schoor, P. Lips, G. Eiriksdottir, V. Gudnason, M. F. Cotch, S. Koskinen, C. Lamberg-Allardt, R. A. Durazo-Arvizu, C. T. Sempos, and M. Kiely, "Vitamin d deficiency in europe: pandemic?" *The American Journal of Clinical Nutrition*, vol. 103, no. 4, pp. 1033–1044, Feb. 2016. [Online]. Available: <https://doi.org/10.3945/ajcn.115.120873>
- [196] Z. Jiang, R. Pu, N. Li, C. Chen, J. Li, W. Dai, Y. Wang, J. Hu, D. Zhu, Q. Yu, Y. Shi, and G. Yang, "High prevalence of vitamin d deficiency in asia: A systematic review and meta-analysis," *Critical Reviews*

- in Food Science and Nutrition*, pp. 1–10, Nov. 2021. [Online]. Available: <https://doi.org/10.1080/10408398.2021.1990850>
- [197] R. M. Mogire, A. Mutua, W. Kimita, A. Kamau, P. Bejon, J. M. Pettifor, A. Adeyemo, T. N. Williams, and S. H. Atkinson, “Prevalence of vitamin d deficiency in africa: a systematic review and meta-analysis,” *The Lancet Global Health*, vol. 8, no. 1, pp. e134–e142, Jan. 2020. [Online]. Available: [https://doi.org/10.1016/s2214-109x\(19\)30457-7](https://doi.org/10.1016/s2214-109x(19)30457-7)
- [198] H. M. Pappa, E. Bern, D. Kamin, and R. J. Grand, “Vitamin d status in gastrointestinal and liver disease,” *Current Opinion in Gastroenterology*, vol. 24, no. 2, pp. 176–183, Mar. 2008. [Online]. Available: <https://doi.org/10.1097/mog.0b013e3282f4d2f3>
- [199] S. A. M. Yazdi, M. Abbasi, and S. M. M. Yazdi, “Epilepsy and vitamin d: a comprehensive review of current knowledge,” *Reviews in the Neurosciences*, vol. 28, no. 2, pp. 185–201, Feb. 2017. [Online]. Available: <https://doi.org/10.1515/revneuro-2016-0044>
- [200] R. D. Bell, C. Y. C. Pak, J. Zerwekh, D. E. Barilla, and M. Vasko, “Effect of phenytoin on bone and vitamin d metabolism,” *Annals of Neurology*, vol. 5, no. 4, pp. 374–378, Apr. 1979. [Online]. Available: <https://doi.org/10.1002/ana.410050411>
- [201] M. M. A. Faridi and A. Aggarwal, “Phenytoin induced vitamin d deficiency presenting as proximal muscle weakness,” *Indian Pediatrics*, vol. 47, no. 7, pp. 624–625, Jul. 2010. [Online]. Available: <https://doi.org/10.1007/s13312-010-0121-3>
- [202] J. P. Ekwaru, J. D. Zwicker, M. F. Holick, E. Giovannucci, and P. J. Veugelers, “The importance of body weight for the dose response relationship of oral vitamin d supplementation and serum 25-hydroxyvitamin d in healthy volunteers,” *PLoS ONE*, vol. 9, no. 11, p. e111265, Nov. 2014. [Online]. Available: <https://doi.org/10.1371/journal.pone.0111265>
- [203] A. Drincic, E. Fuller, R. P. Heaney, and L. A. G. Armas, “25-hydroxyvitamin d response to graded vitamin d₃ supplementation among obese adults,” *The Journal of Clinical Endocrinology and Metabolism*, vol. 98, no. 12, pp. 4845–4851, Dec. 2013. [Online]. Available: <https://doi.org/10.1210/jc.2012-4103>

- [204] L. A. Peterson, X. Zeng, C. P. Caufield-Noll, M. A. Schweitzer, T. H. Magnuson, and K. E. Steele, "Vitamin d status and supplementation before and after bariatric surgery: a comprehensive literature review," *Surgery for Obesity and Related Diseases*, vol. 12, no. 3, pp. 693–702, Mar. 2016. [Online]. Available: <https://doi.org/10.1016/j.soard.2016.01.001>
- [205] M. Chakhtoura, M. Rahme, and G. E.-H. Fuleihan, "Vitamin d metabolism in bariatric surgery," *Endocrinology and Metabolism Clinics of North America*, vol. 46, no. 4, pp. 947–982, Dec. 2017. [Online]. Available: <https://doi.org/10.1016/j.ecl.2017.07.006>
- [206] K. Galior, S. Grebe, and R. Singh, "Development of vitamin d toxicity from overcorrection of vitamin d deficiency: A review of case reports," *Nutrients*, vol. 10, no. 8, p. 953, Jul. 2018. [Online]. Available: <https://doi.org/10.3390/nu10080953>
- [207] B. L. Auguste, C. Avila-Casado, and J. M. Bargman, "Use of vitamin d drops leading to kidney failure in a 54-year-old man," *Canadian Medical Association Journal*, vol. 191, no. 14, pp. E390–E394, Apr. 2019. [Online]. Available: <https://doi.org/10.1503/cmaj.180465>
- [208] M. G. Vogiatzi, E. Jacobson-Dickman, and M. D. D. and, "Vitamin d supplementation and risk of toxicity in pediatrics: A review of current literature," *The Journal of Clinical Endocrinology and Metabolism*, vol. 99, no. 4, pp. 1132–1141, Apr. 2014. [Online]. Available: <https://doi.org/10.1210/jc.2013-3655>
- [209] P. Singh and N. Trivedi, "Tanning beds and hypervitaminosis d: A case report," *Annals of Internal Medicine*, vol. 160, no. 11, p. 810, Jun. 2014. [Online]. Available: <https://doi.org/10.7326/I14-5011-8>
- [210] M. R. Laurent, E. Gielen, S. Pauwels, D. Vandershueren, and R. Bouillon, "Hypervitaminosis d associated with tanning bed use: A case report," *Annals of Internal Medicine*, vol. 166, no. 2, p. 155, Jan. 2017. [Online]. Available: <https://doi.org/10.7326/I16-0138>
- [211] J. L. Pérez-Castrillón, G. Vega, L. Abad, A. Sanz, J. Chaves, G. Hernandez, and A. Dueñas, "Effects of atorvastatin on vitamin d levels in patients with acute ischemic heart disease," *The American Journal of Cardiology*, vol. 99, no. 7, pp. 903–905, Apr. 2007. [Online]. Available: <https://doi.org/10.1016/j.amjcard.2006.11.036>

- [212] Institute of Medicine, Food and Nutrition Board, and Committee to Review Dietary Reference Intakes for Vitamin D and Calcium, *Dietary reference intakes for calcium and vitamin D*, H. B. Del Valle, A. L. Yaktine, C. L. Taylor, and A. C. Ross, Eds. Washington, D.C., DC: National Academies Press, Mar. 2011.
- [213] R. D. Jackson, A. Z. LaCroix, M. Gass, R. B. Wallace, J. Robbins, C. E. Lewis, T. Bassford, S. A. Beresford, H. R. Black, P. Blanchette, D. E. Bonds, R. L. Brunner, R. G. Brzyski, B. Caan, J. A. Cauley, R. T. Chlebowski, S. R. Cummings, I. Granek, J. Hays, G. Heiss, S. L. Hendrix, B. V. Howard, J. Hsia, F. A. Hubbell, K. C. Johnson, H. Judd, J. M. Kotchen, L. H. Kuller, R. D. Langer, N. L. Lasser, M. C. Limacher, S. Ludlam, J. E. Manson, K. L. Margolis, J. McGowan, J. K. Ockene, M. J. O'Sullivan, L. Phillips, R. L. Prentice, G. E. Sarto, M. L. Stefanick, L. V. Horn, J. Wactawski-Wende, E. Whitlock, G. L. Anderson, A. R. Assaf, and D. Barad, "Calcium plus vitamin d supplementation and the risk of fractures," *New England Journal of Medicine*, vol. 354, no. 7, pp. 669–683, Feb. 2006. [Online]. Available: <https://doi.org/10.1056/nejmoa055218>
- [214] Z. Malihi, C. M. M. Lawes, Z. Wu, Y. Huang, D. Waayer, L. Toop, K.-T. Khaw, C. A. Camargo, and R. Scragg, "Monthly high-dose vitamin d supplementation does not increase kidney stone risk or serum calcium: results from a randomized controlled trial," *The American Journal of Clinical Nutrition*, vol. 109, no. 6, pp. 1578–1587, Apr. 2019. [Online]. Available: <https://doi.org/10.1093/ajcn/nqy378>
- [215] Z. Malihi, Z. Wu, A. W. Stewart, C. M. Lawes, and R. Scragg, "Hypercalcemia, hypercalciuria, and kidney stones in long-term studies of vitamin d supplementation: a systematic review and meta-analysis," *The American Journal of Clinical Nutrition*, vol. 104, no. 4, pp. 1039–1051, Sep. 2016. [Online]. Available: <https://doi.org/10.3945/ajcn.116.134981>
- [216] A. Gotfredsen, H. W. Hendel, and T. Andersen, "Influence of orlistat on bone turnover and body composition," *International Journal of Obesity*, vol. 25, no. 8, pp. 1154–1160, Aug. 2001. [Online]. Available: <https://doi.org/10.1038/sj.ijo.0801639>
- [217] W. P. James, A. Avenell, J. Broom, and J. Whitehead, "A one-year trial to assess the value of orlistat in the management of obesity," *Int. J. Obes. Relat. Metab. Disord.*, vol. 21 Suppl 3, pp. S24–30, Jun. 1997.

- [218] J. R. McDuffie, K. A. Calis, S. L. Booth, G. I. Uwaifo, and J. A. Yanovski, "Effects of orlistat on fat-soluble vitamins in obese adolescents," *Pharmacotherapy*, vol. 22, no. 7, pp. 814–822, Jul. 2002. [Online]. Available: <https://doi.org/10.1592/phco.22.11.814.33627>
- [219] K. Robien, S. J. Oppeneer, J. A. Kelly, and J. M. Hamilton-Reeves, "Drug-vitamin d interactions," *Nutrition in Clinical Practice*, vol. 28, no. 2, pp. 194–208, Jan. 2013. [Online]. Available: <https://doi.org/10.1177/0884533612467824>
- [220] J. Schwartz, "Effects of vitamin d supplementation in atorvastatin-treated patients: A new drug interaction with an unexpected consequence," *Clinical Pharmacology and Therapeutics*, vol. 85, no. 2, pp. 198–203, Aug. 2008. [Online]. Available: <https://doi.org/10.1038/clpt.2008.165>
- [221] J. F. Aloia, M. Li-Ng, and S. Pollack, "Statins and vitamin d," *The American Journal of Cardiology*, vol. 100, no. 8, p. 1329, Oct. 2007. [Online]. Available: <https://doi.org/10.1016/j.amjcard.2007.05.024>
- [222] M. Crowe, L. Wollner, and R. A. Griffiths, "Hypercalcaemia following vitamin D and thiazide therapy in the elderly," *Practitioner*, vol. 228, no. 1389, pp. 312–313, Mar. 1984.
- [223] P. J. DRINKA and W. E. NOLTEN, "Hazards of treating osteoporosis and hypertension concurrently with calcium, vitamin d, and distal diuretics," *Journal of the American Geriatrics Society*, vol. 32, no. 5, pp. 405–407, May 1984. [Online]. Available: <https://doi.org/10.1111/j.1532-5415.1984.tb02050.x>
- [224] A. L. Skversky, J. Kumar, M. K. Abramowitz, F. J. Kaskel, and M. L. Melamed, "Association of glucocorticoid use and low 25-hydroxyvitamin d levels: Results from the national health and nutrition examination survey (NHANES): 2001–2006," *The Journal of Clinical Endocrinology and Metabolism*, vol. 96, no. 12, pp. 3838–3845, Dec. 2011. [Online]. Available: <https://doi.org/10.1210/jc.2011-1600>
- [225] B. P. Lukert, "Glucocorticoid-induced osteoporosis: Pathogenesis and management," *Annals of Internal Medicine*, vol. 112, no. 5, p. 352, Mar. 1990. [Online]. Available: <https://doi.org/10.7326/0003-4819-112-5-352>
- [226] R. G. L. de Sévaux, A. J. Hoitsma, F. H. M. Corstens, and J. F. M.

- Wetzels, "Treatment with vitamin d and calcium reduces bone loss after renal transplantation: A randomized study," *Journal of the American Society of Nephrology*, vol. 13, no. 6, pp. 1608–1614, Jun. 2002. [Online]. Available: <https://doi.org/10.1097/01.asn.0000016082.70875.36>
- [227] L. M. Buckley, "Calcium and vitamin d3 supplementation prevents bone loss in the spine secondary to low-dose corticosteroids in patients with rheumatoid arthritis," *Annals of Internal Medicine*, vol. 125, no. 12, p. 961, Dec. 1996. [Online]. Available: <https://doi.org/10.7326/0003-4819-125-12-199612150-00004>
- [228] R. Bouillon, "Comparative analysis of nutritional guidelines for vitamin d," *Nature Reviews Endocrinology*, vol. 13, no. 8, pp. 466–479, Apr. 2017. [Online]. Available: <https://doi.org/10.1038/nrendo.2017.31>
- [229] C. T. Sempos, A. C. Heijboer, D. D. Bikle, J. Bollerslev, R. Bouillon, P. M. Brannon, H. F. DeLuca, G. Jones, C. F. Munns, J. P. Bilezikian, A. Giustina, and N. Binkley, "Vitamin d assays and the definition of hypovitaminosis d: results from the first international conference on controversies in vitamin d," *British Journal of Clinical Pharmacology*, vol. 84, no. 10, pp. 2194–2207, Jul. 2018. [Online]. Available: <https://doi.org/10.1111/bcp.13652>
- [230] C. T. Sempos, G. D. Carter, and N. C. Binkley, "25-hydroxyvitamin d assays," in *Vitamin D*. Elsevier, 2018, pp. 939–957. [Online]. Available: <https://doi.org/10.1016/b978-0-12-809965-0.00052-5>
- [231] C. J. Elder and N. J. Bishop, "Rickets," *The Lancet*, vol. 383, no. 9929, pp. 1665–1676, May 2014. [Online]. Available: [https://doi.org/10.1016/s0140-6736\(13\)61650-5](https://doi.org/10.1016/s0140-6736(13)61650-5)
- [232] B. A. D. S. Sankararaman, "Rickets," <https://www.ncbi.nlm.nih.gov/books/NBK562285/>, 2022, accessed on 2022-09-07.
- [233] S. Uday and W. Höglér, "Nutritional rickets and osteomalacia in the twenty-first century: Revised concepts, public health, and prevention strategies," *Current Osteoporosis Reports*, vol. 15, no. 4, pp. 293–302, Jun. 2017. [Online]. Available: <https://doi.org/10.1007/s11914-017-0383-y>
- [234] C. F. Munns, N. Shaw, M. Kiely, B. L. Specker, T. D. Thacher, K. Ozono, T. Michigami, D. Tiosano, M. Z. Mughal, O. Mäkitie, L. Ramos-Abad, L. Ward,

- L. A. DiMeglio, N. Atapattu, H. Cassinelli, C. Braegger, J. M. Pettifor, A. Seth, H. W. Idris, V. Bhatia, J. Fu, G. Goldberg, L. Sävendahl, R. Khadgawat, P. Pludowski, J. Maddock, E. Hyppönen, A. Oduwole, E. Frew, M. Aguiar, T. Tulchinsky, G. Butler, and W. Höglér, "Global consensus recommendations on prevention and management of nutritional rickets," *The Journal of Clinical Endocrinology and Metabolism*, vol. 101, no. 2, pp. 394–415, Feb. 2016. [Online]. Available: <https://doi.org/10.1210/jc.2015-2175>
- [235] T. D. Thacher, P. R. Fischer, P. J. Tebben, R. J. Singh, S. S. Cha, J. A. Maxson, and B. P. Yawn, "Increasing incidence of nutritional rickets: A population-based study in olmsted county, minnesota," *Mayo Clinic Proceedings*, vol. 88, no. 2, pp. 176–183, Feb. 2013. [Online]. Available: <https://doi.org/10.1016/j.mayocp.2012.10.018>
- [236] G. CAMPBELL, "HOW COMMON IS OSTEOMALACIA IN THE ELDERLY?" *The Lancet*, vol. 324, no. 8399, pp. 386–388, Aug. 1984. [Online]. Available: [https://doi.org/10.1016/s0140-6736\(84\)90553-1](https://doi.org/10.1016/s0140-6736(84)90553-1)
- [237] W. H. Organization, "Chronic rheumatic conditions," <https://web.archive.org/web/20150427154245/http://www.who.int/chp/topics/rheumatic/en>, 2015, accessed on 2022-09-07.
- [238] M. Chung, E. M. Balk, M. Brendel, S. Ip, J. Lau, J. Lee, A. Lichtenstein, K. Patel, G. Raman, A. Tatsioni, T. Terasawa, and T. A. Trikalinos, "Vitamin D and calcium: a systematic review of health outcomes," *Evid. Rep. Technol. Assess. (Full Rep.)*, no. 183, pp. 1–420, Aug. 2009.
- [239] P. Yao, D. Bennett, M. Mafham, X. Lin, Z. Chen, J. Armitage, and R. Clarke, "Vitamin d and calcium for the prevention of fracture," *JAMA Network Open*, vol. 2, no. 12, p. e1917789, Dec. 2019. [Online]. Available: <https://doi.org/10.1001/jamanetworkopen.2019.17789>
- [240] R. Bouillon, D. Manousaki, C. Rosen, K. Trajanoska, F. Rivadeneira, and J. B. Richards, "The health effects of vitamin d supplementation: evidence from human studies," *Nature Reviews Endocrinology*, vol. 18, no. 2, pp. 96–110, Nov. 2021. [Online]. Available: <https://doi.org/10.1038/s41574-021-00593-z>
- [241] L. C. Kahwati, R. P. Weber, H. Pan, M. Gourlay, E. LeBlanc, M. Coker-Schwimmer, and M. Viswanathan, "Vitamin d, calcium, or combined

- supplementation for the primary prevention of fractures in community-dwelling adults," *JAMA*, vol. 319, no. 15, p. 1600, Apr. 2018. [Online]. Available: <https://doi.org/10.1001/jama.2017.21640>
- [242] and David C. Grossman, S. J. Curry, D. K. Owens, M. J. Barry, A. B. Caughey, K. W. Davidson, C. A. Doubeni, J. W. Epling, A. R. Kemper, A. H. Krist, M. Kubik, S. Landefeld, C. M. Mangione, M. Silverstein, M. A. Simon, and C.-W. Tseng, "Vitamin d, calcium, or combined supplementation for the primary prevention of fractures in community-dwelling adults," *JAMA*, vol. 319, no. 15, p. 1592, Apr. 2018. [Online]. Available: <https://doi.org/10.1001/jama.2018.3185>
- [243] J. C. Gallagher, "Vitamin d and bone density, fractures, and falls: the end of the story?" *The Lancet Diabetes and Endocrinology*, vol. 6, no. 11, pp. 834–835, Nov. 2018. [Online]. Available: [https://doi.org/10.1016/s2213-8587\(18\)30269-9](https://doi.org/10.1016/s2213-8587(18)30269-9)
- [244] M. J. Bolland, A. Grey, and A. Avenell, "Effects of vitamin d supplementation on musculoskeletal health: a systematic review, meta-analysis, and trial sequential analysis," *The Lancet Diabetes and Endocrinology*, vol. 6, no. 11, pp. 847–858, Nov. 2018. [Online]. Available: [https://doi.org/10.1016/s2213-8587\(18\)30265-1](https://doi.org/10.1016/s2213-8587(18)30265-1)
- [245] and David C. Grossman, S. J. Curry, D. K. Owens, M. J. Barry, A. B. Caughey, K. W. Davidson, C. A. Doubeni, J. W. Epling, A. R. Kemper, A. H. Krist, M. Kubik, S. Landefeld, C. M. Mangione, M. Pignone, M. Silverstein, M. A. Simon, and C.-W. Tseng, "Interventions to prevent falls in community-dwelling older adults," *JAMA*, vol. 319, no. 16, p. 1696, Apr. 2018. [Online]. Available: <https://doi.org/10.1001/jama.2018.3097>
- [246] J. M. Guirguis-Blake, Y. L. Michael, L. A. Perdue, E. L. Coppola, and T. L. Beil, "Interventions to prevent falls in older adults," *JAMA*, vol. 319, no. 16, p. 1705, Apr. 2018. [Online]. Available: <https://doi.org/10.1001/jama.2017.21962>
- [247] M. S. LeBoff, S. H. Chou, K. A. Ratliff, N. R. Cook, B. Khurana, E. Kim, P. M. Cawthon, D. C. Bauer, D. Black, J. C. Gallagher, I.-M. Lee, J. E. Buring, and J. E. Manson, "Supplemental vitamin d and incident fractures in midlife and older adults," *New England Journal of Medicine*, vol. 387, no. 4, pp. 299–309, Jul. 2022. [Online]. Available: <https://doi.org/10.1056/nejmoa2202106>
- [248] L. L. Brown, B. Cohen, D. Tabor, G. Zappalà, P. Maruvada, and P. M. Coates, "The vitamin d paradox in black americans: a systems-based approach to

- investigating clinical practice, research, and public health - expert panel meeting report," *BMC Proceedings*, vol. 12, no. S6, May 2018. [Online]. Available: <https://doi.org/10.1186/s12919-018-0102-4>
- [249] J. F. Aloia, R. Rubinova, M. Fazzari, S. Islam, M. Mikhail, and L. Ragolia, "Vitamin d and falls in older african american women: The PODA randomized clinical trial," *Journal of the American Geriatrics Society*, vol. 67, no. 5, pp. 1043–1049, Jan. 2019. [Online]. Available: <https://doi.org/10.1111/jgs.15760>
- [250] J. F. Aloia, S. A. Talwar, S. Pollack, and J. Yeh, "A randomized controlled trial of vitamin d3 supplementation in african american women," *Archives of Internal Medicine*, vol. 165, no. 14, p. 1618, Jul. 2005. [Online]. Available: <https://doi.org/10.1001/archinte.165.14.1618>
- [251] N. Li, G. Zhao, W. Wu, M. Zhang, W. Liu, Q. Chen, and X. Wang, "The efficacy and safety of vitamin c for iron supplementation in adult patients with iron deficiency anemia," *JAMA Network Open*, vol. 3, no. 11, p. e2023644, Nov. 2020. [Online]. Available: <https://doi.org/10.1001/jamanetworkopen.2020.23644>
- [252] S. R. Lynch and J. D. Cook, "INTERACTION OF VITAMIN c AND IRON," *Annals of the New York Academy of Sciences*, vol. 355, no. 1 Micronutrient, pp. 32–44, Dec. 1980. [Online]. Available: <https://doi.org/10.1111/j.1749-6632.1980.tb21325.x>
- [253] M. P. Erin LeBlanc Roger Chou, Bernadette Zakher Monica Daeges, *Screening for Vitamin D Deficiency: Systematic Review for the U.S. Preventive Services Task Force Recommendation*, A. for Healthcare Research and Q. (US), Eds. Washington, D.C., DC: U.S. Preventive Services Task Force Evidence Syntheses, 2014.
- [254] M. K. Shea, R. A. Fielding, and B. Dawson-Hughes, "The effect of vitamin d supplementation on lower-extremity power and function in older adults: a randomized controlled trial," *The American Journal of Clinical Nutrition*, vol. 109, no. 2, pp. 369–379, Feb. 2019. [Online]. Available: <https://doi.org/10.1093/ajcn/nqy290>
- [255] A. M. M. Vaes, M. Tieland, N. Toussaint, R. Nilwik, L. B. Verdijk, L. J. C. van Loon, and L. C. P. G. M. de Groot, "Cholecalciferol or 25-hydroxycholecalciferol supplementation does not affect muscle strength and physical performance in

- prefrail and frail older adults," *The Journal of Nutrition*, vol. 148, no. 5, pp. 712–720, May 2018. [Online]. Available: <https://doi.org/10.1093/jn/nxy024>
- [256] J. E. Manson, S. S. Bassuk, and J. E. Buring, "Vitamin d, calcium, and cancer," *JAMA*, vol. 317, no. 12, p. 1217, Mar. 2017. [Online]. Available: <https://doi.org/10.1001/jama.2017.2155>
- [257] L. Yin, J. M. Ordóñez-Mena, T. Chen, B. Schöttker, V. Arndt, and H. Brenner, "Circulating 25-hydroxyvitamin d serum concentration and total cancer incidence and mortality: A systematic review and meta-analysis," *Preventive Medicine*, vol. 57, no. 6, pp. 753–764, Dec. 2013. [Online]. Available: <https://doi.org/10.1016/j.ypmed.2013.08.026>
- [258] J. Han, X. Guo, X. Yu, S. Liu, X. Cui, B. Zhang, and H. Liang, "25-hydroxyvitamin d and total cancer incidence and mortality: A meta-analysis of prospective cohort studies," *Nutrients*, vol. 11, no. 10, p. 2295, Sep. 2019. [Online]. Available: <https://doi.org/10.3390/nu11102295>
- [259] N. Keum and E. Giovannucci, "Vitamin d supplements and cancer incidence and mortality: a meta-analysis," *British Journal of Cancer*, vol. 111, no. 5, pp. 976–980, Jun. 2014. [Online]. Available: <https://doi.org/10.1038/bjc.2014.294>
- [260] G. Bjelakovic, L. L. Gluud, D. Nikolova, K. Whitfield, G. Krstic, J. Wetterslev, and C. Gluud, "Vitamin d supplementation for prevention of cancer in adults," *Cochrane Database of Systematic Reviews*, Jun. 2014. [Online]. Available: <https://doi.org/10.1002/14651858.cd007469.pub2>
- [261] N. Keum, D. Lee, D. Greenwood, J. Manson, and E. Giovannucci, "Vitamin d supplementation and total cancer incidence and mortality: a meta-analysis of randomized controlled trials," *Annals of Oncology*, vol. 30, no. 5, pp. 733–743, May 2019. [Online]. Available: <https://doi.org/10.1093/annonc/mdz059>
- [262] J. E. Manson, N. R. Cook, I.-M. Lee, W. Christen, S. S. Bassuk, S. Mora, H. Gibson, D. Gordon, T. Copeland, D. D'Agostino, G. Friedenberg, C. Ridge, V. Bubes, E. L. Giovannucci, W. C. Willett, and J. E. Buring, "Vitamin d supplements and prevention of cancer and cardiovascular disease," *New England Journal of Medicine*, vol. 380, no. 1, pp. 33–44, Jan. 2019. [Online]. Available: <https://doi.org/10.1056/nejmoa1809944>

- [263] V. A. Moyer, "Vitamin, mineral, and multivitamin supplements for the primary prevention of cardiovascular disease and cancer: U.s. preventive services task force recommendation statement," *Annals of Internal Medicine*, vol. 160, no. 8, p. 558, Apr. 2014. [Online]. Available: <https://doi.org/10.7326/m14-0198>
- [264] J. A. Cauley, R. T. Chlebowski, J. Wactawski-Wende, J. A. Robbins, R. J. Rodabough, Z. Chen, K. C. Johnson, M. J. O'Sullivan, R. D. Jackson, and J. E. Manson, "Calcium plus vitamin d supplementation and health outcomes five years after active intervention ended: The women's health initiative," *Journal of Women's Health*, vol. 22, no. 11, pp. 915–929, Nov. 2013. [Online]. Available: <https://doi.org/10.1089/jwh.2013.4270>
- [265] R. T. Chlebowski, K. C. Johnson, C. Kooperberg, M. Pettinger, J. Wactawski-Wende, T. Rohan, J. Rossouw, D. Lane, M. J. O'Sullivan, S. Yasmeen, R. A. Hiatt, J. M. Shikany, M. Vitolins, J. Khandekar, and F. A. H. and, "Calcium plus vitamin d supplementation and the risk of breast cancer," *JNCI: Journal of the National Cancer Institute*, vol. 100, no. 22, pp. 1581–1591, Nov. 2008. [Online]. Available: <https://doi.org/10.1093/jnci/djn360>
- [266] J. Wactawski-Wende, J. M. Kotchen, G. L. Anderson, A. R. Assaf, R. L. Brunner, M. J. O'Sullivan, K. L. Margolis, J. K. Ockene, L. Phillips, L. Pottern, R. L. Prentice, J. Robbins, T. E. Rohan, G. E. Sarto, S. Sharma, M. L. Stefanick, L. V. Horn, R. B. Wallace, E. Whitlock, T. Bassford, S. A. Beresford, H. R. Black, D. E. Bonds, R. G. Brzyski, B. Caan, R. T. Chlebowski, B. Cochrane, C. Garland, M. Gass, J. Hays, G. Heiss, S. L. Hendrix, B. V. Howard, J. Hsia, F. A. Hubbell, R. D. Jackson, K. C. Johnson, H. Judd, C. L. Kooperberg, L. H. Kuller, A. Z. LaCroix, D. S. Lane, R. D. Langer, N. L. Lasser, C. E. Lewis, M. C. Limacher, and J. E. Manson, "Calcium plus vitamin d supplementation and the risk of colorectal cancer," *New England Journal of Medicine*, vol. 354, no. 7, pp. 684–696, Feb. 2006. [Online]. Available: <https://doi.org/10.1056/nejmoa055222>
- [267] S. Yao, M. L. Kwan, I. J. Ergas, J. M. Roh, T.-Y. D. Cheng, C.-C. Hong, S. E. McCann, L. Tang, W. Davis, S. Liu, C. P. Quesenberry, M. M. Lee, C. B. Ambrosone, and L. H. Kushi, "Association of serum level of vitamin d at diagnosis with breast cancer survival," *JAMA Oncology*, vol. 3, no. 3, p. 351, Mar. 2017. [Online]. Available: <https://doi.org/10.1001/jamaoncol.2016.4188>
- [268] T. Skaaby, L. L. N. Husemoen, B. H. Thuesen, C. Pisinger, T. Jørgensen,

- N. Roswall, S. C. Larsen, and A. Linneberg, "Prospective population-based study of the association between serum 25-hydroxyvitamin-d levels and the incidence of specific types of cancer," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 23, no. 7, pp. 1220–1229, Jul. 2014. [Online]. Available: <https://doi.org/10.1158/1055-9965.epi-14-0007>
- [269] K. M. O'Brien, D. P. Sandler, J. A. Taylor, and C. R. Weinberg, "Serum vitamin d and risk of breast cancer within five years," *Environmental Health Perspectives*, vol. 125, no. 7, p. 077004, Jul. 2017. [Online]. Available: <https://doi.org/10.1289/ehp943>
- [270] M. McNamara and K. D. Rosenberger, "The significance of vitamin d status in breast cancer: A state of the science review," *Journal of Midwifery and Women's Health*, vol. 64, no. 3, pp. 276–288, Apr. 2019. [Online]. Available: <https://doi.org/10.1111/jmwh.12968>
- [271] D. Muller, A. Hodge, A. Fanidi, D. Albanes, X. Mai, X. Shu, S. Weinstein, T. Larose, X. Zhang, J. Han, M. Stampfer, S. Smith-Warner, J. Ma, J. Gaziano, H. Sesso, V. Stevens, M. McCullough, T. Layne, R. Prentice, M. Pettinger, C. Thomson, W. Zheng, Y. Gao, N. Rothman, Y. Xiang, H. Cai, R. Wang, J. Yuan, W. Koh, L. Butler, Q. Cai, W. Blot, J. Wu, P. Ueland, Ø. Midttun, A. Langhammer, K. Hveem, M. Johansson, J. Hultdin, K. Grankvist, A. Arslan, L. L. Marchand, G. Severi, M. Johansson, and P. Brennan, "No association between circulating concentrations of vitamin d and risk of lung cancer: an analysis in 20 prospective studies in the lung cancer cohort consortium (LC3)," *Annals of Oncology*, vol. 29, no. 6, pp. 1468–1475, Jun. 2018. [Online]. Available: <https://doi.org/10.1093/annonc/mdy104>
- [272] K. J. H. and, "Overview of the cohort consortium vitamin d pooling project of rarer cancers," *American Journal of Epidemiology*, vol. 172, no. 1, pp. 4–9, Jun. 2010. [Online]. Available: <https://doi.org/10.1093/aje/kwq119>
- [273] R. Z. Stolzenberg-Solomon, R. Vieth, A. Azad, P. Pietinen, P. R. Taylor, J. Virtamo, and D. Albanes, "A prospective nested case-control study of vitamin d status and pancreatic cancer risk in male smokers," *Cancer Research*, vol. 66, no. 20, pp. 10213–10219, Oct. 2006. [Online]. Available: <https://doi.org/10.1158/0008-5472.can-06-1876>

- [274] F. J. van Duijnhoven, M. Jenab, K. Hveem, P. D. Siersema, V. Fedirko, E. J. Duell, E. Kampman, A. Halfweeg, H. J. van Kranen, J. M. van den Ouwelander, E. Weiderpass, N. Murphy, A. Langhammer, E. Ness-Jensen, A. Olsen, A. Tjønneland, K. Overvad, C. Cadeau, M. Kvaskoff, M.-C. Boutron-Ruault, V. A. Katzke, T. Kühn, H. Boeing, A. Trichopoulou, A. Kotanidou, M. Kritikou, D. Palli, C. Agnoli, R. Tumino, S. Panico, G. Matullo, P. Peeters, M. Brustad, K. S. Olsen, C. Lasheras, M. Obón-Santacana, M.-J. Sánchez, M. Dorronsoro, M.-D. Chirlaque, A. Barricarte, J. Manjer, M. Almquist, F. Renström, W. Ye, N. Wareham, K.-T. Khaw, K. E. Bradbury, H. Freisling, D. Aune, T. Norat, E. Riboli, and H. B. B. de Mesquita, "Circulating concentrations of vitamin d in relation to pancreatic cancer risk in european populations," *International Journal of Cancer*, vol. 142, no. 6, pp. 1189–1201, Nov. 2017. [Online]. Available: <https://doi.org/10.1002/ijc.31146>
- [275] M. Song, I.-M. Lee, J. E. Manson, J. E. Buring, R. Dushkes, D. Gordon, J. Walter, K. Wu, A. T. Chan, S. Ogino, C. S. Fuchs, J. A. Meyerhardt, and E. L. Giovannucci, "No association between vitamin d supplementation and risk of colorectal adenomas or serrated polyps in a randomized trial," *Clinical Gastroenterology and Hepatology*, vol. 19, no. 1, pp. 128–135.e6, Jan. 2021. [Online]. Available: <https://doi.org/10.1016/j.cgh.2020.02.013>
- [276] S. D. Crockett, E. L. Barry, L. A. Mott, D. J. Ahnen, D. J. Robertson, J. C. Anderson, K. Wallace, C. A. Burke, R. S. Bresalier, J. C. Figueiredo, D. C. Snover, and J. A. Baron, "Calcium and vitamin d supplementation and increased risk of serrated polyps: results from a randomised clinical trial," *Gut*, vol. 68, no. 3, pp. 475–486, Mar. 2018. [Online]. Available: <https://doi.org/10.1136/gutjnl-2017-315242>
- [277] M. L. McCullough, E. S. Zoltick, S. J. Weinstein, V. Fedirko, M. Wang, N. R. Cook, A. H. Eliassen, A. Zeleniuch-Jacquotte, C. Agnoli, D. Albanes, M. J. Barnett, J. E. Buring, P. T. Campbell, T. V. Clendenen, N. D. Freedman, S. M. Gapstur, E. L. Giovannucci, G. G. Goodman, C. A. Haiman, G. Y. F. Ho, R. L. Horst, T. Hou, W.-Y. Huang, M. Jenab, M. E. Jones, C. E. Joshu, V. Krogh, I.-M. Lee, J. E. Lee, S. Männistö, L. L. Marchand, A. M. Mondul, M. L. Neuhouser, E. A. Platz, M. P. Purdue, E. Riboli, T. E. Robsahm, T. E. Rohan, S. Sasazuki, M. J. Schoemaker, S. Sieri, M. J. Stampfer, A. J. Swerdlow, C. A. Thomson, S. Tretli, S. Tsugane, G. Ursin, K. Visvanathan, K. K. White,

- K. Wu, S.-S. Yaun, X. Zhang, W. C. Willett, M. H. Gail, R. G. Ziegler, and S. A. Smith-Warner, "Circulating vitamin d and colorectal cancer risk: An international pooling project of 17 cohorts," *JNCI: Journal of the National Cancer Institute*, vol. 111, no. 2, pp. 158–169, Jun. 2018. [Online]. Available: <https://doi.org/10.1093/jnci/djy087>
- [278] S. Shahvazi, S. Soltani, S. Ahmadi, R. de Souza, and A. Salehi-Abargouei, "The effect of vitamin d supplementation on prostate cancer: A systematic review and meta-analysis of clinical trials," *Hormone and Metabolic Research*, vol. 51, no. 01, pp. 11–21, Dec. 2018. [Online]. Available: <https://doi.org/10.1055/a-0774-8809>
- [279] Z. yu Song, Q. Yao, Z. Zhuo, Z. Ma, and G. Chen, "Circulating vitamin d level and mortality in prostate cancer patients: a dose–response meta-analysis," *Endocrine Connections*, vol. 7, no. 12, pp. R294–R303, Dec. 2018. [Online]. Available: <https://doi.org/10.1530/ec-18-0283>
- [280] V. Nair-Shalliker, A. Bang, S. Egger, M. Clements, R. A. Gardiner, A. Krieker, M. J. Seibel, S. K. Chambers, M. G. Kimlin, B. K. Armstrong, and D. P. Smith, "Post-treatment levels of plasma 25- and 1, 25-dihydroxy vitamin d and mortality in men with aggressive prostate cancer," *Scientific Reports*, vol. 10, no. 1, May 2020. [Online]. Available: <https://doi.org/10.1038/s41598-020-62182-w>
- [281] R. C. Travis, A. Perez-Cornago, P. N. Appleby, D. Albanes, C. E. Joshu, P. L. Lutsey, A. M. Mondul, E. A. Platz, S. J. Weinstein, T. M. Layne, K. J. Helzlsouer, K. Visvanathan, D. Palli, P. H. Peeters, B. B. de Mesquita, A. Trichopoulou, M. J. Gunter, K. K. Tsilidis, M.-J. Sánchez, A. Olsen, H. Brenner, B. Schöttker, L. Perna, B. Holleczek, P. Knekt, H. Rissanen, B. B. Yeap, L. Flicker, O. P. Almeida, Y. Y. E. Wong, J. M. Chan, E. L. Giovannucci, M. J. Stampfer, G. Ursin, R. E. Gislevfoss, T. Bjørge, H. E. Meyer, R. Blomhoff, S. Tsugane, N. Sawada, D. R. English, D. W. Eyles, A. K. Heath, E. J. Williamson, J. Manjer, J. Malm, M. Almquist, L. L. Marchand, C. A. Haiman, L. R. Wilkens, J. M. Schenk, C. M. Tangen, A. Black, M. B. Cook, W.-Y. Huang, R. G. Ziegler, R. M. Martin, F. C. Hamdy, J. L. Donovan, D. E. Neal, M. Touvier, S. Hercberg, P. Galan, M. Deschamps, T. J. Key, and N. E. Allen, "A collaborative analysis of individual participant data from 19 prospective studies assesses circulating vitamin d and prostate cancer risk," *Cancer Research*, vol. 79, no. 1, pp. 274–285, Jan. 2019. [Online]. Available: <https://doi.org/10.1158/0008-5472.can-18-2318>

- [282] X. Jiang, N. L. Dimou, K. Al-Dabholani, S. J. Lewis, R. M. Martin, P. C. Haycock, M. J. Gunter, T. J. Key, R. A. Eeles, K. Muir, D. Neal, G. G. Giles, E. L. Giovannucci, M. Stampfer, B. L. Pierce, J. M. Schildkraut, S. W. Andersen, D. Thompson, W. Zheng, P. Kraft, and K. K. T. and, "Circulating vitamin d concentrations and risk of breast and prostate cancer: a mendelian randomization study," *International Journal of Epidemiology*, vol. 48, no. 5, pp. 1416–1424, Dec. 2018. [Online]. Available: <https://doi.org/10.1093/ije/dyy284>
- [283] A. K. Heath, A. M. Hodge, P. R. Ebeling, D. W. Eyles, D. Kvaskoff, D. D. Buchanan, G. G. Giles, E. J. Williamson, and D. R. English, "Circulating 25-hydroxyvitamin d concentration and risk of breast, prostate, and colorectal cancers: The melbourne collaborative cohort study," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 28, no. 5, pp. 900–908, May 2019. [Online]. Available: <https://doi.org/10.1158/1055-9965.epi-18-1155>
- [284] J. M. Schenk, C. A. Till, C. M. Tangen, P. J. Goodman, X. Song, K. C. Torkko, A. R. Kristal, U. Peters, and M. L. Neuhauser, "Serum 25-hydroxyvitamin d concentrations and risk of prostate cancer: Results from the prostate cancer prevention trial," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 23, no. 8, pp. 1484–1493, Jul. 2014. [Online]. Available: <https://doi.org/10.1158/1055-9965.epi-13-1340>
- [285] A. R. Kristal, C. Till, X. Song, C. M. Tangen, P. J. Goodman, M. L. Neuhauser, J. M. Schenk, I. M. Thompson, F. L. Meyskens, G. E. Goodman, L. M. Minasian, H. L. Parnes, and E. A. Klein, "Plasma vitamin d and prostate cancer risk: Results from the selenium and vitamin e cancer prevention trial," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 23, no. 8, pp. 1494–1504, Jul. 2014. [Online]. Available: <https://doi.org/10.1158/1055-9965.epi-14-0115>
- [286] Y. Xu, X. Shao, Y. Yao, L. Xu, L. Chang, Z. Jiang, and Z. Lin, "Positive association between circulating 25-hydroxyvitamin d levels and prostate cancer risk: new findings from an updated meta-analysis," *Journal of Cancer Research and Clinical Oncology*, vol. 140, no. 9, pp. 1465–1477, May 2014. [Online]. Available: <https://doi.org/10.1007/s00432-014-1706-3>
- [287] E. Kassi, C. Adamopoulos, E. K. Basdra, and A. G. Papavassiliou, "Role of vitamin d in atherosclerosis," *Circulation*, vol. 128, no. 23, pp. 2517–2531, Dec. 2013. [Online]. Available: <https://doi.org/10.1161/circulationaha.113.002654>

- [288] I. A. Mheid and A. A. Quyyumi, "Vitamin d and cardiovascular disease," *Journal of the American College of Cardiology*, vol. 70, no. 1, pp. 89–100, Jul. 2017. [Online]. Available: <https://doi.org/10.1016/j.jacc.2017.05.031>
- [289] R. Zhang, B. Li, X. Gao, R. Tian, Y. Pan, Y. Jiang, H. Gu, Y. Wang, Y. Wang, and G. Liu, "Serum 25-hydroxyvitamin d and the risk of cardiovascular disease: dose-response meta-analysis of prospective studies," *The American Journal of Clinical Nutrition*, vol. 105, no. 4, pp. 810–819, Mar. 2017. [Online]. Available: <https://doi.org/10.3945/ajcn.116.140392>
- [290] D. Durup, H. L. Jørgensen, J. Christensen, A. Tjønneland, A. Olsen, J. Halkjær, B. Lind, A.-M. Heegaard, and P. Schwarz, "A reverse j-shaped association between serum 25-hydroxyvitamin d and cardiovascular disease mortality: The CopD study," *The Journal of Clinical Endocrinology and Metabolism*, vol. 100, no. 6, pp. 2339–2346, Jun. 2015. [Online]. Available: <https://doi.org/10.1210/jc.2014-4551>
- [291] R. Zhou, M. Wang, H. Huang, W. Li, Y. Hu, and T. Wu, "Lower vitamin d status is associated with an increased risk of ischemic stroke: A systematic review and meta-analysis," *Nutrients*, vol. 10, no. 3, p. 277, Feb. 2018. [Online]. Available: <https://doi.org/10.3390/nu10030277>
- [292] P. Brøndum-Jacobsen, M. Benn, G. B. Jensen, and B. G. Nordestgaard, "25-hydroxyvitamin d levels and risk of ischemic heart disease, myocardial infarction, and early death," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 32, no. 11, pp. 2794–2802, Nov. 2012. [Online]. Available: <https://doi.org/10.1161/atvaha.112.248039>
- [293] R. Scragg, A. W. Stewart, D. Waayer, C. M. M. Lawes, L. Toop, J. Sluyter, J. Murphy, K.-T. Khaw, and C. A. Camargo, "Effect of monthly high-dose vitamin d supplementation on cardiovascular disease in the vitamin d assessment study," *JAMA Cardiology*, vol. 2, no. 6, p. 608, Jun. 2017. [Online]. Available: <https://doi.org/10.1001/jamacardio.2017.0175>
- [294] J. A. Ford, G. S. MacLennan, A. Avenell, M. Bolland, A. Grey, and M. W. and, "Cardiovascular disease and vitamin d supplementation: trial analysis, systematic review, and meta-analysis," *The American Journal of Clinical Nutrition*, vol. 100, no. 3, pp. 746–755, Jul. 2014. [Online]. Available:

- <https://doi.org/10.3945/ajcn.113.082602>
- [295] D. T. Dibaba, "Effect of vitamin d supplementation on serum lipid profiles: a systematic review and meta-analysis," *Nutrition Reviews*, vol. 77, no. 12, pp. 890–902, Aug. 2019. [Online]. Available: <https://doi.org/10.1093/nutrit/nuz037>
- [296] M. Golzarand, S. Shab-Bidar, G. Koochakpoor, R. S. J., and K. Djafarian, "Effect of vitamin d₃ supplementation on blood pressure in adults: An updated meta-analysis," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 26, no. 8, pp. 663–673, Aug. 2016. [Online]. Available: <https://doi.org/10.1016/j.numecd.2016.04.011>
- [297] L. A. Beveridge, A. D. Struthers, F. Khan, R. Jorde, R. Scragg, H. M. Macdonald, J. A. Alvarez, R. S. Boxer, A. Dalbeni, A. D. Gepner, N. M. Isbel, T. Larsen, J. Nagpal, W. G. Petchey, H. Stricker, F. Strobel, V. Tangpricha, L. Toxqui, M. P. Vaquero, L. Wamberg, A. Zittermann, and M. D. Witham, "Effect of vitamin d supplementation on blood pressure," *JAMA Internal Medicine*, vol. 175, no. 5, p. 745, May 2015. [Online]. Available: <https://doi.org/10.1001/jamainternmed.2015.0237>
- [298] K. S. Vimaleswaran, A. Cavadino, D. J. Berry, R. Jorde, A. K. Dieffenbach, C. Lu, A. C. Alves, H. J. L. Heerspink, E. Tikkanen, J. Eriksson, A. Wong, M. Mangino, K. A. Jablonski, I. M. Nolte, D. K. Houston, T. S. Ahluwalia, P. J. van der Most, D. Pasko, L. Zgaga, E. Thiering, V. Vitart, R. M. Fraser, J. E. Huffman, R. A. de Boer, B. Schöttker, K.-U. Saum, M. I. McCarthy, J. Dupuis, K.-H. Herzig, S. Sebert, A. Pouta, J. Laitinen, M. E. Kleber, G. Navis, M. Lorentzon, K. Jameson, N. Arden, J. A. Cooper, J. Acharya, R. Hardy, O. Raitakari, S. Ripatti, L. K. Billings, J. Lahti, C. Osmond, B. W. Penninx, L. Rejnmark, K. K. Lohman, L. Paternoster, R. P. Stolk, D. G. Hernandez, L. Byberg, E. Hagström, H. Melhus, E. Ingelsson, D. Mellström, Östen Ljunggren, I. Tzoulaki, S. McLachlan, E. Theodoratou, C. M. T. Tiesler, A. Jula, P. Navarro, A. F. Wright, O. Polasek, J. F. Wilson, I. Rudan, V. Salomaa, J. Heinrich, H. Campbell, J. F. Price, M. Karlsson, L. Lind, K. Michaëlsson, S. Bandinelli, T. M. Frayling, C. A. Hartman, T. I. A. Sørensen, S. B. Kritchevsky, B. L. Langdahl, J. G. Eriksson, J. C. Florez, T. D. Spector, T. Lehtimäki, D. Kuh, S. E. Humphries, C. Cooper, C. Ohlsson, W. März, M. H. de Borst, M. Kumari, M. Kivimaki, T. J. Wang, C. Power, H. Brenner, G. Grimnes, P. van der Harst, H. Snieder, A. D. Hingorani, S. Pilz, J. C. Whittaker, M.-R. Järvelin,

- and E. Hyppönen, "Association of vitamin d status with arterial blood pressure and hypertension risk: a mendelian randomisation study," *The Lancet Diabetes and Endocrinology*, vol. 2, no. 9, pp. 719–729, Sep. 2014. [Online]. Available: [https://doi.org/10.1016/s2213-8587\(14\)70113-5](https://doi.org/10.1016/s2213-8587(14)70113-5)
- [299] I. Anjum, S. S. Jaffery, M. Fayyaz, Z. Samoo, and S. Anjum, "The role of vitamin d in brain health: A mini literature review," *Cureus*, Jul. 2018. [Online]. Available: <https://doi.org/10.7759/cureus.2960>
- [300] M. F. Holick, "Vitamin d and brain health: the need for vitamin d supplementation and sensible sun exposure," *Journal of Internal Medicine*, vol. 277, no. 1, pp. 90–93, Nov. 2014. [Online]. Available: <https://doi.org/10.1111/joim.12308>
- [301] D. W. Eyles, S. Smith, R. Kinobe, M. Hewison, and J. J. McGrath, "Distribution of the vitamin d receptor and 1-alpha-hydroxylase in human brain," *Journal of Chemical Neuroanatomy*, vol. 29, no. 1, pp. 21–30, Jan. 2005. [Online]. Available: <https://doi.org/10.1016/j.jchemneu.2004.08.006>
- [302] L. R. Harms, T. H. Burne, D. W. Eyles, and J. J. McGrath, "Vitamin d and the brain," *Best Practice and Research Clinical Endocrinology and Metabolism*, vol. 25, no. 4, pp. 657–669, Aug. 2011. [Online]. Available: <https://doi.org/10.1016/j.beem.2011.05.009>
- [303] R. E. S. Anglin, Z. Samaan, S. D. Walter, and S. D. McDonald, "Vitamin d deficiency and depression in adults: systematic review and meta-analysis," *British Journal of Psychiatry*, vol. 202, no. 2, pp. 100–107, Feb. 2013. [Online]. Available: <https://doi.org/10.1192/bjp.bp.111.106666>
- [304] U. Gowda, M. P. Mutowo, B. J. Smith, A. E. Wluka, and A. M. Renzaho, "Vitamin d supplementation to reduce depression in adults: Meta-analysis of randomized controlled trials," *Nutrition*, vol. 31, no. 3, pp. 421–429, Mar. 2015. [Online]. Available: <https://doi.org/10.1016/j.nut.2014.06.017>
- [305] R. Jorde and J. Kubiak, "No improvement in depressive symptoms by vitamin d supplementation: results from a randomised controlled trial," *Journal of Nutritional Science*, vol. 7, 2018. [Online]. Available: <https://doi.org/10.1017/jns.2018.19>
- [306] E. J. de Koning, P. Lips, B. W. J. H. Penninx, P. J. M. Elders, A. C. Heijboer,

- M. den Heijer, P. M. Bet, H. W. J. van Marwijk, and N. M. van Schoor, "Vitamin d supplementation for the prevention of depression and poor physical function in older persons: the d-vitaal study, a randomized clinical trial," *The American Journal of Clinical Nutrition*, vol. 110, no. 5, pp. 1119–1130, Jul. 2019. [Online]. Available: <https://doi.org/10.1093/ajcn/nqz141>
- [307] R. Jorde and G. Grimnes, "Vitamin d: no cure for depression," *The American Journal of Clinical Nutrition*, vol. 110, no. 5, pp. 1043–1044, Aug. 2019. [Online]. Available: <https://doi.org/10.1093/ajcn/nqz186>
- [308] O. I. Okereke, C. F. Reynolds, D. Mischoulon, G. Chang, C. M. Vyas, N. R. Cook, A. Weinberg, V. Bubes, T. Copeland, G. Friedenberg, I.-M. Lee, J. E. Buring, and J. E. Manson, "Effect of long-term vitamin d₃ supplementation vs placebo on risk of depression or clinically relevant depressive symptoms and on change in mood scores," *JAMA*, vol. 324, no. 5, p. 471, Aug. 2020. [Online]. Available: <https://doi.org/10.1001/jama.2020.10224>
- [309] J. Nakahara, M. Maeda, S. Aiso, and N. Suzuki, "Current concepts in multiple sclerosis: Autoimmunity versus oligodendroglialopathy," *Clinical Reviews in Allergy and Immunology*, vol. 42, no. 1, pp. 26–34, Dec. 2011. [Online]. Available: <https://doi.org/10.1007/s12016-011-8287-6>
- [310] V. A. Jagannath, G. Filippini, C. D. Pietrantonj, G. V. Asokan, E. W. Robak, L. Whamond, and S. A. Robinson, "Vitamin d for the management of multiple sclerosis," *Cochrane Database of Systematic Reviews*, Sep. 2018. [Online]. Available: <https://doi.org/10.1002/14651858.cd008422.pub3>
- [311] K. L. Munger, K. Hongell, J. Åivo, M. Soilu-Hänninen, H.-M. Surcel, and A. Ascherio, "25-hydroxyvitamin d deficiency and risk of MS among women in the finnish maternity cohort," *Neurology*, vol. 89, no. 15, pp. 1578–1583, Sep. 2017. [Online]. Available: <https://doi.org/10.1212/WNL.0000000000004489>
- [312] J. Salzer, G. Hallmans, M. Nystrom, H. Stenlund, G. Wadell, and P. Sundstrom, "Vitamin d as a protective factor in multiple sclerosis," *Neurology*, vol. 79, no. 21, pp. 2140–2145, Nov. 2012. [Online]. Available: <https://doi.org/10.1212/WNL.0b013e3182752ea8>
- [313] K. L. Munger, L. I. Levin, B. W. Hollis, N. S. Howard, and A. Ascherio, "Serum 25-hydroxyvitamin d levels and risk of multiple sclerosis," *JAMA*, vol. 296, no. 23, p.

- 2832, Dec. 2006. [Online]. Available: <https://doi.org/10.1001/jama.296.23.2832>
- [314] M. B. Sintzel, M. Rametta, and A. T. Reder, "Vitamin d and multiple sclerosis: A comprehensive review," *Neurology and Therapy*, vol. 7, no. 1, pp. 59–85, Dec. 2017. [Online]. Available: <https://doi.org/10.1007/s40120-017-0086-4>
- [315] R. A. Marrie and C. A. Beck, "Preventing multiple sclerosis," *Neurology*, vol. 89, no. 15, pp. 1538–1539, Sep. 2017. [Online]. Available: <https://doi.org/10.1212/wnl.0000000000004506>
- [316] A. G. Pittas, B. Dawson-Hughes, P. Sheehan, J. H. Ware, W. C. Knowler, V. R. Aroda, I. Brodsky, L. Ceglia, C. Chadha, R. Chatterjee, C. Desouza, R. Dolor, J. Foreyt, P. Fuss, A. Ghazi, D. S. Hsia, K. C. Johnson, S. R. Kashyap, S. Kim, E. S. LeBlanc, M. R. Lewis, E. Liao, L. M. Neff, J. Nelson, P. O'Neil, J. Park, A. Peters, L. S. Phillips, R. Pratley, P. Raskin, N. Rasouli, D. Robbins, C. Rosen, E. M. Vickery, and M. Staten, "Vitamin d supplementation and prevention of type 2 diabetes," *New England Journal of Medicine*, vol. 381, no. 6, pp. 520–530, Aug. 2019. [Online]. Available: <https://doi.org/10.1056/nejmoa1900906>
- [317] A. Mousa, N. Naderpoor, H. Teede, R. Scragg, and B. de Courten, "Vitamin d supplementation for improvement of chronic low-grade inflammation in patients with type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials," *Nutrition Reviews*, vol. 76, no. 5, pp. 380–394, Feb. 2018. [Online]. Available: <https://doi.org/10.1093/nutrit/nux077>
- [318] X. Li, Y. Liu, Y. Zheng, P. Wang, and Y. Zhang, "The effect of vitamin d supplementation on glycemic control in type 2 diabetes patients: A systematic review and meta-analysis," *Nutrients*, vol. 10, no. 3, p. 375, Mar. 2018. [Online]. Available: <https://doi.org/10.3390/nu10030375>
- [319] S. Rafiq and P. Jeppesen, "Is hypovitaminosis d related to incidence of type 2 diabetes and high fasting glucose level in healthy subjects: A systematic review and meta-analysis of observational studies," *Nutrients*, vol. 10, no. 1, p. 59, Jan. 2018. [Online]. Available: <https://doi.org/10.3390/nu10010059>
- [320] A. Mousa, N. Naderpoor, M. P. de Courten, H. Teede, N. Kellow, K. Walker, R. Scragg, and B. de Courten, "Vitamin d supplementation has no effect on insulin sensitivity or secretion in vitamin d deficient, overweight or obese adults: a randomized placebo-controlled trial," *The American*

- Journal of Clinical Nutrition*, p. ajcn152736, May 2017. [Online]. Available: <https://doi.org/10.3945/ajcn.117.152736>
- [321] J. C. Seida, J. Mitri, I. N. Colmers, S. R. Majumdar, M. B. Davidson, A. L. Edwards, D. A. Hanley, A. G. Pittas, L. Tjosvold, and J. A. Johnson, "Effect of vitamin d₃ supplementation on improving glucose homeostasis and preventing diabetes: A systematic review and meta-analysis," *The Journal of Clinical Endocrinology and Metabolism*, vol. 99, no. 10, pp. 3551–3560, Oct. 2014. [Online]. Available: <https://doi.org/10.1210/jc.2014-2136>
- [322] R. Jorde, S. T. Sollid, J. Svartberg, H. Schirmer, R. M. Joakimsen, I. Njølstad, O. M. Fuskevåg, Y. Figenschau, and M. Y. S. Hutchinson, "Vitamin d 20 000 IU per week for five years does not prevent progression from prediabetes to diabetes," *The Journal of Clinical Endocrinology and Metabolism*, vol. 101, no. 4, pp. 1647–1655, Apr. 2016. [Online]. Available: <https://doi.org/10.1210/jc.2015-4013>
- [323] "Vitamin d supplementation and prevention of type 2 diabetes," *New England Journal of Medicine*, vol. 381, no. 18, pp. 1784–1786, Oct. 2019. [Online]. Available: <https://doi.org/10.1056/nejmc1912185>
- [324] C. P. Earthman, L. M. Beckman, K. Masodkar, and S. D. Sibley, "The link between obesity and low circulating 25-hydroxyvitamin d concentrations: considerations and implications," *International Journal of Obesity*, vol. 36, no. 3, pp. 387–396, Jun. 2011. [Online]. Available: <https://doi.org/10.1038/ijo.2011.119>
- [325] S. R. Mallard, A. S. Howe, and L. A. Houghton, "Vitamin d status and weight loss: a systematic review and meta-analysis of randomized and nonrandomized controlled weight-loss trials," *The American Journal of Clinical Nutrition*, vol. 104, no. 4, pp. 1151–1159, Sep. 2016. [Online]. Available: <https://doi.org/10.3945/ajcn.116.136879>
- [326] B. Caan, "Calcium plus vitamin d supplementation and the risk of postmenopausal weight gain," *Archives of Internal Medicine*, vol. 167, no. 9, p. 893, May 2007. [Online]. Available: <https://doi.org/10.1001/archinte.167.9.893>
- [327] K. Pathak, M. J. Soares, E. K. Calton, Y. Zhao, and J. Hallett, "Vitamin d supplementation and body weight status: a systematic review and meta-analysis of randomized controlled trials," *Obesity Reviews*, vol. 15, no. 6, pp. 528–537, Feb. 2014. [Online]. Available: <https://doi.org/10.1111/obr.12162>

- [328] S. Pilz, S. Frisch, H. Koertke, J. Kuhn, J. Dreier, B. Obermayer-Pietsch, E. Wehr, and A. Zittermann, "Effect of vitamin d supplementation on testosterone levels in men," *Hormone and Metabolic Research*, vol. 43, no. 03, pp. 223–225, Dec. 2010. [Online]. Available: <https://doi.org/10.1055/s-0030-1269854>
- [329] F. Limongi, M. Noale, A. Marseglia, A. Gesmundo, M. Mele, S. Banni, G. Crepaldi, and S. Maggi, "Impact of cheese rich in conjugated linoleic acid on low density lipoproteins cholesterol: Dietary intervention in older people (CLADIS study)," *J. Food Nutr. Res.*, vol. 6, no. 1, pp. 1–7, Dec. 2017.
- [330] H. O. Santos, S. Howell, K. Nichols, and F. J. Teixeira, "Reviewing the evidence on vitamin d supplementation in the management of testosterone status and its effects on male reproductive system (testis and prostate): Mechanistically dazzling but clinically disappointing," *Clinical Therapeutics*, vol. 42, no. 6, pp. e101–e114, Jun. 2020. [Online]. Available: <https://doi.org/10.1016/j.clinthera.2020.03.016>
- [331] E. Lerchbaum, S. Pilz, C. Trummer, V. Schwetz, O. Pachernegg, A. C. Heijboer, and B. Obermayer-Pietsch, "Vitamin d and testosterone in healthy men: A randomized controlled trial," *The Journal of Clinical Endocrinology and Metabolism*, vol. 102, no. 11, pp. 4292–4302, Aug. 2017. [Online]. Available: <https://doi.org/10.1210/jc.2017-01428>
- [332] A. Zittermann, J. B. Ernst, S. Prokop, U. Fuchs, J. Dreier, J. Kuhn, C. Knabbe, H. K. Berthold, I. Gouni-Berthold, J. F. Gummert, J. Börgermann, and S. Pilz, "Vitamin d supplementation does not prevent the testosterone decline in males with advanced heart failure: the EVITA trial," *European Journal of Nutrition*, vol. 58, no. 2, pp. 673–680, Mar. 2018. [Online]. Available: <https://doi.org/10.1007/s00394-018-1666-5>
- [333] M. Algarni, M. A. Hadi, A. Yahyouche, S. Mahmood, and Z. Jalal, "A mixed-methods systematic review of the prevalence, reasons, associated harms and risk-reduction interventions of over-the-counter (OTC) medicines misuse, abuse and dependence in adults," *Journal of Pharmaceutical Policy and Practice*, vol. 14, no. 1, Sep. 2021. [Online]. Available: <https://doi.org/10.1186/s40545-021-00350-7>
- [334] G. C. Richards, J. K. Aronson, B. MacKenna, B. Goldacre, F. D. R. Hobbs, and C. Heneghan, "Sales of over-the-counter products containing

- codeine in 31 countries, 2013–2019: A retrospective observational study," *Drug Safety*, vol. 45, no. 3, p. 237–247, Jan. 2022. [Online]. Available: <http://dx.doi.org/10.1007/s40264-021-01143-2>
- [335] C. H. Basch, K. J. Roberts, E. A. Zagnit, and A. Martinez, "Marketing strategies used to promote children's medicine sold on internet sites of pharmaceutical stores," *Journal of Community Health*, vol. 41, no. 6, p. 1212–1216, May 2016. [Online]. Available: <http://dx.doi.org/10.1007/s10900-016-0205-7>
- [336] E. D. McCoul, "Direct-to-consumer advertising of over-the-counter sinonasal remedies: A history of mixed messages," *The Laryngoscope*, vol. 130, no. 9, p. 2114–2119, Oct. 2019. [Online]. Available: <http://dx.doi.org/10.1002/LARY.28366>
- [337] E. Sánchez-Sánchez, F. L. Fernández-Cerezo, J. Díaz-Jimenez, M. Rosety-Rodriguez, A. J. Díaz, F. J. Ordonez, M. A. Rosety, and I. Rosety, "Consumption of over-the-counter drugs: Prevalence and type of drugs," *International Journal of Environmental Research and Public Health*, vol. 18, no. 11, 2021. [Online]. Available: <https://www.mdpi.com/1660-4601/18/11/5530>
- [338] M. Wójta-Kempa and D. M. Krzyżanowski, "Correlates of abusing and misusing over-the-counter pain relievers among adult population of wrocław (poland)," *Advances in Clinical and Experimental Medicine*, vol. 25, no. 2, pp. 349–360, Apr. 2016. [Online]. Available: <https://www.doi.org/10.17219/ACEM/58887>
- [339] P. J. Lein and P. N. Bernardino, "Investigating OTC drugs: Are over-the-counter drugs an under-appreciated toxic danger?" *Open Access Government*, vol. 39, no. 1, pp. 26–27, Jun. 2023. [Online]. Available: <https://doi.org/10.56367/oag-039-10036>
- [340] C. Lionis, E. Petelos, S. Shea, G. Bagiartaki, I. G. Tsiligianni, A. Kamekis, V. Tsiantou, M. Papadakaki, A. Tatsioni, J. Moschandreas, A. Saridaki, A. Bertsias, T. Faresjo, A. Faresjo, L. Martinez, D. Agius, Y. Uncu, G. Samoutis, J. Vlcek, A. Abasaeed, and B. Merkouris, "Irrational prescribing of over-the-counter (otc) medicines in general practice: testing the feasibility of an educational intervention among physicians in five european countries," *BMC Family Practice*, vol. 15, no. 1, Feb. 2014. [Online]. Available: <http://dx.doi.org/10.1186/1471-2296-15-34>

- [341] S. Lorentzen, B. Lorentzen, and B.-M. Wikström, "Gender differences in usage of over-the-counter analgesics among norwegian adolescents," *Open Journal of Nursing*, vol. 08, no. 11, pp. 860–878, 2018. [Online]. Available: <https://doi.org/10.4236/ojn.2018.811065>
- [342] X. Qian, H. Yin, and X. Li, "Marketing investment and pharmaceutical product diffusion," *Marketing Intelligence and Planning*, vol. 39, no. 3, p. 394–407, Sep. 2020. [Online]. Available: <http://dx.doi.org/10.1108/MIP-07-2020-0283>
- [343] K. Brune, B. Renner, and G. Tiegs, "Acetaminophen/paracetamol: A history of errors, failures and false decisions," *European Journal of Pain*, vol. 19, no. 7, pp. 953–965, Nov. 2014. [Online]. Available: <https://doi.org/10.1002/ejp.621>
- [344] S. Schulz, D. Mayer, M. Pfeiffer, R. Stumm, T. Koch, and V. Höllt, "Morphine induces terminal mu-opioid receptor desensitization by sustained phosphorylation of serine-375," *The EMBO Journal*, vol. 23, no. 16, pp. 3282–3289, Jul. 2004. [Online]. Available: <https://doi.org/10.1038/sj.emboj.7600334>
- [345] "Misuse of prescription and over-the-counter drugs to obtain illicit highs: how pharmacists can prevent abuse," *The Pharmaceutical Journal*, 2020. [Online]. Available: <https://doi.org/10.1211/pj.2020.20208538>
- [346] R. Vardanyan and V. Hruby, "Antihistamine drugs," in *Synthesis of Best-Seller Drugs*. Elsevier, 2016, pp. 247–263. [Online]. Available: <https://doi.org/10.1016/b978-0-12-411492-0.00016-x>
- [347] G. M. Walsh, "Antihistamines (h₁ receptor antagonists)," in *Side Effects of Drugs Annual 28*. Elsevier, 2005, pp. 177–183. [Online]. Available: [https://doi.org/10.1016/s0378-6080\(05\)80437-4](https://doi.org/10.1016/s0378-6080(05)80437-4)
- [348] Y. Faki, , and A. Er, "Different chemical structures and physiological/pathological roles of cyclooxygenases," *Rambam Maimonides Medical Journal*, vol. 12, no. 1, p. e0003, Jan. 2021. [Online]. Available: <https://doi.org/10.5041/rmmj.10426>
- [349] S. K. Bardal, J. E. Waechter, and D. S. Martin, "Cough, cold, and allergy," in *Applied Pharmacology*. Elsevier, 2011, pp. 127–134. [Online]. Available: <https://doi.org/10.1016/b978-1-4377-0310-8.00012-9>
- [350] S. Beharry and S. Gibbons, "An overview of emerging and new psychoactive

- substances in the united kingdom," *Forensic Science International*, vol. 267, pp. 25–34, Oct. 2016. [Online]. Available: <https://doi.org/10.1016/j.forsciint.2016.08.013>
- [351] C. A. Malinky, C. W. Lindsley, and C. Han, "DARK classics in chemical neuroscience: Loperamide," *ACS Chemical Neuroscience*, vol. 12, no. 16, pp. 2964–2973, Aug. 2021. [Online]. Available: <https://doi.org/10.1021/acscchemneuro.1c00382>
- [352] A. N. Edinoff, C. R. Swinford, A. S. Odisho, C. R. Burroughs, C. W. Stark, W. A. Raslan, E. M. Cornett, A. M. Kaye, and A. D. Kaye, "Clinically relevant drug interactions with monoamine oxidase inhibitors," *Health Psychology Research*, vol. 10, no. 4, Nov. 2022. [Online]. Available: <http://dx.doi.org/10.52965/001c.39576>
- [353] K. Linde, M. M. Berner, and L. Kriston, "St john's wort for major depression," *Cochrane Database of Systematic Reviews*, Oct. 2008. [Online]. Available: <http://dx.doi.org/10.1002/14651858.CD000448.pub3>
- [354] B. J. Komoroski, S. Zhang, H. Cai, J. M. Hutzler, R. Frye, T. S. Tracy, S. C. Strom, T. Lehmann, C. Y. W. Ang, Y. Y. Cui, and R. Venkataramanan, "Induction and inhibition of cytochromes p450 by the st. john's wort constituent hyperforin in human hepatocyte cultures," *Drug Metabolism and Disposition*, vol. 32, no. 5, p. 512–518, Apr. 2004. [Online]. Available: <http://dx.doi.org/10.1124/dmd.32.5.512>
- [355] E. D. López Vila, C. Buts, and M. Jegers, "A quantitative classification of otc medicines regulations in 30 european countries: dispensing restrictions, distribution, pharmacy ownership, and pricing systems," *Journal of Pharmaceutical Policy and Practice*, vol. 16, no. 1, Jan. 2023. [Online]. Available: <http://dx.doi.org/10.1186/s40545-023-00522-7>
- [356] H. Ino and E. Nakazawa, "Over-the-counter drug misuse and dependence: Public health ethics' foray into fight against the codeine crisis," *Pharmacy*, vol. 10, no. 6, p. 155, Nov. 2022. [Online]. Available: <http://dx.doi.org/10.3390/pharmacy10060155>
- [357] S. Prabhuswamy and P. Lakshminarayana, "A cross-sectional study on knowledge and practice of over-the-counter drugs among 2nd year medical

- students in a tertiary care teaching hospital," *National Journal of Physiology, Pharmacy and Pharmacology*, vol. 12, no. 8, p. 1, 2022. [Online]. Available: <http://dx.doi.org/10.5455/njppp.2022.12.062821202217062022>
- [358] A. Kamekis, C. Lionis, M. Papadakaki, E. Pitelos, and N. Papadakis, "Guidelines and educational interventions for optimal use of over the counter medicines: information, training, university education (full text in greek)," *Academia*, p. No 14 (2019), 2019. [Online]. Available: <https://pasithee.library.upatras.gr/academia/article/view/2981>
- [359] J. Gardiner and A. S. Kesselheim, "Over-the-counter monograph safety, innovation, and reform act," *Journal of Law, Medicine and Ethics*, vol. 49, no. 2, p. 321–327, 2021. [Online]. Available: <http://dx.doi.org/10.1017/JME.2021.46>
- [360] A. C. Wallace, "Managing use of over-the-counter medications in the school setting: Keeping kids in school and ready to learn," *NASN School Nurse*, vol. 31, no. 4, p. 210–213, Mar. 2016. [Online]. Available: <http://dx.doi.org/10.1177/1942602X16636123>
- [361] "Fit Futures study," <https://uit.no/research/fitfutures>, accessed: 2023-05-30.
- [362] A. Winther, E. Dennison, L. A. Ahmed, A.-S. Furberg, G. Grimnes, R. Jorde, C. G. Gjesdal, and N. Emaus, "The tromsø study: Fit futures: a study of norwegian adolescents' lifestyle and bone health," *Archives of Osteoporosis*, vol. 9, no. 1, Jun. 2014. [Online]. Available: <https://doi.org/10.1007/s11657-014-0185-0>
- [363] K. Olsen, B. M. Falch, K. Danielsen, M. Johannessen, J. U. E. Sollid, I. Thune, G. Grimnes, R. Jorde, G. S. Simonsen, and A.-S. Furberg, "Staphylococcus aureus nasal carriage is associated with serum 25-hydroxyvitamin d levels, gender and smoking status. the tromsø staph and skin study," *European Journal of Clinical Microbiology and Infectious Diseases*, vol. 31, no. 4, pp. 465–473, Aug. 2011. [Online]. Available: <https://doi.org/10.1007/s10096-011-1331-x>
- [364] D. B. Stensen, L. Småbrekke, K. Olsen, G. Grimnes, C. S. Nielsen, G. S. Simonsen, J. U. E. Sollid, and A.-S. Furberg, "Hormonal contraceptive use and staphylococcus aureus nasal and throat carriage in a norwegian youth population," *PLOS ONE*, vol. 14, no. 7, p. e0218511, Jul. 2019. [Online]. Available: <https://doi.org/10.1371/journal.pone.0218511>

- [365] J. L. Nouwen, A. Ott, M. F. Q. Kluytmans-Vandenbergh, H. A. M. Boelens, A. Hofman, A. van Belkum, and H. A. Verbrugh, "Predicting the staphylococcus aureus nasal carrier state: Derivation and validation of a "culture rule"," *Clinical Infectious Diseases*, vol. 39, no. 6, pp. 806–811, Sep. 2004. [Online]. Available: <https://doi.org/10.1086/423376>
- [366] A. van Belkum, N. J. Verkaik, C. P. de Vogel, H. A. Boelens, J. Verveer, J. L. Nouwen, H. A. Verbrugh, and H. F. L. Wertheim, "Reclassification of staphylococcus aureus nasal carriage types," *The Journal of Infectious Diseases*, vol. 199, no. 12, pp. 1820–1826, Jun. 2009. [Online]. Available: <https://doi.org/10.1086/599119>
- [367] M. Hallin, A. W. Friedrich, and M. J. Struelens, "spa typing for epidemiological surveillance of staphylococcus aureus," in *Methods in Molecular Biology*. Humana Press, 2009, pp. 189–202. [Online]. Available: https://doi.org/10.1007/978-1-60327-999-4_15
- [368] M. Sangvik, R. S. Olsen, K. Olsen, G. S. Simonsen, A.-S. Furberg, and J. U. E. Sollid, "Age- and gender-associated staphylococcus aureus spa types found among nasal carriers in a general population: the tromsø staph and skin study," *Journal of Clinical Microbiology*, vol. 49, no. 12, pp. 4213–4218, Dec. 2011. [Online]. Available: <https://doi.org/10.1128/jcm.05290-11>
- [369] S. report, *World Health Organization. (2005). Surveillance of chronic disease risk factors : country level data and comparable estimates. World Health Organization.*, ser. SuRF report. World Health Organization, 2005.
- [370] "Ods vitamin d initiative (2004-2018)," <https://ods.od.nih.gov/Research/VitaminD.aspx>, accessed: 2023-05-30.
- [371] K. D. Cashman, K. G. Dowling, Z. Škrabáková, M. Kiely, C. Lamberg-Allardt, R. A. Durazo-Arvizu, C. T. Sempos, S. Koskinen, A. Lundqvist, J. Sundvall, A. Linneberg, B. Thuesen, L. L. N. Husemoen, H. E. Meyer, K. Holvik, I. M. Grønborg, I. Tetens, and R. Andersen, "Standardizing serum 25-hydroxyvitamin d data from four nordic population samples using the vitamin d standardization program protocols: Shedding new light on vitamin d status in nordic individuals," *Scandinavian Journal of Clinical and Laboratory Investigation*, vol. 75, no. 7, pp. 549–561, Aug. 2015. [Online]. Available: <https://doi.org/10.1080/09054555.2015.1018300>

- <https://doi.org/10.3109/00365513.2015.1057898>
- [372] "Olink target 96 inflammation," <https://www.olink.com/products-services/target/inflammation/>, accessed: 2023-05-30.
- [373] H. Köhler and S. Link, "Sql schema design: foundations, normal forms, and normalization," *Information Systems*, vol. 76, pp. 88–113, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306437917305069>
- [374] J. Domingo-Ferrer, "Inference control in statistical databases," in *Encyclopedia of Database Systems*. Springer US, 2009, pp. 1472–1476. [Online]. Available: https://doi.org/10.1007/978-0-387-39940-9_203
- [375] T. A. S. S. P. Borgatti, *Non-Parametric Standard Errors and Tests for Network Statistics*, INSNA, Ed. INSNA, 1999. [Online]. Available: http://www.analytictech.com/borgatti/papers/snijders_borgatti-density_significance.pdf Framingham Heart Study (FHS)
- [376] J. Knox, A.-C. Uhlemann, and F. D. Lowy, "Staphylococcus aureus infections: transmission within households and the community," *Trends in Microbiology*, vol. 23, no. 7, pp. 437–444, Jul. 2015. [Online]. Available: <https://doi.org/10.1016/j.tim.2015.03.007>
- [377] C. L. Bihan, J.-R. Zahar, and J.-F. Timsit, "Staphylococcus aureus transmission in the intensive care unit: the potential role of the healthcare worker carriage," *Annals of Infection*, vol. 1, pp. 1–1, 2017. [Online]. Available: <https://doi.org/10.21037/aoi.2017.08.01>
- [378] C. O. Solberg, "Spread of staphylococcus aureus in hospitals: Causes and prevention," *Scandinavian Journal of Infectious Diseases*, vol. 32, no. 6, pp. 587–595, Jan. 2000. [Online]. Available: <https://doi.org/10.1080/003655400459478>
- [379] G. A. J. Ayliffe, "The progressive intercontinental spread of methicillin-resistant staphylococcus aureus," *Clinical Infectious Diseases*, vol. 24, no. Supplement_1, pp. S74–S79, Jan. 1997. [Online]. Available: https://doi.org/10.1093/clinids/24.supplement_1.s74
- [380] R. B. Rothenberg, J. J. Potterat, D. E. Woodhouse, S. Q. Muth, W. W.

- Darrow, and A. S. Klov Dahl, "Social network dynamics and HIV transmission," *AIDS*, vol. 12, no. 12, pp. 1529–1536, Aug. 1998. [Online]. Available: <https://doi.org/10.1097/00002030-199812000-00016>
- [381] S. J. Song, C. Lauber, E. K. Costello, C. A. Lozupone, G. Humphrey, D. Berg-Lyons, J. G. Caporaso, D. Knights, J. C. Clemente, S. Nakielny, J. I. Gordon, N. Fierer, and R. Knight, "Cohabiting family members share microbiota with one another and with their dogs," *eLife*, vol. 2, Apr. 2013. [Online]. Available: <https://doi.org/10.7554/elife.00458>
- [382] E. Lund, M. Holden, L.-T. R. Busund, I. Snapkov, N. Shvetsov, and L. Holden, "Trajectories of gene expression, seasonal influenza, and within-host seasonal immunity: transfer value to covid-19," *medRxiv*, Mar. 2022. [Online]. Available: <https://doi.org/10.1101/2022.03.01.22271679>
- [383] G. G. Jensen, F. Uekermann, K. Sneppen, and L. Simonsen, "Could short-lasting non-specific immunity explain seasonal spacing of epidemic diseases?" 2019. [Online]. Available: <https://arxiv.org/abs/1908.00925>
- [384] S. Thukral and V. Rana, "Versatility of fuzzy logic in chronic diseases: A review," *Medical Hypotheses*, vol. 122, pp. 150–156, Jan. 2019. [Online]. Available: <https://doi.org/10.1016/j.mehy.2018.11.017>
- [385] A. Amirkhani, E. I. Papageorgiou, A. Mohseni, and M. R. Mosavi, "A review of fuzzy cognitive maps in medicine: Taxonomy, methods, and applications," *Computer Methods and Programs in Biomedicine*, vol. 142, pp. 129–145, Apr. 2017. [Online]. Available: <https://doi.org/10.1016/j.cmpb.2017.02.021>
- [386] S. Mollerup, P. Worning, A. Petersen, and M. D. Bartels, "spa typing of methicillin-resistant staphylococcus aureus based on whole-genome sequencing: the impact of the assembler," *Microbiology Spectrum*, vol. 10, no. 6, Dec. 2022. [Online]. Available: <http://dx.doi.org/10.1128/spectrum.02189-22>
- [387] P. Asadollahi, N. N. Farahani, M. Mirzaii, S. S. Khoramrooz, A. van Belkum, K. Asadollahi, M. Dadashi, and D. Darban-Sarokhalil, "Distribution of the most prevalent spa types among clinical isolates of methicillin-resistant and -susceptible staphylococcus aureus around the world: A review," *Frontiers in Microbiology*, vol. 9, Feb. 2018. [Online]. Available: <http://dx.doi.org/10.3389/fmicb.2018.00163>

- [388] A. Blomfeldt, K. W. Larssen, A. Moghen, C. Gabrielsen, P. Elstrøm, H. V. Aamot, and S. B. Jørgensen, "Emerging multidrug-resistant bengal bay clone st772-mrsa-v in norway: molecular epidemiology 2004–2014," *European Journal of Clinical Microbiology and Infectious Diseases*, vol. 36, no. 10, p. 1911–1921, May 2017. [Online]. Available: <http://dx.doi.org/10.1007/s10096-017-3014-8>
- [389] K. Holvik, H. E. Meyer, E. Haug, and L. Brunvand, "Prevalence and predictors of vitamin d deficiency in five immigrant groups living in oslo, norway: the oslo immigrant health study," *European Journal of Clinical Nutrition*, vol. 59, no. 1, pp. 57–63, Jul. 2004. [Online]. Available: <https://doi.org/10.1038/sj.ejcn.1602033>
- [390] A. Björk, Å. Andersson, G. Johansson, K. Björkegren, A. Bardel, and P. Kristiansson, "Evaluation of sun holiday, diet habits, origin and other factors as determinants of vitamin d status in swedish primary health care patients: a cross-sectional study with regression analysis of ethnic swedish and immigrant women," *BMC Family Practice*, vol. 14, no. 1, Sep. 2013. [Online]. Available: <https://doi.org/10.1186/1471-2296-14-129>
- [391] C. A. Martin, U. Gowda, and A. M. Renzaho, "The prevalence of vitamin d deficiency among dark-skinned populations according to their stage of migration and region of birth: A meta-analysis," *Nutrition*, vol. 32, no. 1, pp. 21–32, Jan. 2016. [Online]. Available: <https://doi.org/10.1016/j.nut.2015.07.007>
- [392] D. S. Nielsen, D. N. Rasmussen, and M. Sodemann, "Vitamin d status in patients attending a danish migrant health clinic: A clinical retrospective study," *Journal of Immigrant and Minority Health*, vol. 17, no. 2, pp. 474–481, May 2014. [Online]. Available: <https://doi.org/10.1007/s10903-014-0031-x>
- [393] A. Chernet, N.-P. Hensch, K. Kling, V. Sydow, C. Hatz, D. H. Paris, K. Rentsch, B. Nickel, A. Neumayr, and N. D. Labhardt, "Serum 25-hydroxyvitamin d levels and intramuscular vitamin d₃ supplementation among eritrean migrants recently arrived in switzerland," *Swiss Medical Weekly*, vol. 147, no. 5152, Dec. 2017. [Online]. Available: <https://doi.org/10.4414/smw.2017.14568>
- [394] M. Ceccarelli, E. Chiappini, R. Arancio, M. Zaffaroni, S. L. Placa, M. D'Andrea, C. de Waure, R. M. D. Riol, and P. V. and, "Vitamin d deficiency in a population of migrant children: an italian retrospective cross-sectional multicentric study," *European Journal of Public Health*, vol. 30, no. 3, pp. 522–527, Oct. 2019.

- [Online]. Available: <https://doi.org/10.1093/eurpub/ckz182>
- [395] N. Smith, L. L. Sievert, S. Muttukrishna, K. Begum, L. Murphy, T. Sharmin, R. Gunu, O. Chowdhury, and G. R. Bentley, "Mismatch: a comparative study of vitamin d status in british-bangladeshi migrants," *Evolution, Medicine, and Public Health*, vol. 9, no. 1, pp. 164–173, Jan. 2021. [Online]. Available: <https://doi.org/10.1093/emph/eoab001>
- [396] M. del Carmen Navarro, P. Saavedra, E. Jódar, M. J. G. de Tejada, A. Mirallave, and M. Sosa, "Osteoporosis and metabolic syndrome according to socio-economic status, contribution of PTH, vitamin d and body weight: The canarian osteoporosis poverty study (COPS)," *Clinical Endocrinology*, vol. 78, no. 5, pp. 681–686, Apr. 2013. [Online]. Available: <https://doi.org/10.1111/cen.12051>
- [397] J. Öberg, R. Jorde, B. Almås, N. Emaus, and G. Grimnes, "Vitamin d deficiency and lifestyle risk factors in a norwegian adolescent population," *Scandinavian Journal of Public Health*, vol. 42, no. 7, pp. 593–602, Jul. 2014. [Online]. Available: <https://doi.org/10.1177/1403494814541593>
- [398] R. Jorde, E. B. Mathiesen, S. Rogne, T. Wilsgaard, M. Kjærgaard, G. Grimnes, and H. Schirmer, "Vitamin d and cognitive function: The tromsø study," *Journal of the Neurological Sciences*, vol. 355, no. 1-2, pp. 155–161, Aug. 2015. [Online]. Available: <https://doi.org/10.1016/j.jns.2015.06.009>
- [399] J. Öberg, R. Jorde, Y. Figenschau, P. M. Thorsby, S. R. Dahl, A. Winther, and G. Grimnes, "100 YEARS OF VITAMIN d: Combined hormonal contraceptives and vitamin d metabolism in adolescent girls," *Endocrine Connections*, vol. 11, no. 3, Mar. 2022. [Online]. Available: <https://doi.org/10.1530/ec-21-0395>
- [400] J. A. Partridge, R. J. Brustad, and M. B. Stellino, *Social influence in sport*, I. T. S. Horn, Ed. Human Kinetics, 2008. [Online]. Available: <https://psycnet.apa.org/record/2008-01583-012>
- [401] J. R. Chalcraft, L. M. Cardinal, P. J. Wechsler, B. W. Hollis, K. G. Gerow, B. M. Alexander, J. F. Keith, and D. E. Larson-Meyer, "Vitamin d synthesis following a single bout of sun exposure in older and younger men and women," *Nutrients*, vol. 12, no. 8, p. 2237, Jul. 2020. [Online]. Available: <https://doi.org/10.3390/nu12082237>

- [402] E. Theodoratou, I. Tzoulaki, L. Zgaga, and J. P. A. Ioannidis, "Vitamin d and multiple health outcomes: umbrella review of systematic reviews and meta-analyses of observational studies and randomised trials," *BMJ*, vol. 348, no. apr01 2, pp. g2035–g2035, Apr. 2014. [Online]. Available: <https://doi.org/10.1136/bmj.g2035>
- [403] R. Chowdhury, S. Kunutsor, A. Vitezova, C. Oliver-Williams, S. Chowdhury, J. C. K. de Jong, H. Khan, C. P. Baena, D. Prabhakaran, M. B. Hoshen, B. S. Feldman, A. Pan, L. Johnson, F. Crowe, F. B. Hu, and O. H. Franco, "Vitamin d and risk of cause specific death: systematic review and meta-analysis of observational cohort and randomised intervention studies," *BMJ*, vol. 348, no. apr01 2, pp. g1903–g1903, Apr. 2014. [Online]. Available: <https://doi.org/10.1136/bmj.g1903>
- [404] P. Welsh and N. Sattar, "Vitamin d and chronic disease prevention," *BMJ*, vol. 348, no. apr01 2, pp. g2280–g2280, Apr. 2014. [Online]. Available: <https://doi.org/10.1136/bmj.g2280>
- [405] C. M. Lloyd and R. J. Snelgrove, "Type 2 immunity: Expanding our view," *Science Immunology*, vol. 3, no. 25, Jul. 2018. [Online]. Available: <https://doi.org/10.1126/sciimmunol.aat1604>
- [406] B. L. Perry, K. C. Yang, P. Kaminski, M. Odabas, J. Park, M. Martel, C. B. Oser, P. R. Freeman, Y.-Y. Ahn, and J. Talbert, "Co-prescription network reveals social dynamics of opioid doctor shopping," *PLOS ONE*, vol. 14, no. 10, p. e0223849, Oct. 2019. [Online]. Available: <https://doi.org/10.1371/journal.pone.0223849>
- [407] J. B. Rice, A. G. White, H. G. Birnbaum, M. Schiller, D. A. Brown, and C. L. Roland, "A model to identify patients at risk for prescription opioid abuse, dependence, and misuse," *Pain Medicine*, vol. 13, no. 9, pp. 1162–1173, Sep. 2012. [Online]. Available: <https://doi.org/10.1111/j.1526-4637.2012.01450.x>
- [408] G. Miglio, L. Basso, L. G. Armando, S. Traina, E. Benetti, A. Diarassouba, R. B. Parisi, M. Esiliato, C. Rolando, E. Remani, and C. Cena, "A network approach for the study of drug prescriptions: Analysis of administrative records from a local health unit (ASL TO4, regione piemonte, italy)," *International Journal of Environmental Research and Public Health*, vol. 18, no. 9, p. 4859, May 2021. [Online]. Available: <https://doi.org/10.3390/ijerph18094859>

- [409] J. Zhang, L. Tong, P. Lamberson, R. Durazo-Arvizu, A. Luke, and D. Shoham, "Leveraging social influence to address overweight and obesity using agent-based models: The role of adolescent social networks," *Social Science and Medicine*, vol. 125, pp. 203–213, Jan. 2015. [Online]. Available: <https://doi.org/10.1016/j.socscimed.2014.05.049>
- [410] D. R. Schaefer and S. D. Simpkins, "Using social network analysis to clarify the role of obesity in selection of adolescent friends," *American Journal of Public Health*, vol. 104, no. 7, pp. 1223–1229, Jul. 2014. [Online]. Available: <https://doi.org/10.2105/ajph.2013.301768>
- [411] H. Fonseca and M. G. de Matos, "Perception of overweight and obesity among portuguese adolescents: an overview of associated factors," *European Journal of Public Health*, vol. 15, no. 3, pp. 323–328, May 2005. [Online]. Available: <https://doi.org/10.1093/eurpub/cki071>
- [412] T. W. Valente, K. Fujimoto, C.-P. Chou, and D. Spruijt-Metz, "Adolescent affiliations and adiposity: A social network analysis of friendships and obesity," *Journal of Adolescent Health*, vol. 45, no. 2, pp. 202–204, Aug. 2009. [Online]. Available: <https://doi.org/10.1016/j.jadohealth.2009.01.007>
- [413] K. de la Haye, G. Robins, P. Mohr, and C. Wilson, "Homophily and contagion as explanations for weight similarities among adolescent friends," *Journal of Adolescent Health*, vol. 49, no. 4, pp. 421–427, Oct. 2011. [Online]. Available: <https://doi.org/10.1016/j.jadohealth.2011.02.008>
- [414] X.-B. Cui, J. Fei, S. Chen, G. L. Edwards, and S.-Y. Chen, "ADAR1 deficiency protects against high-fat diet-induced obesity and insulin resistance in mice," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 320, no. 1, pp. E131–E138, Jan. 2021. [Online]. Available: <https://doi.org/10.1152/ajpendo.00175.2020>
- [415] Y. Yue, C. Zhang, X. Zhang, S. Zhang, Q. Liu, F. Hu, X. Lv, H. Li, J. Yang, X. Wang, L. Chen, Z. Yao, H. Duan, and W. Niu, "An AMPK/axin1-rac1 signaling pathway mediates contraction-regulated glucose uptake in skeletal muscle cells," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 318, no. 3, pp. E330–E342, Mar. 2020. [Online]. Available: <https://doi.org/10.1152/ajpendo.00272.2019>

- [416] G. Thorleifsson, G. B. Walters, D. F. Gudbjartsson, V. Steinthorsdottir, P. Sulem, A. Helgadottir, U. Styrkarsdottir, S. Gretarsdottir, S. Thorlacius, I. Jonsdottir, T. Jonsdottir, E. J. Olafsdottir, G. H. Olafsdottir, T. Jonsson, F. Jonsson, K. Borch-Johnsen, T. Hansen, G. Andersen, T. Jorgensen, T. Lauritzen, K. K. Aben, A. L. Verbeek, N. Roeleveld, E. Kampman, L. R. Yanek, L. C. Becker, L. Tryggvadottir, T. Rafnar, D. M. Becker, J. Gulcher, L. A. Kiemeney, O. Pedersen, A. Kong, U. Thorsteinsdottir, and K. Stefansson, "Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity," *Nature Genetics*, vol. 41, no. 1, pp. 18–24, Dec. 2008. [Online]. Available: <https://doi.org/10.1038/ng.274>
- [417] R. M. C. Ignacio, C. R. Gibbs, E.-S. Lee, and D.-S. Son, "Differential chemokine signature between human preadipocytes and adipocytes," *Immune Network*, vol. 16, no. 3, p. 189, 2016. [Online]. Available: <https://doi.org/10.4110/in.2016.16.3.189>
- [418] I. J. B. van der Zalm, E. S. van der Valk, V. L. Wester, N. M. A. Nagtzaam, E. F. C. van Rossum, P. J. M. Leenen, and W. A. Dik, "Obesity-associated t-cell and macrophage activation improve partly after a lifestyle intervention," *International Journal of Obesity*, vol. 44, no. 9, pp. 1838–1850, Jul. 2020. [Online]. Available: <https://doi.org/10.1038/s41366-020-0615-6>
- [419] A. M. El-Sayed, L. Seemann, P. Scarborough, and S. Galea, "Are network-based interventions a useful antiobesity strategy? an application of simulation models for causal inference in epidemiology," *American Journal of Epidemiology*, vol. 178, no. 2, pp. 287–295, May 2013. [Online]. Available: <https://doi.org/10.1093/aje/kws455>
- [420] A. Andersen, B. E. Holstein, L. Berntsson, and E. H. Hansen, "Parental symptoms and children's use of medicine for headache: data reported by parents from five nordic countries," *International Journal of Public Health*, vol. 57, no. 1, pp. 217–223, Aug. 2011. [Online]. Available: <https://doi.org/10.1007/s00038-011-0284-2>
- [421] H.-E. Sung, L. Richter, R. Vaughan, P. B. Johnson, and B. Thom, "Nonmedical use of prescription opioids among teenagers in the united states: Trends and correlates," *Journal of Adolescent Health*, vol. 37, no. 1, pp. 44–51, Jul. 2005. [Online]. Available: <https://doi.org/10.1016/j.jadohealth.2005.02.013>

- [422] E. I. Schistad, X. Y. Kong, A.-S. Furberg, E. Bäckryd, G. Grimnes, N. Emaus, L. A. Rosseland, T. Gordh, A. Stubhaug, B. Engdahl, B. Halvorsen, and C. S. Nielsen, "A population-based study of inflammatory mechanisms and pain sensitivity," *Pain*, vol. 161, no. 2, pp. 338–350, Oct. 2019. [Online]. Available: <https://doi.org/10.1097/j.pain.0000000000001731>
- [423] M. H. Baumann, G. Tocco, D. M. Papsun, A. L. Mohr, M. F. Fogarty, and A. J. Krotulski, "U-47700 and its analogs: Non-fentanyl synthetic opioids impacting the recreational drug market," *Brain Sciences*, vol. 10, no. 11, p. 895, Nov. 2020. [Online]. Available: <https://doi.org/10.3390/brainsci10110895>
- [424] I. D. Moldovan, K. Suh, E. Y. Liu, and A. Jolly, "Network analysis of cases with methicillin-resistant staphylococcus aureus and controls in a large tertiary care facility," *Am. J. Infect. Control*, vol. 47, no. 12, pp. 1420–1425, Dec. 2019.
- [425] B. Sowmiya, V. S. Abhijith, S. Sudersan, R. Sakthi Jaya Sundar, M. Thangavel, and P. Varalakshmi, "A survey on security and privacy issues in contact tracing application of covid-19," *SN Comput. Sci.*, vol. 2, no. 3, p. 136, Mar. 2021.
- [426] S. Lalmuhanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for covid-19 (SARS-CoV-2) pandemic: A review," *Chaos Solitons Fractals*, vol. 139, no. 110059, p. 110059, Oct. 2020.
- [427] M. Whaiduzzaman, M. R. Hossain, A. R. Shovon, S. Roy, A. Laszka, R. Buyya, and A. Barros, "A privacy-preserving mobile and fog computing framework to trace and prevent COVID-19 community transmission," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 12, pp. 3564–3575, Dec. 2020.
- [428] W. Lv, S. Wu, C. Jiang, Y. Cui, X. Qiu, and Y. Zhang, "Towards large-scale and privacy-preserving contact tracing in COVID-19 pandemic: A blockchain perspective," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 282–298, Jan. 2022.
- [429] L. Garg, E. Chukwu, N. Nasser, C. Chakraborty, and G. Garg, "Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model," *IEEE Access*, vol. 8, pp. 159 402–159 414, Aug. 2020.
- [430] D. J. Leith and S. Farrell, "Measurement-based evaluation of Google/Apple exposure notification API for proximity detection in a commuter bus," *PLoS One*, vol. 16, no. 4, p. e0250826, Apr. 2021.

- [431] H. Xu, L. Zhang, O. Onireti, Y. Fang, W. J. Buchanan, and M. A. Imran, "Beep-Trace: Blockchain-enabled privacy-preserving contact tracing for COVID-19 pandemic and beyond," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3915–3929, Mar. 2021.
- [432] F. R. de Grujil, "UV adaptation: Pigmentation and protection against overexposure," *Experimental Dermatology*, vol. 26, no. 7, pp. 557–562, May 2017. [Online]. Available: <https://doi.org/10.1111/exd.13332>
- [433] N. Egami, "Spillover effects in the presence of unobserved networks," *Political Analysis*, vol. 29, no. 3, pp. 287–316, Nov. 2020. [Online]. Available: <https://doi.org/10.1017/pan.2020.28>
- [434] A. Steptoe and A. V. D. Roux, "Happiness, social networks, and health," *BMJ*, vol. 337, no. dec04 2, pp. a2781–a2781, Dec. 2008. [Online]. Available: <https://doi.org/10.1136/bmj.a2781>
- [435] L. L. Yan, J. Peng, and Y. Tan, "Network dynamics: How can we find patients like us?" *Information Systems Research*, vol. 26, no. 3, pp. 496–512, Sep. 2015. [Online]. Available: <https://doi.org/10.1287/isre.2015.0585>
- [436] J. M. Fletcher, "Social networks and health outcomes," in *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK, 2018, pp. 12568–12572. [Online]. Available: https://doi.org/10.1057/978-1-349-95189-5_2926
- [437] (2022, Jun.) Affero gpl3.0. <https://www.gnu.org/licenses/agpl-3.0.html>.
- [438] D. nasjonale forskningsetiske komiteene. Relevant lovverk. [Online]. Available: <https://www.forskningsetikk.no/lovverk/>
- [439] ——. Artikler fra forskningsetisk bibliotek (fbib). [Online]. Available: <https://www.forskningsetikk.no/tema/>
- [440] J. Johansson, K. M. Kanstad, O. Løvsletten, E. Skog, and S. Grimsgaard. The tromsø study. [Online]. Available: https://uit.no/research/tromsostudy?p_document_id=708030&Baseurl=/research/#region_708042
- [441] A. Rowhani-Farid, M. Allen, and A. G. Barnett, "What incentives increase data sharing in health and medical research? a systematic review," *Research*

- Integrity and Peer Review*, vol. 2, no. 1, May 2017. [Online]. Available: <http://dx.doi.org/10.1186/s41073-017-0028-9>
- [442] T. Hulsen, "Sharing is caring—data sharing initiatives in healthcare," *International Journal of Environmental Research and Public Health*, vol. 17, no. 9, p. 3046, Apr. 2020. [Online]. Available: <http://dx.doi.org/10.3390/ijerph17093046>
- [443] P. Burns, *The R Inferno*. Barking, England: Lulu.com, Jan. 2012.
- [444] D. Eddelbuettel, *Seamless R and C++ Integration with Rcpp*. Springer New York, 2013. [Online]. Available: <https://doi.org/10.1007/978-1-4614-6868-4>
- [445] J. M. Chambers, *Programming with data*, 1st ed. New York, NY: Springer, Jun. 2004.
- [446] C. M. Gries, T. Biddle, J. L. Bose, T. Kielian, and D. D. Lo, "Staphylococcus aureus fibronectin binding protein a mediates biofilm development and infection," *Infection and Immunity*, vol. 88, no. 5, Apr. 2020. [Online]. Available: <https://doi.org/10.1128/iai.00859-19>
- [447] G. Provenza, M. Provenzano, L. Visai, F. M. Burke, J. A. Geoghegan, M. Stravalaci, M. Gobbi, G. Mazzini, C. R. Arciola, T. J. Foster, and P. Speziale, "Functional analysis of a murine monoclonal antibody against the repetitive region of the fibronectin-binding adhesins fibronectin-binding protein a and fibronectin-binding protein b from staphylococcus aureus," *FEBS Journal*, vol. 277, no. 21, pp. 4490–4505, Sep. 2010. [Online]. Available: <https://doi.org/10.1111/j.1742-4658.2010.07835.x>
- [448] W. Zimmerli, "Clinical presentation and treatment of orthopaedic implant-associated infection," *Journal of Internal Medicine*, vol. 276, no. 2, pp. 111–119, Jul. 2014. [Online]. Available: <https://doi.org/10.1111/joim.12233>
- [449] O. Denis, "Route of transmission of staphylococcus aureus," *The Lancet Infectious Diseases*, vol. 17, no. 2, pp. 124–125, Feb. 2017. [Online]. Available: [https://doi.org/10.1016/s1473-3099\(16\)30512-6](https://doi.org/10.1016/s1473-3099(16)30512-6)
- [450] A. J. Oliveira, M. Rostila, A. P. de Leon, and C. S. Lopes, "The influence of social relationships on obesity: Sex differences in a longitudinal study," *Obesity*, vol. 21, no. 8, pp. 1540–1547, Jul. 2013. [Online]. Available:

- <https://doi.org/10.1002/oby.20286>
- [451] K. Powell, J. Wilcox, A. Clonan, P. Bissell, L. Preston, M. Peacock, and M. Holdsworth, "The role of social networks in the development of overweight and obesity among adults: a scoping review," *BMC Public Health*, vol. 15, no. 1, Sep. 2015. [Online]. Available: <https://doi.org/10.1186/s12889-015-2314-0>
- [452] J. Fowler and N. Christakis, "Estimating peer effects on health in social networks: A response to cohen-cole and fletcher and trogdon, nonnemaker, and pais," *Journal of Health Economics*, vol. 27, no. 5, pp. 1400–1405, Sep. 2008. [Online]. Available: <https://doi.org/10.1016/j.jhealeco.2008.07.001>
- [453] H. G. Dixon, M. L. Scully, M. A. Wakefield, V. M. White, and D. A. Crawford, "The effects of television advertisements for junk food versus nutritious food on children's food attitudes and preferences," *Social Science and Medicine*, vol. 65, no. 7, pp. 1311–1323, Oct. 2007. [Online]. Available: <https://doi.org/10.1016/j.socscimed.2007.05.011>
- [454] P. G. Turner and C. E. Lefevre, "Instagram use is linked to increased symptoms of orthorexia nervosa," *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, vol. 22, no. 2, pp. 277–284, Mar. 2017. [Online]. Available: <https://doi.org/10.1007/s40519-017-0364-2>
- [455] H. G. Dixon, M. L. Scully, M. A. Wakefield, V. M. White, and D. A. Crawford, "The effects of television advertisements for junk food versus nutritious food on children's food attitudes and preferences," *Social Science and Medicine*, vol. 65, no. 7, pp. 1311–1323, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0277953607002730>
- [456] A. Deoke, S. Hajare, and A. Saoji, "Prevalence of overweight in high school students with special reference to cardiovascular efficiency," *Global Journal of Health Science*, vol. 4, no. 2, Feb. 2012. [Online]. Available: <https://doi.org/10.5539/gjhs.v4n2p147>
- [457] A. Lasserre, "Worldwide trends in childhood obesity," *Swiss Medical Weekly*, Mar. 2007. [Online]. Available: <https://doi.org/10.4414/smw.2007.11707>
- [458] J. G. Woo, N. Zhang, M. Fenchel, D. R. Jacobs, T. Hu, E. M. Urbina, T. L. Burns, O. Raitakari, J. Steinberger, L. Bazzano, R. J. Prineas, C. Jaquish,

- M. Juonala, J. R. Ryder, S. R. Daniels, A. Sinaiko, T. Dwyer, and A. Venn, "Prediction of adult class II/III obesity from childhood BMI: the i3c consortium," *International Journal of Obesity*, vol. 44, no. 5, pp. 1164–1172, Oct. 2019. [Online]. Available: <https://doi.org/10.1038/s41366-019-0461-6>
- [459] J. Karczewski, E. Śledzińska, A. Baturo, I. Jończyk, A. Maleszko, A. Maleszko, P. Samborski, B. Begier-Krasińska, and A. Dobrowolska, "Obesity and inflammation," *European Cytokine Network*, vol. 29, no. 3, pp. 83–94, Sep. 2018. [Online]. Available: <https://doi.org/10.1684/ecn.2018.0415>
- [460] M. Safaei, E. A. Sundararajan, M. Driss, W. Boulila, and A. Shapi'i, "A systematic literature review on obesity: Understanding the causes and consequences of obesity and reviewing various machine learning approaches used to predict obesity," *Computers in Biology and Medicine*, vol. 136, p. 104754, Sep. 2021. [Online]. Available: <https://doi.org/10.1016/j.combiomed.2021.104754>
- [461] N. I. Eisenberger, M. Moieni, T. K. Inagaki, K. A. Muscatell, and M. R. Irwin, "In sickness and in health: The co-regulation of inflammation and social behavior," *Neuropsychopharmacology*, vol. 42, no. 1, pp. 242–253, Aug. 2016. [Online]. Available: <https://doi.org/10.1038/npp.2016.141>
- [462] M. Bang, J. Kim, S. K. An, Y. Youm, J. Chey, H. C. Kim, K. Park, K. Namkoong, and E. Lee, "Associations of systemic inflammation with frontotemporal functional network connectivity and out-degree social-network size in community-dwelling older adults," *Brain, Behavior, and Immunity*, vol. 79, pp. 309–313, Jul. 2019. [Online]. Available: <https://doi.org/10.1016/j.bbi.2019.01.025>
- [463] A. R. Henriquez, S. J. Snow, T. W. Jackson, J. S. House, D. I. Alewel, M. C. Schladweiler, M. C. Valdez, D. L. Freeborn, C. N. Miller, R. Grindstaff, P. R. S. Kodavanti, and U. P. Kodavanti, "Social isolation exacerbates acute ozone inhalation induced pulmonary and systemic health outcomes," *Toxicology and Applied Pharmacology*, vol. 457, p. 116295, Dec. 2022. [Online]. Available: <https://doi.org/10.1016/j.taap.2022.116295>
- [464] Y. Koyama, N. Nawa, Y. Yamaoka, H. Nishimura, S. Sonoda, J. Kuramochi, Y. Miyazaki, and T. Fujiwara, "Interplay between social isolation and loneliness and chronic systemic inflammation during the COVID-19 pandemic in japan: Results from u-CORONA study," *Brain, Behavior, and Immunity*, vol. 94, pp.

- 51–59, May 2021. [Online]. Available: <https://doi.org/10.1016/j.bbi.2021.03.007>
- [465] H. Lee, I. S. Lee, and R. Choue, "Obesity, inflammation and diet," *Pediatric Gastroenterology, Hepatology and Nutrition*, vol. 16, no. 3, p. 143, 2013. [Online]. Available: <https://doi.org/10.5223/pghn.2013.16.3.143>
- [466] N. V. Dhurandhar, "Countering the effects of chronic inflammation in obesity," *Korean Journal of Ophthalmology*, vol. 23, p. 73, 2014. [Online]. Available: <https://api.semanticscholar.org/CorpusID:76326638>
- [467] H. Marcos-Pasero, G. Colmenarejo, E. Aguilar-Aguilar, A. R. de Molina, G. Reglero, and V. Loria-Kohen, "Ranking of a wide multidomain set of predictor variables of children obesity by machine learning variable importance techniques," *Scientific Reports*, vol. 11, no. 1, Jan. 2021. [Online]. Available: <https://doi.org/10.1038/s41598-021-81205-8>
- [468] E. O. R. Kari Furu, Svetlana Skurtveit. (2005, Oct.) Selvrapportert legemiddelbruk hos 15–16-åringar i norge. <https://tidsskriftet.no/2005/10/originalartikkel/selvrapportert-legemiddelbruk-hos-15-16-aringer-i-norge>.
- [469] M. Lein. (2021, Dec.) Noen unge bruker paracet mot "alt". <https://nhi.no/forskning-og-intervju/Noen-unge-bruker-Paracet-mot-alt/>.

Appendix A: Main papers

A.1 Paper A



Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid



Social network analysis of *Staphylococcus aureus* carriage in a general youth population

Dina B. Stensen ^{1,2,*}, Rafael A. Nozal Cañadas ^{3,#}, Lars Småbrekke ⁴, Karina Olsen ⁵, Christopher Sivert Nielsen ^{6,7}, Kristian Svendsen ⁴, Anne Merethe Hanssen ⁸, Johanna U. Ericson ⁸, Gunnar Skov Simonsen ^{5,8}, Lars Ailo Bongo ³, Anne-Sofie Furberg ^{5,9}

¹ Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, 9019 Tromsø, Norway

² Division of Internal Medicine, University Hospital of North Norway, 9019, Tromsø, Norway

³ Department of Computer Science, UiT The Arctic University of Norway, 9019, Tromsø, Norway

⁴ Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, 9019, Tromsø, Norway

⁵ Department of Microbiology and Infection Control, Division of Internal Medicine, University Hospital of North Norway, 9019, Tromsø, Norway

⁶ Department of Chronic Disease and Ageing, Norwegian Institute of Public Health, 0473 Oslo, Norway

⁷ Department of Pain Management and Research, Division of Emergencies and Critical Care, Oslo University Hospital, 0424 Oslo, Norway

⁸ Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, 9019 Tromsø, Norway

⁹ Faculty of Health and Social Sciences, Molde University College, 6410 Molde, Norway

ARTICLE INFO

Article history:

Received 6 April 2022

Revised 26 July 2022

Accepted 19 August 2022

Keywords:

Staphylococcus aureus

Population-based study

Spa-type

Network analysis

ABSTRACT

Objectives: *Staphylococcus aureus* carriage increases the risk of infection. We used social network analysis to evaluate whether contacts have the same *S. aureus* genotype indicating direct transmission or whether contagiousness is an indirect effect of contacts sharing the same lifestyle or characteristics.

Methods: The Fit Futures 1 study collected data on social contact among 1038 high school students. *S. aureus* carriage was determined from two nasal swab cultures and the genotype was determined by spa-typing of positive throat swabs.

Results: *S. aureus* carriage and spa-type were transmitted in the social network ($P < 0.001$). The probability of carriage increased by 5% for each *S. aureus* positive contact. Male sex was associated with a 15% lower risk of transmission compared to the female sex, although the carriage prevalence was higher for men (36% vs 24%). Students with medium physical activity levels, medium/high alcohol use, or normal weight had a higher number of contacts and an increased risk of transmission ($P < 0.002$).

Conclusion: We demonstrated the direct social transmission of *S. aureus*. Lifestyle factors are associated with the risk of transmission, suggesting the effects of indirect social groups on *S. aureus* carriage, such as friends having more similar environmental exposures. The male predominance in the carriage is determined by sex-specific predisposing host characteristics as the social transmission is less frequent in males than females. Information on social networks may add to a better understanding of *S. aureus* epidemiology.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.

This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Introduction

Nasal carriage of *Staphylococcus aureus* has a prevalence of 20–30% in the general adult population (Olsen *et al.*, 2012; Stensen *et al.*, 2021)

and 40–50% in older children and adolescents (Stensen *et al.*, 2021) and is more common among men than women (Olsen *et al.*, 2012). Carriers have an increased risk of autoinfection (Bode *et al.*, 2010; Wertheim *et al.*, 2005). Therefore, prevention of nasal carriage may reduce the disease burden of *S. aureus* (Bode *et al.*, 2010). Epidemiologic studies have searched for modifiable risk factors for *S. aureus* nasal carriage as potential targets for interventions, including body mass index (BMI), serum glucose and vitamin D, exogenous and endogenous hormones, and smoking (Johannessen *et al.*, 2012; Olsen *et al.*, 2012; Olsen *et al.*, 2013;

* Corresponding author: Dina B. Stensen, Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Hansine Hansens veg 18, 9019 Tromsø, Norway. Fax: (0047) 77627015.

E-mail address: dina.b.stensen@uit.no (D.B. Stensen).

These authors have contributed equally to the development of the manuscript.

[Sangvik et al., 2011](#); [Stensen et al., 2019](#); [Wertheim et al., 2005](#)). However, these studies did not adjust for social contact.

Direct transmission of *S. aureus* is primarily through physical contact ([Knox et al., 2015](#)); however, no other study has evaluated the direct social transmission of *S. aureus* carriage in young populations. In studies that involve transmissible pathogens, an extensive problem with identifying the risk factors is the lack of adjustment for social contact. Biological host risk factors for *S. aureus* carriage may also be determinants of friendship, thereby producing an association by confounding. Predisposing lifestyle risk factors may be contagious with the consequence of researchers incorrectly assuming the transmission of the pathogen. Prevention of *S. aureus* carriage is dependent on identifying key transmission pathways and causal risk factors to correctly evaluate targets for interventions.

Infectious diseases like tuberculosis, HIV infection, and sexually transmitted diseases have been strongly connected to social networks ([Jolly and Wylie, 2002](#); [Klovadala et al., 2001](#); [Rothenberg et al., 1998](#)). These studies demonstrate that the degree and type of contact between individuals play a significant role in disease incidence. One study showed that the introduction of *S. aureus* into a social network of active drug users created a reservoir for the bacteria linked to the general population ([Gwizdala et al., 2011](#)). A recent case-control study used network analysis to reveal the transmission of methicillin-resistant *S. aureus* (MRSA) through social network in healthcare ([Moldovan et al., 2019](#)). Social group effects also occur in humans, as unrelated individuals living in the same household are found to have more similar microbiota than relatives living in different households ([Song et al., 2013](#)). Transmission of *S. aureus* has also been observed within households ([Miller et al., 2009](#)).

Therefore, the aim of this study is to estimate the extent to which *S. aureus* carriage follows friendship ties and whether the data support the concept of direct social transmission. We also aim to identify host risk factors for *S. aureus* carriage and differentiate between the risk attributable to social contact among similar individuals compared to biologic or lifestyle-related risk.

Methods

Population and study design

The Fit Futures 1 study (FF1) was a youth health survey conducted from September 2010 to April 2011, inviting all first-year students registered at 8 high schools in the municipalities of Tromsø and Balsfjord, North Norway. Altogether, 508 female and 530 male students participated in the survey (93% participation) ([Winther et al., 2014](#)). Participants visited the Clinical Research Unit of the University Hospital of North Norway (UNN) for interviews, questionnaires, clinical examinations, and microbiological samples. Trained research nurses performed all the procedures according to a standard protocol.

Host risk factors

Participants wore light clothing and no shoes. Height (cm) and weight (kg) were measured on an electronic scale with participants wearing light clothing and no shoes, and BMI (kg/m^2) was calculated. The participants reported their sex, age, study program, tobacco use, alcohol use, and recreational physical activity through a web-based questionnaire.

The interview covered current hormonal contraceptive use. We categorized hormonal contraceptive use into progestin-only contraceptives and combination contraceptives with high or low ethinylestradiol dosage ([Stensen et al., 2019](#)).

Assessment of *S. aureus* carriage

The research nurses took a first set of nasal and throat swab samples at the hospital, and a second set at school after a mean interval of 17 days. Nasal vestibules were sampled using the same 0.9 % NaCl-moistened sterile rayon-tipped swab, and both tonsillar regions were sampled with another swab. The swabs were immediately placed in a transport medium (Amies Copan, Brescia, Italy) and stored at 4°C for a maximum of 3 days. All samples were analyzed at the Department of Microbiology and Infection Control, UNN, both by direct culture ([Olsen et al., 2012](#)) and enrichment culture ([Stensen et al., 2019](#)) (Bacto Staphylococcus medium broth, Difco Laboratories, Sparks, Maryland, USA), using blood agar for growth control (Oxoid, UK) and chromID-plates for *S. aureus* detection (SAID, bioMérieux, Marcy l'Etoile, France) and MRSA agar plates SmithMed AS/Microbiological media production, Department of Microbiology and Infection Control, UNN). The growth of any bacterial colonies on agar plates was registered as a valid sample. The dominating *S. aureus* colony type was frozen at -70°C in glycerol-containing liquid medium after confirmation by Staphaurex plus agglutination test (bioMérieux, Marcy l'Etoile, France).

We used *S. aureus* persistent nasal carriage as the main outcome variable in the present analysis as this has been the major phenotype of interest in infection control and epidemiologic studies ([van Belkum et al., 2009](#)). We defined persistent carriage as having either two positive direct cultures or two positive enrichment cultures (Supplementary Figure 1). All *S. aureus* isolates from the first throat swab sample taken at the hospital were spa-typed (staphylococcal protein A) as part of another study by [Sangvik et al. \(2011\)](#).

Social network

We constructed the social network based on the interview question: "Which first level high school students have you had most contact with the last week? Name up to five students at your own school or other schools in Tromsø and Balsfjord." Reciprocity in the nomination was not mandatory. For each nomination, five "yes/no" questions assessed the type of contact: "Did you have physical contact?", "Have you been together at school?", "Have you been together at sports?", "Have you been together at home?", "Have you been together at other places?". This gave five social networks depending on the setting: "physical contact", "school", "sport", "home", and "other" networks. Adding all the relationships together formed a sixth network that was called the "overall" network. To evaluate if the friends mentioned were representative for the participants' social network, the following question was asked: "To what degree does this list of friends give an overview of your social network? Please indicate on a scale from zero (small degree) to ten (high degree)." We excluded 134 nominated friends that did not participate in FF1.

Statistical analysis

We used R version 3.6.3 and R Studio 1.3.1093 for the statistical analysis. To evaluate univariable associations between host factors and *S. aureus* persistent carriage we used Student's *t*-test and chi-square test, with Yates's correction for 2 × 2 tables and Fisher's exact test, when applicable.

In the social network analysis, nodes refer to participants in the network while edges refer to lines representing relationships between participants. To evaluate transmission of *S. aureus* through the social network, we analyzed edges between nodes using Exponential Random Graph Models or additive and multiplicative effects models. We analyzed patterns of connections (non-carriers

Table 1

Characteristics of the study population by *Staphylococcus aureus* persistent nasal carriage determined by direct and enrichment culture. The Fit Futures 1 study (N = 1038).

	Direct culture			Enrichment culture		
	Positive ^c	Negative ^c	Prevalence	Positive ^c	Negative ^c	Prevalence
Sex	< 0.001			< 0.001		
Male	193	337	36.4 %	255	275	48.1 %
Female	122	386	24.0 %	187	321	36.8 %
Study program	0.99			0.08		
General	118	272	30.3 %	163	227	41.8 %
Sports	31	73	29.8 %	55	49	52.9 %
Vocational	166	378	30.5 %	224	320	41.2 %
Smoking	0.93			0.48		
Daily	14	34	29.2 %	24	24	50.0 %
Sometimes	59	129	31.4 %	76	112	40.4 %
Never	236	546	30.2 %	333	449	42.6 %
Snuff use	0.79			0.30		
Daily	73	172	29.8 %	107	138	43.7 %
Sometimes	43	88	32.8 %	63	68	48.1 %
Never	192	450	29.9 %	263	379	41.0 %
Body mass index category	0.21			0.22		
< 18.5 kg/m ²	35	75	31.8 %	55	55	50.0 %
18.5–<25 kg/m ²	201	509	28.3 %	289	421	40.7 %
25–<30 kg/m ²	54	93	36.7 %	68	79	46.3 %
≥30 kg/m ²	22	45	32.8 %	27	40	40.3 %
Physical activity^a	0.15			0.07		
None	80	149	34.9 %	107	122	46.7 %
Light	99	239	29.3 %	129	209	38.2 %
Medium	67	192	25.9 %	105	154	40.5 %
Hard	63	131	32.5 %	93	101	47.9 %
Alcohol intake	0.32			0.780		
Never	88	192	31.4 %	115	165	41.1 %
≤ 1 Month	134	286	31.9 %	183	237	43.6 %
≥ 2 Month	86	232	27.0 %	134	184	42.1 %
Hormonal contraceptives^b	0.76			0.68		
Non-user	78	249	23.9 %	121	206	37.2 %
Progestin-only	3	17	15.0 %	5	15	25.0 %
Combination contraceptives, low estradiol	12	38	24.0 %	19	31	38.0 %
Combination contraceptives, high estradiol	26	73	27.1 %	39	60	39.4 %

^a Physical activity in leisure time: None = reading, watching TV, or other sedentary activity; Low level = walking, cycling, or other forms of exercise at least 4 hours a week; Medium level = participation in recreational sports, heavy outdoor activities with minimum duration of 4 hours a week; High level = Participation in heavy training or sports competitions regularly several times a week.

^b Hormonal contraceptives: Non-user = No current use of hormonal contraceptives (women only); Progestin-only = Use of hormonal contraceptives with progestin (Cerazette, Neplanon, Depo-provera, Implanon); Combination contraceptives, low estradiol = Use of hormonal contraceptives with progestin and ethinyl estradiol less than or equal to 20 µg (Mercilon, Yasminelle, Loette 28, Nuvaring). Combination contraceptives, high estradiol = Use of hormonal contraceptives with progestin and ethinyl estradiol greater than or equal to 30 µg (Marvelon, Yasmin, Microgynon, Oralcon, Diane, Synfase, Evra, Zyrona). Women taking contraceptives, but who were unable to recognize the brand were removed from the analysis.

^c Positive = two consecutive nasal swab cultures positive for *S. aureus* Negative = one or none of two consecutive nasal swab cultures positive for *S. aureus*.

connected to non-carriers, non-carriers connected to carriers, carriers connected to carriers) using the autocorrelation model Simulation Investigation for Empirical Network Analysis (O'Malley and Marsden, 2008). In further analysis, we used bootstrapping of simulated networks against the observed network, descriptive analysis, and logistic regression to evaluate the effect of host risk factors. The statistical background for our methods is described in Supplementary material.

Results

Transmission of *S. aureus* carriage in a general population

In the general population with a mean age of 16.4 years (SD = 1.24, range 15–28), the prevalence of *S. aureus* persistent nasal carriage determined by direct culture was 30.3%, compared with 42.6% when using enrichment culture. No MRSA isolates were detected. Prevalence of persistent carriage was higher in male

compared to female participants; 36.4% versus 24.0% for direct culture, and 48.1% versus 36.8% for enrichment culture. We found no other significant differences in carrier prevalence between groups according to host characteristics (Table 1).

We first evaluated the FF1 social network structure based on all relationships between students in the five subnetworks (Supplementary Figure 2) and information about relationships and persistent carriage status of nodes in the "overall" network diagram (Figure 1). As the population was recruited from two neighboring municipalities, there were two distinct clusters of students. The number of edges within the high school cluster was higher than outside the cluster demonstrating high school as a strong driver of friendship (homophily of 87.8) (Figure 2). Likewise, participants tended to bond with similar students with respect to sex and lifestyle factors (Supplementary Table 1 and Supplementary Figure 3).

To evaluate the effect of the social network on transmission, we assessed relationships that shared the same *S. aureus* spa-type in

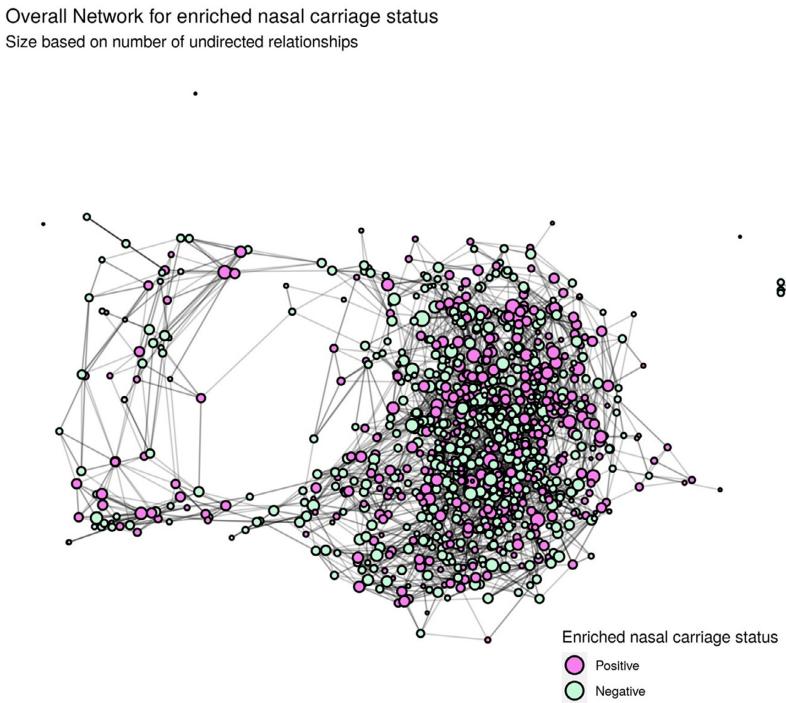


Figure 1. Overall network. The Fit Futures 1 study (N = 1038). *Staphylococcus aureus* persistent nasal carriage status determined by enrichment culture is highlighted for each student (Positive = *S. aureus* detected in two nasal swab samples; Negative = *S. aureus* detected in one or none of two nasal swab samples). Node size is proportional to the number of connections (undirected friendship).

the “overall” network. We registered 212 unique spa types among *S. aureus* throat isolates from 746 students. The 15 most prevalent spa types are listed in Table 2. The network analysis demonstrates that each high school has a unique distribution of spa types. Overall 126 edges of *S. aureus* carriers shared the same spa type. A total of 105 edges (83.3%) connected within the same high school, while 21 edges (16.7%) connected across different high schools (Figure 3). This suggests that spa-type is associated with friendship. The inclusion of non-typeable *S. aureus* strains did not affect the results and was therefore excluded from the analysis.

To test if there was a statistically significant influence of social networks on the transmission of *S. aureus*, we performed simulations of the current population for the six networks (Table 3). In the “overall” network, the transmission of *S. aureus* could be demonstrated for the persistent carriage determined by enrichment culture ($P = 0.02$). Transmission could also be demonstrated in the “school” network (direct culture: $P = 0.02$; enrichment culture: $P = 0.01$) and in the “physical contact” network (direct culture: $P = 0.06$; enrichment culture: $P = 0.04$). The same simulation-based analysis for spa-types showed transmission of *S. aureus* genotypes in all six social networks ($P < 0.001$).

The role of host risk factors in *S. aureus* transmission

In the logistic regression analysis, female participants had the highest risk of being exposed to *S. aureus* through their social interaction (Table 4). Men had a relatively low risk of transmission compared to women (0.85, 95% CI = 0.805–0.884). Also, students using alcohol twice or more per month had a higher risk of transmission of *S. aureus* compared to students using alcohol once per month or less ($P = 0.035$; direct culture). There was a higher probability of transmission among participants doing medium-level physical activity ($P = 0.008$) compared with the light physical activity group.

The mean number of friends for female students was 3.46, which was significantly higher than the average of 3.46 friends among male students ($P = 0.008$) (Supplementary Table 3). Students consuming alcohol more than twice a month had a higher number of friends compared to those consuming less or no alcohol ($P < 0.001$).

We demonstrate that students with a higher number of friends being persistent carriers were more likely to be persistent carriers themselves. This was significant for persistent carriage defined

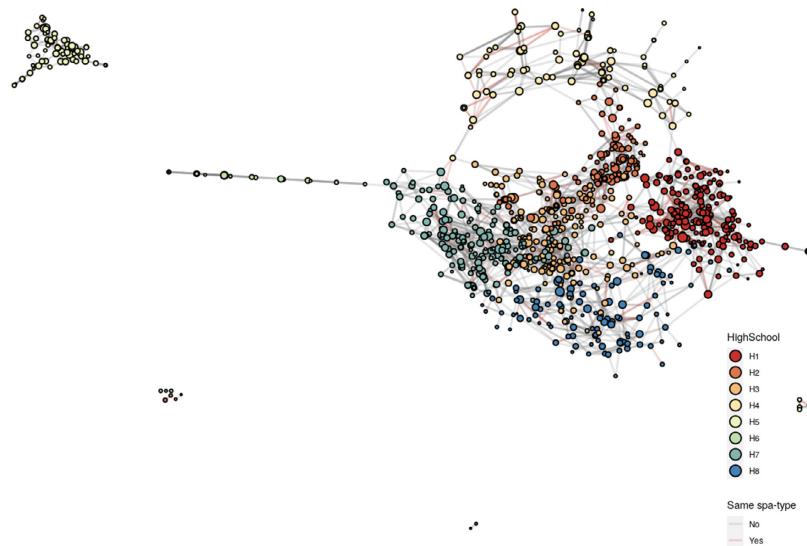


Figure 2. The overall social network within and between high schools, with a multidimensional scaling layout. The Fit Futures 1 study. Edges (lines) connecting nodes (students) with the same *Staphylococcus aureus* spa-type in throat culture are drawn in red. Edges connecting nodes with different spa-type are drawn in gray. High school ID (H1-H8) represents the eight high schools included in Fit Futures 1. H5 represents students at the high school in Balsfjord municipality (isolated cluster, upper left). All other high schools (H1-H4, H6-H8) are in Tromsø municipality. Only students with *S. aureus* isolated by direct or enrichment culture from the first throat swab sample are shown ($N = 746$). Unconnected students are not included ($N = 21$).

Table 2

The most prevalent spa-types for *Staphylococcus aureus* throat carriage. The Fit Futures 1 study ($N = 746$). Only persistent carriers are shown. The plots are the results for enrichment culture.

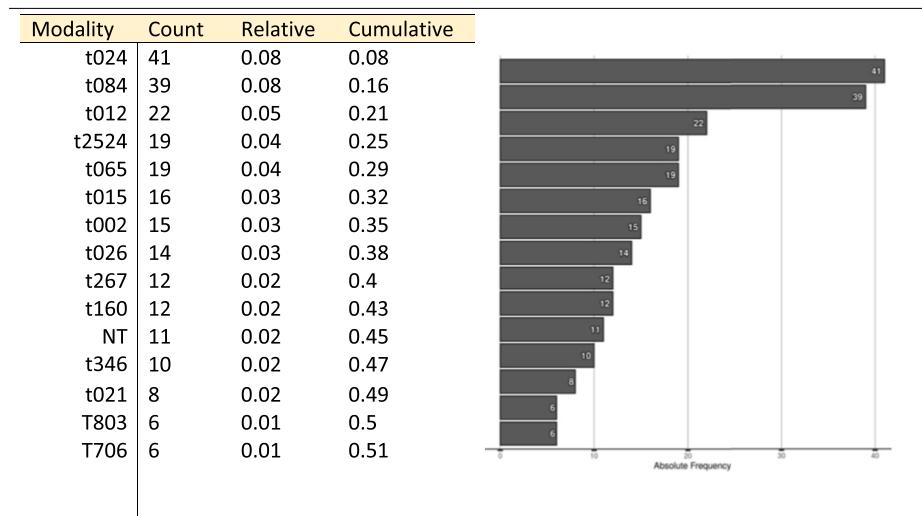




Figure 3. The overall social network with students grouped by high school. The Fit Futures 1 study. Edges (lines) connecting nodes (students) with the same *Staphylococcus aureus* spa-type in throat culture are drawn in red. Edges connecting nodes with different spa types are hidden. High school ID (H1-H8) represents the eight high schools included in the Tromsø Study Fit Futures 1. H5 represents students at the high school in Balsfjord municipality. All other high schools (H1-H4, H6-H8) are in Tromsø municipality. Only students with *S. aureus* isolated by direct or enrichment culture from the first throat swab sample are shown ($N = 746$). Unconnected students are not included ($N = 21$).

Table 3
Summary of 1000 simulations for each type of social network with respect to *Staphylococcus aureus* persistent nasal carriage and *S. aureus* spa-type. The Fit Futures 1 study. A detailed summary of this analysis is presented in Supplementary Table 4.

Network	<i>S. aureus</i> persistent nasal carriage		<i>S. aureus</i> throat colonization
	Direct culture ($N = 1038$)	Enrichment culture ($N = 1038$)	Spa-type ($n = 746$)
Overall	0.07	0.02	< 0.001
Physical	0.06	0.04	< 0.001
School	0.02	0.01	< 0.001
Sports	0.12	0.29	< 0.001
Home	0.34	0.39	< 0.001
Other	0.08	0.06	< 0.001

Numbers represent P -values from t -tests. Statistically significant values highlighted in bold.

by both direct culture ($P = 0.002$) and enrichment culture ($P < 0.001$) (Figure 3). The probability of being a carrier increased by 3.7% (95% CI = 3.52–3.94; univariable logistic regression, result not presented) on increasing the friend circle by one *S. aureus* positive friend (defined by direct culture), as illustrated in Figure 4 and Supplementary Table 4. Similarly, the probability increased by 3.4% (95% CI = 3.33–3.45) for persistent carriage defined by enrichment culture (result not presented).

An adapted linear autocorrelation analysis gave results comparable to the logistic regression analysis (Table 5). The probabil-

ity of persistent carriage increased by 4.8% ($P < 0.001$) for each additional *S. aureus* positive friend defined by direct culture after adjusting for host risk factors. A similar increase of 6.0% ($P < 0.001$) was observed for the enrichment culture. The autocorrelation model also assessed the risk factors that made the participants friends significantly more contagious. For direct culture there was an association between sex, BMI and physical activity ($P = 0.001$ –0.008), and for enrichment culture there was an association between study program, BMI, and physical activity ($P < 0.001$ for all). Because of the assumptions of the autocorrelation model,

Table 4

Associations between host risk factors and transmission of *Staphylococcus aureus* persistent nasal carriage in the overall social network. Results from two different regression analyses producing "P-value" for the social effect of each characteristic and Relative risk for the comparison of risk of transmission between groups. Persistent nasal carriage determined by both direct culture and enrichment culture. The Fit Futures 1 study (N = 1038).

Risk factor (categories)	Direct culture			Enrichment culture		
	P-value	Relative risk	95% CI	P-value	Relative risk	95% CI
Sex						
Female	0.0002	1		0.027	1	
Male	0.999	0.845	0.805 - 0.884	0.843	0.937	0.900 - 0.974
Study program						
Vocational	0.548	1		0.353	1	
General	0.510	1.002	0.950 - 1.054	0.410	0.996	0.954 - 1.038
Sport	0.403	1.009	0.960 - 1.059	0.811	0.974	0.935 - 1.013
BMI ^a						
Underweight	0.793	0.953	0.905 - 1.001	0.841	0.968	0.928 - 1.008
Healthy	0.150	1		0.301	1	
Overweight	1	0.901	0.860 - 0.943	0.763	0.974	0.935 - 1.012
Obese	0.914	0.941	0.896 - 0.987	0.246	1.003	0.959 - 1.047
Smoke						
Daily	0.294	1		0.855	1	
Never	0.503	0.986	0.937 - 1.034	0.486	1.022	0.978 - 1.066
Sometimes	0.723	0.971	0.922 - 1.020	0.262	1.036	0.991 - 1.081
Snuff						
Daily	0.406	1		0.596	1	
Never	0.414	0.999	0.949 - 1.049	0.310	1.017	0.973 - 1.060
Sometimes	0.900	0.962	0.915 - 1.010	0.821	0.986	0.947 - 1.025
Alcohol						
≥ 2 per month	0.035	1		0.434	1	
≤ 1 month	0.806	0.933	0.885 - 0.980	0.602	0.991	0.948 - 1.034
Never	0.739	0.938	0.890 - 0.986	0.323	1.006	0.964 - 1.049
Physical activity ^b						
Light	0.301	1		0.089	1	
None	0.994	0.930	0.886 - 0.975	0.803	0.952	0.914 - 0.990
Medium	0.008	1.053	0.999 - 1.107	0.267	0.982	0.941 - 1.023
Hard	0.883	0.959	0.913 - 1.005	0.817	0.951	0.913 - 0.989
Hormonal contra-ceptives ^c						
Non-user	0.444	1		0.494	1	
Progestin	0.369	1.126	-0.230 - 2.482	0.392	1.239	-0.576 - 3.054
Low Estrogen	0.430	1.024	-0.414 - 2.264	0.475	1.046	-0.840 - 2.932
High Estrogen	0.476	0.940	-0.546 - 2.425	0.483	1.027	-0.862 - 2.916

P-values from comparison between random network against a random network with only that particular category. Participants with missing values are excluded from the analysis. Statistically significant values highlighted in bold.

Relative risk and 95% confidence interval (95% CI) from univariable logistic regression analysis.

^a BMI by kg/m². Underweight = <18.5; Healthy = 18.5-24.9; Overweight = 25.0-29.9; ≥ 30

^b Physical activity: None = reading, watching TV, or other sedentary activity; Low level = walking, cycling, or other forms of exercise at least 4 hours a week; Medium level = participation in recreational sports, heavy outdoor activities with minimum duration of 4 hours a week; High level = Participation in heavy training or sports competitions regularly several times a week.

^c Hormonal contraceptives: Non-user = No current use of hormonal contraceptives (women only); Progestin-only = Use of hormonal contraceptives with progestin (Cerazette, Nexplanon, Depo-provera, Implanon); Combination contraceptives low estradiol = Use of hormonal contraceptives with progestin and ethynodiol less than or equal to 20µg (Mercilon, Yasminelle, Loette 28, Nuvaring). Combination contraceptives high estradiol = Use of hormonal contraceptives with progestin and ethynodiol greater than or equal to 30µg (Marvelon, Yasmin, Micronyclon, Oralcon, Diane, Synfase, Evra, Zyrona). Women taking contraceptives, but who were unable to recognize the brand were removed from the analysis

Table 5
Correlation between host risk factors and *Staphylococcus aureus* carrier status. Fit Futures 1 (N = 1038). Adapted multivariable linear autocorrelation model.

	Estimate ^a	SE	P-value
Direct culture			
ρ	0.048	0.011	<0.001
Sex	-0.048	0.028	0.0016
Study program	0.043	0.022	0.0542
BMI ^b	0.107	0.018	<0.001
Smoke	-0.012	0.027	0.650
Snuff	-0.001	0.020	0.968
Alcohol	0.038	0.021	0.066
Physical activity	0.033	0.012	0.008
Enrichment culture			
ρ	0.060	0.010	<0.001
Sex	-0.017	0.030	0.578
Study program	0.087	0.024	<0.001
Body mass index	0.085	0.019	<0.001
Smoke	-0.019	0.029	0.517
Snuff	0.001	0.022	0.950
Alcohol	0.070	0.022	0.002
Physical activity	0.046	0.014	<0.001

Significant values highlighted in bold.

^a Only estimates for the total model are valid. Beta estimates for individual host factors cannot be interpreted.

^b BMI = body mass index.

beta estimates for individual host factors could not be interpreted and sex-specific host factors (hormonal contraceptives) could not be included in the model. Females tended to have more relationships than males, which was also true for participants with normal BMI and participants with both medium and hard-level physical activity (Supplementary Table 3).

Discussion

In the present study, we demonstrated that social network is associated with *S. aureus* persistent-carrier status and spa type in a young population. This is, to our knowledge, the first study to analyze the transmission of *S. aureus* using social network analysis in a young population. We demonstrated that the probability of being a persistent carrier correlates with the number of close friends colonized with *S. aureus*. The autocorrelation analysis showed a 5-6% increased probability of *S. aureus* carriage with each additional *S. aureus* carrier friend. We also showed that friends tend to have the same spa types, indicating that the social network effect is partly driven by direct transmission of *S. aureus*. Our results coincide with former research that demonstrated comparable results in different cohorts (Gwizdala et al., 2011; Moldovan et al., 2019).

We analyzed different types of networks and found an association between transmission of *S. aureus* in the social network

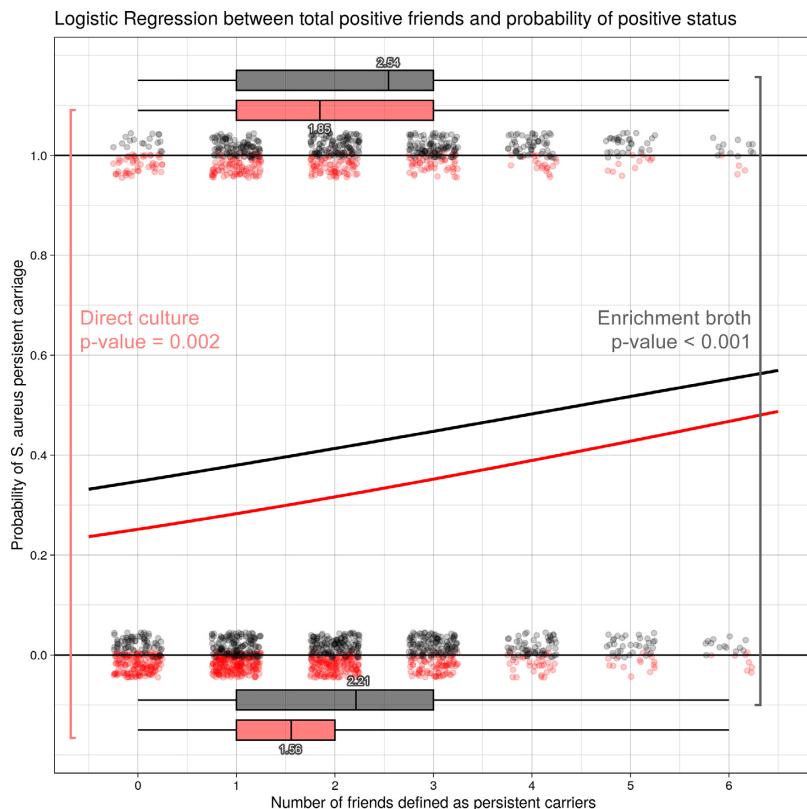


Figure 4. Probability of *Staphylococcus aureus* persistent nasal carriage with respect to friends' carrier status. Univariable logistic regression analysis. Fit Futures 1 ($n=1038$). Red color represents persistent nasal carriage defined by direct culture. Black color represents persistent nasal carriage defined by enrichment culture. The scatterplots show the distribution of persistent nasal carriers (along $Y=1$) and non-carriers (along $Y=0$) for sub-populations of students having from 0 to 6 *S. aureus* positive friends. Boxplots show the mean (middle line) and interquartile range (box limits) of *S. aureus* positive friends for persistent nasal carriers (at the top of the diagram) and non-carriers (at the bottom of the diagram). Outliers with more than 6 *S. aureus* positive friends are excluded from the figure, but did not affect the result ($N=3$). For more information see Supplementary Tables 6 and 7.

where participants confirmed they had physical contact (direct culture: $P = 0.07$; enrichment culture: $P = 0.04$). This might indicate that social contact is a key pathway for the spread of *S. aureus* in the community (Hogea et al., 2014), which is in line with former studies on transmission of *S. aureus* through a social network (Moldovan et al., 2019). Being together at school was also significantly associated with *S. aureus* transmission. We did not find any significant spread of *S. aureus* among students who were at home or participated in sports together.

There is also a substantial social dimension for several of the known host risk factors for *S. aureus* carriage, which suggests that social network effects may have contributed to associations observed in former studies. In a univariable logistic regression model, the risk of being a persistent carrier increased by 3.4–3.7% on increasing the friend circle by one *S. aureus* positive friend. A similar analysis (autocorrelation model) adjusted for host risk factors, showed an increase of 4.8–6.0%. The difference between the logistic

regression model and the autocorrelation model may partly represent the effect of the individual risk factors.

In our study, males had a higher prevalence of *S. aureus* persistent carriage compared to females which corresponds with previous studies (Knox et al., 2015; Mascaro et al., 2019). The social network analysis demonstrates that the female sex is the predominant social risk factor for carriage because of more relationships among females. This may substantiate the hypothesis of sex as a true biological risk factor for *S. aureus* carriage, as the male population has a higher prevalence of carriage while the relative risk of transmission is lower compared to the female population.

We also demonstrated increased transmission of *S. aureus* among students engaged in medium level physical activity in leisure time compared to those with sedentary leisure time. A former study showed an increased risk of *S. aureus* carriage in athletes doing contact sports (Mascaro et al., 2019). Many of the physical activities in youth are contact sports or close-counter train-

ing. In our population a higher percentage of women engaged in medium physical activity compared to men (women = 27% and men = 23%). The increased risk of transmission related to medium physical activity could therefore be partly attributed to the observed sex differences.

The use of alcohol more than twice a month was a social factor associated with the carriage of *S. aureus* by direct culture. This may reflect increased social contact with multiple friends at parties and social gatherings. Participants consuming alcohol more than twice a month had a higher number of friends than participants consuming less or no alcohol. We do not have information about the amount of alcohol consumed, and the alcohol variable is therefore lacking some precision. We also have some outliers that may have affected the results. We found no association between alcohol use and carriage defined by enrichment culture, this may be a result of a large number of statistical tests and the more homogenous variable with a high prevalence of *S. aureus* carriage for enrichment culture.

The autoregression model indicates an association between BMI and transmission of *S. aureus* in addition to sex, alcohol use, physical activity, and study program. The effect of friendship density might be partially related to body size, as students with normal BMI had more friends.

Excluding older outliers above 20 years ($n = 36$) from the network analysis did not affect the results, therefore all participants were included in the analysis. None of the interview questions on social networks provide information about the type or amount of physical contact. We also lack information on the total social network of the participants, including family and other social interactions outside school. Our model also lacks animal contacts and pets, that are known to influence transmission (Loeffler and Lloyd, 2010).

This will give a bias of unknown magnitude and direction. The social networks were constructed by self-reported information on social contacts one week before the study, and this could be misrepresenting of the participants' social contact over long periods of time. We therefore asked all participants to score the representativeness of the nominated friends, and 76 % of the participants claimed a score of five or above (on a scale from one to ten). We therefore believe the representativeness of the nominations to be high (Supplementary Figure 4).

We had complete spa-type data only from throat isolates, while nasal carriage is generally considered as the most clinically relevant phenotype. In a validation study of 100 participants with *S. aureus* isolated from cultures of two nasal and two throat swabs in FF1, 82 participants had the same spa type (data not shown). Therefore, we believe that our findings from social network analysis based on spa type of throat isolates also represent the transmission of nasal *S. aureus*. Another limitation is that we had 10 invalid nasal samples from the first swab and 51 invalid samples from the second swab. These were re-classified as negative for *S. aureus*. Because of the analysis of social networks, we believe that it would have introduced a larger bias in excluding parts of the social network compared with the bias of including potentially misclassified samples. The study was conducted between 2010–2011, and there may be some unknown bias in comparing results to present day. Although we believe that the prevalence of *S. aureus* is quite stable in the general population.

One limitation of defining the outcome of persistent nasal carriage is the number of swabs taken and use of Staphaurex plus agglutination test for *S. aureus* confirmation. Although the Staphaurex plus agglutination test could give false positive reactions for other staphylococci (van Griethuysen et al., 2001), we believe the combination of *S. aureus* selective CHROMagars and agglutination test increase sensitivity and specificity. Nouwen et al. proposed a "culture rule" that concludes that two nasal swabs taken at a week

interval can accurately classify *S. aureus* nasal carriage (Nouwen et al., 2004), but van Belkum et al. demonstrated a median survival of *S. aureus* of more than 154 days among persistent carriers, compared to 14 days among intermittent carriers and 4 days among non-carriers (van Belkum et al., 2009). Thus, it is likely that some participants were misclassified. Both limitations are classified as non-differential bias and therefore are more likely to give underestimations of results.

We reported results using both direct culture and enrichment. When enrichment broth is used, low bacterial loads are also detected, thereby giving an increased prevalence of *S. aureus* positive tests (Antri et al., 2018). In studies of decolonization, enrichment is recommended to prevent possible eradication failure (Diekema et al., 2011). The relevance of low bacterial load carriage in *S. aureus* epidemiology is not known, and most studies have used only direct culture. In this study, the results were similar for both definitions and might demonstrate the robustness of the findings.

Our analysis was modeled by using one time point, while interviews with the different participants were conducted at multiple time points. Most participants nominated friends who had the same attendance date as themselves, e.g., from their own school class (Supplementary Table 5). Furthermore, persistent nasal carriage is a relatively stable phenotype (van Belkum et al., 2009), and we therefore assume that time will not affect the present analysis.

In summary, our data from a general youth population supports social effects on *S. aureus* carriage and these result from both direct social transmission and shared lifestyle risk factors for carriage among friends. We demonstrated relationships between different social networks (i.e., overall, physical contact, school) and *S. aureus* persistent carriage and specific spa-types. We also showed that risk of transmission differs by host lifestyle factors. The male predominance in carriage is determined by sex-specific predisposing host characteristics, as social interactions among men are weak drivers of transmission compared with women. More studies are needed to further evaluate the interplay between the social environment and host risk factors in *S. aureus* carriage and should include household transmission and contact with animals.

Funding

This work was supported by The Northern Norway regional Health Authorities [grant number HNF1457-19]. The funding source had no role in study design, collection, analysis, interpretation of data, writing the report or the decision to submit the article for publication.

Ethical approval statement

Each participant signed a declaration of consent. Participants younger than 16 years had to bring written consent from a parent or guardian. FF1 was approved by The Regional Committee of Medical and Health Research Ethics North Norway (REK North, reference 2009/1282) and the Norwegian Data Protection Authority. The present study was approved by REK North (reference 2011/1710) and was conducted in accordance with the Declaration of Helsinki and national and institutional standards.

Author contributions

Anne-Sofie Furberg, Christopher Sievert Nielsen, Gunnar Skov Simonsen and Lars Ailo Bongo contributed with the conceptualization and design of the work. Anne-Sofie Furberg and Lars Ailo Bongo supervised the work. Johanna UE Sollid performed microbiological analysis of nasal and throat samples. Rafael A. Nozal Cañadas contributed with statistical analysis and statistical methods. Karina Olsen, Lars Småbrekke, Kristian Svendsen, Dina Stensen

and Anne Merethe Hanssen contributed in interpretation of data. Dina B. Stensen and Rafael A. Nozal Canadas wrote the original draft. All authors reviewed and approved the final manuscript. Rafael A. Nozal Canadas and Lars Ailo Bongo verify the underlying data.

Data availability

The data that support the findings of this study are available from The Fit Futures study but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon request and with permission of The Fit Futures study. Proposals for data should be directed to fitfutures@uit.no. Statistical analysis and consent form will be available on request. Proposals should be directed to dina.b.stensen@uit.no.

Declaration of Competing Interest

The authors have no competing interests to declare

Acknowledgments

We are grateful for the contribution by the participants of The Fit Futures 1 study. We also thank the Clinical Research Department at the University Hospital of North Norway, Tromsø, for data collection. Additionally, we thank Maria Sangvik and Trine Tessen for spa-typing. We also want to thank the Department of Medical Microbiology and the Department of Community Medicine at the Faculty of Health Sciences.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ijid.2022.08.018](https://doi.org/10.1016/j.ijid.2022.08.018).

References

- Antri K, Akkou M, Bouchiat C, Bes M, Martins-Simoes P, Dauwalder O, Tristant A, Meugnier H, Rasigade Jp, Etienne J, Vandenesch F, Laurent F, Ramdani-Bougessa N. High levels of *Staphylococcus aureus* and MRSA carriage in healthy population of Algiers revealed by additional enrichment and multilocus screening. *Eur J Clin Microbiol Infect Dis* 2018;37:1521–9.
- Bode LG, Kluytmans JA, Wertheim HF, Bogaars D, Vandenbroucke-Grauls CM, Roosendal R, Troelstra A, Box AT, Voss A, van der Tweel I, van Belkum A, Verbrugh HA, Vos MC. Preventing surgical-site infections in nasal carriers of *Staphylococcus aureus*. *N Engl J Med* 2010;362:9–17.
- Diekema D, Johannsson B, Herwaldt L, Beekmann S, Jernigan J, Kallen A, BERRIOS-Torres S, Polgreen P. Current practice in *Staphylococcus aureus* screening and decolonization. *Infect Control Hosp Epidemiol* 2011;32:1042–4.
- Gwizdala RA, Miller M, Bhat M, Vavagiakis P, Henry C, Neagis A, Shi Q, Lowy FD. *Staphylococcus aureus* colonization and infection among drug users: identification of hidden networks. *Am J Public Health* 2011;101:1268–76.
- Hogea C, Van Effelterre T, Acosta CJ. A basic dynamic transmission model of *Staphylococcus aureus* in the US population. *Epidemiol Infect* 2014;142:468–78.
- Johannessen M, Solid JE, Hanssen AM. Host- and microbe determinants that may influence the success of *S. aureus* colonization. *Front Cell Infect Microbiol* 2012;2:56.
- Jolly AM, Wylie JL. Gonorrhoea and chlamydia core groups and sexual networks in Manitoba. *Sex Transm Infect* 2002;78:145–51.
- Klovadahl AS, Graviss EA, Yaganehdoost A, Ross MW, Wanger A, Adams GJ, Musser JM. Networks and tuberculosis: an undetected community outbreak involving public places. *Soc Sci Med* 2001;52:681–94.
- Knox J, Uhlemann AC, Lowy FD. *Staphylococcus aureus* infections: transmission within households and community. *Trends Microbiol* 2015;23:437–44.
- Loeffler A, Lloyd DH. Companion animals: a reservoir for methicillin-resistant *Staphylococcus aureus* in the community? *Epidemiol Infect* 2010;138:595–605.
- Mascaro V, Capano MS, Iona T, Nobile CGA, Ammendola A, Pavia M. Prevalence of *Staphylococcus aureus* carriage and pattern of antibiotic resistance, including methicillin resistance, among contact sport athletes in Italy. *Infect Drug Resist* 2019;12:1161–70.
- Miller M, Cook HA, Furuya EY, Bhat M, Lee MH, Vavagiakis P, Visintainer P, Vasquez G, Larson E, Lowy FD. *Staphylococcus aureus* in the community: colonization versus infection. *PLoS One* 2009;4:e6708.
- Moldovan ID, Suh K, Liu EY, Jolly A. Network analysis of cases with methicillin-resistant *Staphylococcus aureus* and controls in a large tertiary care facility. *Am J Infect Control* 2019;47:1420–5.
- Nouwen JL, Ott A, Kluytmans-Vandenbergh MF, Boelens HA, Hofman A, van Belkum A, Verbrugh HA. Predicting the *Staphylococcus aureus* nasal carrier state: derivation and validation of a "culture rule". *Clin Infect Dis* 2004;39:806–11.
- Olsen K, Danielsen K, Wilsgaard T, Sangvik M, Solid JU, Thune I, Eggen AE, Simonsen GS, Furberg AS. Obesity and *Staphylococcus aureus* nasal colonization among women and men in a general population. *PLoS One* 2013;8:e63716.
- Olsen K, Falch BM, Danielsen K, Johannessen M, Ericson Solid JU, Thune I, Grimnes G, Jorde R, Simonsen GS, Furberg AS. *Staphylococcus aureus* nasal carriage is associated with serum 25-hydroxyvitamin D levels, gender and smoking status. The Tromsø staph and skin Study. *Eur J Clin Microbiol Infect Dis* 2012;31:465–73.
- O'Malley AJ, Marsden PV. The analysis of social networks. *Health Serv Outcomes Res Methodol* 2008;8:222–69.
- Rothenberg RR, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, Klovadahl AS. Social network dynamics and HIV transmission. *AIDS* 1998;12:1529–36.
- Sangvik M, Olsen RS, Olsen K, Simonsen GS, Furberg AS, Solid JU. Age- and gender-associated *Staphylococcus aureus* spa types found among nasal carriers in a general population: the Tromsø staph and Skin Study. *J Clin Microbiol* 2011;49:4213–18.
- Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, Gordon JI, Fierer N, Knight R. Co-habiting family members share microbiota with one another and with their dogs. *eLife* 2013;2:e00458.
- Stensen DB, Småbrekke L, Olsen K, Grimnes G, Nielsen CS, Simonsen GS, Solid JUE, Furberg AS. Hormonal contraceptive use and *Staphylococcus aureus* nasal and throat carriage in a Norwegian youth population. *PLoS One* 2019;14.
- Stensen DB, Småbrekke L, Olsen K, Grimnes G, Nielsen CS, Solid JUE, Simonsen GS, Almás B, Furberg AS. Circulating sex-steroids and *Staphylococcus aureus* nasal carriage in a general female population. *Eur J Endocrinol* 2021;184:337–46.
- van Belkum A, Verkaik NJ, De Vogel CP, Boelens HA, Verveer J, Nouwen JL, Verbrugh HA, Wertheim HF. Reclassification of *Staphylococcus aureus* nasal carriage types. *J Infect Dis* 2009;199:1820–6.
- Van Griethuysen A, Bes M, Etienne J, Zbinden R, Kluytmans J. International multicenter evaluation of latex agglutination tests for identification of *Staphylococcus aureus*. *J Clin Microbiol* 2001;39:86–9.
- Wertheim HF, Melles DC, Vos MC, van Leeuwen W, van Belkum A, Verbrugh HA, Nouwen JL. The role of nasal carriage in *Staphylococcus aureus* infections. *Lancet Infect Dis* 2005;5:751–62.
- Winther A, Dennison E, Ahmed IA, Furberg AS, Grimnes G, Jorde R, Gjesdal CG, Ennaus N. The Tromsø Study: Fit Futures: a study of Norwegian adolescents' lifestyle and bone health. *Arch Osteoporos* 2014;9:185.

Social network analysis of *Staphylococcus aureus* carriage in a general youth population

Supplementary Material

Dina B. Stensen, MD *^{1,2}, Rafael A. Nozal Cañadas, MSc *³, Lars Småbrekke, PhD ⁴, Karina Olsen, PhD ⁵, Christopher Sivert Nielsen, PhD ^{6,7}, Kristian Svendsen, PhD ⁴, Anne Merethe Hanssen, PhD ⁸, Johanna U. Ericson, PhD ⁸, Gunnar Skov Simonsen, PhD ^{5,8}, Lars Ailo Bongo, PhD ³, Anne-Sofie Furberg, PhD ^{5,9}

¹Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Hansine Hansens veg 18, 9019 Tromsø, Norway

²Division of Internal Medicine, University Hospital of North Norway, Sykehusvegen 38, 9019 Tromsø, Norway

³Department of Computer Science, UiT The Arctic University of Norway, Hansine Hansens veg 54, 9019 Tromsø, Norway

⁴Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Hansine Hansens veg 18, 9019 Tromsø, Norway

⁵Department of Microbiology and Infection Control, Division of Internal Medicine, University Hospital of North Norway, Sykehusvegen 38, 9019 Tromsø, Norway

⁶Department of Chronic Diseases and Ageing, Norwegian Institute of Public Health, Marcus Thranae gate 6, 0473 Oslo, Norway

⁷Department of Pain Management and Research, Division of Emergencies and Critical Care, Oslo University Hospital, Postboks 4956 Nydalen, 0424 Oslo, Norway

⁸Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, Hansine Hansens veg 18, 9019 Tromsø, Norway

⁹Faculty of Health and Social Sciences, Molde University College, Britvegen 2, 6410 Molde, Norway.

*These authors have contributed equally to the development of the manuscript

Corresponding author: Dina B. Stensen.

Institution: Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway.

Postal address: Hansine Hansens veg 18, 9019 Tromsø, Norway. E-mail: dina.b.stensen@uit.no

Table of contents

Statistical background	4
Supplementary Figure 1 Flowchart with inclusion criteria for definitions	7
Supplementary Figure 2 Overview of the different social networks.....	8
Supplementary Figure 3 Goodness of fit for the ERGM	9
Supplementary Figure 4 Histogram of representativeness of the social network.....	10
Supplementary Table 1 ERGM analysis of relationships within groups	11
Supplementary Table 2 Detailed summary of 1000 simulations for each social network.....	12
Supplementary Table 3 Average popularity in the overall network for each host risk factor..	13
Supplementary Table 4 Logistic regression model of carrier status with respect to friends....	15
Supplementary Table 5 Attendance dates for each high school	16

Statistical background

The use of Network analysis has increased exponentially over the last few years. In this brief introduction we include the statistical background on random graph analysis and autocorrelation networks models, and some included references to provide a deeper understanding of the topic.

In statistics, a general rule to solve problems is to find all possibilities and compare all the scenarios in which something happens against all scenarios in which something does not happen. Such a ratio will give you the probability of something occurring. In our case, it is impossible to compare random graphs with all possible random graphs to get the real probability. Such calculations are unattainable, and it is necessary to constrain the amount of possible random graphs based on some assumptions (1). The constraints added to the random graph will give a model which is similar enough to reality. In our case, we use the same frequency tables with a network with the same topology as constriction. We also assume that high contagiousness would cluster positives together with positives, and negatives together with negatives.

In this context, identical topology means having the same nodes (participants) and same edges (relationships) as the original network; but each node has randomly assigned attributes based on the probability distribution of each category (i.e., *S. aureus* persistent carrier status is assigned randomly to each node, following an arbitrary 30% prevalence probability, instead of using the original value).

Our bootstrapping (2) consists of counting how many relationships connect two nodes with the same attributes in our network (i.e., persistent carrier with persistent carrier or same *spa*-

types) in 1000 simulations. This gives us a distribution of 1000 values (with a mean and standard deviation) which we can compare to the real number of homophilic relationships in our network. We then perform simple hypothesis testing like t-test, where we consider a p-value of 0.05 or less to be statistically significant. As the numbers of these tests are low, there is no need for p-value correction for false positives.

Further, we can use the random network average number of relationships that we just created to compare with our network. We can create, again, a random network with the same topology, but using the conditional probability for each host factor independently. In this way we can check how much each of the categories deviates with respect to each other, and we can identify which category has higher or lower risk for the outcome variable.

Network autocorrelation models (3) are a special case of autoregression analysis in time series (4) where we want to find how much influence your neighbors (typically your social network of friends) have over you. We aim to find the ρ coefficient in the formula:

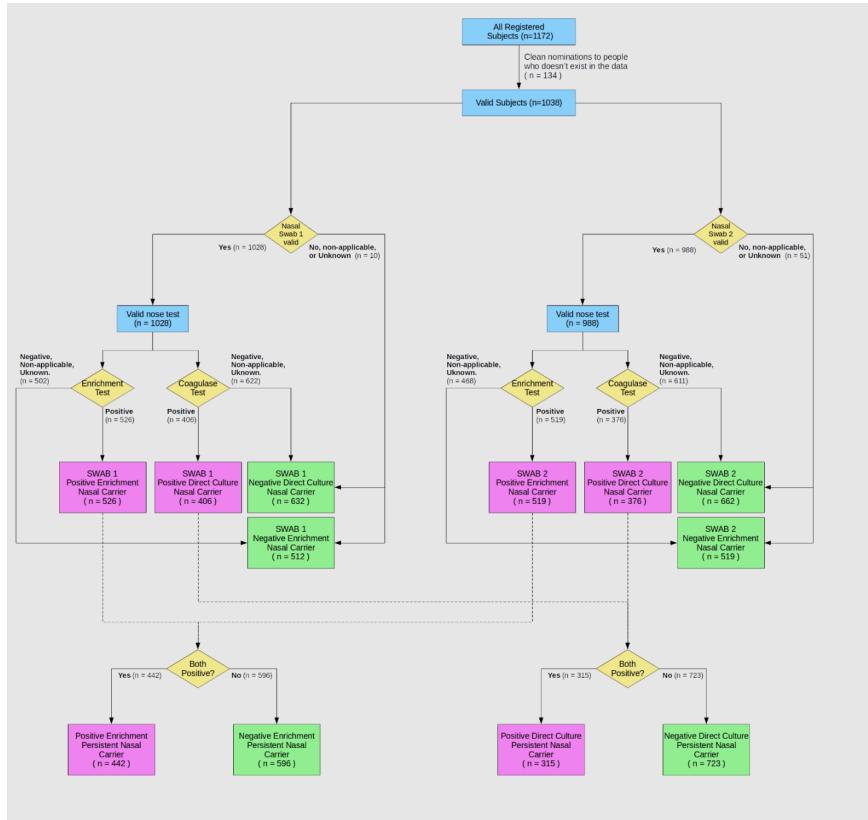
$$Y^{(t+1)} = \rho W_n Y^{(t)} + X\beta + \epsilon$$

W is a weighted matrix (typically normalized to 1) indicating which neighbors have influence over you, X is your explanatory variable (in our case, sex, BMI, smoke, and so on), β is a vector of coefficients (similar to linear regression) and ϵ is a random noise vector. Y is your dependent variable vector (in our case persistent carrier status), which over time (t), will converge to a common value. The ρ coefficient represents how much you are following the pressure of your neighbor influence and ranges typically from 0 to infinity, although negative values are also valid depending on your context. A value close to 0 would mean that you are completely ignoring your neighbors and the explanatory variables really do not have any influence on you. Positive values indicate that people can exert influence over you, and in our case, your friends will increase your risk of being a carrier.

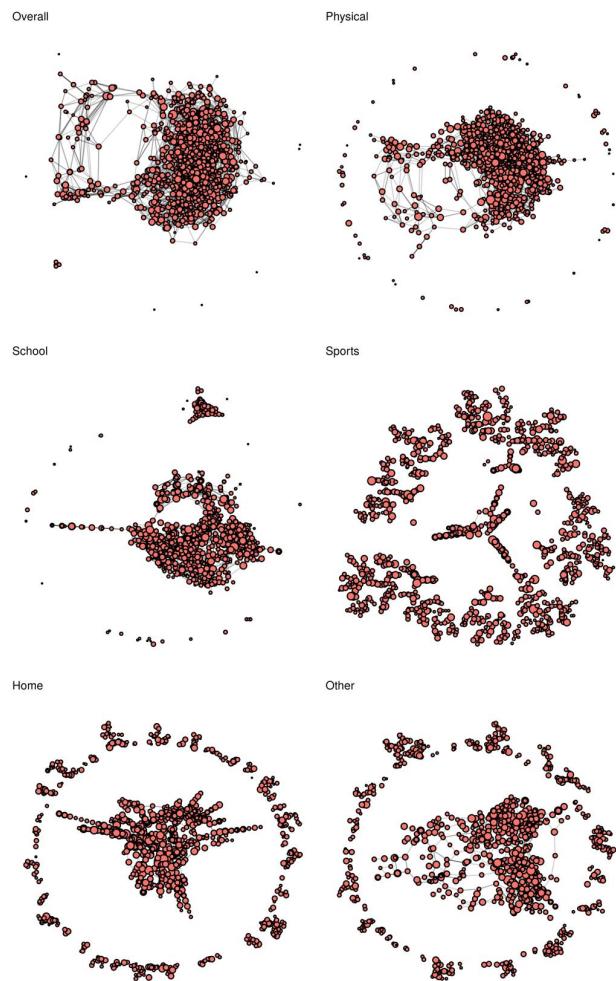
Significant negative values would indicate that you dislike your neighbors so much, that you will do the complete opposite of what he tells you to do. In our case, this is not a valid case for the ρ coefficient as you cannot protect someone from carriage, you simply will not transmit the bacteria. However, for the explanatory variables, negative values of ρ coefficient are valid in our case because we encode categorical variables with dummy variables.

References:

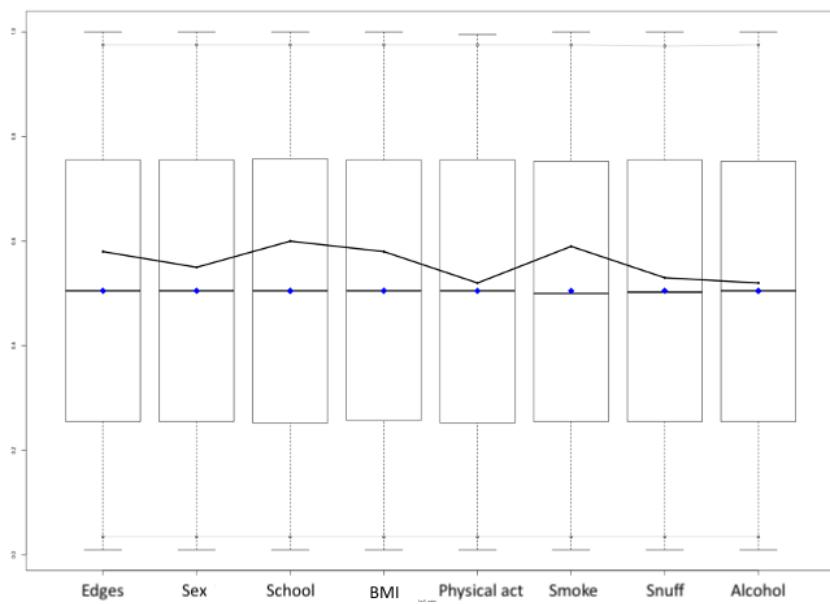
1. Bevan, David. Introduction to random graphs by Alan Frieze and Michał Karoński, Theorem 11.6. The Mathematical Gazette, Volume 101, Issue 551. Cambridge University Press; 2016. 464 p.
DOI: <https://doi.org/10.1017/mag.2017.110>
2. Snijders, Tom A B, Stephen P Borgatti. Non-Parametric Standard Errors and Tests for Network Statistics. Connections 22(2): 161-70. 1999. Available from:
http://www.analytictech.com/borgatti/papers/snijders_borgatti-density_significance.pdf
3. O'Malley AJ, Marsden PV. The Analysis of Social Networks. Health Serv Outcomes Res Methodol. 2008 Dec 1;8(4):222-269. doi: [10.1007/s10742-008-0041-z](https://doi.org/10.1007/s10742-008-0041-z). PMID: 20046802; PMCID: PMC2799303.
4. Anselin, L. Spatial Econometrics: Methods and Models. Netherland: Kluwer Academic; 1988.



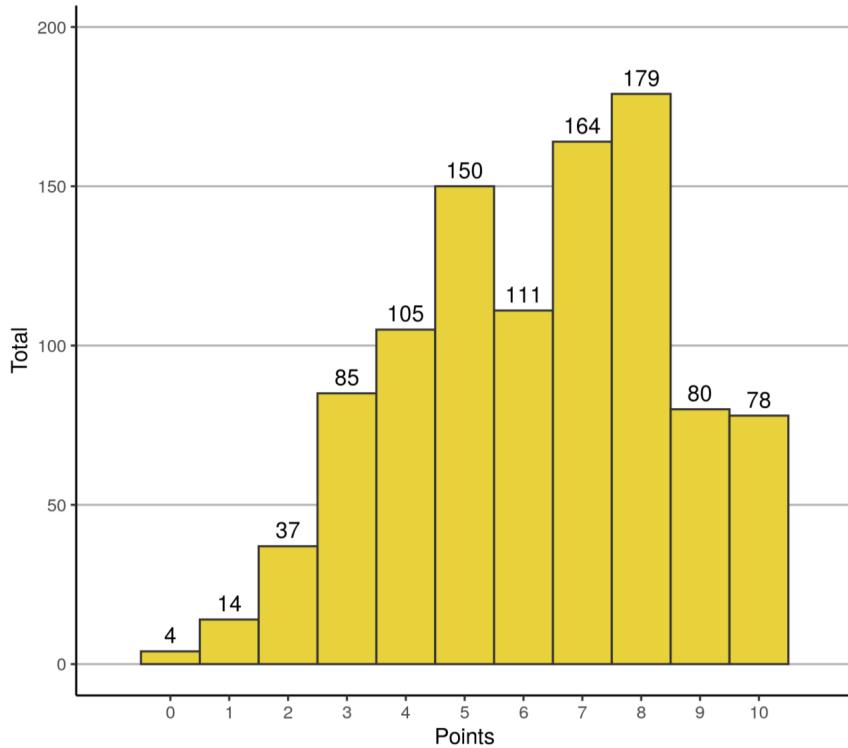
Supplementary Figure 1 Flowchart with inclusion criteria for definitions of *Staphylococcus aureus* persistent nasal carriage. The Fit Futures 1 study (N = 1038). Direct culture (bottom right) and enrichment culture (bottom left).



Supplementary Figure 2 Overview of the different social networks. The Fit Futures 1 study ($N = 1038$). From top to bottom and left to right: overall social network, physical contact, together at school, together in sports, together at home, and together in other settings. Each node represents a student. Each edge represents an undirected nomination (connection). The size of the node is proportional to the number of connections.



Supplementary Figure 3 Goodness of fit for the ERGM (Exponential Random Graph Model) analysis. The Fit Futures 1 study ($N = 1038$). Y-axis = proportion of statistics, X-axis = model statistics.



Supplementary Figure 4 Histogram of representativeness of the social network. The Fit Futures 1 study ($N = 1038$). 0 being not representative at all, and 10 being a perfect overview. For specific networks, having a mean score from top to bottom of school (6.61), other (6.52), physical (6.42) overall (6.29) sports (6.24) and home (6.13) (information not included in figure).

Supplementary Table 1 ERGM (Exponential Random Graph Model) analysis of relationships within groups of participants with the same characteristics. The Fit Futures 1 study (N = 1038).

	Homophily (%)	Estimate (logit)	Std Error	P-value
Edges	--	-8.41	0.08	< 0.001
Sex	84.05	1.47	0.05	< 0.001
School	87.85	2.16	0.06	< 0.001
BMI ^a	54.23	0.18	0.04	< 0.001
Smoke	68.06	0.22	0.04	< 0.001
Snuff	57.71	0.31	0.04	< 0.001
Alcohol	45.26	0.42	0.04	< 0.001
Physical activity	40.22	0.43	0.04	< 0.001

^a BMI = body mass index

Supplementary Table 2 Detailed summary of 1000 simulations for each social network. The Fit Futures 1 study (N = 1038).

Network	Total Relationships	Equal relationships	MIN	Q1	Median	Q3	MAX	SD	Direct culture P-value
Overall	3767	2260	2012	2136	2177	2214	2353	57	0.07
Physical	2823	1698	1492	1596	1628	1658	1756	45	0.06
School	2979	1814	1559	1687	1718	1747	1866	46	0.02
Sports	598	365	285	333	345	357	404	17	0.12
Home	1247	731	644	703	720	737	812	25	0.34
Others	1095	663	567	616	632	648	705	22	0.08
Network	Total relationships	Equal relationships	MIN	Q1	Median	Q3	MAX	SD	Enrichment culture P-value
Overall	3767	2013	1784	1899	1926	1953	2040	40	0.02
Physical	2823	1502	1339	1418	1442	1465	1576	34	0.04
School	2979	1610	1401	1499	1524	1548	1647	36	0.01
Sports	598	314	257	296	306	315	367	15	0.29
Home	1247	644	570	623	638	652	717	22	0.39
Others	1095	588	507	545	558	571	625	19	0.06
Network	Total relationships	Equal relationships	MIN	Q1	Median	Q3	MAX	SD	Spa-type P-value
Overall	1948	136	20	45	51	58	90	9.6	< 0.001
Physical	1459	111	16	33	38	44	84	8.0	< 0.001
School	1539	100	15	35	41	46	76	8.2	< 0.001
Sports	335	21	0	7	9	12	22	3.7	< 0.001
Home	664	63	4	14	17	21	38	5.1	< 0.001
Others	563	45	4	12	15	18	30	4.5	< 0.001

Columns 4-9^a contain the simulation summary statistics of the 1000 simulation result, in order, the minimum value of same-to-same relationships, first quartile, median rounded to the nearest integer, third quartile, maximum value, and standard deviation rounded to the nearest integer. The last column is the result of applying a t-test with the equal relationship against a distribution formed with the average of the 1000 simulations, and the standard deviation of the 1000 simulations. Significant p-values are highlighted in bold.

Supplementary Table 3 Average popularity in the overall network for each host risk factor.
The Fit Futures 1 study (N = 1038).

	Average Popularity ^a (3.62)	Relative physical isolation ^b (%)	Relative frequency all (%)
Sex	0.008		
Male	3.46	70	51.1
Female	3.81	30	48.9
BMI-category^c	0.001		
< 18.5 kg/m ²	3.61	9.23	10.60
18.5-<25 kg/m ²	3.72	62.31	68.40
25-<30 kg/m ²	3.63	14.62	14.16
> 30 kg/m ²	2.64	13.08	6.45
Smoking	0.003		
Daily	2.75	10	4.62
Sometimes	3.90	13.85	18.11
Never	3.60	72.31	75.34
Snuff use	0.003		
Daily	3.82	19.23	23.60
Sometimes	4.05	7.69	12.62
Never	3.45	69.23	61.85
Study program	0.004		
General	3.83	33.08	37.57
Sports	3.95	5.38	10.02
Vocational	3.43	62.31	52.41
Physical activity^d	0.254		
None	3.45	29.23	22.06
Light	3.56	30.77	32.56
Medium	3.73	21.54	24.95
Hard	3.80	14.62	18.69

Alcohol intake	< 0.001		
Never	3.05	41.54	26.97
<= 1 Month	3.76	30.77	40.46
> 2 Month	3.93	23.85	30.64
Direct culture persistent carriage	0.347		
Positive	3.72	30	30.35
Negative	3.59	70	69.65
Enrichment culture persistent carriage	0.007		
Positive	3.87	38.46	42.58
Negative	3.47	61.54	57.42
Hormonal Contraceptives (Women only, n = 505) ^e	Average Popularity (3.81)	Relative physical isolation (%)	Relative frequency all (%)
	0.006		
Non-user	4.00	58.97	64.88
Progestin only	2.65	13.16	3.97
Low Estrogen	3.44	13.16	9.92
High Estrogen	3.63	15.79	19.64
P-values are given next to variable names and represent a significant difference from the popularity average. P-values are calculated from t-test for two categories or ANOVA for more than two categories.			
^a Average popularity = Average number of friends nominating a participant as their friend			
^b Relative physical isolation = Number of participants not being nominated at all			
^c BMI = body mass index			
^d Physical activity: None = reading, watching TV, or other sedentary activity; Low level = walking, cycling, or other forms of exercise at least 4 hours a week; Medium level = participation in recreational sports, heavy outdoor activities with minimum duration of 4 hours a week; High level = Participation in heavy training or sports competitions regularly several times a week.			
^e Hormonal contraceptives: Non-user = No current use of hormonal contraceptives (women only); Progestin-only = Use of hormonal contraceptives with progestin (Cerazette, Nexplanon, Depo-provera, Implanon); Combination contraceptives low estradiol = Use of hormonal contraceptives with progestin and ethinyl estradiol less than or equal to 20µg (Mercilon, Yasminelle, Loette 28, Nuvaring). Combination contraceptives high estradiol = Use of hormonal contraceptives with progestin and ethinyl estradiol greater than or equal to 30µg (Marvelon, Yasmin, Microgynon, Oralcon, Diane, Synfase, Evra, Zyrcona). Women taking contraceptives, but who were unable to recognize the brand were removed from the analysis			

Supplementary Table 4 Logistic regression model of *Staphylococcus aureus* persistent nasal carrier status with respect to positive friends. The Fit Futures 1 study (N = 1038).

	Estimate	Std Error	P-value
Direct culture			
Intercept	- 1.09	0.11	<0.001
Number of friends that are persistent carriers	0.16	0.05	0.0016
Enrichment culture			
Intercept	- 0.63	0.12	<0.001
Number of friends that are persistent carriers	0.14	0.04	<0.001

Supplementary Table 5 Attendance dates for each high school. The Fit Futures 1 study (N = 1038).

Week	Year	H1	H2	H3	H4	H5	H6	H7	H8	Friends
38	2010	32	0	0	0	0	0	0	0	64.79 %
39	2010	24	0	0	0	0	0	0	0	56.39 %
40	2010	36	0	0	0	0	0	0	0	47.82 %
41	2010	36	0	0	0	0	0	0	0	57.22 %
42	2010	35	0	0	0	0	0	0	0	49.29 %
43	2010	30	0	0	0	0	0	0	0	54.56 %
44	2010	6	16	0	0	0	0	0	0	42.80 %
45	2010	0	40	0	0	0	0	0	0	57.79 %
46	2010	0	42	0	0	0	0	0	0	62.78 %
47	2010	0	32	0	0	0	0	0	0	60.57 %
48	2010	0	6	0	0	0	0	0	28	60.69 %
49	2010	4	0	0	0	0	0	0	34	48.51 %
50	2010	4	4	0	0	0	0	0	31	39.06 %
51	2010	0	0	0	0	0	0	0	0	100.00 %
52	2010	0	0	0	0	0	0	0	0	100.00 %
1	2011	0	2	0	0	0	0	6	27	32.14 %
2	2011	0	0	0	0	0	0	43	0	61.51 %
3	2011	0	0	0	0	0	0	45	0	47.00 %
4	2011	0	0	0	0	0	0	40	0	47.87 %
5	2011	0	0	0	0	0	0	46	0	54.24 %
6	2011	0	0	30	0	0	0	10	0	45.17 %

7	2011	0	0	41	0	0	0	0	0	50.57 %
8	2011	0	0	44	0	0	0	2	0	56.52 %
9	2011	0	0	43	0	0	0	0	0	53.10 %
10	2011	0	0	0	0	0	0	0	0	100 %
11	2011	0	0	8	12	17	0	0	0	69.10 %
12	2011	0	0	0	4	18	19	0	0	54.51 %
13	2011	0	0	0	15	24	5	0	0	45.04 %
14	2011	0	0	0	22	26	0	0	0	43.44 %
15	2011	0	0	2	31	0	2	0	0	43.76 %
16	2011	0	0	0	0	0	0	0	0	100.00 %
17	2011	0	0	0	14	0	0	0	0	33.10 %

The first attendance date was 2010-September-20th, which corresponds to Week 38 of 2010. The last attendance date was 2011-April-27th, which corresponds to week 17 of 2011. Notice the public holidays in Norway, during weeks 51 and 52 of 2010 is Christmas holidays, and week 16 of 2011 is Easter holiday. H1 to H8 correspond to each of the high school identifiers. The "Friends" column shows the average proportion of friends nominated by each participant who attended the Fit Futures 1 study in the same week as the subject himself/ herself. The weighted average for all weeks is 52.07%.

A.2 Paper B

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

 Rafael A. Nozal Cañadas,¹  Christopher Sivert Nielsen,²  Anne-Sofie Furberg,³  Anne Merethe Hanssen,⁴ and  Lars Ailo Bongo⁵

¹ Department of Computer Science, Faculty of Science and Technology, UiT The Arctic University of Norway
rca015@uit.no (Corresponding Author),

² Department of Chronic Diseases, Norwegian Institute of Public Health and Department of Pain Management, and Research, Division of Emergencies and Critical Care, Oslo University Hospital
christopersivert.nielsen@fhi.no,

³ Department of Microbiology and Infection Control, University Hospital of North Norway HF, Faculty of Health and Social Sciences, and Molde University College anne-sofie.furberg@uit.no

⁴ Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway
anne-merethe.hanssen@uit.no,

⁵ Department of Computer Science, Faculty of Science and Technology, UiT The Arctic University of Norway,
lars.ailo.bongo@uit.no

ABSTRACT

Background: Vitamin D status correlates with 25OHD levels which depends on nutritional intake and UVB exposure. These two factors are influenced by friends, as people tend to participate in the same activities or eat a similar diet to their peers.

Objectives: Investigate how social interactions in a general high school population above the Arctic Circle influence 25OHD levels in the population.

Methods: The Fit Futures 1 study was performed over 8 months and interview data on social contact among 1038 first-level students in 8 high schools in Northern Norway were collected. Serum levels of 25OHD were measured ($n = 890$). The participants filled in a questionnaire about nutritional consumption, solarium habits, ethnicity, and chronic diseases. The participant's BMI was also measured.

Results: Once high schools' social biases were accounted for, only UVB radiation levels explained the differences in 25OHD levels. For non-solarium users, logistic regression analysis showed a positive correlation between a person's 25OHD levels and the 25OHD level of friends in the general population and within the same high school for most of the schools. We saw that women can influence other women into going to the solarium, while this influence was not present for men.

Conclusions: 25OHD levels can be influenced by social networks. This study can help to add weight to current public health recommendations due to the positive spillover effect in the network.

Keywords Network Analysis · Vitamin D · Social Influence

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

1 Introduction

Humans tend to form friendship ties with individuals who are similar to them in certain aspects, referred to as homophily [1]. Although there is vast evidence that homophily shapes social networks, there is limited insight into how homophily functions in these networks. Most studies of homophily have been concerned with demographic variables such as age, sex, and social class. More recent evidence demonstrates deeper similarities among friends in behavior and personality [2]. Christakis and Fowler showed a spread of obesity through social ties [3]. If friends are more similar to one another in terms of healthy or unhealthy behavior, then social network proximity should be associated with similarity in biomarker profiles. Whether or not adolescents tend to associate with others who have similar lifestyles and experience similar lifestyle-related biomarker profiles has yet to be evaluated.

Vitamin D deficiency is emerging as a very common condition worldwide [4] and is associated with unfavorable skeletal outcomes, excess mortality, and a higher risk of infections. Serum 25-hydroxyvitamin D (25(OH)D) is considered to be the best biomarker of the body's vitamin D status and integrates Vitamin D derived from dietary intake and cutaneous synthesis after exposure to ultraviolet B (UVB) radiation of the solar electromagnetic spectrum. It has been estimated that the general European population gets 80-90% of its vitamin D from endogenous production in UVB-exposed skin. There is, however, considerable variation in this proportion across populations, population groups, and between individuals. Importantly, populations living at higher latitudes with periodic lack of photosynthesis may be at higher risk of vitamin D deficiency. Among a general youth population at 69°N participating in the Fit futures study, Tromsø Norway, 60% had vitamin D deficiency, defined by serum 25OHD below 50 nmol/l [5]. In the general adult population participating in the Tromsø Study, 19% had vitamin D deficiency, which is lower than the prevalence found in national data for Norway (28%) and in studies among adult populations further south [6–10].

Data from the Fit Futures and Tromsø study suggest that a sufficient level of serum 25(OH)D reflects several healthy lifestyle factors, such as outdoor activities, lower BMI, and fish-rich diet [5, 11], which may be defined by social interactions.

To our knowledge, no previous studies have been done on the effect of social networks in relation to vitamin D levels. Only one previous study has shown that poor economic factors influence health, because of the lack of vitamin D [12]. There has been a shift in social dynamics with indoor isolation during the COVID-19 pandemic, which showed how the severity of SARS-Covid-19 was linked to vitamin D deficiency [13]. Finally, ethnicity tends to be a strong social cohesion factor [14–26], and people of Middle Eastern, black, and South Asian descent require higher UVB levels and show higher deficiency prevalence than the white population [27–33]. Our objectives are to perform an explorative analysis in a general youth high school population to determine if social dynamics can affect 25OHD levels, and if so, to what extent environmental and lifestyle factors follow the same dynamics.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

2 Methods

2.1 Population and study design

The Fit Futures (FF) study [34] is a cohort with repeated health surveys among students from 8 high schools (H1-H8) in the Norwegian municipalities of Tromsø and Balsfjord (supplementary figure 4). FF1 was conducted from October 2010 to May 2011 (supplementary figure 5). All first-year students in the 8 high schools were invited (supplementary table 3), with consecutive inclusion of the eight schools. A total of 1117 youths were invited 93% attended, 508 girls (48.9%), and 530 boys. The age ranges from 15 to 28 years old, with 822 (79.2%) being 16 years or younger, and 52 (5%) older than 18 years. Students with special educational needs or mental disabilities are allowed to study for several years in high school in Norway.

The participants had a one-day visit to The Clinical Research Unit at the University Hospital of North Norway (UNN Tromsø), which included clinical examinations, microbiological samples, blood samples, an interview (self-reported social network, acute and chronic disease, medication, pregnancy), and a web-based general questionnaire. All procedures were performed by trained research nurses.

2.2 Social network assessment

The social network was constructed based on the following question: “*Which students have you had the most contact with the last week? Name up to 5 students at your own school or other schools in Tromsø and Balsfjord.*”. Reciprocity in the nomination was not mandatory. For each of the nominations, five “yes/no” questions assessed the type of contact they had with their nominations: “*Do you have physical contact?*”, “*Are you together at school?*”, “*Are you together at sports?*”, “*Are you together at home?*”, “*Are you together at other places?*”. This resulted in five social networks: Physical, School, Sport, Home, and Other. Adding all the relationships together formed the Overall Network. Illustrations for each network are presented in the supplementary materials (supplementary figure 6).

To evaluate if the friends mentioned were representative of the participant’s social network, the following question was asked: “*To what degree does this table of friends give an overview of your social network? Please indicate on a scale from 0 (small degree) to 10 (high degree).*” (supplementary figure 7).

2.3 Vitamin D assessment

Non-fasting blood samples were collected from an antecubital vein, and serum was separated and frozen at -70°C in the Fit Futures Biobank at the UiT The Arctic University of Norway. All serum samples (n = 890) were sent to the Hormone Laboratory, Haukeland University Hospital, Bergen, Norway; and 25OHD, 25OHD2, and 25OHD3 were analyzed by high-pressure liquid chromatography-mass spectroscopy (LC-MS/MS). A sample from all blood vials was reanalyzed at University College Cork, Cork, Ireland, by LC-MS/MS again as a part of the Vitamin D Standardization Project (VDSP) [35], and standardization was applied to the rest of the samples [36].

25OHD was used as a marker for vitamin D levels. This combines both sources of provitamin D + UVB, and D2+D3 from diet. It has a longer half-life span in blood than other available metabolites.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

In the analyses where categorical data is necessary, levels were defined as a binary variable dividing into vitamin D deficiency (< 50 nmol/L) or not vitamin D deficiency (≥ 50 nmol/L). In all cases, vitamin D toxicity is defined as greater than 150 nmol/L. [37–40]

25OHD levels are increased during pregnancy. All women who reported a possible pregnancy were given pregnancy tests, which all came back negative. The skin also loses the efficiency of synthesizing vitamin D with age [41], since the population is composed of young adults, no further analysis taking age into consideration was performed.

2.4 Diseases and medicine usage

There is a low prevalence of vitamin D absorption-impairing diseases, which are also heterogeneously distributed among the study population. Nobody in our population reported having bariatric surgery.

We have no significant number of students taking vitamin D-influencing medication such as anti-seizure drugs [42–44], steroid drugs [45–51], fat absorption reduction [45, 52–54], cholesterol metabolism modification [45, 55–57] or diuretics. The only anti-inflammatories drug [45–51], reported is “Ibuprofen 200mg” (n = 120), which does not affect the vitamin D level.

2.5 Melanin levels and ethnicity classification

The participants answered the question “*Do you consider yourself as...*”, with possible answers “Norwegian?”, “Sami”, “Kven / Finnish?” or “Other? (Please specify)”. Also, two questions regarding the country of birth of parents: “*Was your biological mother/father born in Norway? If not specify*”. These questions are summarized into a single variable with the ethnicity, or combinations of ethnicities, for each participant (supplementary table 6).

We combined all answers (n = 1018) into a single melanin quantity binary variable, dividing for assumed “Fair Skin” (996) for those of European, North American, or North Asia background, or mixing of any of these; and “Dark Skin” (22) for anyone with South American, African, South Asian, or any mixed background that included these. No conditions related to albinism or hyperpigmentation were self-reported in this population.

2.6 Solarium assessment

Visits to the solarium were recorded by the question “*Have you used a solarium during the last 4 weeks?*”. In Norway, access to solarium was restricted for teenagers from 2012, but not law-enforced until 2017 [58]. This data was gathered during 2010 and 2011 before any restrictions. Students were divided into solarium and non-solarium users according to their answers.

2.7 Nutritional information

There were 6 relevant questions regarding vitamin D intake. “*How often do you usually eat fat fish (e.g. salmon, trout, mackerel, herring)?*”, “*How often do you usually eat lean fish (e.g. cod, saithe, haddock)?*”, “*How much do you usually drink of whole milk, kefir and yoghurt?*”, “*How often do you usually eat cheese (all kinds)?*”, “*Do you take cod liver*

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

oil, cod liver oil capsules or fish oil capsules?", "Do you use vitamin or mineral supplements?". All answers were classified as categorical variables, ranging from never to every day.

In Norway at the time of the study, fortification of food with vitamin D was only common in low-fat milk and flavored milk [59] (0.4 µg vitamin D per 100g) and for baby formula (0.48–0.72 µg/100 kJ) [60].

Cholesterol is an important precursor in the transformation of 7-dehydro cholesterol (pro-vitamin D3) into pre-vitamin D3 via UVB catalyzation. Our population shows overwhelmingly healthy levels of HLD, and LDL, and no significant number of diseases that might affect the liver cholesterol biosynthetic pathway.

2.8 Anthropometric assessment

Weight and height were measured using an automatic electronic scale (Jenix DS 102 stadiometer, Dong Sahn Jenix, Seoul, Korea) with participants wearing light clothing and no footwear. Body mass index (BMI) is calculated as weight (kg) divided by the squared height (m²) with no correction for sex or age.

2.9 Physical activity assessment

The participants stated their physical activity level according to four hierarchical levels using a slightly modified version of the Saltin-Grimby Physical Activity Level Scale [61].

"Exercise and physical exertion in leisure time. If your activity varies much, for example between summer and winter, then give an average" with possible answers *"Reading, watching TV, or other sedentary activity?"*, *"Walking, cycling, or other forms of exercise at least 4 hours a week? (including walking or cycling to place of school, shopping, Sunday-walking, etc.)"*, *"Participation in recreational sports, heavy outdoor activities, snow clearing etc? (note: duration of activity at least 4 hours a week)"* and *"Participation in hard training or sports competitions, regularly several times a week?"*. The answers are shortened into "None", "Light", "Medium" and "Hard" respectively for convenience.

A related question was *"How many hours per day do you spend by the PC, watch TV, DVD etc. outside school during weekends?"* with answers ranging from "None" to "10 hours or more".

2.10 Recreational drugs assessment

The use of recreational drugs was self-reported via the web-based questionnaire. For alcohol consumption, the question was *"How often do you drink alcohol?"*, with possible answers "Never", "Once per month or less", "2-4 times per month", "2-3 times per week", "4 or more times per week". Due to the low number of answers in some of the categories, "2-4 times per month", "2-3 times per week", "4 or more times per week" are all combined into "Twice per month or more". For smoking the question was *"Do you smoke?"* with possible answers "No, never", "Sometimes" and "Daily". For snuff use the question was *"Do you use snuff?"* with possible answers "No, never", "Sometimes" and "Daily".

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

2.11 Natural UVB light, sun irradiance, and polar night

Tromsø and Balsfjord are located inside the Arctic Circle (60° N). The polar night started on the 23rd of November of 2010 and ended on the 19th of January 2011. During this time there is no sun irradiance. Solar irradiance above 60° N for 2011 is estimated to be about 8000 Wh/m^2 at its peak in July [62] (32.000 Wh/m^2 in the tropics at the same time [63]). After the polar night until the end of April, snow covers the ground; snow reflects UVB with about 86% efficiency. To determine which dates in our timeframe have relevant UVB, we estimated the duration of sun exposure to get 1000 IU of vitamin D synthesis. Assuming a clear day, snowy ground, type II skin, and 10% body exposure [64]: On March 1st it is not possible to acquire that amount. On March 15th it takes 5.22h. April 1st it takes 1.63h. On May 1st it takes 0.58h long. During autumn, assuming grass instead of snow, it is also not possible to acquire such an IU amount from October 1st.

Possible traveling was self-reported with the question: “*Have you been on a beach holiday during the last two months?*”, with only possible answers “Yes” and “No”. Traditionally, traveling might occur before high schools start near the end of August, in the nearly two weeks of winter break during Christmas, and a week during Easter break centered around the 24th of April in 2011. The data regarding traveling does not specify to which latitude, for how long, or estimate the grade of natural UVB irradiance.

2.12 Statistical Analysis

Statistical analyses were performed by using R version 4.1.2 and R Studio built 382.

Homophily, χ^2 tables, and bootstrapping with 1000 simulations were used to evaluate the similarities in solarium habits using a simple t-testing for theoretical same-to-same relationships against the simulated same-to-same relationships; as we have done in previous work [65]. Logistic regression was used to compare each person’s 25OHD levels (high/low) and the number of friends with high 25OHD levels. For both the general population, and for each high school individually.

χ^2 tables, two-sided Welch’s t-test when two categories are present, and a one-way ANOVA for the rest of the variables, were performed to determine statistically significant differences between groups. Bonferroni correction was applied for high numbers of multiple comparisons. Univariate regression models were used to compare relevant blood serum variables.

3 Results

3.1 Preliminary analysis of high school biases

There is a heterogeneous distribution of high or low melanin individuals across different high schools (supplementary table 6) with no significant bias ($p\text{-value} = 0.5$). The diet only has a bias in lean fish consumption ($p\text{-value} = 0.003$), with H1 and H8 consuming less than the expected frequency, and H3 and H4 consuming more than expected.

Blood extraction by date (supplementary figure 5) divides the school into Autumn 2010 (H1, H2), Winter 2010 (H3, H7, H8), and Spring 2011 (H4, H5, H6). Schools H2, H7, and H8 also have a significant number of students

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

with blood samples taken during the polar night. H1 and H6 show a bias towards sunbathing traveling (supplementary table 14). H1 was the first school to be tested so it is expected that students had recent travels due to the summer holiday. H6 students are expected to travel more related to training or competitions. The differences in 25OHD levels between solarium and non-solarium users are described further down.

3.2 Population levels of vitamin D

Previous studies done within the same population have shown an increase of 25OHD levels associated with vitamin and mineral supplements, physical activity, sunbathing holidays, and use of solariums for both sexes, while only men showed increased levels for the use of snuff, consumption of fortified milk, and fish liver oil [5]. Other studies in the general Tromsø population have shown that vitamin D has been positively associated with older age, blood sample time collection, sunbathing holiday, higher alcohol intake, use of fish oil and vitamin supplements, and negatively associated with smoking and obesity [66]. Adolescents in Tromsø have a higher prevalence of vitamin D deficiency compared with a similar-aged population in Spain [67]. We show 25OHD levels for comparison across categories of all variables of interest (table 1), however, the relevant variables in the general population do not seem to be relevant once we analyze high schools one by one.

“Solarium visit in the last 4 weeks” shows a dramatic difference in vitamin D deficiency between people who visit solariums (27.1%) and people who do not (70.7%). As such, all further analysis regarding high school bias is performed using only people who do not go to the solarium (supplementary tables 8, 9, 10, 11, 12 and 13).

“BMI” is known to be a risk factor for vitamin D levels. Fat cells can sequester vitamin D due to being a fat-soluble vitamin, and it is not surprising to find lower levels for higher %fat individuals. This population has a similar BMI for both men (22.51 ± 4.22) and women (22.62 ± 4.24) in the total population and for men (22.61 ± 4.38) and women (22.89 ± 4.48) in the non-solarium population. However, BMI is not evenly distributed across high schools for non-solarium users (p-value < 0.0001). There is a higher prevalence of obesity in H2 and H5, overweight in H8, and underweight in H4. H7 has a lower prevalence of overweight and obese individuals. All H6 students fall into the “Healthy” BMI category. H6 total number of students is low, and the p-value to ascertain a non-random habit is not significant. However, H6 is exclusively a sports high-school so it is likely that they are biased towards a healthier lifestyle even though the binomial test for this variable cannot suggest so.

“Sex” differences are partially explained because women visit the solarium more often than men and this influence will be discussed later. High schools are biased with respect to the sex distribution (p-value < 0.0001), with H1 and H8 leaning toward men, and H2 and H3 leaning toward women.

Recreational drug consumption also shows a strong bias towards high school for alcohol, smoking, and snuff habits (p-value < 0.0001 for all cases). Schools that show a pro-smoking bias are H1, H2, H5, and H8. Schools that show a pro-snuff bias are in H1 and have very strong usage in H8. Schools that show a pro-alcohol bias are H5 and very strong in H8. Anti-alcohol bias is shown in H6. Anti-snuff bias is shown in H7, and very slightly in H3 and H6. There seems to be no “no smoking” bias, but smoke frequency is lower in H3, H4, and H7. In H6 nobody smokes, but once again the p-value is not low enough to suggest a non-random habit, although H6 is a sport school so it is likely that they avoid smoking.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Physical activity also shows a bias with respect to high schools ($p\text{-value} < 0.0001$). With H2 and H8 leaning toward none, H1, H3, and H5 toward light to medium, and H6 and H7 toward hard.

Finally, we stratify all non-solarium users by high schools and run a two-sided Welch's t-test when two categories are present, and a one-way ANOVA for the rest of the variables. After correcting all results for multiple testing with Bonferroni, "Sex" was relevant for H8 ($p\text{-value} = 0.0059$, {men $n = 59$, $x = 28.45 \text{ nmol/l}$; women $n = 19$, $x = 46.34 \text{ nmol/l}$ }), "Physical Activity" for H3 ($p\text{-value} = 0.03$, {none $= 24$, $x = 30.89 \text{ nmol/l}$; light $n = 44$, $x = 40.14 \text{ nmol/l}$; medium $n = 42$, $x = 49.47 \text{ nmol/l}$; hard $n = 19$, $x = 51.2 \text{ nmol/l}$ }). "*Holiday / Sunbathing in the last 2 months*" was only significant in H1, first school to be tested after summer, and H3, first school to be completely tested after Christmas.

3.3 Similarities in solarium habits among friends

Women in the general population have higher vitamin D levels across the year (table 1 and figure 1), this also happens with people going to the solarium (table 1 and figure 2). We checked χ^2 tables for diet and solarium habits to test differences between men and women. The diet table showed no significant differences in diet, but significant differences in solarium habits were found ($p\text{-value} < 0.0001$) (table 2). Men tend to not go to the solarium (33.64% of the solarium population) while women tend to go to the solarium (66.46%). This might indicate that social relationships between men and women groups affect solarium habits.

Sex is a strong homophily variable defining friendship (84.05%), and men are friends with mostly men (72.06%) and women are friends with mostly women (72.89%). Our previous studies [65] on the data also show that men have fewer friends (3.37) on average than women (3.85) ($p\text{-value} = 0.02$). The homophily for solarium visitors is 68.36%, with "yes" having a homophily of 22% but not significant, and "no" having a 66% (8% lower than it should) with $p\text{-value} < 0.0001$. This means that people who do not go to the solarium are less likely to form friendships with each other. Moreover, it correlates with the sex dynamics of men having fewer friends, and men not going into the solarium as often.

We compared the social network against simulated networks to check if people going to the solarium influence other people into going to the solarium. For this analysis, we filtered out people who did not answer the question about the solarium ($n = 33$). Instead of the original 3767 relationships, we were left with 3575 relationships in total. Doing 1000 simulations we got an average of 2365 same-to-same relationships with a significance of $p\text{-value} = 0.008$ for the overall network. This indicates that people are biased and relationships with respect to solarium habits are non-random. We tried the same analysis again using only same-sex friends. Sex has very high homophily, as such we believe that barely any influence from men to women and vice-versa is lost in this analysis. For women, the total same-to-same relationships are 1049 with a $p\text{-value}$ of 0.0002 which seems to indicate that women tend to form groups of friends with the same solarium habits. For men, however, we have 1122 same-to-same relationships, resulting in a $p\text{-value}$ of 0.23, showing no bias in this case.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Table 1: Descriptive statistics with respect to all variables of interest. In the column "Variable" we show the population composition, the column "All" is the analysis of the whole population without missing values, and the column "Non-solarium only" represents the population that did not go to the solarium in the last 4 weeks. In "Variable", we show the categories of each variable (i.e.: Man, Woman in Sex), the absolute frequency of each category (n), and the relative frequency of each category (f). In both "All" and "Non-solarium only", we show the absolute frequency of people with valid 25OHD values (N), the average for each category (Average), how many values are below 50 nmol/l (N<50), and the relative frequency in each category of people with low vitamin D (F). To calculate the p-values we run a two-sided Welch's t-test when two categories are present, and a one-way ANOVA for the rest of the variables. Relevant variables and p-values are highlighted in bold. The Fit Futures 1 study.

Variable	All			No solarium only		
	n	f	N	Average	N <50	F
Solarium visit in the last 4 weeks						
Yes	217	0.209	199	66.62	54	0.271
No	78	0.759	711	40.79	503	0.707
<Missing values>	33	0.002				
Holidays / Sunbathing last 2 months						
Yes	77	0.074	73	62.05	22	0.286
No	928	0.894	836	45.12	533	0.638
<Missing values>	33	0.002				
P-value = <0.0001						
BMI						
Underweight	110	0.106	102	44.02	67	0.657
Healthy	710	0.684	647	49.09	368	0.569
Overweight	147	0.142	129	42.59	85	0.659
Obese	67	0.065	59	33.81	49	0.831
<Missing values>	4	0.004				
P-value = 0.0005						
General Health						
Very bad	7	0.007	7	57.87	3	0.429
Bad	53	0.051	49	48.10	26	0.531
Neither good nor bad	218	0.21	195	48.35	117	0.600
Good	497	0.479	458	45.60	286	0.624
Excellent	241	0.232	210	47.49	124	0.590
<Missing values>	22	0.021				
P-value = 0.4						
Smoke						
Never	782	0.753	716	46.80	430	0.601
Sometimes	188	0.181	165	47.45	99	0.600
Daily	48	0.046	42	42.58	31	0.738
<Missing values>	20	0.019				
P-value = 0.469						
Snuff						
Never	642	0.618	590	46.79	356	0.603
Sometimes	131	0.126	116	51.87	61	0.526
Daily	245	0.236	217	43.81	143	0.659
<Missing values>	20	0.019				
P-value = 0.009						
Alcohol						
Never	280	0.27	260	45.10	166	0.638
Once per month or less	420	0.405	374	47.16	226	0.604
Twice or more per moth	318	0.306	289	47.62	168	0.581
<Missing values>	20	0.019				
P-value = 0.397						
Physical Activity						
None	229	0.221	202	35.30	106	0.822
Light	381	0.372	301	45.23	195	0.646
Moderate	259	0.25	238	50.88	124	0.521
Hard	194	0.187	184	56.08	77	0.418
<Missing values>	18	0.017				
P-value = <0.0001						
Screen Time						
			N	P-value = 0.801		
About half an hour or less	43	0.041	40	56.05	21	0.522
About 1 to 1.5 hours	174	0.168	156	45.77	94	0.603
About 2 to 3 hours	38	0.372	351	47.03	213	0.607
About 4 to 6 hours	324	0.312	280	47.36	174	0.602
About 7 to 9 hours	68	0.066	63	43.83	41	0.651
10 hours or more	22	0.021	20	46.65	12	0.600
<Missing values>	21	0.020				
P-value = <0.0001						
High School						
			N	P-value = 0.0001		
H1	207	0.199	180	42.59	122	0.678
H2	142	0.137	121	42.39	84	0.694
H3	168	0.162	159	47.3	95	0.597
H4	98	0.094	92	47.27	50	0.543
H5	85	0.082	71	46.54	46	0.648
H6	26	0.025	24	63.68	7	0.292
H7	192	0.185	181	53.51	91	0.503
H8	120	0.116	111	41.64	76	0.683
P-value = <0.0001						
Skin type						
			N	P-value = 0.0114		
Fair	996	0.960	904	47.06	543	0.545
Dark	22	0.022	20	30.41	18	0.900
<Missing values>	20	0.019				
P-value = 0.0181						
Variable						
			All			
					No solarium only	

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

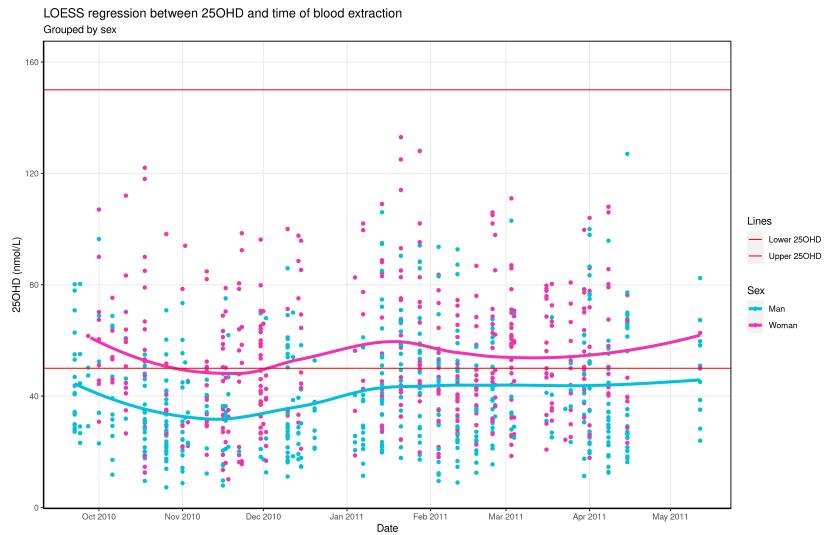


Figure 1: 25OHD levels from October 2010 to May 2011 divided by sex. Horizontal red lines mark the boundaries of healthy 25OHD levels. Women display higher levels than men across the whole year. The Fit Futures I study, $N = 890$.

Table 2: χ^2 table for “Did you go to the solarium in the last 4 weeks?” (yes/no), and sex. The header numbers indicate the total population absolute and relative frequency for men and women. The lower marginals are how many of these have available data with respect to the solarium question. Each inner cell is divided into 3 parts; the left-most is the total number of relationships in this combination, the center one is the expected number of relationships, and the right part contains an arrow indicating over (up) or underrepresented (down) using a two-sided binomial test with at least p -value < 0.1 . Women are biased toward the yes answer, while men are biased towards the no answer. The Fit Futures I study.

$\chi^2 < 0.0001$	Man		Woman		Total	Freq
	n = 530 , f = .51	n = 508 , f = .49				
Yes	73 (106)	↓	144 (103)	↑	217	21%
No	436 (386)	↑	352 (376)	↓	788	76%
Total	509		496		1005	
Frequency	49%		48%			97%

3.4 Social influence in 25OHD

We want to check if non-solarium students with high 25OHD levels influence other non-solarium students to have high 25OHD levels as well by using logistic regression. The result shows that the chances of having high vitamin D increase by 7.25% [5.65%, 8.85%] for each additional friend who also has high vitamin D (figure 3). This result however can be biased. For example, different high schools had different blood extraction dates around the year, and friendship has very high homophily for high school (87.5%), so it is possible that we are just measuring high schools with higher density connections, while those schools coincidentally extracted blood at peak or bottom sun irradiance.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

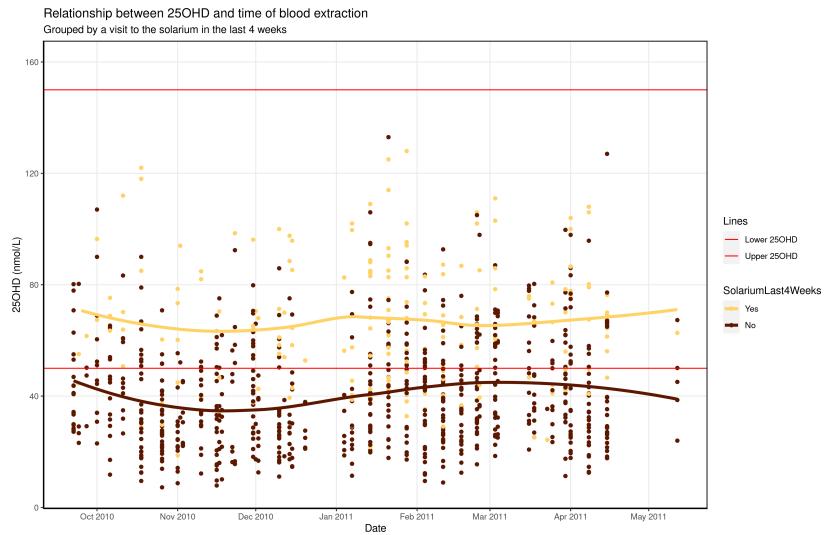


Figure 2: 25OHD levels from October 2010 to May 2011 divided by solarium habit. Horizontal red lines mark the boundaries of healthy 25OHD levels. People going to the solarium display higher levels than people who don't go to the solarium across the whole year. People who don't go to the solarium display a dip in levels in December while a peak of levels in near April. There was a second dip in the middle of May due to a small sample size and low vitamin D levels of students who were tested at that time. The Fit Futures I study. N = 890

We repeated the logistic regression analysis for each high school independently. Five of the eight also showed statistically significant 25OHD differences (figure 3) (H3, H5, H6, H7, and H8), with two having blood extraction during the polar night (H7 and H8). This suggests that friends influence each other into sharing certain lifestyles which makes 25OHD levels similar.

Furthermore, we plotted each student's 25OHD level against his or her friends' average of 25OHD. We see a weak correlation ($R^2= 0.1$, p-value < 0.0001) among the whole population (supplementary figure 17). The same analysis of each individual high school shows more relevant results for H2, H5, and H8 (supplementary figure 17). We also performed another 1000 simulations for each high school individually, for non-solarium users only, and found non-random friendship based on vitamin D status for H1, H2, H3, H4, H5, and H8 (supplementary table 5).

3.5 Levels by nutritional data

Finally, we wanted to check if the diet was influenced by friends. Performing ANOVA in all our nutritional variables showed no 25OHD significance difference between consumption groups for fat fish, cheese, dairy, vitamin pills, or even fish oil complements. Analyzing each high school independently, we found no correlation at all in any of the 48 possible combinations without even applying any post-hoc correction. The only difference present was in the consumption of lean fish in the general population, however, the difference was only between lower-frequency consumption groups and

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

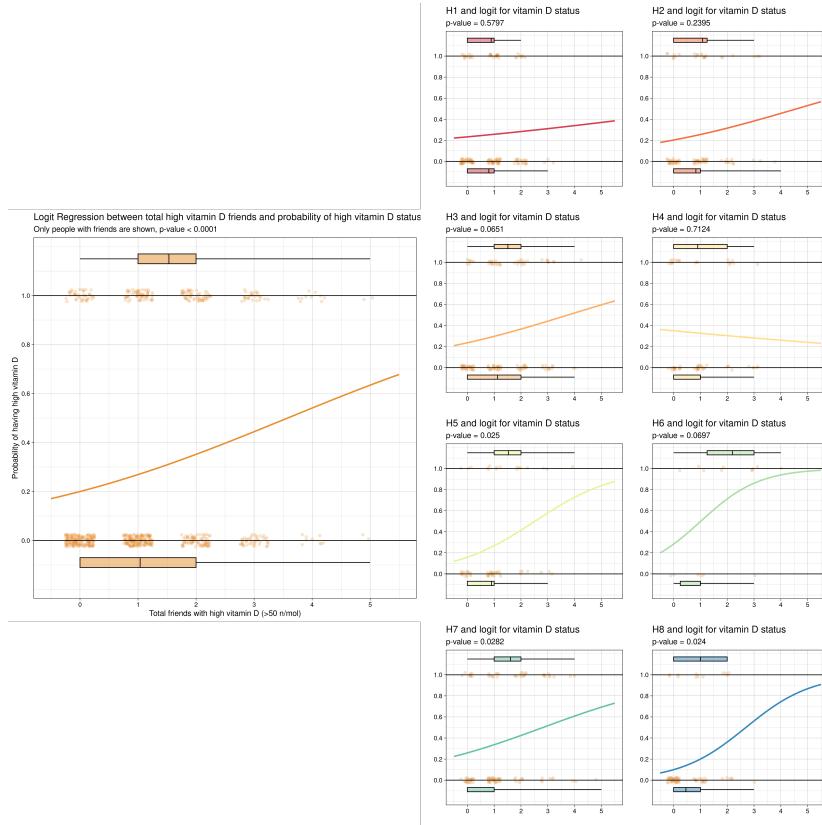


Figure 3: (Left) Logistic regression with respect to 25OHD levels between each person and high 25OHD friends. Each dot represents a person who does not go to the solarium, and who has at least one friend. The dots are laid out on the Y-axis in 1 if the subject has high vitamin D ($25\text{OHD} > 50 \text{ nmol/l}$), or 0 if low vitamin D ($25\text{OHD} < 50 \text{ nmol/l}$). In the X-axis we count how many high vitamin D friends this person has. The boxplots represent the difference in the number of high vitamin D friends between people who have low vitamin D (0) or high vitamin D (1). People with low vitamin D have an average of 1.04 friends with high vitamin D, while people with high vitamin D have an average of 1.53 friends with high vitamin D ($p\text{-value} < 0.0001$). (Right) The logistic regression analysis was performed for each high school independently. Each plot has a $p\text{-value}$ displayed under the title. Relevant levels are for H3, H5, H6, H7 and H8.

the middle-frequency consumption groups, while the high-frequency groups show approximately the same levels of 25OHD as people not eating any lean fish. We believe that this is the effect of confounding variables between lean fish frequency and 25OHD. We also tried to perform the same 1000 simulations for each of the 6 food groups in the overall network and we found no evidence that friends influence dietary habits among each other.

4 Discussion

4.1 Social influence of high school in vitamin D levels

Among the variables of sex, BMI, recreational drugs, health, solarium, and sports habits, high school has the stronger homophily when it comes to friendship (min 70% per school). We see that women tend to form friendships with other women if they share the same solarium habits. In previous works [58] the main motivation of Norwegian teenagers for going to the solarium was “*To get a tan*” with nearly 80% for girls and 60% for boys in both 2016 and 2017, followed by “*To prepare for holidays*” 32% in girls and 16.3% in boys in 2016, and near 23% for both in 2017. “*To make vitamin D*” appears as the third motivation with 20% for both in 2016 and 17.5% for both in 2017. It should be noticed that previous work showed that none of these reasons are justified due to solariums being too powerful, having the wrong UVA/UVB ratio, not increasing protection, and increasing hypervitaminosis D [55, 58, 68–70].

For non-solarium users, it appears that there is a general tendency in the population in which students who have a higher number of friends with high vitamin D levels tend to have higher vitamin D levels themselves. This tendency is also present in 5 of the 8 high schools. The simulations and the regression models also show a tendency to have similar vitamin D levels compared to friends. This accounts for possible bias towards total UVB. Furthermore, skin tone seems to be homogenously distributed across high schools, and the total number of dark-skin individuals is low (2% population) so this effect is not due to having an unbalanced skin type distribution. In our data, there seems to be no bias towards diet in any school other than lean fish consumption. Lean fish does not contain significant levels of vitamin D, and higher 25OHD levels in this consumption group appear to be due to confounding factors.

This seems to indicate that people follow the same healthy or unhealthy habits related to vitamin D absorption as their friends. This might include socio-economic factors such as being able to afford more traveling to more sunny areas, being more physically active and spending more time outdoors, having better diets that are rich in fatty fish, or better education related to vitamin D deficiency and supplementation.

4.2 Confounding of high school in vitamin D levels with other factors

H8 is the school with the lowest average 25OHD levels and is biased towards all the variables with lower-than-average 25OHD levels (males with -9.44 nmol/l 25OHD lower than women, not eating lean fish -5.71, alcohol consumption -3.49, smoking -6.7, snuff -6.91, and no physical activity -10.93). H1 is a similar school with no bias towards alcohol, and bias towards light physical activity; that despite having more traveling than any other school, is tied at second place with lower 25OHD levels. These two schools have 31% of the student population. In contrast, H6 and H7 are the complete opposite, with the highest vitamin D levels of any school, and bias towards no alcohol, no smoke, no snuff, and hard physical activity. Both H6 and H7 alone represent 21% of the population. Studying these variable levels for each high school independently, among students that do not go to the solarium, shows only 4 significant values after Bonferroni correction. H8 for sex, H3 for physical activity, and H1 and H3 for traveling.

Non-solarium women in H8 have +13.77 nmol/l average vitamin D levels than their male counterparts. These women do not show different habits than men in diet, BMI, physical activity, or recent sunbathing. This school has a bias towards the male student population, but neither sex displays different social dynamics that differ significantly from any other school. A network plot from the non-solarium population in H8 (supplementary figure 16), confirms the

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

logistic regression analysis and shows a pattern in which women with high vitamin D are friends with other women with high vitamin D, low vitamin D men are friends with low vitamin D friends, and high vitamin D men are friends with high vitamin D friends. Therefore, a proper follow-up of the students in this school can help to understand what helped these women to gain much better levels than their male counterparts, which is of particular interest given that H8 blood samples were taken in the middle of the polar night. A possible explanation is the use of oral contraceptives with estrogen [71], which are common in this school. However, in this school, we found no 25OHD significant difference between women not going to the solarium using Microgynon ($n = 8$) and those not using oral contraceptives ($n = 9$). The rest of the network plots, for all other high schools, also show clusters of non-solarium men or women with similar 25OHD levels in accordance to the simulation results (supplementary figures 9, 10, 11, 12, 13, 14, 15)

Physical activity (PA) in H3 is the only place where this variable is significant. PA is linked to outdoor sports and activities. However, Tromsø is cold and has a daily average temperature ranging from an average of 7.8°C in September to -3.3°C in February. Due to both clothing and lack of UVB, it is very unlikely that people in this school are exposed to enough skin to get enough UVB on a daily basis. However, it is likely that people doing physical activity have a healthier diet. Evidence suggests that obese people require higher vitamin D [72, 73], so physical activity PA is recommended to lower the vitamin D requirements regardless of whether it helps absorption or not.

Traveling to southern latitudes also helps with the UVB levels in H1 and H3. However, we do not have refined data regarding the amount of time of sun exposure, latitude, or possible skin lesions during that time to make a proper assessment. Further data is needed to evaluate the risk-benefit relationship of sun exposure with respect to gains of vitamin D in this particular population. Food fortification seems to have helped with lowering cancer mortality rates in Europe [59], and will be a safer public health approach than recommending increasing UVA+UVB exposure.

Altogether the results from the present analysis indicate that the only good predictor of vitamin D levels in this population is UVB radiation absorption due to associations with the date of the year, traveling, or solarium usage, and the significance of vitamin D with respect to these variables stems from the biases present in high school social dynamics.

4.3 Contradictory results and studies

We showed that getting closer to people with high vitamin D levels tends to increase vitamin D levels in the individual. This is of course due to similar interests to those lifestyles of peers, and we have shown extensively the effects it has in different high schools. However, due to the limited nutritional data and PA estimates, we cannot ascertain the effects that each possible healthier lifestyle may have on the network.

Vitamin D research presents contradictory evidence in many studies. This might be due to 25OHD tests not being properly standardized and poor experiment design [7, 74, 75] or not accounting for other supplements [76, 77]. Our study also adds to the confusion regarding the effective sources of vitamin D. For dietary habits in each school independently, or in the total population, there are no significant differences in vitamin D levels with neither fat fish consumption, and for daily intake of fish oil pills that typically contain 500-1.000 UI of D₃ per pill. It is very unlikely that people reporting taking these pills do not display higher levels. For example, an interventional study in Northern Ireland with university students showed that daily supplements of 600 IU increased by almost +40nmol/l with respect to

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

the control group [78]. Our dataset uses memory-based dietary assessment methods (M-BMs), described as "*uncritical faith in the validity and value of M-BMs has wasted substantial resources and constitutes the greatest impediment to scientific progress in obesity and nutrition research*" [79]. Similar conclusions can be found in self-reported data in the geriatric residential population [80]. Other studies suggest that the validity of Automated Self-Administered 24-hour recalls (ASA24s) is good enough [81] and multiple ASA24s and 4-d food records (4DFRs) provided the best estimates [82]. Our students report on food-frequency questionnaires (FFQs), which vary from the last 24 hours of food consumption prior to blood extraction, which was not even taken after 8 hours of fasting. It is important to enhance the scrutiny of dietary intake in future epidemiological studies. As such, we believe using FFQ is not a good option, and at the very least, having a nutritional professional oversee the ASA24s. Ideally, 4DFRs should be inputted into a food database (such as <https://www.matportalen.no/>) to retrieve final nutritional values and compare them with biomarkers and metabolites in blood.

Similarly, PA can be collected in a better way. The metabolic equivalent of task (MET) is a unit that measures how active a person is and roughly translates into 1 kcal/kg/hour, 1W/kg, or the energy required to sit down for an hour. In our questionnaires, PA is also self-reported and open to interpretation. Light activity includes walking and cycling, which vary from 2 to 6 METs. Medium activity includes recreational sports, such as light weightlifting (3.5 METs), tennis (5), basketball (8), football (10), jogging (11), or snow cleaning (4-8). Hard includes sports competitions, which again, vary too much for each type. Our categories have potential MET overlapping. METs have some limitations, and it is harder to calculate the exact value from person to person, but while it is also self-reported, it is a more objective measure of PA. This would also keep specific activities separated for better analysis. Alternatively, we could make use of accelerometers as has been reported in previous FF studies [83], but didn't have access to this data.

5 Conclusions

We saw that UVB radiation presents itself as the main influence of 25OHD levels in this population, even though low levels are present due to the Arctic geolocation. Among people not using solarium, those with greater numbers of friends with >50nmol/l tend to have higher levels of 25OHD themselves and vice versa, with an estimated probability of +7.25% per high vitamin D friend. FFQ seems to be a non-reliable tool for nutrition assessment and at the very least needs to be substituted with an ASA24s questionnaire; same with PA and METs.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

6 List of abbreviations

Acronym	Meaning
25OHD	25-hydroxycholecalciferol (Calcifediol)
4DFRs	4-d food records
ASA24s	Automated Self-Administered 24-h recalls
BMI	Body Mass Index
FF	Fit Futures
FFQs	food-frequency questionnaires
LC-MS/MS	High-pressure liquid chromatography-mass spectroscopy
M-BMs	memory-based dietary assessment methods
MET	metabolic equivalent of task
PA	Physical Activity
PTH	Parathyroid hormone
REK	Regional Committee of Medical and Health Research Ethics
UNN	University Hospital of North Norway
UVA	Ultraviolet A
UVB	Ultraviolet B
VDSP	Vitamin D Standardization Program

7 Declarations

7.1 Ethics approval and consent to participate

A declaration of consent was signed by each participant in FF1, participants younger than 16 years of age had to bring written consent from a parent or guardian. FF1 was approved by The Regional Committee of Medical and Health Research Ethics (REK) and the Norwegian Data Protection Authority. The present study was approved by REK North, reference 2011/1710 /REK Nord.

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

The Fit Futures data is not publicly available due to Norwegian privacy laws, but researchers can apply for access at <https://uit.no/research/fitfutures>

The analysis code is open source using the AGPLv3 license and it is available at our GIT repository (<https://github.com/uit-hdl/mimisbrunnr/>)

Our GIT repository also contains all results displayed here. Formats include Latex, CSV, and HTML for all tables, and PNG and PDF for all images. To avoid visual cluttering, some p-values displayed here are in GP Prism 5.04/d format (asterisks instead of numbers). The raw results contain all tables with both GP Prism format and the numerical format.

7.4 Competing interests

All authors state no conflict of interest.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

7.5 Funding

The “Population Studies in the North” (BiN) group at “UiT: The Arctic University of Norway” funded this study. The funders had no role in study design, data collection, analysis, interpretation, or decision to submit the manuscript for publication.

7.6 Authors contributions

RANC performed conceptualization, methodology, software, formal analysis, data curation, visualization, original draft, review, and editing of the paper. ASF and LAB performed review and editing, supervision, project administration, and funding acquisition. CSN participated in the conceptualization, review and editing, funding acquisition, and project administration. AMH collaborated with review editing and supervision.

7.7 Acknowledgements

We are grateful for the contribution of the participants in the Fit Futures 1 study. We also thank the Clinical Research Department at the University Hospital of North Norway, Tromsø, for data collection. We also want to thank the Department of Microbiology and Infection Control and the Department of Community Medicine at the Faculty of Health Sciences for their analysis of the different blood samples.

We would also like to thank the help and collaboration of Guri Grimnes of the Tromsø Endocrine Research Group, Department of Clinical Medicine at the UiT The Arctic University of Norway, and Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

8 Bibliography

References

- [1] P. F. Lazarsfeld, R. K. Merton *et al.*, "Friendship as a social process: A substantive and methodological analysis," *Freedom and control in modern society*, vol. 18, no. 1, pp. 18–66, 1954.
- [2] M. Selfhout, W. Burk, S. Branje, J. Denissen, M. van Aken, and W. Meeus, "Emerging late adolescent friendship networks and big five personality traits: A social network approach," *Journal of Personality*, vol. 78, no. 2, pp. 509–538, Apr. 2010. [Online]. Available: <https://doi.org/10.1111/j.1467-6494.2010.00625.x>
- [3] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, vol. 357, no. 4, pp. 370–379, Jul. 2007. [Online]. Available: <https://doi.org/10.1056/nejmsa066082>
- [4] K. Amrein, M. Scherkl, M. Hoffmann, S. Neuwersch-Sommeregger, M. Köstenberger, A. T. Berisha, G. Martucci, S. Pilz, and O. Malle, "Vitamin d deficiency 2.0: an update on the current status worldwide," *European Journal of Clinical Nutrition*, vol. 74, no. 11, pp. 1498–1513, Jan. 2020. [Online]. Available: <https://doi.org/10.1038/s41430-020-0558-y>
- [5] J. Öberg, R. Jorde, B. Almås, N. Emaus, and G. Grimnes, "Vitamin d deficiency and lifestyle risk factors in a norwegian adolescent population," *Scandinavian Journal of Public Health*, vol. 42, no. 7, pp. 593–602, Jul. 2014. [Online]. Available: <https://doi.org/10.1177/1403494814541593>
- [6] K. Amrein, M. Scherkl, M. Hoffmann, S. Neuwersch-Sommeregger, M. Köstenberger, A. T. Berisha, G. Martucci, S. Pilz, and O. Malle, "Vitamin d deficiency 2.0: an update on the current status worldwide," *European Journal of Clinical Nutrition*, vol. 74, no. 11, pp. 1498–1513, Jan. 2020. [Online]. Available: <https://doi.org/10.1038/s41430-020-0558-y>
- [7] K. D. Cashman, K. G. Dowling, Z. Škrabáková, M. Gonzalez-Gross, J. Valtueña, S. D. Henauw, L. Moreno, C. T. Damsgaard, K. F. Michaelsen, C. Mølgaard, R. Jorde, G. Grimnes, G. Moschonis, C. Mavrogianni, Y. Manios, M. Thamm, G. B. Mensink, M. Rabenberg, M. A. Busch, L. Cox, S. Meadows, G. Goldberg, A. Prentice, J. M. Dekker, G. Nijpels, S. Pilz, K. M. Swart, N. M. van Schoor, P. Lips, G. Eiriksdottir, V. Gudnason, M. F. Cotch, S. Koskinen, C. Lamberg-Allardt, R. A. Durazo-Arvizu, C. T. Sempos, and M. Kiely, "Vitamin d deficiency in europe: pandemic?" *The American Journal of Clinical Nutrition*, vol. 103, no. 4, pp. 1033–1044, Feb. 2016. [Online]. Available: <https://doi.org/10.3945/ajcn.115.120873>
- [8] Z. Jiang, R. Pu, N. Li, C. Chen, J. Li, W. Dai, Y. Wang, J. Hu, D. Zhu, Q. Yu, Y. Shi, and G. Yang, "High prevalence of vitamin d deficiency in asia: A systematic review and meta-analysis," *Critical Reviews in Food Science and Nutrition*, pp. 1–10, Nov. 2021. [Online]. Available: <https://doi.org/10.1080/10408398.2021.1990850>
- [9] R. M. Mogire, A. Mutua, W. Kimita, A. Kamau, P. Bejon, J. M. Pettifor, A. Adeyemo, T. N.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

- Williams, and S. H. Atkinson, "Prevalence of vitamin d deficiency in africa: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 8, no. 1, pp. e134–e142, Jan. 2020. [Online]. Available: [https://doi.org/10.1016/s2214-109x\(19\)30457-7](https://doi.org/10.1016/s2214-109x(19)30457-7)
- [10] M. Kull, R. Kallikorm, A. Tamm, and M. Lember, "Seasonal variance of 25-(OH) vitamin d in the general population of estonia, a northern european country," *BMC Public Health*, vol. 9, no. 1, Jan. 2009. [Online]. Available: <https://doi.org/10.1186/1471-2458-9-22>
- [11] R. Jorde, E. B. Mathiesen, S. Rogne, T. Wilsgaard, M. Kjærgaard, G. Grimnes, and H. Schirmer, "Vitamin d and cognitive function: The tromsø study," *Journal of the Neurological Sciences*, vol. 355, no. 1-2, pp. 155–161, Aug. 2015. [Online]. Available: <https://doi.org/10.1016/j.jns.2015.06.009>
- [12] M. del Carmen Navarro, P. Saavedra, E. Jódar, M. J. G. de Tejada, A. Mirallave, and M. Sosa, "Osteoporosis and metabolic syndrome according to socio-economic status, contribution of PTH, vitamin d and body weight: The canarian osteoporosis poverty study (COPS)," *Clinical Endocrinology*, vol. 78, no. 5, pp. 681–686, Apr. 2013. [Online]. Available: <https://doi.org/10.1111/cen.12051>
- [13] E. Shojaeeefar, N. Malih, and N. Rezaei, "The possible double-edged sword effects of vitamin d on COVID-19: A hypothesis," *Cell Biology International*, vol. 45, no. 1, pp. 54–57, Oct. 2020. [Online]. Available: <https://doi.org/10.1002/cbin.11469>
- [14] J. Stehlé, F. Charbonnier, T. Picard, C. Cattuto, and A. Barrat, "Gender homophily from spatial behavior in a primary school: A sociometric study," *Social Networks*, vol. 35, no. 4, pp. 604–613, Oct. 2013. [Online]. Available: <https://doi.org/10.1016/j.socnet.2013.08.003>
- [15] J. Moody, "Race, school integration, and friendship segregation in america," *American Journal of Sociology*, vol. 107, no. 3, pp. 679–716, Nov. 2001. [Online]. Available: <https://doi.org/10.1086/338954>
- [16] Z. Qian and D. T. Licher, "Social boundaries and marital assimilation: Interpreting trends in racial and ethnic intermarriage," *American Sociological Review*, vol. 72, no. 1, pp. 68–94, Feb. 2007. [Online]. Available: <https://doi.org/10.1177/000312240707200104>
- [17] J. E. Cheadle and P. Schwadel, "The 'friendship dynamics of religion,' or the 'religious dynamics of friendship'? a social network analysis of adolescents who attend small schools," *Social Science Research*, vol. 41, no. 5, pp. 1198–1212, Sep. 2012. [Online]. Available: <https://doi.org/10.1016/j.ssresearch.2012.03.014>
- [18] J. M. McPherson and L. Smith-Lovin, "Homophily in voluntary organizations: Status distance and the composition of face-to-face groups," *American Sociological Review*, vol. 52, no. 3, p. 370, Jun. 1987. [Online]. Available: <https://doi.org/10.2307/2095356>
- [19] "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, vol. 77, no. 4, pp.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

- 1003–1045, 2009. [Online]. Available: <https://doi.org/10.3982/ecta7528>
- [20] G. Kossinets and D. J. Watts, “Origins of homophily in an evolving social network,” *American Journal of Sociology*, vol. 115, no. 2, pp. 405–450, Sep. 2009. [Online]. Available: <https://doi.org/10.1086/599247>
- [21] M. McPherson, L. Smith-Lovin, and J. M. Cook, “Birds of a feather: Homophily in social networks,” *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, Aug. 2001. [Online]. Available: <https://doi.org/10.1146/annurev.soc.27.1.415>
- [22] J. A. Smith, M. McPherson, and L. Smith-Lovin, “Social distance in the united states.” *American Sociological Review*, vol. 79, no. 3, pp. 432–456, Apr. 2014. [Online]. Available: <https://doi.org/10.1177/0003122414531776>
- [23] F. Karimi, M. Génois, C. Wagner, P. Singer, and M. Strohmaier, “Homophily influences ranking of minorities in social networks,” *Scientific Reports*, vol. 8, no. 1, Jul. 2018. [Online]. Available: <https://doi.org/10.1038/s41598-018-29405-7>
- [24] E. Lee, F. Karimi, C. Wagner, H.-H. Jo, M. Strohmaier, and M. Galesic, “Homophily and minority-group size explain perception biases in social networks,” *Nature Human Behaviour*, vol. 3, no. 10, pp. 1078–1087, Aug. 2019. [Online]. Available: <https://doi.org/10.1038/s41562-019-0677-4>
- [25] C. Avin, H. Daltrophe, B. Keller, Z. Lotker, C. Mathieu, D. Peleg, and Y.-A. Pignolet, “Mixed preferential attachment model: Homophily and minorities in social networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 555, p. 124723, Oct. 2020. [Online]. Available: <https://doi.org/10.1016/j.physa.2020.124723>
- [26] A. Asikainen, G. Iñiguez, J. Ureña-Carrión, K. Kaski, and M. Kivelä, “Cumulative effects of triadic closure and homophily in social networks,” *Science Advances*, vol. 6, no. 19, May 2020. [Online]. Available: <https://doi.org/10.1126/sciadv.aax7310>
- [27] K. Holvik, H. E. Meyer, E. Haug, and L. Brunvand, “Prevalence and predictors of vitamin d deficiency in five immigrant groups living in oslo, norway: the oslo immigrant health study.” *European Journal of Clinical Nutrition*, vol. 59, no. 1, pp. 57–63, Jul. 2004. [Online]. Available: <https://doi.org/10.1038/sj.ejcn.1602033>
- [28] A. Björk, Å. Andersson, G. Johansson, K. Björkegren, A. Bardel, and P. Kristiansson, “Evaluation of sun holiday, diet habits, origin and other factors as determinants of vitamin d status in swedish primary health care patients: a cross-sectional study with regression analysis of ethnic swedish and immigrant women,” *BMC Family Practice*, vol. 14, no. 1, Sep. 2013. [Online]. Available: <https://doi.org/10.1186/1471-2296-14-129>
- [29] C. A. Martin, U. Gowda, and A. M. Renzaho, “The prevalence of vitamin d deficiency among dark-skinned populations according to their stage of migration and region of birth: A meta-analysis,” *Nutrition*, vol. 32, no. 1, pp. 21–32, Jan. 2016. [Online]. Available: <https://doi.org/10.1016/j.nut.2015.07.007>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

- [30] D. S. Nielsen, D. N. Rasmussen, and M. Sodemann, "Vitamin d status in patients attending a danish migrant health clinic: A clinical retrospective study," *Journal of Immigrant and Minority Health*, vol. 17, no. 2, pp. 474–481, May 2014. [Online]. Available: <https://doi.org/10.1007/s10903-014-0031-x>
- [31] "Serum 25-hydroxyvitamin d levels and intramuscular vitamin d3 supplementation among eritrean migrants recently arrived in switzerland," *Swiss Medical Weekly*, vol. 147, no. 5152, Dec. 2017. [Online]. Available: <https://doi.org/10.4414/smw.2017.14568>
- [32] M. Ceccarelli, E. Chiappini, R. Arancio, M. Zaffaroni, S. L. Placa, M. D'Andrea, C. de Waure, R. M. D. Rioli, and P. V. and, "Vitamin d deficiency in a population of migrant children: an italian retrospective cross-sectional multicentric study," *European Journal of Public Health*, vol. 30, no. 3, pp. 522–527, Oct. 2019. [Online]. Available: <https://doi.org/10.1093/eurpub/ckz182>
- [33] N. Smith, L. L. Sievert, S. Muttukrishna, K. Begum, L. Murphy, T. Sharmin, R. Gunu, O. Chowdhury, and G. R. Bentley, "Mismatch: a comparative study of vitamin d status in british-bangladeshi migrants," *Evolution, Medicine, and Public Health*, vol. 9, no. 1, pp. 164–173, Jan. 2021. [Online]. Available: <https://doi.org/10.1093/emph/eoab001>
- [34] "Fit Futures study," <https://uit.no/research/fitfutures>, accessed: 2023-05-30.
- [35] S. A. Wise, J. E. Camara, C. T. Sempos, P. Lukas, C. L. Goff, S. Peeters, C. Q. Burdette, F. Nalin, G. Hahm, R. A. Durazo-Arvizu, A. J. Kuszak, J. Merkel, and É. Cavalier, "Vitamin d standardization program (VDSP) intralaboratory study for the assessment of 25-hydroxyvitamin d assay variability and bias," *The Journal of Steroid Biochemistry and Molecular Biology*, vol. 212, p. 105917, Sep. 2021. [Online]. Available: <https://doi.org/10.1016/j.jsbmb.2021.105917>
- [36] K. D. Cashman, K. G. Dowling, Z. Škrabáková, M. Kiely, C. Lamberg-Allardt, R. A. Durazo-Arvizu, C. T. Sempos, S. Koskinen, A. Lundqvist, J. Sundvall, A. Linneberg, B. Thuesen, L. L. N. Husemoen, H. E. Meyer, K. Holvik, I. M. Grønborg, I. Tetens, and R. Andersen, "Standardizing serum 25-hydroxyvitamin d data from four nordic population samples using the vitamin d standardization program protocols: Shedding new light on vitamin d status in nordic individuals," *Scandinavian Journal of Clinical and Laboratory Investigation*, vol. 75, no. 7, pp. 549–561, Aug. 2015. [Online]. Available: <https://doi.org/10.3109/00365513.2015.1057898>
- [37] M. F. Holick, N. C. Binkley, H. A. Bischoff-Ferrari, C. M. Gordon, D. A. Hanley, R. P. Heaney, M. H. Murad, and C. M. Weaver, "Evaluation, treatment, and prevention of vitamin d deficiency: an endocrine society clinical practice guideline," *The Journal of Clinical Endocrinology and Metabolism*, vol. 96, no. 7, pp. 1911–1930, Jul. 2011. [Online]. Available: <https://doi.org/10.1210/jc.2011-0385>
- [38] M. F. Holick, "Vitamin d deficiency," *New England Journal of Medicine*, vol. 357, no. 3, pp. 266–281, Jul. 2007. [Online]. Available: <https://doi.org/10.1056/nejmra070553>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

- [39] C. J. Rosen, S. A. Abrams, J. F. Aloia, P. M. Brannon, S. K. Clinton, R. A. Durazo-Arvizu, J. C. Gallagher, R. L. Gallo, G. Jones, C. S. Kovacs, J. E. Manson, S. T. Mayne, A. C. Ross, S. A. Shapses, and C. L. Taylor, "IOM committee members respond to endocrine society vitamin d guideline," *The Journal of Clinical Endocrinology and Metabolism*, vol. 97, no. 4, pp. 1146–1152, Apr. 2012. [Online]. Available: <https://doi.org/10.1210/jc.2011-2218>
- [40] L. L. Brown, B. Cohen, D. Tabor, G. Zappalà, P. Maruvada, and P. M. Coates, "The vitamin d paradox in black americans: a systems-based approach to investigating clinical practice, research, and public health - expert panel meeting report," *BMC Proceedings*, vol. 12, no. S6, May 2018. [Online]. Available: <https://doi.org/10.1186/s12919-018-0102-4>
- [41] J. R. Chalcraft, L. M. Cardinal, P. J. Wechsler, B. W. Hollis, K. G. Gerow, B. M. Alexander, J. F. Keith, and D. E. Larson-Meyer, "Vitamin d synthesis following a single bout of sun exposure in older and younger men and women," *Nutrients*, vol. 12, no. 8, p. 2237, Jul. 2020. [Online]. Available: <https://doi.org/10.3390/nu12082237>
- [42] S. A. M. Yazdi, M. Abbasi, and S. M. M. Yazdi, "Epilepsy and vitamin d: a comprehensive review of current knowledge," *Reviews in the Neurosciences*, vol. 28, no. 2, pp. 185–201, Feb. 2017. [Online]. Available: <https://doi.org/10.1515/revneuro-2016-0044>
- [43] R. D. Bell, C. Y. C. Pak, J. Zerwekh, D. E. Barilla, and M. Vasko, "Effect of phenytoin on bone and vitamin d metabolism," *Annals of Neurology*, vol. 5, no. 4, pp. 374–378, Apr. 1979. [Online]. Available: <https://doi.org/10.1002/ana.410050411>
- [44] M. M. A. Faridi and A. Aggarwal, "Phenytoin induced vitamin d deficiency presenting as proximal muscle weakness," *Indian Pediatrics*, vol. 47, no. 7, pp. 624–625, Jul. 2010. [Online]. Available: <https://doi.org/10.1007/s13312-010-0121-3>
- [45] K. Robien, S. J. Oppeneer, J. A. Kelly, and J. M. Hamilton-Reeves, "Drug-vitamin d interactions," *Nutrition in Clinical Practice*, vol. 28, no. 2, pp. 194–208, Jan. 2013. [Online]. Available: <https://doi.org/10.1177/0884533612467824>
- [46] M. Crowe, L. Wollner, and R. A. Griffiths, "Hypercalcaemia following vitamin D and thiazide therapy in the elderly," *Practitioner*, vol. 228, no. 1389, pp. 312–313, Mar. 1984.
- [47] P. J. DRINKA and W. E. NOLTEN, "Hazards of treating osteoporosis and hypertension concurrently with calcium, vitamin d, and distal diuretics," *Journal of the American Geriatrics Society*, vol. 32, no. 5, pp. 405–407, May 1984. [Online]. Available: <https://doi.org/10.1111/j.1532-5415.1984.tb02050.x>
- [48] A. L. Skversky, J. Kumar, M. K. Abramowitz, F. J. Kaskel, and M. L. Melamed, "Association of glucocorticoid use and low 25-hydroxyvitamin d levels: Results from the national health and nutrition examination survey (NHANES): 2001–2006," *The Journal of Clinical Endocrinology and Metabolism*, vol. 96, no. 12, pp. 3838–3845, Dec. 2011. [Online]. Available: <https://doi.org/10.1210/jc.2011-1600>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

- [49] B. P. Lukert, "Glucocorticoid-induced osteoporosis: Pathogenesis and management," *Annals of Internal Medicine*, vol. 112, no. 5, p. 352, Mar. 1990. [Online]. Available: <https://doi.org/10.7326/0003-4819-112-5-352>
- [50] R. G. L. de Sévaux, A. J. Hoitsma, F. H. M. Corstens, and J. F. M. Wetzels, "Treatment with vitamin d and calcium reduces bone loss after renal transplantation: A randomized study," *Journal of the American Society of Nephrology*, vol. 13, no. 6, pp. 1608–1614, Jun. 2002. [Online]. Available: <https://doi.org/10.1097/01.asn.0000016082.70875.36>
- [51] L. M. Buckley, "Calcium and vitamin d3 supplementation prevents bone loss in the spine secondary to low-dose corticosteroids in patients with rheumatoid arthritis," *Annals of Internal Medicine*, vol. 125, no. 12, p. 961, Dec. 1996. [Online]. Available: <https://doi.org/10.7326/0003-4819-125-12-199612150-00004>
- [52] A. Gotfredsen, H. W. Hendel, and T. Andersen, "Influence of orlistat on bone turnover and body composition," *International Journal of Obesity*, vol. 25, no. 8, pp. 1154–1160, Aug. 2001. [Online]. Available: <https://doi.org/10.1038/sj.ijo.0801639>
- [53] W. P. James, A. Avenell, J. Broom, and J. Whitehead, "A one-year trial to assess the value of orlistat in the management of obesity," *Int. J. Obes. Relat. Metab. Disord.*, vol. 21 Suppl 3, pp. S24–30, Jun. 1997.
- [54] J. R. McDuffie, K. A. Calis, S. L. Booth, G. I. Uwaifo, and J. A. Yanovski, "Effects of orlistat on fat-soluble vitamins in obese adolescents," *Pharmacotherapy*, vol. 22, no. 7, pp. 814–822, Jul. 2002. [Online]. Available: <https://doi.org/10.1592/phco.22.11.814.33627>
- [55] J. L. Pérez-Castrillón, G. Vega, L. Abad, A. Sanz, J. Chaves, G. Hernandez, and A. Dueñas, "Effects of atorvastatin on vitamin d levels in patients with acute ischemic heart disease," *The American Journal of Cardiology*, vol. 99, no. 7, pp. 903–905, Apr. 2007. [Online]. Available: <https://doi.org/10.1016/j.amjcard.2006.11.036>
- [56] J. Schwartz, "Effects of vitamin d supplementation in atorvastatin-treated patients: A new drug interaction with an unexpected consequence," *Clinical Pharmacology and Therapeutics*, vol. 85, no. 2, pp. 198–203, Aug. 2008. [Online]. Available: <https://doi.org/10.1038/clpt.2008.165>
- [57] J. F. Aloia, M. Li-Ng, and S. Pollack, "Statins and vitamin d," *The American Journal of Cardiology*, vol. 100, no. 8, p. 1329, Oct. 2007. [Online]. Available: <https://doi.org/10.1016/j.amjcard.2007.05.024>
- [58] L. A. B. Trude E. Robsahm, Jo S. Stenehjem and M. B. Veierød, "Prevalence of indoor tanning among teenagers in norway before and after enforcement of ban for ages under 18 years," 2020, last accessed 30 September 2022. [Online]. Available: <https://www.medicaljournals.se/acta/content/html/10.2340/00015555-3482>
- [59] T. Niedermaier, T. Gredner, S. Kuznia, B. Schöttker, U. Mons, J. Lakerveld, W. Ahrens, and H. B. and, "Vitamin d food fortification in european countries: the underused potential to prevent cancer deaths," *European Journal of Epidemiology*, vol. 37, no. 4, pp. 309–320, Apr. 2022. [Online]. Available: <https://doi.org/10.1007/s10654-022-00638-0>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

<https://doi.org/10.1007/s10654-022-00867-4>

- [60] H. og omsorgsdepartementet, “Forskrift om næringsmidler til særskilte grupper,” 2014, last accessed 30 September 2022. [Online]. Available: <https://lovdata.no/dokument/SF/forskrift/2014-01-10-21?q=s%C3%A6rskilte%20grupper>
- [61] S. K. Beldo, N. A. Aars, T. Christoffersen, A.-S. Furberg, P. A. Halvorsen, B. H. Hansen, A. Horsch, E. H. Saglev, S. Syed, and B. Morseth, “Criterion validity of the saltin-grimby physical activity level scale in adolescents. the fit futures study,” *PLOS ONE*, vol. 17, no. 9, p. e0273480, Sep. 2022. [Online]. Available: <https://doi.org/10.1371/journal.pone.0273480>
- [62] B. Babar and T. Boström, “Estimating solar irradiation in the arctic,” *Renewable Energy and Environmental Sustainability*, vol. 1, p. 34, 2016. [Online]. Available: <https://doi.org/10.1051/rees/2016048>
- [63] Solargis, “Solar resource map, direct normal irradiation,” 2020, last accessed 30 September 2022. [Online]. Available: https://en.wikipedia.org/wiki/File:World_DNI_Solar-resource-map_GlobalSolarAtlas_World-Bank-Esmap-Solargis.png
- [64] O. Engelsen, M. Brustad, L. Aksnes, and E. Lund, “Daily duration of vitamin d synthesis in human skin with relation to latitude, total ozone, altitude, ground cover, aerosols and cloud thickness,” *Photochemistry and Photobiology*, vol. 81, no. 6, p. 1287, 2005. [Online]. Available: <https://doi.org/10.1562/2004-11-19-mn-375>
- [65] D. B. Stensen, R. A. N. Cañadas, L. Småbrekke, K. Olsen, C. S. Nielsen, K. Svendsen, A. M. Hanssen, J. U. Ericson, G. S. Simonsen, L. A. Bongo, and A.-S. Furberg, “Social network analysis of *staphylococcus aureus* carriage in a general youth population,” *International Journal of Infectious Diseases*, vol. 123, pp. 200–209, Oct. 2022. [Online]. Available: <https://doi.org/10.1016/j.ijid.2022.08.018>
- [66] N. Petrenya, C. Lamberg-Allardt, M. Melhus, A. R. Broderstad, and M. Brustad, “Vitamin d status in a multi-ethnic population of northern norway: the SAMINOR 2 clinical survey,” *Public Health Nutrition*, vol. 23, no. 7, pp. 1186–1200, Feb. 2019. [Online]. Available: <https://doi.org/10.1017/s1368980018003816>
- [67] T. Durá-Travé, “Deficiencia de vitamina d en escolares y adolescentes con un estado,” *NUTRICION HOSPITALARIA*, no. 3, p. 1061–1066, Sep. 2015. [Online]. Available: <https://doi.org/10.3305/nh.2015.32.3.9316>
- [68] L. T. N. Nilsen, T. N. Aalerud, M. Hannevik, and M. B. Veierød, “UVB and UVA irradiances from indoor tanning devices,” *Photochemical and Photobiological Sciences*, vol. 10, no. 7, pp. 1129–1136, Jul. 2011. [Online]. Available: <https://doi.org/10.1039/c1pp05029j>
- [69] P. Singh and N. Trivedi, “Tanning beds and hypervitaminosis d: A case report,” *Annals of Internal Medicine*, vol. 160, no. 11, p. 810, Jun. 2014. [Online]. Available: <https://doi.org/10.7326/l14-5011-8>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

- [70] M. R. Laurent, E. Gielen, S. Pauwels, D. Vanderschueren, and R. Bouillon, "Hypervitaminosis d associated with tanning bed use: A case report," *Annals of Internal Medicine*, vol. 166, no. 2, p. 155, Jan. 2017. [Online]. Available: <https://doi.org/10.7326/l16-0138>
- [71] J. Öberg, R. Jorde, Y. Figenschau, P. M. Thorsby, S. R. Dahl, A. Winther, and G. Grimnes, "100 YEARS OF VITAMIN d: Combined hormonal contraceptives and vitamin d metabolism in adolescent girls," *Endocrine Connections*, vol. 11, no. 3, Mar. 2022. [Online]. Available: <https://doi.org/10.1530/ec-21-0395>
- [72] J. P. Ekwaru, J. D. Zwicker, M. F. Holick, E. Giovannucci, and P. J. Veugelers, "The importance of body weight for the dose response relationship of oral vitamin d supplementation and serum 25-hydroxyvitamin d in healthy volunteers," *PLoS ONE*, vol. 9, no. 11, p. e111265, Nov. 2014. [Online]. Available: <https://doi.org/10.1371/journal.pone.0111265>
- [73] A. Drincic, E. Fuller, R. P. Heaney, and L. A. G. Armas, "25-hydroxyvitamin d response to graded vitamin d3 supplementation among obese adults," *The Journal of Clinical Endocrinology and Metabolism*, vol. 98, no. 12, pp. 4845–4851, Dec. 2013. [Online]. Available: <https://doi.org/10.1210/jc.2012-4103>
- [74] C. T. Sempos, A. C. Heijboer, D. D. Bikle, J. Bollerslev, R. Bouillon, P. M. Brannon, H. F. DeLuca, G. Jones, C. F. Munns, J. P. Bilezikian, A. Giustina, and N. Binkley, "Vitamin d assays and the definition of hypovitaminosis d: results from the first international conference on controversies in vitamin d," *British Journal of Clinical Pharmacology*, vol. 84, no. 10, pp. 2194–2207, Jul. 2018. [Online]. Available: <https://doi.org/10.1111/bcp.13652>
- [75] C. T. Sempos, G. D. Carter, and N. C. Binkley, "25-hydroxyvitamin d assays," in *Vitamin D*. Elsevier, 2018, pp. 939–957. [Online]. Available: <https://doi.org/10.1016/b978-0-12-809965-0.00052-5>
- [76] N. Li, G. Zhao, W. Wu, M. Zhang, W. Liu, Q. Chen, and X. Wang, "The efficacy and safety of vitamin c for iron supplementation in adult patients with iron deficiency anemia," *JAMA Network Open*, vol. 3, no. 11, p. e2023644, Nov. 2020. [Online]. Available: <https://doi.org/10.1001/jamanetworkopen.2020.23644>
- [77] S. R. Lynch and J. D. Cook, "INTERACTION OF VITAMIN c AND IRON," *Annals of the New York Academy of Sciences*, vol. 355, no. 1 Micronutrient, pp. 32–44, Dec. 1980. [Online]. Available: <https://doi.org/10.1111/j.1749-6632.1980.tb21325.x>
- [78] M. S. Barnes, P. J. Robson, M. P. Bonham, J. J. Strain, and J. M. W. Wallace, "Effect of vitamin d supplementation on vitamin d status and bone turnover markers in young adults," *European Journal of Clinical Nutrition*, vol. 60, no. 6, pp. 727–733, Jan. 2006. [Online]. Available: <https://doi.org/10.1038/sj.ejcn.1602374>
- [79] E. Archer, G. Pavella, and C. J. Lavie, "The inadmissibility of what we eat in america and NHANES dietary data in nutrition and obesity research and the scientific formulation of national dietary guidelines," *Mayo Clinic Proceedings*, vol. 90, no. 7, pp. 911–926, Jul. 2015. [Online]. Available: <https://doi.org/10.1016/j.mayocp.2015.04.009>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

-
- [80] S. Whybrow, R. J. Stubbs, A. M. Johnstone, L. M. O'Reilly, Z. Fuller, M. B. E. Livingstone, and G. W. Horgan, "Plausible self-reported dietary intakes in a residential facility are not necessarily reliable," *European Journal of Clinical Nutrition*, vol. 70, no. 1, pp. 130–135, Jul. 2015. [Online]. Available: <https://doi.org/10.1038/ejcn.2015.124>
 - [81] J. R. Hébert, T. G. Hurley, S. E. Steck, D. R. Miller, F. K. Tabung, K. E. Peterson, L. H. Kushi, and E. A. Frongillo, "Considering the value of dietary assessment data in informing nutrition-related health policy," *Advances in Nutrition*, vol. 5, no. 4, pp. 447–455, Jul. 2014. [Online]. Available: <https://doi.org/10.3945/an.114.006189>
 - [82] Y. Park, K. W. Dodd, V. Kipnis, F. E. Thompson, N. Potischman, D. A. Schoeller, D. J. Baer, D. Midthune, R. P. Troiano, H. Bowles, and A. F. Subar, "Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers," *The American Journal of Clinical Nutrition*, vol. 107, no. 1, pp. 80–93, Jan. 2018. [Online]. Available: <https://doi.org/10.1093/ajcn/nqx002>
 - [83] S. K. Beldo, B. Morseth, T. Christoffersen, P. A. Halvorsen, B. H. Hansen, A.-S. Furberg, U. Ekelund, and A. Horsch, "Prevalence of accelerometer-measured physical activity in adolescents in fit futures – part of the tromsø study," *BMC Public Health*, vol. 20, no. 1, Jul. 2020. [Online]. Available: <https://doi.org/10.1186/s12889-020-09171-w>

9 Supplementary materials

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

10 High-schools information

10.1 Map overview

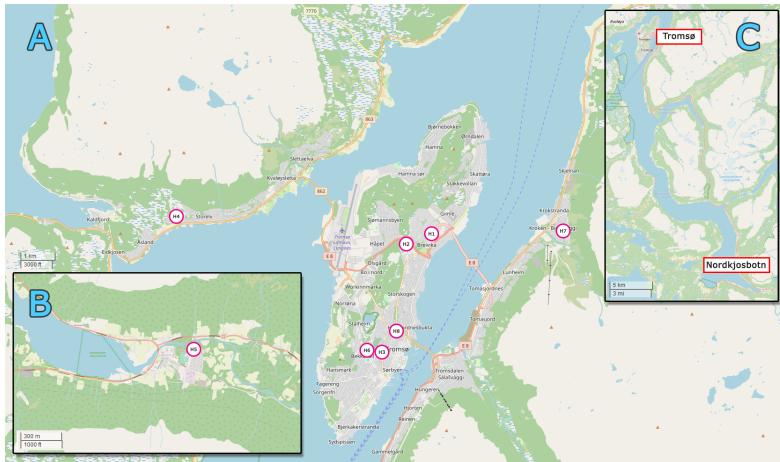


Figure 4: Overview map with the different high schools which were studied. Each high school location is highlighted with a circle with the ID of the school in the center. Area "A" is the view of the Tromsø area, schools H1 to H4 and H6 to H8 are located here. Area "B" is the view of Nordkjosbotn where H5 is located. Area "C" shows the distance and scale between A and B.

10.2 High-schools summary table

Table 3: Information with the school names, study program, the total number of students in FF1, astronomical season, and whether a significant part of the student's samples was taken during the polar night.

ID	Name	Studies program	FF1 Students	Extraction	Polar night
H1	Breivika videregående skole	Vocational	207	Autumn	No
H2	Brevang videregående skole	Vocational and General	142	Autumn	Yes
H3	Kongsbakken videregående skole	Vocational and General	168	Winter	No
H4	Kvaløya videregående skole	Vocational and General	98	Spring	No
H5	Nordkjosbotn videregående skole	Vocational and General	85	Spring	No
H6	Norges Toppidrettsgymnas Tromsø	Sports	26	Spring	No
H7	Tromsdalen videregående skole	Sports and General	192	Winter	Yes
H8	Tromsø maritime skole	Vocational	120	Winter	Yes

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

10.3 High-schools dates of blood extraction

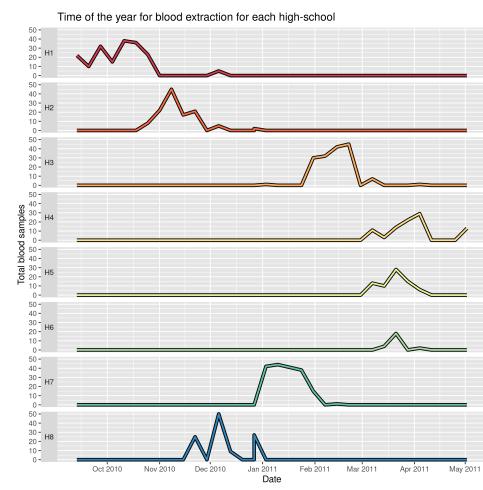


Figure 5: Total blood extractions per high school across time of the year for each high school.

11 Networks information

11.1 All graphs

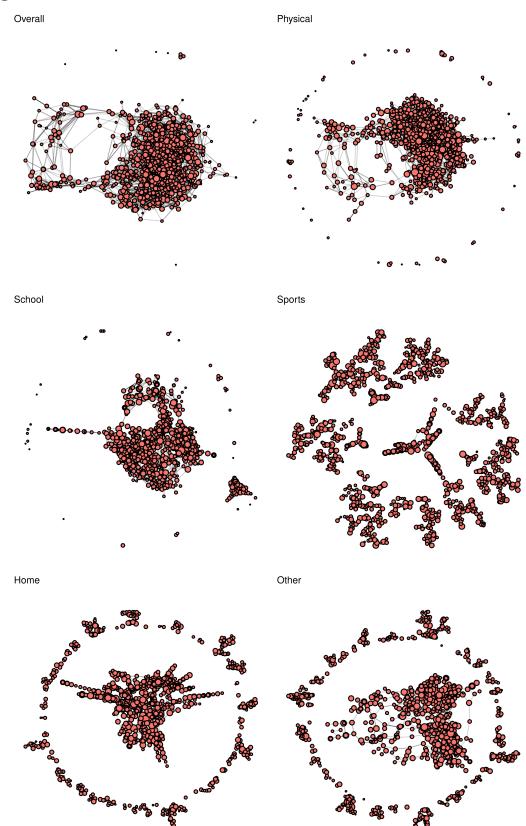


Figure 6: Overview of all friendship networks. The figure is reproduced from "Social network analysis of *Staphylococcus aureus* carriage in a general youth population" with permissions. DOI: <https://doi.org/10.1016/j.ijid.2022.08.018>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

11.2 Network relevance overview

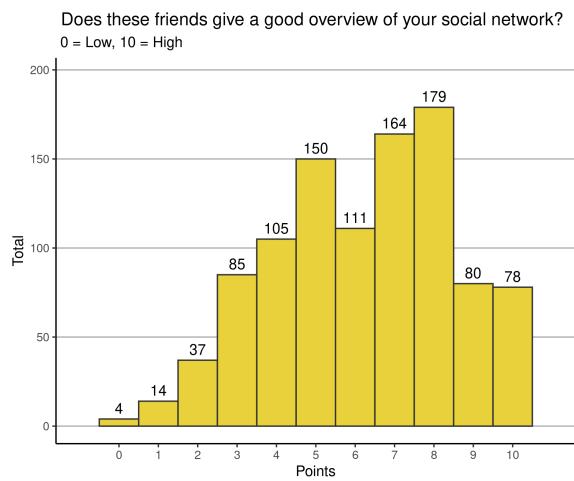


Figure 7: Histogram with how good is the self-reported network (0 - 10, x-axis) by each student (y-axis). The figure is reproduced from "Social network analysis of *Staphylococcus aureus* carriage in a general youth population" with permissions. DOI: <https://doi.org/10.1016/j.ijid.2022.08.018>

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

11.3 Network relevance by high school

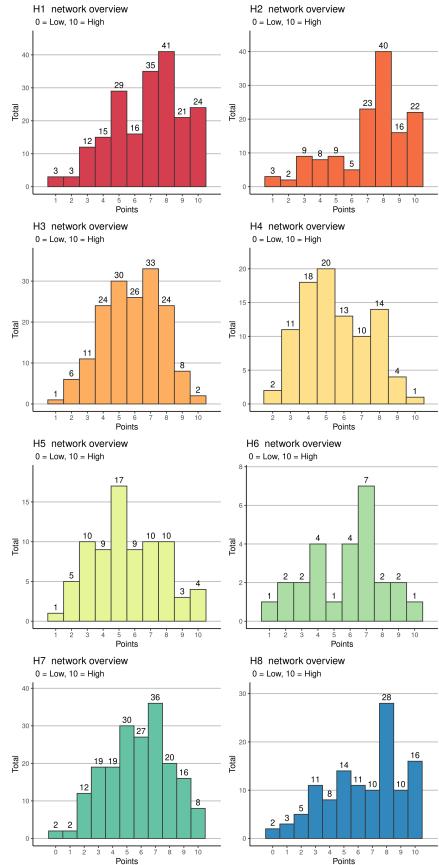


Figure 8: For each of the high schools (H1 to H8), histogram with how good is the self-reported network (0 - 10, x-axis) by each student (y-axis)

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

11.4 Network relevance table

Table 4: Highschools and friendship overview. The first row represents all high schools combined. H1 to H8 rows represent each high school separately. Homophily represents how many students of this school form friendships with a student of the same school. Frequency is the relative frequency of students in each high school. Significance is the p-value of a two-sided binomial test of relationships within the same high school, total relationships, and relative frequency of students in each high school. Average and Median are the values of how good is the self-reported network (0 - 10) by each student in each high school.

High School	Homophily	Frequency	Significance	Average	Median
All	0.87			6.22	6
H1	0.76	0.2	****	6.77	7
H2	0.78	0.14	****	7.2	8
H3	0.76	0.16	****	5.84	6
H4	0.77	0.09	****	5.54	5
H5	0.96	0.08	****	5.55	5
H6	0.76	0.03	****	5.73	6
H7	0.79	0.18	****	5.8	6
H8	0.72	0.12	****	6.42	7

11.5 Non-solarium networks by high school

Non solarium relationships in Highschool 1
Fruchterman - Reingold layout with node size proportional to 25OHD levels

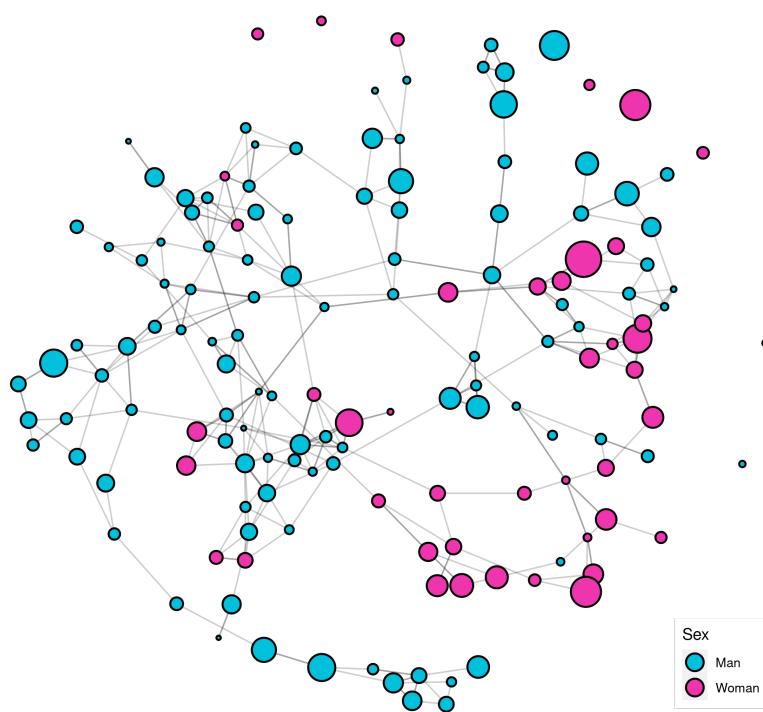


Figure 9: Relationships in H1 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 150 students and 255 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 2
Fruchterman - Reingold layout with node size proportional to 25OHD levels

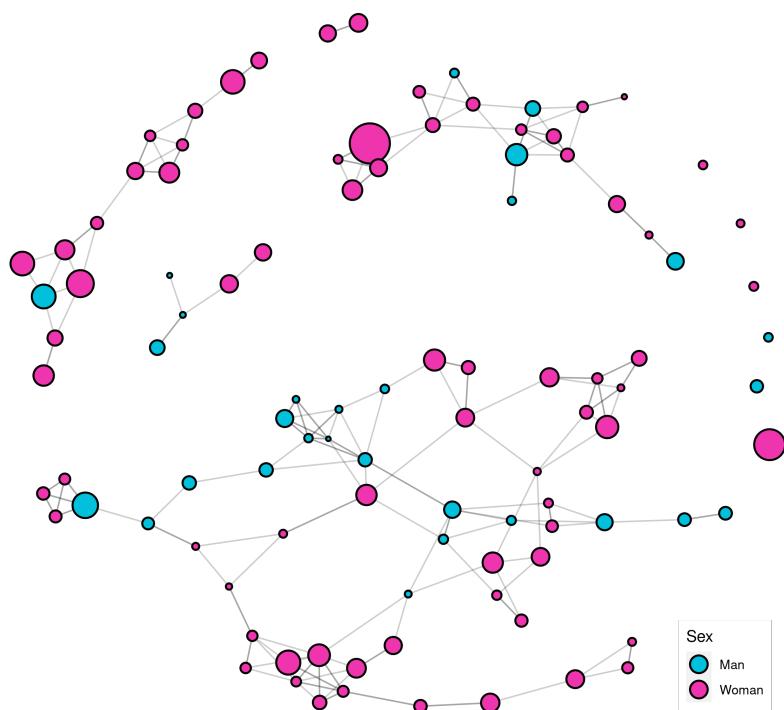


Figure 10: Relationships in H2 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 101 students and 157 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 3
Fruchterman - Reingold layout with node size proportional to 25OHD levels

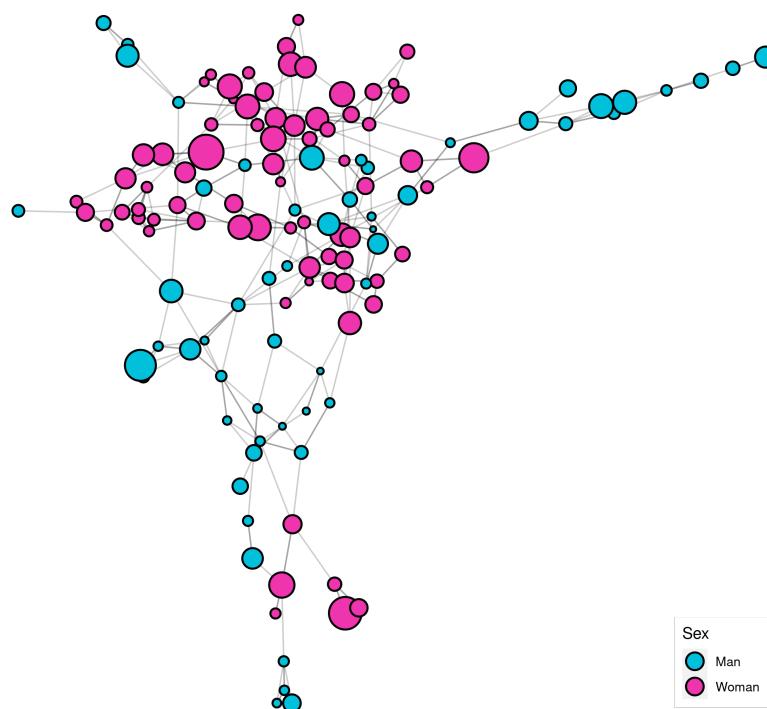


Figure 11: Relationships in H3 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 129 students and 247 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 4
Fruchterman - Reingold layout with node size proportional to 25OHD levels

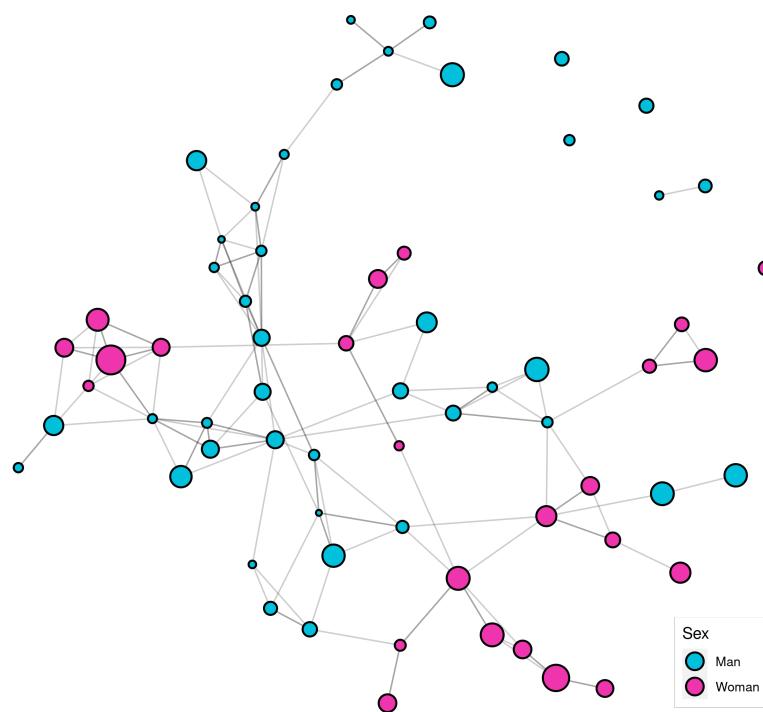


Figure 12: Relationships in H4 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 65 students and 110 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 5
Fruchterman - Reingold layout with node size proportional to 25OHD levels

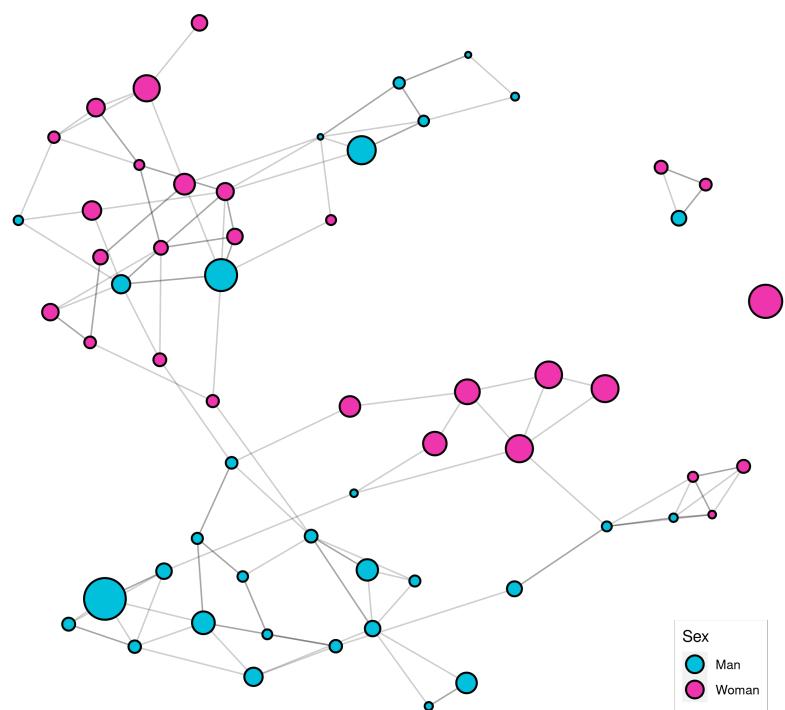


Figure 13: Relationships in H5 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 59 students and 102 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 6
Fruchterman - Reingold layout with node size proportional to 25OHD levels



Figure 14: Relationships in H7 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 16 students and 19 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 7
Fruchterman - Reingold layout with node size proportional to 25OHD levels



Figure 15: Relationships in H7 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 113 students and 180 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Non solarium relationships in Highschool 8
Fruchterman - Reingold layout with node size proportional to 25OHD levels

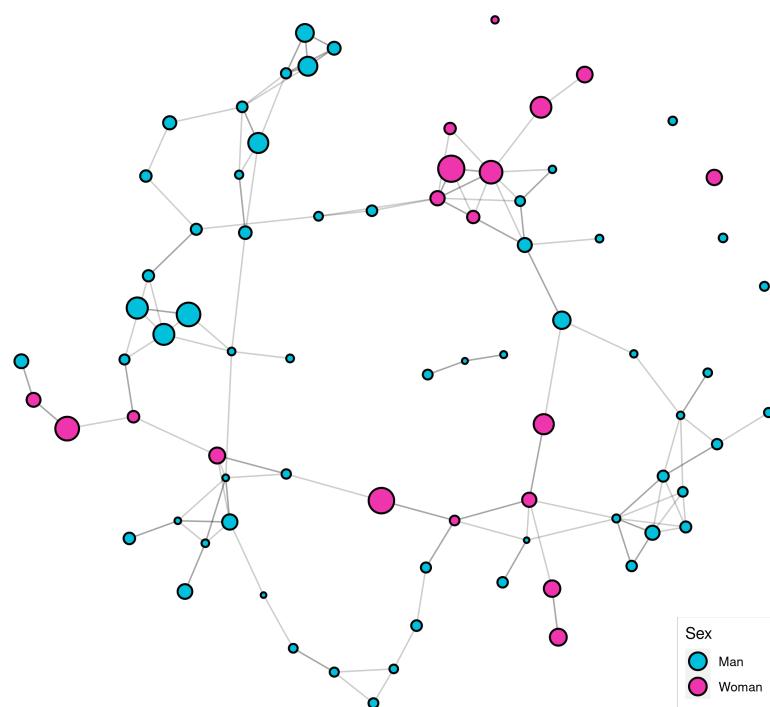


Figure 16: Relationships in H8 for non-solarium users highlighted by sex. Node size is proportional to the 25OHD level. Layout of the nodes using Fruchterman - Reingold. A total of 78 students and 112 undirected relationships are displayed.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

12 Other social influences

12.1 Non-solarium simulations by high school

Table 5: Simulations performed for each high school independently for non-solarium users.

Highschool	Relationships		Simulations statistics							P-value
	Total	Equal	MIN	Q1	Median	Average	Q3	MAX	SD	
H1	341	223	151	169	175	177.16	186	211	12.45	***
H2	221	137	92	109	116.5	116.2	123	154	10.26	*
H3	348	212	156	170	181	180.99	189	227	13.88	*
H4	153	101	57	74	81	81.24	88	104	9.75	*
H5	135	88	47	65	71	70.22	76	95	8.35	*
H6	30	20	9	13	16	15.96	19	23	3.42	ns
H7	251	141	102	123	132	131.05	138	159	10.52	ns
H8	152	114	63	72	78	78.46	85	98	8.28	****

12.2 25OHD and Friends' average 25OHD by high school

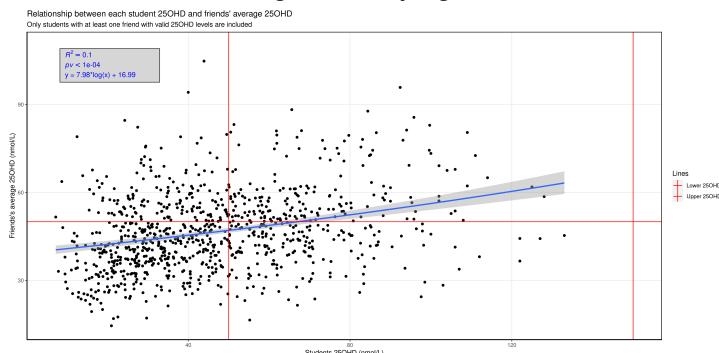


Figure 17: Relationship between each student 25OHD levels (X-axis) in comparison with the students' friends average 25OHD levels (Y-axis). Only students with a valid 25OHD value with at least one friend who also has a valid 25OHD value are included (n=930)

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

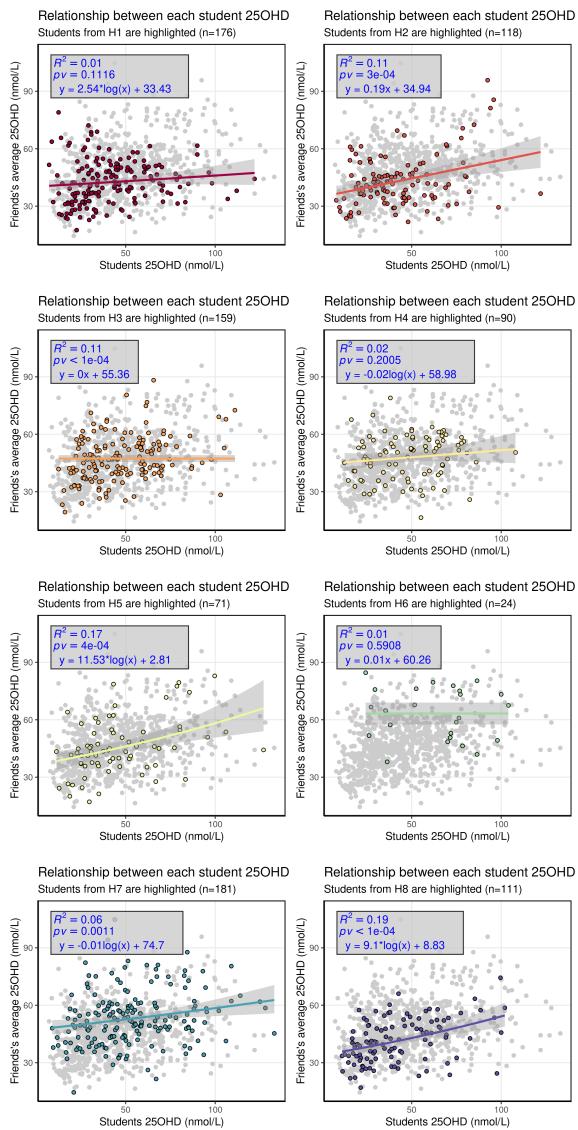


Figure 18: Relationship between each student 25OHD levels in comparison with the students' friends average 25OHD levels, for each high school. Students from other high schools are grayed out.

13 Ethnicity

Table 6: Absolute frequency for each ethnicity sorted by frequency

Modality	Count	Relative	Cumulative
Norwegian	900	0.867	0.867
Norwegian-Sami	27	0.026	0.893
Didn't Answer	20	0.019	0.912
Norwegian-Kven	10	0.01	0.922
Sami	7	0.007	0.929
Norwegian-Swedish	7	0.007	0.935
Russian	4	0.004	0.939
Norwegian-Sami-Kven	4	0.004	0.943
African	3	0.003	0.946
Afghan	3	0.003	0.949
Swedish	3	0.003	0.952
Norwegian-Somalian	3	0.003	0.955
Norwegian-Spanish	3	0.003	0.958
Norwegian-Danish	3	0.003	0.961
Other	2	0.002	0.962
Eritrean	2	0.002	0.964
Thai	2	0.002	0.966
German	2	0.002	0.968
Polish	2	0.002	0.97
Kven	2	0.002	0.972
Norwegian-Russian	2	0.002	0.974
Canadian	1	0.001	0.975
Tamil	1	0.001	0.976
Palestinian	1	0.001	0.977
Italian	1	0.001	0.978
Bulgarian	1	0.001	0.979
Belgian	1	0.001	0.98
Dutch	1	0.001	0.981
Norwegian-Other	1	0.001	0.982
Norwegian-Colombian	1	0.001	0.983
Norwegian-Brasilian	1	0.001	0.984
Norwegian-Canadian	1	0.001	0.985
Norwegian-Ghanaian	1	0.001	0.986
Norwegian-Gambian	1	0.001	0.987
Norwegian-African	1	0.001	0.987
Norwegian-Philipine	1	0.001	0.988
Norwegian-Tamil	1	0.001	0.989
Norwegian-Thai	1	0.001	0.99
Norwegian-Chinese	1	0.001	0.991
Norwegian-Turquish	1	0.001	0.992
Norwegian-Portuguese	1	0.001	0.993
Norwegian-Italian	1	0.001	0.994
Norwegian-Belgian	1	0.001	0.995
Norwegian-German	1	0.001	0.996
Norwegian-Finnish	1	0.001	0.997
Norwegian-Swedish-Dutch	1	0.001	0.998
Norwegian-Sami-Icelandic	1	0.001	0.999
Norwegian-Sami-Swedish	1	0.001	1

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

Table 7: *Chi-square test skin group and high-school*

0.5	Fair	Dark	Total	Freq
H1	190/192	7/4	197	0.19
H2	135/136	5/3	140	0.14
H3	164/163	3/3	167	0.16
H4	95/94	2/2	97	0.1
H5	81/79	0/1	81	0.08
H6	26/25	0/0	26	0.03
H7	188/186	3/4	191	0.19
H8	117/116	2/2	119	0.12
Total	996	22	1018	
Freq	0.98	0.02		1

14 Non solarium distributions by high schools

14.1 Sex

Table 8: *Xi² table with respect to sex and high school*

7e-13	Man	Woman	Total	Freq
H1	107/85 ++	43/64 —	150	0.21
H2	29/57 —	72/43 +++++	101	0.14
H3	56/73 —	73/55 ++	129	0.18
H4	41/37	24/27	65	0.09
H5	31/33	28/25	59	0.08
H6	12/9 4/6		16	0.02
H7	71/64	42/48	113	0.16
H8	59/44 ++	19/33 —	78	0.11
Total	406	305	711	
Freq	0.57	0.43		1

14.2 BMI

Table 9: *Xi² table with respect to BMI and high school*

9e-04	Underweight	Healthy	Overweight	Obese	Total	Freq
H1	16/16	91/100	26/20	16/11	149	0.21
H2	14/11	57/67	16/13	13/7 +	100	0.14
H3	13/14	96/86	15/17	5/10 -	129	0.18
H4	12/7 +	39/43	9/8	5/5	65	0.09
H5	4/6	40/39	6/8	9/4 ++	59	0.08
H6	0/1	16/10	0/2	0/1	16	0.02
H7	15/12	88/75	8/15	— 2/8	113	0.16
H8	6/8	49/52	17/10	++ 6/6	78	0.11
Total	80	476	97	56	709	
Freq	0.11	0.67	0.14	0.08		1

14.3 Alcohol

Table 10: *Xi² table with respect to alcohol and high school for non-solarium users*

0.001	Never	Once per month or less	Twice or more per month	Total	Freq
H1	44/45	60/56	37/38	141	0.2
H2	36/32	38/40	26/27	100	0.14
H3	40/41	61/52	28/35	129	0.19
H4	25/20	23/25	16/17	64	0.09
H5	17/18	17/22	23/15 +	57	0.08
H6	11/5 ++	5/6	0/4 -	16	0.02
H7	38/36	48/45	27/30	113	0.16
H8	14/24 —	29/31	34/21 +++	77	0.11
Total	225	281	191	697	
Freq	0.32	0.4	0.27		1

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

14.4 Smoke

Table 11: *Xi2 table with respect to smoke and high school for non-solarium users*

8e-06	Never	Sometimes	Daily	Total	Freq
H1	102/112	31/22	++ 8/6	141	0.2
H2	76/79	15/15	8/4 +	99	0.14
H3	111/103	18/20	0/5 —	129	0.19
H4	59/51	4/10	- 1/2	64	0.09
H5	37/45	15/8	+ 5/2	57	0.08
H6	16/12	0/2	0/0	16	0.02
H7	102/90	10/17	- 1/5	113	0.16
H8	54/62	16/12	8/3 ++	78	0.11
Total	557	109	31	697	
Freq	0.8	0.16	0.04		1

14.5 Snuff

Table 12: *Xi2 table with respect to snuff and high school for non-solarium users*

3e-07	Never	Sometimes	Daily	Total	Freq
H1	85/96	15/15	42/30 ++	142	0.2
H2	63/67	12/10	24/21	99	0.14
H3	99/87	15/13	15/27 —	129	0.19
H4	47/43	3/6	14/13	64	0.09
H5	38/38	6/6	13/12	57	0.08
H6	15/10	1/1	0/3 -	16	0.02
H7	92/76 +	12/11	9/24 —	113	0.16
H8	34/52 —	10/8	33/16 ++++	77	0.11
Total	473	74	150	697	
Freq	0.68	0.11	0.22		1

14.6 Sports

Table 13: *Xi² table with respect to sport and high school for non-solarium users*

9e-34	None	Light	Medium	Hard	Total	Freq
H1	44/36	63/47	++	30/34	5/23	—
H2	38/25	++	40/33	14/24	—	8/16
H3	24/33	-	44/42	42/31	++	19/21
H4	16/16		14/21	21/15		13/10
H5	15/14		19/18	21/13	+	2/9
H6	0/4	-	0/5	—	0/3	—
H7	12/29	—	22/37	—	33/27	46/18
H8	32/20	+++	30/25	10/19	—	6/12
Total	181		232	171	115	699
Freq	0.26		0.33	0.24	0.16	1

14.7 Sunbathing

Table 14: *Xi-square test for sunbathing and high school for non-solarium users.*

4e-07	Yes	No	Total	Freq
H1	25/10	++++	124/138	149
H2	5/7		96/93	101
H3	5/9		124/119	129
H4	1/4	-	64/60	65
H5	2/4		57/54	59
H6	5/1	+++	11/14	16
H7	6/8		107/104	113
H8	3/5		74/71	77
Total	52		657	709
Freq	0.07		0.93	1

15 Other figures

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

15.1 PTH predictors

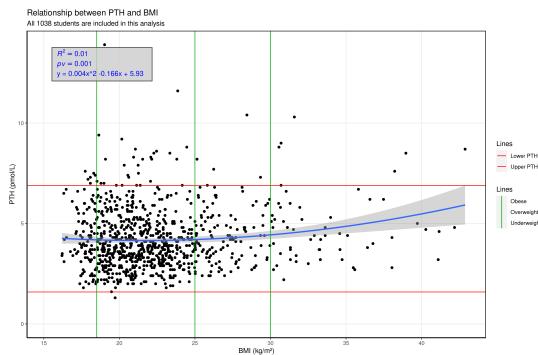


Figure 19: Relationship between PTH and BMI. All the students are included in this analysis. Vertical lines represent the threshold between BMI categories. Horizontal lines represent the healthy boundaries of PTH levels.

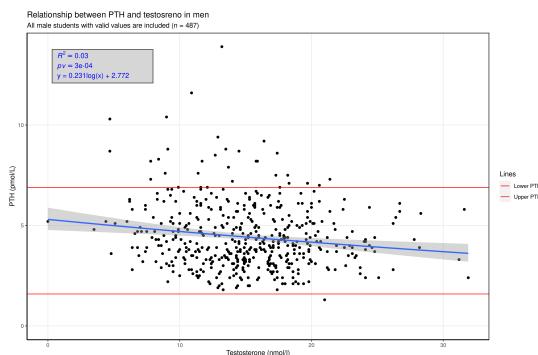


Figure 20: Relationship between PTH and testosterone in men. All the students are included in this analysis. Horizontal lines represent the healthy boundaries of PTH levels.

The Social Sunshine of the Arctic Youth: Exploring friendship's influence on Vitamin D levels.

PREPRINT

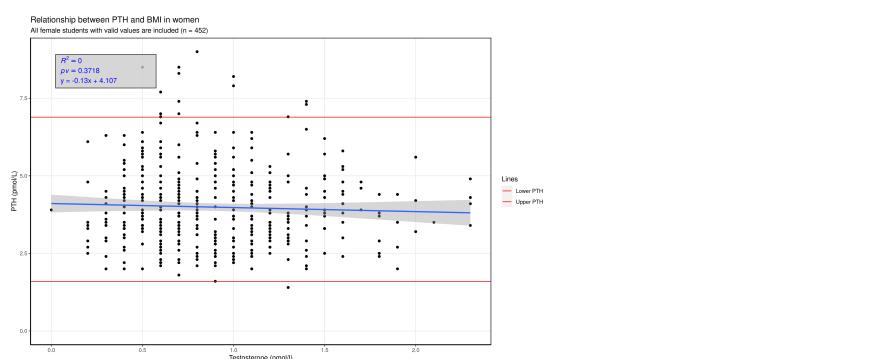


Figure 21: Relationship between PTH and testosterone in women. All the students are included in this analysis. Horizontal lines represent the healthy boundaries of PTH levels.

A.3 Paper C



An introduction to network analysis for studies of medication use

Mohsen Askar^a, Raphael Nozal Cañadas^b, Kristian Svendsen^{a,*}

^a Department of Pharmacy, Faculty of Health Sciences, UiT The Arctic University of Norway, Norway

^b Department of Informatics, Faculty of Science and Technology, UiT The Arctic University of Norway, Norway

ARTICLE INFO

Keywords:
Network analysis
Co-medication
Prescriptions
Drug interactions
Registries

ABSTRACT

Background: Network Analysis (NA) is a method that has been used in various disciplines such as Social sciences and Ecology for decades. So far, NA has not been used extensively in studies of medication use. Only a handful of papers have used NA in Drug Prescription Networks (DPN). We provide an introduction to NA terminology alongside a guide to creating and extracting results from the medication networks.

Objective: To introduce the readers to NA as a tool to study medication use by demonstrating how to apply different NA measures on 3 generated medication networks.

Methods: We used the Norwegian Prescription Database (NorPD) to create a network that describes the co-medication in elderly persons in Norway on January 1, 2013. We used the Norwegian Electronic Prescription Support System (FEST) to create another network of severe drug-drug interactions (DDIs). Lastly, we created a network combining the two networks to show the actual use of drugs with severe DDIs. We used these networks to elucidate how to apply and interpret different network measures in medication networks.

Results: Interactive network graphs are made available online, Stata and R syntaxes are provided. Various useful network measures for medication networks were applied such as network topological features, modularity analysis and centrality measures. Edge lists data used to generate the networks are openly available for readers in an open data repository to explore and use.

Conclusion: We believe that NA can be a useful tool in medication use studies. We have provided information and hopefully inspiration for other researchers to use NA in their own projects. While network analyses are useful for exploring and discovering structures in medication use studies, it also has limitations. It can be challenging to interpret and it is not suitable for hypothesis testing.

Introduction

Studies in social pharmacy and pharmacoepidemiology often utilize highly complex data and require the use of sophisticated methods to discern important patterns. Data used for quantitative studies in social pharmacy and pharmacoepidemiology can be described as attribute data and relational data. Attribute data includes the characteristics of the studied objects (e.g. sex, age, medication use, sociodemographic information, etc.) while relational data contains the various relationships between subjects. The suitable way of studying attributes data is quantitative analyses, whereas, for relational data, Network Analysis (NA) is the appropriate approach.¹ The subjects studied in network analyses can take many different forms.

A network can be described as a graph that shows the interconnections between a set of actors. Each actor is represented by a node and each connection between these nodes is represented by an

edge.² NA is a mathematical approach to study the relationships among nodes.³ The mathematical background of NA are summarized elsewhere.^{4,5}

Network Analysis has its roots in many research disciplines.⁶ Network analysis is used, among others, in social studies,⁷ ecological studies,⁸ genetics⁹ and systems pharmacology.¹⁰

As seen in Fig. 1, a network can be undirected (a and b) or directed (c and d). In a directed network, arrows show the direction of the relationship between nodes. In an undirected network, the relationship does not have a specific direction. The network edges can be weighted (b and d) or unweighted (a and c). In an unweighted network, the two nodes either have a relationship or not, while a weighed network considers the strength of the relationship.

* Corresponding author. Faculty of Pharmacy, UiT Norges arktiske universitet, Postboks 6050 Langnes, 9037, Tromsø, Norway.

E-mail address: kristian.svendsen@uit.no (K. Svendsen).

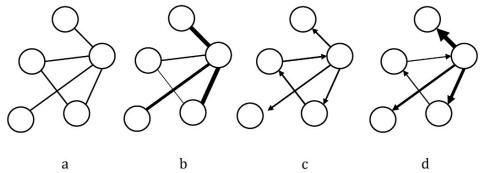


Fig. 1. Different types of networks, a) undirected and unweighted network, b) undirected and weighted network, c) directed and unweighted, and d) directed and weighted.

Use of Network Analysis in public health

Transmission networks have been used to examine the risk of disease transmission by investigating the relations between the infected people and healthy ones.^{11–13} Another form of transmission is the transmission of information. NA has been used to visualize the dissemination of public health information to different organizations and consumers. Some network characteristics reveal the pattern and the main actors contributing the most to information spread. Simulated networks can be used to suggest how to accelerate information spread.¹⁴ An example of this type of networks, the diffusion of information among physicians regarding a new drug. The study showed that more socially integrated physicians introduced the drug months before corresponding isolated physicians. NA was also used to study how health workers' professional and personal behavior impact health services.^{16,17}

Drug prescription network (DPN)

Pharmacoepidemiological studies of medications that are prescribed or dispensed is a relatively new application of NA. To our knowledge, Cavallo et al. were the first to study a drug prescription network in 2013. They used medications as the nodes and the number of patients being prescribed these medications as weighted edges. They aimed mainly at describing the topology of the co-prescription network to demonstrate which drug classes are most co-prescribed. They also compared the male/female networks and networks from different age strata and found that women in general were co-prescribed more drug classes.¹⁸

Bazzoni et al. were the first to use the term Drug Prescription Networks (DPN) in their paper published in 2015. They concluded that the DPNs are dense, highly clustered, modular and assortative. In this specific study, density reflected frequent co-prescribing. Modularity suggested that the network could be subdivided into clusters. The study also showed that it is possible to highlight spatial and temporal changes by

comparing different networks.¹⁹

Network Analysis terminology

We organized the key measures that are useful in studies of medication use under 4 main categories: (1) Topology analysis (2) Modularity analysis (3) Network comparison (4) Bipartite networks.²⁰ (Fig. 2).

1. Topology analysis

Network topological features refer to a group of characteristics, which either describe the network as a whole (network-level) or define individual actors of the network (node-level). There are many topological measures and each of them gives information about a specific network attribute, which then may warrant further investigation.

a. Global network description (network-level): A group of measures that describe the network as a whole.

- Number of nodes: the total number of drugs in the network. The network nodes can be grouped to show the number of drugs in each drug class. Different networks of different populations will have different distributions of drugs in the drug classes.
- Density: the density of a network is the number of actual edges divided by the total number of edges that would exist if all the nodes in the network were connected. This potential number can be calculated by the formula below where n is the number of nodes:

$$\frac{n \times (n - 1)}{2}$$

The network density can be useful in terms of comparison between different networks that describe the same type of drug-drug relation.

Assortativity: a network is assortative when the nodes that share a similar trait tend to connect. This trait can be many characteristics such as the nodes' degree. In this case, the assortativity means that nodes with a high number of edges tend to connect. Assortativity can be examined in terms of other common characteristics between the nodes as well. Assortativity coefficient is measured using Pearson correlation coefficient. Assortativity coefficient is scaled between -1 and 1, where 1 is most assortative.²¹

b. Node-level measures

Node-level measures describe the features of the different nodes across the network.

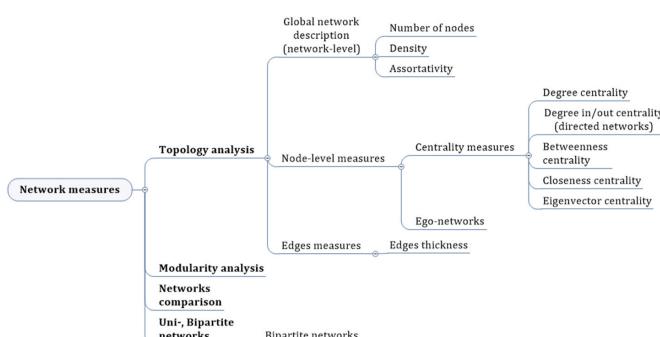


Fig. 2. Summarizing some of the Network Analysis measures that can be useful in the studies of medication use.

Centrality measures

Centrality measures indicate the importance of the network nodes by assigning a score to each of them. There are many different centrality measures and each of them can be used to describe a specific type of importance. By comparing the different centrality measures of a node, we can understand the different ways a node is influential to the network. This paper will discuss 4 of the most common types of centrality measures: degree, betweenness, closeness, and eigenvector centrality. The mathematical explanations of these measures are mentioned here.^{22,23}

Degree centrality

Degree centrality is the number of edges that are connected to a node. A higher score indicates that the node is connected to many other nodes. Node A in Fig. 3 has a degree score of 4. In a directed network, the degree is split into In-degree, which is the number of edges that direct to a node and Out-degree, which is the number of edges that originate from the node. In- and Out-degrees will therefore show the directions of relationships in a directed network. In Fig. 3, nodes C and D have an in-degree score of 3, while nodes A and G have an out-degree score of 3.

Betweenness centrality

The betweenness centrality of a node indicates how many times this node was used to connect two other nodes by the shortest possible path. Increasing the number of shortest paths will increase the betweenness centrality score.²² In Fig. 3, node A has the highest betweenness centrality score of 1.5.

Closeness centrality

It is a measure of the average distance between the node and all other nodes in the network. Nodes with the highest closeness score have the shortest distances to all other network nodes. The nodes A, B and F have the highest closeness centrality score of 1.

Eigenvector centrality

It is a measure of the importance of a node in a network based on the node's connections with other vital nodes. Relative scores are given to all nodes in the network based on the concept that connections to high-scored nodes give a higher score to the node than equal connections to

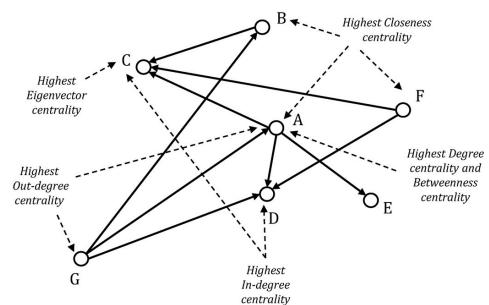


Fig. 3. Illustrating the different types of centrality. Node (A) represents the highest score of Degree and Betweenness centralities. The highest Eigenvector centrality score is assigned to the node (C). Nodes (A, B, F) have similar closeness Centrality. Nodes with the most in-degree edges are (C, D), while (A, G) have the most out-degree edges.

low-scored nodes. In other words, a high eigenvector score means that a node is connected to many nodes, which themselves are connected to important nodes in the network and have high scores of eigenvector centrality. This means that a node with a high eigenvector centrality score is not necessarily connected to the highest number of nodes in the network but is connected to the nodes with a high number of edges.²⁴ Node C in Fig. 3 has the highest eigenvector centrality score of 1. Assigning the centrality of each node in the network may lead us to visualize the network from a single specific important node perspective; this is called an *Ego-network* and it visualizes the part of the network that has the node of interest and the nodes that are directly connected to it.

c. Edge-level measures

Edge-thickness: in a weighted network, the edge-thickness represents a quantitative measure of the strength of the connection between two nodes. This representation is unique for NA and can be used to study many research questions. We will show an example where the number of users that co-medicated a pair of medications are used to represent the edge-thickness. In this context, thicker edges represent more frequently used pairs of medications.

2. Modularity analysis (Community detection)

One key feature of the network structure is its modularity. A module is a group of nodes that have many connections between each other and few(er) connections to the other nodes in the network.²⁵ There are many techniques of community detection including density-based, centrality-based, partition-based and hierarchical clustering techniques^{20,26,27}

3. Network comparison

It is possible to compare two or more networks to show the changes over time (temporal), between different areas (spatial), or between different groups of patients. These comparisons can be done by comparing the characteristics of the networks to highlight the differences in numbers and influences of the nodes. Another way to compare different networks is to subtract or divide the values of the edges between two networks. This will create edges representing the differences between the networks. By comparing many networks, dynamic graphs can be created showing the topological changes from a network to the next. Nodes will appear, disappear or change their locations as the dynamic graph moves through the different networks.²⁸

4. Bipartite networks

A network can be uni- or multipartite. We will only discuss uni- and bipartite networks. Unipartite networks have one set of nodes, while in bipartite networks the nodes belong to two disjointed sets (such as prescribers and patients). In a bipartite network, edges connect the nodes from different sets^{29,30} (Fig. 4).

The aim of this paper is to introduce the readers to NA as a tool to

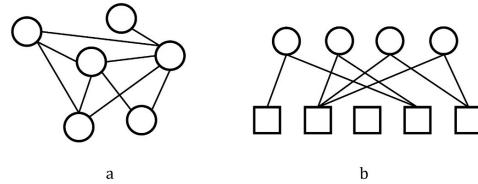


Fig. 4. (a) Unipartite network consisting of one type of nodes; (b) bipartite network consisting of two different types of nodes (circles and squares) in which the edges link between nodes from different types.

study medication use by demonstrating a practical real-life example of medication use in the elderly in Norway whenever it is possible, otherwise by giving an example from other studies.

Methods

We created a network of co-medication in elderly persons in Norway. We also created a network describing the severe drug-drug interactions (DDIs). Finally, we generated a network with the actual use of drugs with severe DDIs by combining the previous two networks.

Data sources

Co-medication network

The dataset used comes from the Norwegian Prescription Database (NorPD). It covers all dispensed prescriptions to elderly persons (≥ 65 years) in Norway between 2012 and 2014. The NorPD collects data from all pharmacies in Norway and covers all outpatient dispensing for the entire Norwegian population. Details on the NorPD are published elsewhere.³¹ In total, the dataset included 765,383 patients, 344,285 men (45%) and 421,098 women (55%) with 75 years as mean patient age. Edges in this network represent the number of patients who combined pairs of medications. In order to define the co-medication, we created treatment episodes using the Proportion of Days Covered (PDC) approach.³² We assumed that patients used one Defined Daily Dose (DDD)³³ per day and added 20% to each prescription duration to account for imperfect adherence. We also allowed a medication-free gap of 14 days before ending a treatment episode and starting another. This means that if the patient exceeds 14 days without the medication, the treatment episode for this medication ends and a new episode starts if the patient picks up a new prescription. Finally, co-medication was defined as the overlapping drug treatment episodes at the index date, January 1, 2013.

For each pair of nodes (drugs), we summed up the number of co-medication occurrences (i.e. number of patients combining these two drugs) to create a weighted and undirected network.

We excluded the medications that have no defined DDD such as the medications for topical use, vaccines, and ophthalmologicals. In total, we excluded 357 medications (217 local and 140 systematic drugs). The co-medication network is shown here: <https://mohsenaskar.github.io/co-medication/network/>. The network is searchable by substance name. Clicking on any node shows the ego-network of this node as well as some network measures.

Severe drug interactions network

To create this network, we used a dataset derived from the Norwegian Electronic Prescription Support System (FEST). FEST is a national information service that provides common pharmaceutical data to the IT-systems that are involved in the drug prescribing process including systems used by physicians, hospitals and pharmacies.³⁴ Drug-drug interactions is a part of the FEST database. In FEST, the DDIs are divided into 3 categories; interactions that should be avoided (i.e. severe), interactions where precautions should be taken and interactions that do not require any action. Only severe DDIs were included in the study. There were 57,151 unique severe interactions. The edges in this network represent the presence of a severe interaction between the two nodes.

The network is undirected and unweighted. The severe DDIs network is shown here: <https://mohsenaskar.github.io/DDI/network/>

Combining co-medication and DDIs networks

Both DDI and co-medication network has drugs as nodes. When combining the two networks only edges that exists in both networks are included (only edges with any users combining the medications and where there is a severe DDI). The number of users for each edge from the co-medication network becomes the weight of the edges in the combined network.

This network is shown here: <https://mohsenaskar.github.io/DDI-i-n-co-medication-network/network/>

Preparing the data to create a network

The data from the NorPD contains attributable data including a patient identity number, sex, year of birth, and data about each individual dispensed drug. To create a network, this data needed to be reshaped. The first step was to create a file with only medications that were used on the index date. Secondly, the file was aggregated such that an edge list was created. The edge list contains 2 variables defining the pairs of drugs and one variable with the number of users co-medicating with each pair of drugs. This edge list can be used by various software as described below. The process of data preparation is summed up in Fig. 5 and the edge list is openly available at the UiT The Arctic University of Norway open data repository here: <https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/1OUTYI>. The Stata syntax for creating the edge list is supplied in supplementary 2.

Software to use for network analysis

There are many available tools to use for NA. We will focus on how to use the *nwcommands* suite of commands that can be downloaded into *Stata* and the *igraph* package in *R* as well as visualization in *Gephi*. Other packages like “*igraph*” or “*NetworkX*” for *Python* are popular as well. All these packages can be used for visualizing and computing different network measures with differences in their integrated features and performance.^{35,36}

Using Stata (*nwcommands*, *nwANND*)

Using the edge list, *nwcommands* will create an adjacency matrix.³⁷ The adjacency matrix is a square matrix that contains the relationships between every pair of nodes in the network. The adjacency matrix can be saved as *Pajek* format that can be later imported and used by *Gephi*. In addition, *nwcommands* can display some network measures on both the network and node-level. *NwANND* is used for calculating the assortativity coefficient.³⁸ The syntax can be found in supplementary 2.

Using R (*igraph*)

Igraph (<https://igraph.org/>) is a library for creating and analyzing graphs. It is widely used by network researchers to analyze graphs and networks. It is currently available for *C*, *C++*, *Python*, *R* and *Mathematica*.

One of the strengths of *igraph* is that it can be programmed with a high-level programming language and still be very efficient when handling large networks. In our *R* context, *igraph* integrates well with the visualization package (*ggplot2*) via the *ggraph* library.

Igraph uses an edge list and can link it with attribute data for each node as well. An example code for network visualization using *igraph* and *ggraph*, is given in Supplementary 2.

Using gephi

Gephi (<https://gephi.org/users/>) is an open-source and free stand-alone software. The software can handle small to medium-sized networks (up to 150000 nodes). *Gephi* is user-friendly and requires no programming experience.³⁹ With many visualizing layouts and network measures, *Gephi* can provide a good starting point for the drug network study.³⁹ After importing the adjacency matrix to *Gephi*, we can process the network by applying different visualizing layouts, adding filters and colors. The structure of most drug networks can be complex and the unprocessed form of the network is often uninformative. By using different attributes (e.g. sex, modularity, etc.) the network can become more easily interpretable.²⁸

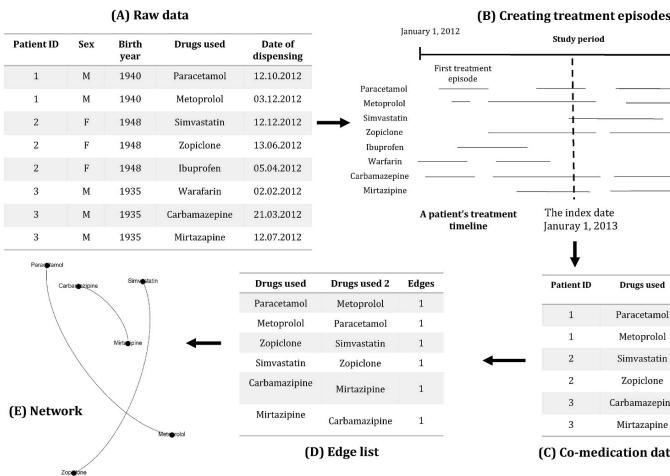


Fig. 5. Data preparation process. A) Raw clinical data in the long-form. B) Creating treatment episodes, a line indicates an episode. Gaps between lines indicate a medication-free period of more than 14 days. C) Data including only medications used at the index date D) Creating the edge list including 3 variables; two variables represent the nodes (Drug used, Drug used 2) and the third variable is the edge weight between the pair of nodes. E) Using the software to generate the network.

Results

We will present results from our networks using the same terms and order as in the introduction (Fig. 2). Table 1 provides the main topological features of the co-medication and the severe DDIs networks. The co-medication network is denser than the severe DDIs network, indicating that the drugs in the co-medication network are more connected. The assortativity coefficient shows that the co-medication network is non-assortative in terms of degree similarity, while the severe DDIs network is more assortative. Centrality measures in the co-medication network revealed that the same 5 drugs are the most central in all measures, while in the severe DDIs network; there is more variation in the top 5 drugs in each centrality measure. The results also showed that both networks are modular.

Fig. 6 shows that the majority of anatomical drug classes were assortative. This means that the drugs from the same anatomical group tend to be more co-prescribed. We also investigated the assortativity of the drugs on the pharmacological level (3rd level Anatomical Therapeutic Chemical classification) in supplementary 3.

Ego-networks as a measure can be seen by accessing the online networks we created and selecting individual nodes. The different network links can be found in the method section.

The top 10 edge weights for the severe DDIs in the co-medication network and co-medication only network are shown in Tables 2 and 3 respectively. We see in Table 2 that the number of patients using drugs causing severe DDIs are relatively low (less than 1000 users for all) while the most commonly co-medicated drugs seen in Table 3 is much higher with acetylsalicylic Acid (aspirin) and simvastatin having around 83000 users representing almost 11% of the population.

Modularity analysis

We found 4 modules in the co-medication network and 11 modules in the severe DDI network. For the co-medication network, there was one large community and 3 other smaller communities. Nervous and Respiratory system groups (N- and R-groups) drugs are just found in module 0, while Cardiac-, Alimentary-, Blood groups(C, A-, B- groups)

are common groups between modules 1 and 2, but with considering the number of users in each module we can locate in which module these ATC groups represent the most importance. Drugs used for diabetes, (A10) group, present only in module 2. The complete tables of modules are listed in supplementary 1. For the severe DDIs network, the modules found are shown below in Fig. 7.

Discussion

A Network visualizes the relationships of a dataset in one graph. This unique ability of data representation is combined with many measures that are helpful for many research disciplines. A starting point for generating any network is to select the nodes and define the edges. A precise definition of the edges allows the researcher to extract the correct information. NA is a well-suited approach to study complex systems. Although the approach has been widely used in many fields of research, only a few studies studied the drug-relations in a network.^{18,19}

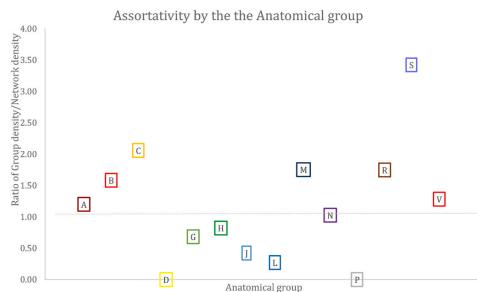
Our results show that many network outcomes can be useful in the studies of medication use. Moreover, some results are unique measures that only NA can perform such as edge measure and modules detection. Employing centrality measures in the drug study introduce an opportunity to observe the influence of the different drugs in the drug-network. Determining this influence can be useful for clinicians and decision-makers.

After generating a network, some topological features have to be reported first to get a general idea about the network content and its basic characteristics. Network-level measures such as assortativity and density reveal many clues for further investigation. Centrality measures show how influential each node is in the network. It is possible to have high centrality of one type and a low of another for the same node.⁸ In order to study the importance of the nodes, it is necessary to use more than one measure of centrality. Recent studies suggest using centrality measures as an alternative approach for variables selection. Lutz et al. used the centrality measures to identify 4 additional variables contributing to the predicting of treatment dropout in patients with anxiety disorders.⁴⁰ Valenzuela et al. described a methodology based on degree and centrality measures to obtain the most representative variables for

Table 1
The topological measures of co-medication and severe DDIs networks.

Outcome	Co-medication network	DDIs network	Indicates
1. Topology analysis			
a) Network-level measures			
Number of nodes	762	1699	Number of drugs present in the network.
Number of edges	75052	57151	Number of connections between the network nodes
Density	0.26	0.04	The extent of connections between the network's nodes
Average degree	99	34	The average number of connections that each node has.
Assortativity coefficient	-0.26	0.4	To what extent nodes with higher degree tends to correlate.
b) Node-level measures			
<i>Centrality measures</i>			
Nodes with the highest Degree centrality scores	Acetylsalicylic acid Simvastatin Zopiclone Paracetamol Metoprolol	Typhoid vaccine Erythromycin Prilkerium Clarithromycin Moxifloxacin	Combining these centrality measures can be used to assign the importance of each drug to the network.
Nodes with the highest Betweenness centrality scores	Acetylsalicylic acid Simvastatin Zopiclone Paracetamol	Typhoid vaccine Padelipofin Hyperici herba (St John's-wort) Tuberculosis vaccine Ginkgo leaves	
Nodes with the highest Closeness centrality scores	Metoprolol Acetylsalicylic acid Simvastatin Zopiclone Paracetamol Metoprolol	Bromelains Telbivudine Peg interferon alfa-2a Diazepam Oxazepam	
Nodes with highest Eigenvector centrality scores	Acetylsalicylic acid Simvastatin Zopiclone Metoprolol Paracetamol	Typhoid vaccine Erythromycin Clarithromycin Chloramphenicol Moxifloxacin	
c) Edge-level measures			
Average path length	1.77	3.09	Average shortest path between two nodes.
Thickest edge weight	82948	1	For the weighted co-medication network the number reflects the highest number of patients co-medication. This highlights clinically important combinations.
Edges range	1–82948	0–1	
2. Modularity			
Modularity	0.088	0.54	Indicates presence of modules in the network.
Number of modules (communities)	4	11	
Number of nodes in largest module	530 (module 0)	372 (module 4)	

Assortativity by the the Anatomical group

**Fig. 6.** Assortativity of network nodes in terms of similarity by the anatomical group. Squares above 1 represent a drug group with a higher density than the general density of the network (0.26). The S (Sensory organs) anatomical group had the highest assortativity. D (Dermatologicals) and P (Antiparasitic products) groups had no edges because these drug classes were excluded from the study.**Table 2**

The top 10 clinically relevant severe DDIs in the co-medication network.

The severe DDI drug pair		No. of patients co-medicating
1	Codeine and paracetamol	Tramadol 855
2	Esomeprazole	Clopidogrel 823
3	Simvastatin	Carbamazepine 480
4	Metoprolol	Paroxetine 454
5	Metoprolol	Verapamil 380
6	Lansoprazole	Clopidogrel 308
7	Diclofenac	Ibuprofen 305
8	Diazepam	Oxazepam 300
9	Carbamazepine	Zopiclone 280
10	Omeprazole	Clopidogrel 277

Table 3

The top 10 combined drugs in the co-medication network.

Most combined drugs		No. of patients co-medicated
1	Acetylsalicylic acid	Simvastatin 82948
2	Acetylsalicylic acid	Metoprolol 52577
3	Acetylsalicylic acid	Atorvastatin 42753
4	Metoprolol	Simvastatin 36792
5	Acetylsalicylic acid	Amlodipine 32628
6	Acetylsalicylic acid	Zopiclone 29173
7	Amlodipine	Simvastatin 22554
8	Acetylsalicylic acid	Ramipril 19660
9	Simvastatin	Zopiclone 18845
10	Metformin	Acetylsalicylic acid 18507

predicting successful aging.⁴¹ These approaches are interesting and represent an alternative method to the other variable selection methods. Edge-level measures are the core of the networks and the principal for many network measures.

Modularity analysis exposes the network structure. This measure is believed to introduce special importance in the drug study. Bazzoni et al. found the DPN to be modular,¹⁹ which is consistent with what we found in our networks. Further investigation is needed to assess the underlying patterns in the modules we found in our co-medication network. Modules can be interpreted as clusters of patients with similar diagnoses using the same medications. In our initial analysis of modularity, we identified 4 modules, further work could be done to identify smaller groups by detecting the sub-clusters inside each module. Modules in the DDI network could be connected to pharmacological data to see the importance of pharmacokinetic interactions through systems such as the cytochrome P450 system. We have not explored this but there is a great

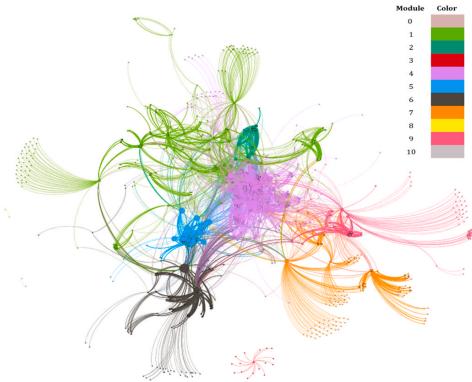


Fig. 7. The 11 modules that were detected in the severe DDIs network. Different colors indicate different modules. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

potential in using modularity analyses in order to understand networks.

Comparing different networks can reveal the change in patterns over time, place and different populations. Networks that describe the relation between drug-use and morbidities for a patient or a group of similar patients over time may identify the development of co-morbidities and drug use.

Bipartite networks provide a variety of possibilities to study many situations in which drugs are involved with other network actors such as physicians or diseases. Dasgupta & Chawla created a bipartite drug-disease network to study the interactions between drugs and co-morbidities.⁴² Hu et al. studied the prescribing of some opioids by creating a bipartite network of patients and prescribers and using the network to analyze the relationship between patients and prescribers and detect “doctor shopping” and suspicious network nodes.⁴³ A redrawn example from this study is shown in Fig. 8.

Our study has some limitations. As we used the DDD to outline the treatment episodes, we excluded the medications that have no defined DDD. This reduced the represented co-medication in our networks to the actual co-medication at the index date.

NA also has some important limitations. As a tool, it can be used to explore data, to find unusual structures, group nodes together and find unusual individual nodes. However, it can be hard to interpret results from NA and it is only suited for hypothesis generation. It also cannot explore many sets of relationships between variables at the same time as well as determining causal relationships. For such research questions, other hypothesis testing methodologies will be more needed. However, in research focused on exploration, NA can be a valuable tool.

Conclusion

The main purpose of this paper was to demystify the NA as a method. We have explained the terminology of network analyses and showed, with examples, how network analyses can be used for hypothesis generation. The online links to our networks visualize the data much better than a static picture can and we hope that we have provided enough information, and inspiration, to explore how you can use NA on your own data. We are confident that the future will see many new applications of NA and interesting results for researchers in social pharmacy and pharmacoepidemiology.

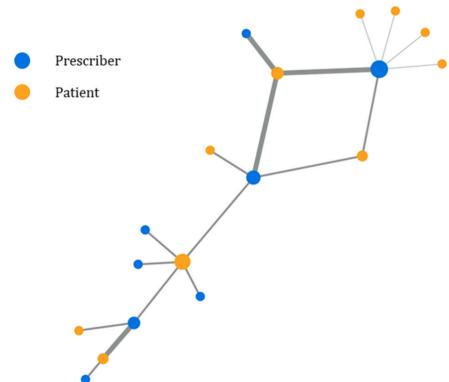


Fig. 8. An example of a bipartite network representing a sub-graph of two types of nodes (i.e. prescribers and patients) linked by the number of Fentanyl® patches prescriptions. The bigger nodes indicate more number of connections. The thicker edges indicate a higher number of prescriptions. (redrawn from “Network analysis and visualization of opioid prescribing data”⁴³).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Mohsen Askar: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Raphael Nozal Cañas:** Software, Writing – review & editing. **Kristian Svendsen:** Conceptualization, Software, Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no conflicts of interest related to this study.

Acknowledgments

We thank Lars Småbrekke and Elin Lehnboim for their useful feedback on the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.sapharm.2021.06.021>.

References

- Scott J. *Social Network Analysis*. SAGE Publications Ltd; 2017. <https://doi.org/10.4135/9781520716597>.
- Prell C. *Social Network Analysis: History, Theory & Methodology*. Sage; 2012.
- Wasserman S, Faust K. *Social Network Analysis*. Cambridge University Press; 1994. <https://doi.org/10.1017/CBO9780511815478>.
- Newman M. *Networks: An Introduction*. Oxford University Press, Inc.; 2010.
- Newman MEJ. Mathematics of networks. In: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK; 2018:8525–8533. https://doi.org/10.1057/978-1-349-95189-5_2565.
- Luke DA, Harris JK. Network analysis in public health history, methods, and applications. *Annu Rev Publ Health*. 2007;28:69–93. <https://doi.org/10.1146/annurev.publhealth.28.021406.144132>.

7. Freeman L. *The Development of Social Network Analysis, A Study in the Sociology of Science*. Vancouver, BC Canada: EP Empirical Press; 2004.
8. Delmas E, Besson M, Brice M-H, et al. Analysing ecological networks of species interactions. *Biol Rev*. 2019;94(1):16–36. <https://doi.org/10.1111/brv.12433>.
9. Ghazalpour A, Doss S, Zhang S, et al. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet*. 2006;2(8):e130. <https://doi.org/10.1371/journal.pgen.0020130>.
10. Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu Rev Pharmacol Toxicol*. 2012;52(1):505–521. <https://doi.org/10.1146/annurev-pharmtox-010611-134520>.
11. Ken TDE, Matt JK. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc Natl Acad Sci U S A*. 2002;99(20):13330. <https://doi.org/10.1073/pnas.002244299>.
12. Keeling MJ, Eames KTD. Networks and epidemic models. *J R Soc Interface*. 2005;2(4):295–307. <https://doi.org/10.1098/rsif.2005.0051>.
13. Grande KM, Stanley M, Redo C, Wergin A, Guilfoyle S, Gasiorowicz M. Social network diagramming as an applied tool for public health: lessons learned from an HCV cluster. *Am J Public Health*. 2015;105(8):1611–1616. <https://doi.org/10.2105/AJPH.2014.302193>.
14. Valente TW, Davis RL. Accelerating the diffusion of innovations using opinion leaders. *Ann Am Acad Polit Soc Sci*. 1999;566(1):55–67. <https://doi.org/10.1177/0002716299566001005>.
15. Coleman J, Katz E, Menzel H. The diffusion of an innovation among physicians. *Sociometry*. 1957;20(4):253–270. <https://doi.org/10.2307/2785979>.
16. Poghosyan I, Luero RJ, Knutson AR, Friedberg M W, Poghosyan H. Social networks in health care teams: evidence from the United States. *J Health Organisat Manag*. 2016;30(7):1119–1139. <https://doi.org/10.1108/JHOM-12-2015-0201>.
17. Scott J, Tallia A, Crosson JC, et al. Social network analysis as an analytic tool for interaction patterns in primary care practices. *Ann Fam Med*. 2005;3(3):443. <https://doi.org/10.1370/afm.344>.
18. Cavallo P, Pagano S, Boccia G, De Caro F, De Santis M, Capunzo M. Network analysis of drug prescriptions. *Pharmacoepidemiol Drug Saf*. 2013;22(2):130–137. <https://doi.org/10.1002/pds.3384>.
19. Bazzoni G, Marangoni A, Tettamanti M, et al. The drug prescription network: a system-level view of drug co-prescription in community-dwelling elderly people. *Rejuvenation Res*. 2015;18(2):153–161. <https://doi.org/10.1089/rej.2014.1628>.
20. Kim E-Y. Knowledge-based bioinformatics: from analysis to interpretation, edited by gil altrovitz & marco ramoni. *Health Inform Res*. 2010;16(4):312. <https://doi.org/10.4258/hir.2010.16.4.312>.
21. Barrenas F, Chavali S, Holme P, Mobini R, Benson M. Network properties of complex human disease genes identified genome-wide association studies. In: Mailund T, ed. *PLoS One*. vol. 4. 2009, e8090. <https://doi.org/10.1371/journal.pone.0008090>, 11.
22. Freeman L, Freeman L. A set of measures of centrality based on betweenness. *Sociometry*. 1977;40(1):35–41. <https://doi.org/10.2307/3033543>.
23. Bonacich P. Power and centrality: a family of measures. *Am J Sociol*. 1987;92(5):1170–1182. <https://doi.org/10.1086/228631>.
24. Ruhnaud B. Eigenvector-centrality — a node-centrality? *Soc Networks*. 2000;22(4):357–365. [https://doi.org/10.1016/S0378-8733\(00\)00031-9](https://doi.org/10.1016/S0378-8733(00)00031-9).
25. Ji X, Machiraju R, Ritter A, Yen PY. Examining the distribution, modularity, and community structure in article networks for systematic reviews. *AMIA. Annu Symp proceedings AMIA Symp*. 2015;2015:1927–1936.
26. Zhang M, Lu LJ. Modules in networks, algorithms and methods. In: *Encyclopedia of Systems Biology*. New York: Springer; 2013:1447–1450. https://doi.org/10.1007/978-1-4419-9863-7_557.
27. Newman MEJ. Modularity and community structure in networks. *Proc Natl Acad Sci Unit States Am*. 2006;103(23):8577. <https://doi.org/10.1073/pnas.0601602103>.
28. Cherven K. *Mastering Gephi Network Visualization Produce Advanced Network Graphs in Gephi and Gain Valuable Insights into Your Network Datasets*. 2015. Published online.
29. Introduction to bipartite graphs. In: Arsanian AS, Häggkvist R, Denley TMJ, eds. *Bipartite Graphs And Their Applications*. Cambridge Tracts in Mathematics. Cambridge University Press; 1998:7–22. <https://doi.org/10.1017/CBO9780511984068.004>.
30. Guillaume J-L, Latapy M. Bipartite structure of all complex networks. *Inf Process Lett*. 2004;90(5):215–221. <https://doi.org/10.1016/j.ipl.2004.03.007>.
31. Furu K. Establishment of the nationwide Norwegian Prescription Database (NorPD); new opportunities for research in pharmacoepidemiology in Norway. *Epidemiology*. 2008;18(2):129–136. <https://doi.org/10.1093/ije/dyn182.23>.
32. Raeber AM, Schmidtmel JJ, Karter LA, Konieczny FJ, Steiner FJ. Standardizing terminology and definitions of medication adherence and persistence in research employing electronic databases. *Med Care*. 2013;51(Suppl 8):S11–S21. <https://doi.org/10.1097/MLR.0b013e31829b1d2a>, 8 Suppl 3.
33. Üstün TB. International classification systems for health. In: *International Encyclopedia of Public Health*. Elsevier; 2017:304–311. <https://doi.org/10.1016/B978-0-12-803678-5.00237-X>.
34. Norwegian Medicines Agency. Fest implementation guidelines. https://lege.midleverket.no/Documents/Andretemaer/FEST/Hvordan.bruke.FEST/Implementation_guide.FEST.v3.0.pdf.
35. Pavlopoulos GA, Paez-Espino D, Kyriakis NC, Illopolos I. Empirical comparison of visualization tools for larger-scale network analysis. *Adv Bioinformatics*. 2017;2017:1–8. <https://doi.org/10.1155/2017/1278932>.
36. Akhtar N. Social network analysis tools. In: *Fourth International Conference on Communication Systems and Network Technologies*. IEEE; 2014. 2014:388–392. <https://doi.org/10.1109/CSNT.2014.83>.
37. Grund T. Network analysis using stata. <https://nwcommands.wordpress.com/tutorials-and-slides/>; 2015.
38. Joyce C. NWANND: Stata module to compute ANND (average nearest neighbor degree) and related measures. *Stata Softw Components*; 2016. Published online December 18 <https://ideas.repec.org/c/boe/bocode/s458261.html>. Accessed October 29, 2020.
39. Bastian M, Heymann S, Jacomy M. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. 2009. <https://doi.org/10.13140/2.1.1341.1520>.
40. Lutz W, Schwartz B, Hofmann SG, Fisher AJ, Husen K, Rubel JA. Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: a methodological proof-of-concept study. *Sci Rep*. 2018;8(1):7819. <https://doi.org/10.1038/s41598-018-25953-0>.
41. Valenzuela JFB, Monterola C, Tong VJC, Filipó T, Ng TP, Larbi A. Degree and centrality-based approaches in network-based variable selection: insights from the Singapore Longitudinal Aging Study. In: Marinò IP, ed. *PLoS One*. vol. 14. 2019, e0219186. <https://doi.org/10.1371/journal.pone.0219186>, 7.
42. Dasgupta D, Chawla NV. *Disease and Medication Networks: An Insight into Disease-Drug Interactions*. 2014:1.
43. Hu X, Gallagher M, Loveday W, Dev A, Connor JP. Network analysis and visualisation of opioid prescribing data. *IEEE J Biomed Health Informatics*. 2020;24(5):1447–1455. <https://doi.org/10.1109/JBHI.2019.2939028>.

Appendix B: Preliminary Results

In this chapter, we explore the latest findings and ongoing research that form a crucial part of my PhD work. These results bear significant relevance to this thesis and enrich our understanding of the research domain, although still in progress and yet to be formally published. By including these preliminary findings, we aim to provide a comprehensive and up-to-date exploration of the topic. It is important to note that these findings are presented as drafts and are subject to further refinement and validation before publication.

B.1 Result I

Social network influences on obesity in a general youth population.

In recent years studies have evaluated the effect of peer pressure on obesity, via close friend networks in adults [11, 450, 451] and adolescents [452], or direct advertising of junk food [453] or healthy habits [454]. These studies have shown that people's health, including obesity, tends to become similar to those in their social network. We would like to check if these results also apply to our student population and how these evolve over time.

In these results (tables B.1, B.2 figures B.1, B.2), first, we showed the social dynamics with respect to BMI (table B.1). We show that this population behaves similarly to other student populations [455–457], and there is a bias in forming friendships with respect to BMI. "Overweight" and "Obese" students tend to form connections with each other, while at the same "Obese" students have lower connectivity and are more isolated in the network, and seem to have a negative bias towards "Healthy" students. On the other hand, "Healthy" form connections mostly among themselves, with a negative bias toward both "Overweight" and "Obese". Using the same simulation technique as in Paper A (section 4.3.2), we also show a bias toward obesity spread in all networks (table B.2), being stronger in the "Sports" network, and weaker in the "School" network.

Table B.1: Bias with respect to the total number of relationships between each BMI category. The top column shows the absolute and relative frequencies of the population ($n = 1034$). People with unknown BMI are excluded from the analysis ($n = 4$). Each combination of categories represents people who nominate a friend (rows) and people who are nominated (columns). Each cell is divided into three parts, the left-most is the total number of relationships in this combination, the center one is the expected number of relationships (only given if a bias was found), the right part contains an arrow indicating over (up ↑) or underrepresented (down ↓) using a two-sided binomial test with at least $p\text{-value} < 0.1$. The table χ^2 test is $< 10^{-10}$. In the bottom and to the left, we have the marginal absolute and relative frequencies for each combination.

	Underweight $n = 110, f = .106$	Healthy $n = 710, f = .684$	Overweight $n = 147, f = .142$	Obese $n = 67, f = .064$	Total	Freq
Underweight	47	285	48	24	404	10.8%
Healthy	282	1870 (1822)	345 (367) ↑	90 (122) ↓	2587	69.1%
Overweight	46	347	97 (74) ↑	34 (24) ↑	524	14.0%
Obese	22	135 (160) ↓	42 (32) ↑	29 (10) ↑	228	6.1%
Total	397	2637	532	177	3743	
Frequency	10.6%	70.5%	14.2%	4.7%		100%

Table B.2: Results of the spread of BMI across all networks simulations. The first column is the name of the network. Second is how many relationships are in that network. The third column is how many of those relationships share the same BMI category. The next 7 columns are the details of the simulation results, with the important one being the "Average" and "SD (Standard Deviation)", which shows how many same-to-same relationships in average we had in the 1000 simulations, using a network with the same topology but randomizing the BMI according to the BMI probability density data of the original network. The last column is the p-value, rounded to 3 decimals, which shows if there is a significant difference between the averaged same-to-same simulated relationship and the real same-to-same relationships.

Network	Real networks		Simulated 1000 networks							P-value
	Total Relationships	Equal Relationships	MIN	Q1	Median	Average	Q3	MAX	SD	
Overall	3767	2043	1761	1853	1893	1894.1	1934	2077	59.5	0.006
Physical	2823	1584	1233	1368	1402	1406.1	1445	1561	59.6	0.001
School	2979	1590	1337	1459	1490	1493.7	1536	1647	59.9	0.054
Sports	598	415	257	289	303	301.9	314	355	19.1	<0.0001
Home	1247	722	563	603	621	624.3	645	709	29.9	0.001
Other	1095	612	488	532	552	552.7	575	623	28.8	0.020

We checked for the general evolution of BMI over time (figure B.1). We observed that roughly half of the "Underweights" go into the "Healthy" group while almost none of the "Healthy" descent into the "Underweights". However, a similar trend is seen with a big group of "Healthy" going into "Overweight" without barely "Overweight" descend-

ing into “Healthy”. There is also another group going from “Overweight” to “Obese”; this group is larger than the group descending to “Healthy”. Both “Underweight” and “Healthy” groups decrease in size while both “Overweight” and “Obese” increase. Some from the “Obese” group descent into the “Overweight” group, but not enough to compensate for those new students entering the group. The average BMI (kg/m^2) in FF1 was 22.57 ± 4.23 , and in FF2 was 23.32 ± 4.25 . For men, in FF1 was 22.51 ± 4.22 and 23.55 ± 4.19 in FF2. For women was 22.62 ± 4.24 in FF1 and 23.12 ± 4.30 in FF2.

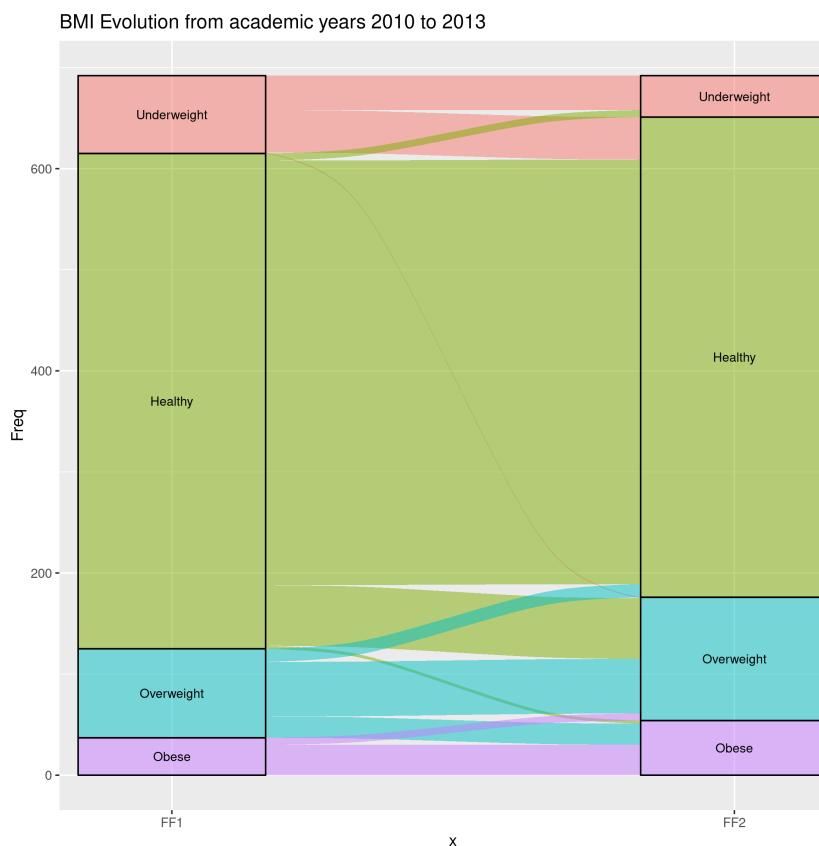


Figure B.1: An overview of the evolution of BMI from Fit Futures 1 (2010-2011) to Fit Futures 2 (2012-2013) for every student with valid BMI data in both studies ($n = 692$). On the left column, we have the FF1 BMI categories. In the right column, we have the FF2 categories two years later. The threads from left to right indicate the evolution of each person with respect to the BMI category.

We also tested if the number of high BMI friends is related to the BMI value in FF2 (figure B.2). On average, the “Healthy” and “Underweight” groups tend to go down in FF2, the greater the number of friends with $\text{BMI} > 25$ they had during FF1. We only see a sharp decrease in “Overweight” for the number of friends equal to 5, but they go into the “Obese” category instead which is a sharp increase in this group. Using logistic regression we estimated an 8.5% increase in the risk of high BMI with respect

to each additional "Overweight" or "Obese" friend. We also show that as the number of friends in FF1 with $BMI > 25$ increases, there is a higher probability of not landing in the "Healthy" group in FF2, and is also likely that the student will land in "Overweight" or "Obese" instead.

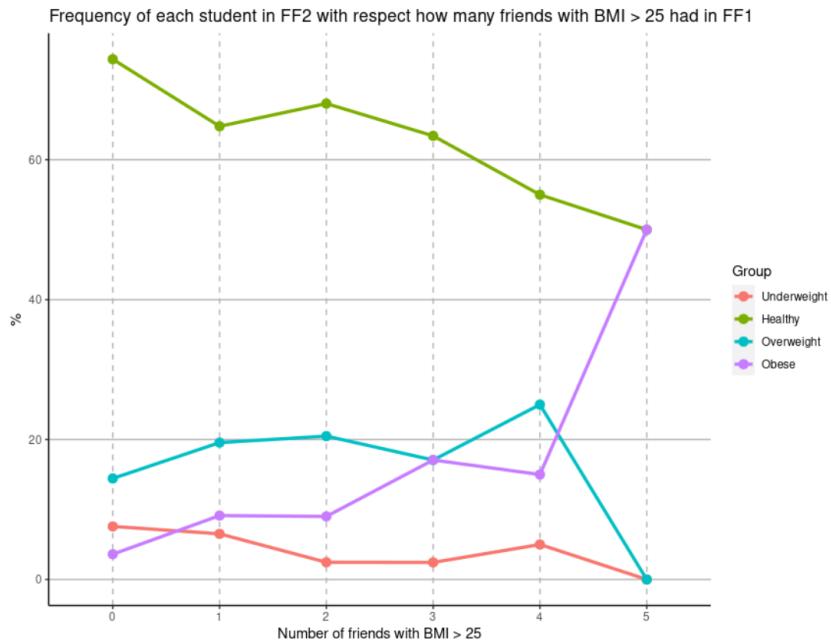


Figure B.2: The influence of friends from FF1 have with respect to FF2 BMI. The X-axis is the number of friends in FF1 with $BMI > 25 \text{ kg/m}^2$ The Y-axis is the relative number of students belonging to each BMI category in FF2.

Overall, these results seem to indicate that friendship is stagnated and students tend to keep within the same BMI group. And more importantly, an individual with "Healthy" friends has a better chance of being "Healthy" in the future, and an individual with "Overweight" or "Obese" friends has also a better chance of being "Overweight" or "Obese" in the future as well.

B.2 Result II

Social network influences on inflammatory response in a general youth population.

Chronic inflammation is a health concern with evidence suggesting that it contributes to the pathogenesis of numerous diseases, including cancer, cardiovascular disease, diabetes, and neurodegenerative disorders. Obesity has been shown to be correlated from childhood to adulthood [458], while also being responsible for a range of

adverse health conditions. Inflammation is shown to vary as a function of social isolation; in particular social behavior and the inflammatory process seem to be regulating one another [409, 459–464]. Obesity is also an underlying condition in both chronic inflammation and metabolic diseases [459, 465, 466]. In this explorative study, we investigate how the three concepts of inflammation in the form of a 92 proteomic assay, obesity in the form of anthropometric variables, and social interaction in the form of our network of friends, are related to one another. In particular, we look if there is a similarity of inflammation among social contacts. We found some results in the data that suggest that inflammatory markers are common among friends of the same sex and the same high school suggesting that social influence has an underlying link.

To evaluate if there is a correlation between each person and his/her friends' biomarker levels, we tried regression models (linear, quadratic, logarithmic, or exponential) using each person's friends' average biomarker level as the independent variable, and the person's biomarker levels as the dependent variable. Without any stratification by sex or by high school, we found no statistically relevant results for any biomarker. We found no correlation for sex stratification using all high schools. But biomarker levels are influenced by sex and social influence is driven by high school. If we stratify by these two variables; we found a total of 50 biomarkers for men and 46 for women that are significant with respect to some high schools (tables B.3 and B.4). This suggests that some biomarker levels are influenced by the student's social network, and in particular by high school, which has a relevant homophily coefficient (87.85% p-value <0.0001). Furthermore, high schools where blood samples were extracted later during the academic year (March / April 2011) show a higher amount of biomarkers correlation in contrast with other schools (October 2010 to March 2011). For High School 5, we found 13 markers for men, 11 for women; for High School 6, 23 for men, 17 for women; and for the rest of high schools combined 14 for men, 18 for women.

We wanted to check not only that biomarkers are similar among friends, but also different from non-friends. We defined biomarker distance as the ratio of the average square difference between each person's biomarker level and levels among students who are not his/her friends, and the average square difference between each person's biomarker level and levels among students who are his/her friends. We analyzed using same-sex friends, only for people with at least two friends of the same sex with valid biomarker levels. Values closer to 1 indicate no difference between each person's friends and non-friends biomarker level average. Values greater than 1 indicate that the distance to the friends' biomarkers is shorter (more similar) than to the non-friends biomarkers.

Values smaller than 1 indicate that non-friends are closer and more similar than friends. We arbitrarily defined distances ± 0.1 from 1 as relevant. We found for both men and women 14 proteins in each group with a ratio higher than 1.1, for men we found three proteins lower than 0.9, and for women, we found five proteins lower than 0.9 (tables B.5, B.6, B.7, B.8).

There is also a correlation between anthropometry and inflammation (table B.9 and table B.10); Result I and Result III suggest that social contact also influences the spread of obesity which is determined by anthropometrical variables. Not all the biomarkers presented in the anthropometric variables are necessarily expressed in each high school or vice versa. Inflammation driven by obesity only, or social influence only, would have similar results in both groups, so this would hint that we have a mix of the two factors. Furthermore, we see that high schools 5 and 6 have the greatest number of biomarkers being significant compared with the rest. This seems to indicate that biomarker levels gain similarity between friends over time. We also saw that the biomarkers that were negatively correlated with respect to friends' influence do not correlate with anthropometry and very little with high schools.

These results indicate, in the context of inflammation, that friendship either influence directly the immune system among a cluster of friends, or clusters of friends share some common habits which influence their immune system the same way.

Table B.3: Overview of all biomarkers with respect to each high school (n=8) for men. All biomarkers with either R² < 0.2 and p-value > 0.05, are hidden from the tables. P-values are expressed in GP Prism 5.04/d format.

Highschool	Protein	R2	P-value
H2	C-X-C motif chemokine 9	0.26	*
	Eotaxin	0.33	**
	Interleukin-10	0.55	****
	Interleukin-10 receptor subunit alpha	0.25	**
	Interleukin-13	0.66	****
	Interleukin-33	0.27	**
	Interleukin-6	0.28	*
	Interleukin-8	0.24	**
H3	Macrophage colony-stimulating factor 1	0.49	****
	Oncostatin-M	0.26	*
H4	Cystatin D	0.21	**
	STAM-binding protein	0.23	***
H5	Leukemia inhibitory factor receptor	0.27	***
	Adenosine Deaminase	0.44	****
	Beta-nerve growth factor	0.22	**
	C-C motif chemokine 28	0.28	**
	C-X-C motif chemokine 11	0.21	*
	CD40L receptor	0.29	**
	CUB domain-containing protein 1	0.23	*
	Fractalkine	0.48	***
	Interleukin-10	0.21	**
	Interleukin-10 receptor subunit alpha	0.34	**
H6	Interleukin-17C	0.24	**
	Macrophage colony-stimulating factor 1	0.24	*
	Monocyte chemoattractant protein 3	0.21	**
	Tumor necrosis factor receptor superfamily member 9	0.4	***
	Artemin	0.32	*
	Beta-nerve growth factor	0.27	*
	C-C motif chemokine 20	0.57	**
	C-C motif chemokine 25	0.45	**
	C-C motif chemokine 28	0.62	***
	C-X-C motif chemokine 5	0.32	*
	C-X-C motif chemokine 6	0.4	*
	Caspase-8	0.49	**
	CD40L receptor	0.72	***
	CUB domain-containing protein 1	0.63	***
H7	Delta and Notch-like epidermal growth factor-related receptor	0.56	**
	Fibroblast growth factor 23	0.49	**
	Fractalkine	0.41	*
	Interferon gamma	0.4	*
	Interleukin-33	0.28	*
	Interleukin-8	0.57	**
	Leukemia inhibitory factor receptor	0.49	**
	Neurotrophin-3	0.51	**
	Oncostatin-M	0.28	*
	Osteoprotegerin	0.53	**
H8	Programmed cell death 1 ligand 1	0.47	**
	TNF-related apoptosis-inducing ligand	0.48	**
	Urokinase-type plasminogen activator	0.65	***

Table B.4: Overview of all biomarkers with respect to each high school (n=8) for women. All biomarkers with either R₂ < 0.2 and p-value > 0.05, are hidden from the tables. P-values are expressed in GP Prism 5.04/d format.

Highschool	Protein	R ₂	P-value
H1	Glial cell line-derived neurotrophic factor	0.26	****
	Interferon gamma	0.24	***
H2	Tumor necrosis factor	0.4	****
H4	Fibroblast growth factor 19	0.23	**
	Interleukin-10 receptor subunit beta	0.25	*
	Interleukin-33	0.29	**
	Matrix metalloproteinase-10	0.24	**
	Neurturin	0.31	***
	Programmed cell death 1 ligand 1	0.32	**
	STAM-binding protein	0.24	**
	Sulfotransferase 1A1	0.22	*
H5	Urokinase-type plasminogen activator	0.28	**
	C-C motif chemokine 28	0.27	**
	C-X-C motif chemokine 5	0.31	**
	C-X-C motif chemokine 9	0.21	*
	Interleukin-10 receptor subunit beta	0.27	***
	Interleukin-6	0.32	***
	Interleukin-7	0.22	**
	Latency-associated peptide transforming growth factor beta-1	0.26	***
H6	Matrix metalloproteinase-10	0.22	**
	TNF-related activation-induced cytokine	0.24	**
	Tumor necrosis factor	0.36	****
	Vascular endothelial growth factor A	0.24	**
	Brain-derived neurotrophic factor	0.76	**
	C-C motif chemokine 23	0.95	***
	Delta and Notch-like epidermal growth factor-related receptor	0.85	***
	Fibroblast growth factor 21	0.66	**
H8	Fms-related tyrosine kinase 3 ligand	0.76	*
	Interleukin-10 receptor subunit alpha	0.69	**
	Interleukin-17A	0.64	*
	Interleukin-2 receptor subunit beta	0.75	*
	Interleukin-20 receptor subunit alpha	0.73	**
	Interleukin-33	0.64	**
	Interleukin-6	0.9	***
	Leukemia inhibitory factor receptor	0.89	***
	Monocyte chemoattractant protein 3	0.78	**
	Neurturin	0.45	*
	Osteoprotegerin	0.59	*
	Stem cell factor	0.67	**
	T cell surface glycoprotein CD6 isoform	0.45	*
	C-C motif chemokine 4	0.27	***
	Caspase-8	0.47	****
	Interleukin-17C	0.21	**
	Interleukin-24	0.36	****
	Osteoprotegerin	0.21	**
	Tumor necrosis factor receptor superfamily member 9	0.29	***

Table B.5: Ratio of average square distances between each person's biomarker level and friend's biomarker levels, and average square distance between each person's biomarker levels and non-friends biomarker levels. Values greater than 1 suggest that clusters of friends have similar biomarkers levels in comparison with the rest of the non-friend population. Relevant values are highlighted in bold, with green for >1.1 and red for <0.9. Values are rounded to two decimals but highlighted according to the original values. (Table 1 of 4)

Protein	Men	Women
Adenosine Deaminase	1.12	1.03
Artemin	0.98	1.04
Axin-1	1.06	1.02
Brain-derived neurotrophic factor	1.06	1.14
Beta-nerve growth factor	1.29	0.91
Caspase-8	1.01	1.04
Eotaxin	0.94	1.04
C-C motif chemokine 19	1	0.96
C-C motif chemokine 20	1.09	0.97
C-C motif chemokine 23	1.06	1.1
C-C motif chemokine 25	1.04	1.02
C-C motif chemokine 28	0.76	0.84
C-C motif chemokine 3	1	0.96
C-C motif chemokine 4	0.98	1.02
Natural killer cell receptor 2B4	1.02	1
CD40L receptor	1.03	1.04
T-cell surface glycoprotein CD5	1.09	1.07
T cell surface glycoprotein CD6 isoform	1.04	0.97
CUB domain-containing protein 1	1.04	1.14
Macrophage colony-stimulating factor 1	1.01	1.13
Cystatin D	1.04	0.97
Fractalkine	1.12	1.03
C-X-C motif chemokine 1	1.07	1.08
C-X-C motif chemokine 10	1.05	0.97

Table B.6: Ratio of average square distances (Table 2 of 4)

Protein	Men	Women
C-X-C motif chemokine 11	1.05	1.02
C-X-C motif chemokine 5	1.01	1.09
C-X-C motif chemokine 6	0.95	1
C-X-C motif chemokine 9	1.11	1
Delta and Notch-like epidermal growth factor-related receptor	1.14	1.09
Eukaryotic translation initiation factor 4E-binding protein 1	1.21	1.03
Protein S100-A12	1.03	0.99
Fibroblast growth factor 19	1.08	1.06
Fibroblast growth factor 21	1.07	1.12
Fibroblast growth factor 23	1.08	1.06
Fibroblast growth factor 5	1.03	0.9
Fms-related tyrosine kinase 3 ligand	1.11	0.99
Glial cell line-derived neurotrophic factor	1.08	1.05
Hepatocyte growth factor	1.04	1
Interferon gamma	1.04	0.79
Interleukin-10	0.99	1.16
Interleukin-10 receptor subunit alpha	1.03	0.96
Interleukin-10 receptor subunit beta	1.1	1.06
Interleukin-12 subunit beta	1.08	1
Interleukin-13	1.04	1.11
Interleukin-15 receptor subunit alpha	1.02	1
Interleukin-17A	0.95	1
Interleukin-17C	1.06	1.04
Interleukin-18	0.97	1.05

Table B.7: Ratio of average square distances (Table 3 of 4)

Protein	Men	Women
Interleukin-18 receptor 1	0.94	1.08
Interleukin-1 alpha	1.09	1.12
Interleukin-2	1.09	0.97
Interleukin-20	1.22	0.98
Interleukin-20 receptor subunit alpha	1.01	0.98
Interleukin-22 receptor subunit alpha-1	1	0.92
Interleukin-24	0.98	1.03
Interleukin-2 receptor subunit beta	1.05	0.93
Interleukin-33	1.03	0.99
Interleukin-4	0.98	1.09
Interleukin-5	0.87	1.04
Interleukin-6	0.98	1.3
Interleukin-7	1.03	1
Interleukin-8	1.05	1.01
Leukemia inhibitory factor	0.92	0.85
Leukemia inhibitory factor receptor	1.08	0.96
Monocyte chemotactic protein 1	1.05	1.18
Monocyte chemotactic protein 2	0.92	1.06
Monocyte chemotactic protein 3	1.01	0.92
Monocyte chemotactic protein 4	1	1.12
Matrix metalloproteinase-1	1.05	0.97
Matrix metalloproteinase-10	1.11	1.1
Neurturin	0.91	1.06
Neurotrophin-3	0.98	1.02

Table B.8: Ratio of average square distances (Table 4 of 4)

Protein		Men	Women
Osteoprotegerin		1.07	0.98
Oncostatin-M		1.09	1.03
Programmed cell death 1 ligand 1		0.93	0.96
Stem cell factor		1.05	1.08
SIR2-like protein 2		1.04	1.02
Signaling lymphocytic activation molecule		1.02	0.88
Sulfotransferase 1A1		1.01	1.07
STAM-binding protein		1.1	0.99
Transforming growth factor alpha		1.09	1.02
Latency-associated peptide transforming growth factor beta-1		1.08	0.99
Tumor necrosis factor		0.87	0.69
TNF-beta		1.04	1.02
Tumor necrosis factor receptor superfamily member 9		1.24	1.08
Tumor necrosis factor ligand superfamily member 14		1.08	1.01
TNF-related apoptosis-inducing ligand		1.17	1.01
TNF-related activation-induced cytokine		1.1	1.11
Thymic stromal lymphopoietin		1.01	0.98
Tumor necrosis factor		1.02	1.02
Urokinase-type plasminogen activator		1.24	1.08
Vascular endothelial growth factor A		1	1.16

Table B.9: Anthropometric variables with respect to biomarkers levels in men. Each cell is a p-value in GP Prism 5.04/d format. Biomarkers rows with no statistically significant p-values are hidden. All p-values are corrected for Bonferroni.

Protein	Waist	Hip	Height	Weight	BMI	HR	SYSBP	DIABP
C-C motif chemokine 3	***	**	ns	**	**	ns	ns	ns
C-C motif chemokine 4	**	ns	ns	ns	ns	ns	ns	ns
CUB domain-containing protein 1	****	****	ns	****	****	ns	ns	ns
Macrophage colony-stimulating factor 1	**	****	ns	***	***	ns	ns	ns
Delta and Notch-like epidermal growth factor-related receptor	ns	ns	ns	*	ns	ns	ns	ns
Fibroblast growth factor 19	ns	ns	ns	ns	*	ns	ns	ns
Fibroblast growth factor 21	*	ns	ns	ns	ns	ns	ns	ns
Glial cell line-derived neurotrophic factor	**	ns	ns	*	**	ns	ns	ns
Hepatocyte growth factor	****	***	ns	**	****	ns	ns	ns
Interleukin-18	***	***	ns	***	***	ns	ns	ns
Interleukin-18 receptor 1	****	****	ns	****	****	ns	ns	ns
Interleukin-6	****	***	ns	***	****	ns	ns	ns
Monocyte chemoattractant protein 3	****	****	ns	****	****	ns	ns	ns
Stem cell factor	****	****	ns	****	****	ns	ns	ns
Tumor necrosis factor receptor superfamily member 9	***	ns	ns	*	**	ns	ns	ns

Table B.10: Anthropometric variables with respect to biomarkers levels in women. Each cell is a p-value in GP Prism 5.04/d format. Biomarkers rows with no statistically significant p-values are hidden. All p-values are corrected for Bonferroni.

Protein	Waist	Hip	Height	Weight	BMI	HR	SYSBP	DIABP
Caspase-8	*	***	ns	***	**	ns	ns	ns
C-C motif chemokine 3	*	ns	ns	ns	ns	ns	ns	ns
CUB domain-containing protein 1	****	****	ns	****	****	ns	ns	ns
Macrophage colony-stimulating factor 1	****	***	ns	**	**	ns	ns	ns
Delta and Notch-like epidermal growth factor-related receptor	ns	ns	ns	*	*	ns	ns	ns
Fibroblast growth factor 21	*	ns	ns	ns	*	ns	ns	ns
Hepatocyte growth factor	****	***	ns	**	***	ns	ns	ns
Interleukin-10 receptor subunit beta	****	*	ns	**	**	ns	ns	ns
Interleukin-18	**	*	ns	ns	**	ns	ns	ns
Interleukin-18 receptor 1	****	***	ns	***	****	ns	ns	ns
Interleukin-6	****	****	ns	****	****	ns	ns	ns
Interleukin-7	**	**	ns	**	*	ns	ns	ns
Monocyte chemotactic protein 3	****	****	ns	****	****	ns	ns	ns
Monocyte chemotactic protein 4	*	ns	ns	ns	ns	ns	ns	ns
Latency-associated peptide transforming growth factor beta-1	*	*	ns	ns	ns	ns	ns	ns
TNF-related apoptosis-inducing ligand	**	*	ns	ns	*	ns	ns	ns
TNF-related activation-induced cytokine	*	**	ns	*	ns	ns	ns	ns
Vascular endothelial growth factor A	**	*	ns	*	***	ns	ns	ns

B.3 Result III

Measuring social influence with random forest regression and artificial neural networks

There are plenty of studies and machine learning models that predict obesity given a multivariate dataset, some using up to 190 multidomain variables [467]. But to our knowledge, no machine learning method has been used to predict changes in BMI using social networks and compare how well the model evaluate those variables' score to classical lifestyle factors.

We want to determine which variables are more important to predict BMI in FF2. We measured the influence of host factors, and the influence of the total number of friends grouped by BMI category, using RF (easier explainability) and ANNs (higher accuracy). We measure variable influence using SHAP for both models and MDI for RF.

We split our subjects into six different datasets, organized by whether or not the student's BMI increased or decreased, and which was the original BMI in FF1.

- **(A) “Getting worse or staying bad”.** Cases for students who are Healthy in FF1 and end up Overweight or Obese in FF2. Or students who are Overweight in FF1 and end up Overweight or Obese in FF2. Or students who are Obese in FF1 and stayed Obese in FF2.
- **(B) “Getting better or staying good”.** Cases for students who are Healthy, Overweight, or Obese BMI in FF1 and end up Healthy in FF2. Or Obese BMI in FF2 and end up Overweight in FF2.
- **(C) “Stay Healthy”.** Cases for students who have a Healthy BMI in both FF1 and FF2.
- **(D) “Bad cases get strictly better”.** Cases for students who have FF1 BMI > 25 , and their FF2 BMI $<$ FF1 BMI.
- **(E) “Healthy to worse”.** Cases for students with a Healthy BMI in FF1, who have a FF2 BMI > 25 .
- **(F) “Overweight or worse get strictly worse”.** Cases for students with FF1 BMI > 25 , who have a FF1 BMI $<$ FF2 BMI

We run both models against every dataset (table B.11). We also measured all the mean SHAP absolute values with respect to every dataset, every model, and every variable (figure B.3). Initial BMI in FF1 is always ranked as a very important variable in every model. In general, individual social influence variables are also highly ranked among sex and sport frequency. Dataset E seems to stand up with respect to the rest due to variable importance being quite different and more balanced across each variable. RF tends to lower the importance of any variable that is not BMI, while ANN tends to give more importance to the rest of the variables. Furthermore, we also tested all variables MDI with respect to all RF models (figure B.4).

Table B.11: Summary of all models and datasets used. From left to right, the ID of each dataset is represented by a letter, the short name of the dataset, the total samples in the dataset, Mean Absolute Error (MAE) for each model (RF or ANN) performing in this dataset, mean BMI in FF1 and mean BMI in FF2 in this dataset.

ID	Name	Samples	MAE		Mean BMI	
			RF	ANN	FF1	FF2
A	Healthy or worse to Overweight or worse	168	1.71	1.81	26.88	29.07
B	Healthy or worse to Healthy or better	440	0.99	1.00	21.43	21.98
C	Healthy to Healthy	420	1.05	1.07	21.12	21.81
D	Overweight or worse to lower BMI	44	0.96	1.39	28.69	26.79
E	Healthy to Overweight or worse BMI	63	0.90	1.60	23.32	26.52
F	Overweight or worse to higher BMI	80	1.86	1.95	28.99	31.46

On average, the models presented a Mean Absolute Error (MAE) of 1.35, and evaluated either "Total Healthy Friends" or "Total Overweight Friends" as either the most important variable or among the top 3 more important, for all models and datasets, after the variable BMI in FF1.

It is possible to evaluate specific individuals to check what makes them gain or lose BMI in FF2 with respect to FF1. Here we present two examples using RF in dataset A (figures B.5 and B.6). Individual cases should not be extrapolated as variable weights for the whole dataset.

Partial dependencies plots indicate how changing a particular variable changes the output of the model. All plots produced in this section are done using the RF models.

All datasets except C (Healthy to Healthy) show a decrease in BMI with respect to the total number of healthy friends (figure B.7). Datasets B (Overweight or Obese to

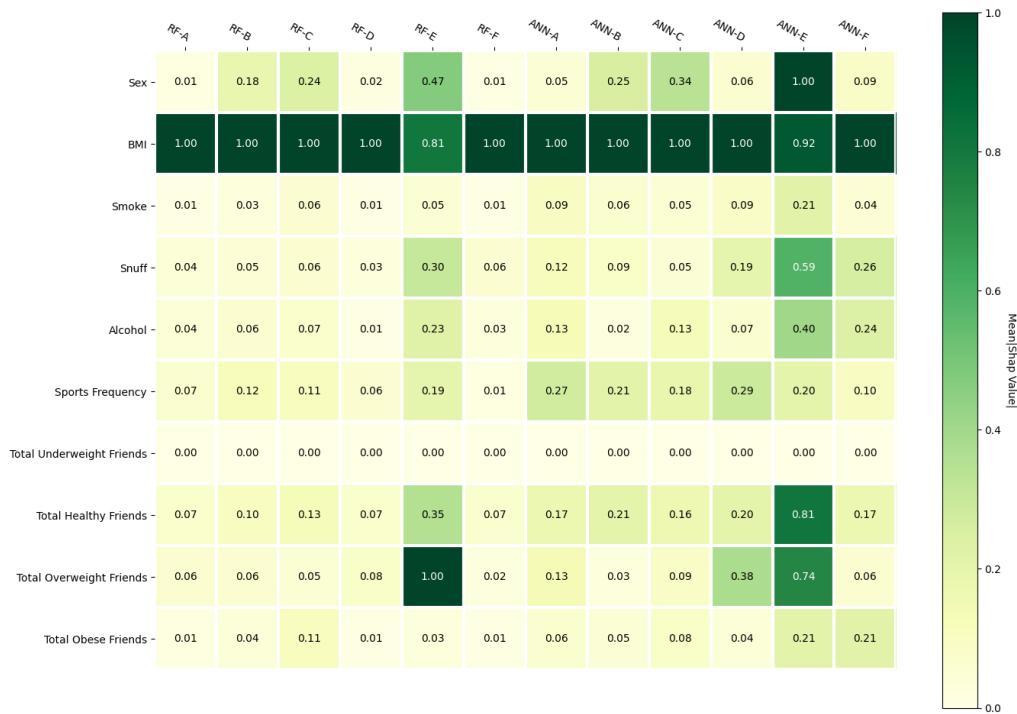


Figure B.3: Heatmap with the normalized mean $|\text{SHAP values}|$ for the model with respect to each dataset (top names) against each variable (left names). Values closer to 1 indicate strong importance in the model.

better) and C also show a slight increase in BMI with sports, this might be due to an increase in muscle mass and reduced total percentage of fat due to exercise. Increasing the BMI is not necessarily bad if the increase is due to a higher weight due to increased muscle mass. So, in the C dataset, this increase seems to be justified due to staying healthy. In the B dataset, it would indicate that sports also help by decreasing total body fat.

All datasets except for D (Overweight or worse got better BMI) showed an increase in BMI with respect to the total number of overweight friends (figure B.8).

The dependency plots show on average a very slight increase ($+0.12$) in the final BMI. The weight of this variable is also low in the RF models. We show in the friendship bias section that obese individuals have low popularity and are not well connected with the healthy group. For all combined 692 valid samples, the average of total overweight friends is 0.28 ± 0.56 . It would appear that increasing connectivity with obese individuals did not have a meaningful effect. However, since the connectivity is low, we can't extrapolate on how the effect would be in a better-connected population.

The E dataset (Healthy BMI that ends up in Overweight or worse) seems to have a completely different weight of SHAP values according to both RF and ANN (figure B.3), and to MDI in RF (figure B.4). Is also the group with the bigger BMI increase ($+3.2 \pm 1.6$). The variable Total Overweight friends get a very high impact on the model despite people not having that many overweight friends. In general, Total Healthy friends decrease the BMI but it shows a spike in BMI for specifically "3" healthy friends (figure B.7). It shows an increase with smoke but a decrease with snuff and alcohol frequency (figure B.9). Females have slightly more risk than males. We tried limiting the dataset from Healthy to Overweight only in case the Healthy to Obese jump was too much of an outlier, but it had barely any effect on the outputs. Analyzing all individuals one by one using the waterfall plots did not show any pattern. Dataset C shows a spike in BMI for "2" total numbers of healthy friends (figure B.7), but the model is more predictable. E is just a subset of the A dataset which does not show any strange particularities either.

Similarly to Result I, here we evaluate the influence of friendship on obesity and it appears that there's also an influence depending on the number of "Healthy", "Overweight", or "Obese" friends. And our machine learning models give a similar explanation compared with previous results.

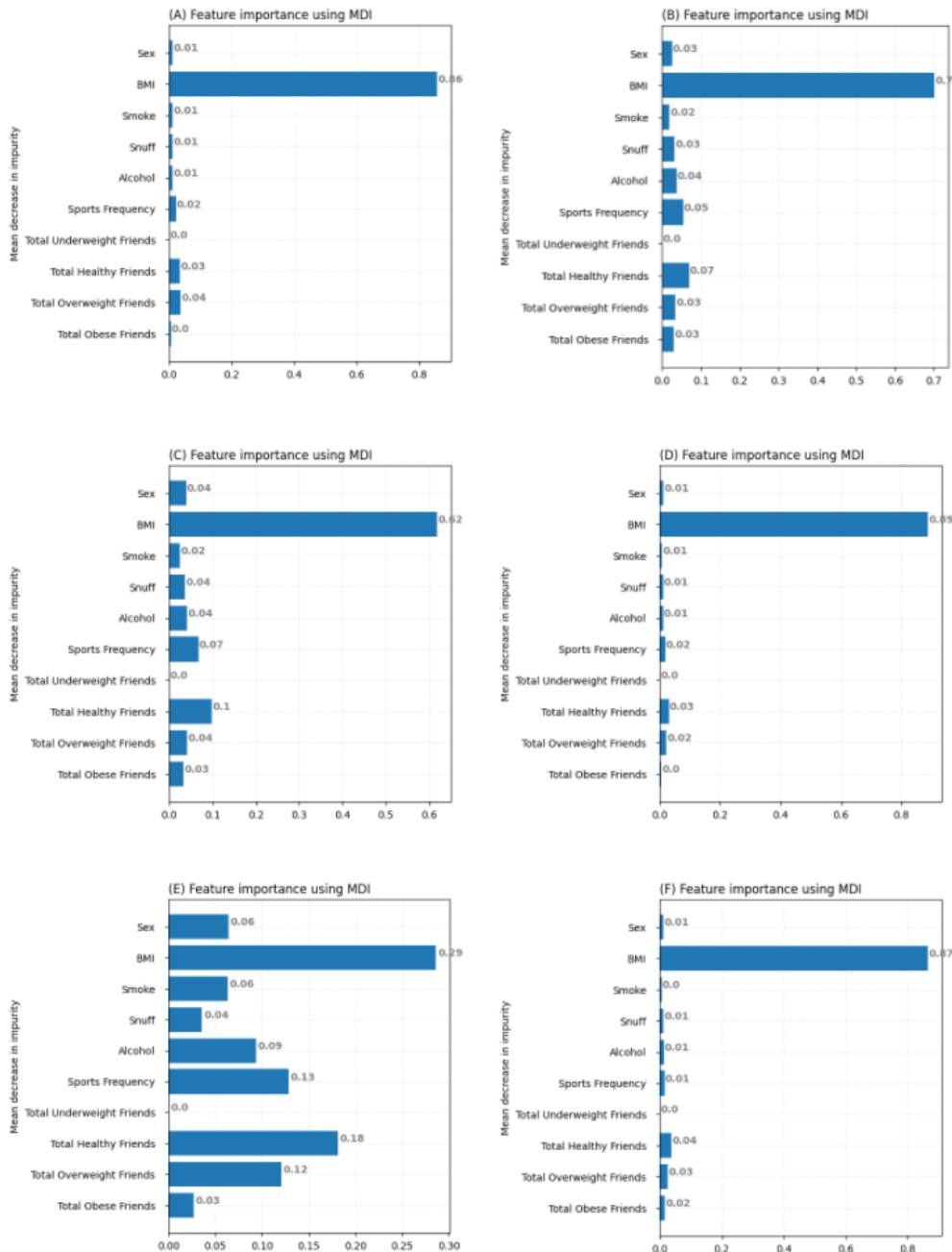


Figure B.4: Barplots using MDI in RF for each of the datasets. The x-axis represents the MDI value, the greater the value the greater the importance. The y-axis represents each of the studied variables.

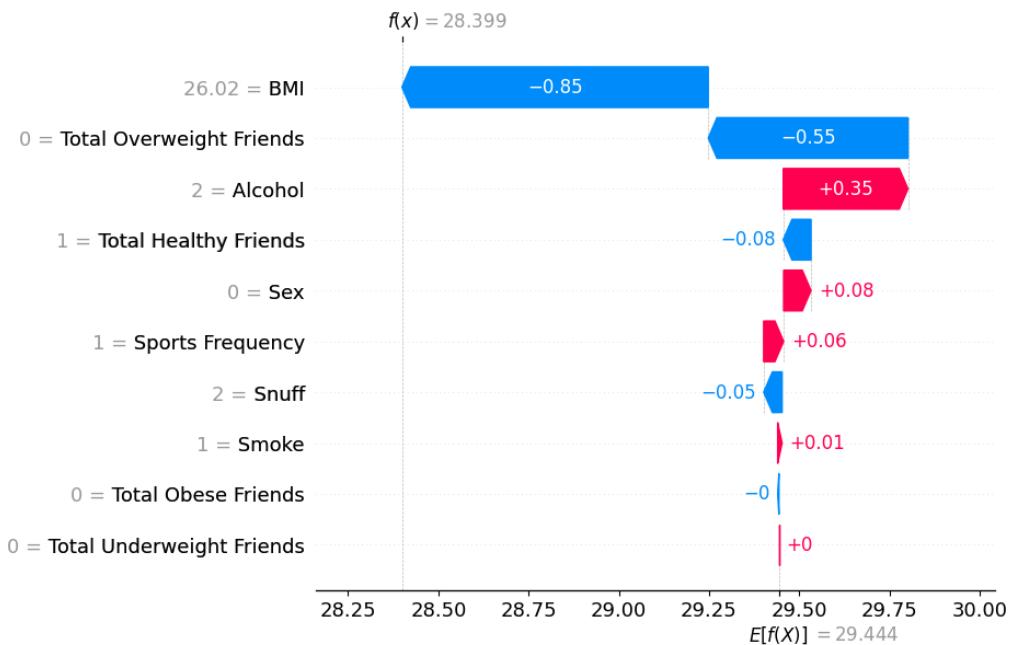


Figure B.5: Waterfall plot with an individual case in the (A) dataset. The first row is the initial BMI in FF1, which was 26.09 (overweight) but is relatively close to 25 and this person is almost classified as healthy. This contributed the most (-0.85) to the final BMI in FF2 displayed on the figure's top “ $f(x) = 28.399$ ”, which increases below average in comparison with the rest of the samples in this dataset; with the average being 29.444 displayed on the bottom of the figure as “ $E[f(X)] = 29.2444$ ”. After that, we see that having 0 overweight friends also contributed (-0.55) in favor of lowering the final BMI. The alcohol consumption of 2, equivalent to “Twice or more per month”, contributed in the opposite direction, increasing the final BMI to +0.35. One healthy friend contributed a little bit to decrease the final BMI. Being a man (sex = 0) contributed a little bit to worse BMI. And so on for the rest of the variables until the change becomes inconsequential.

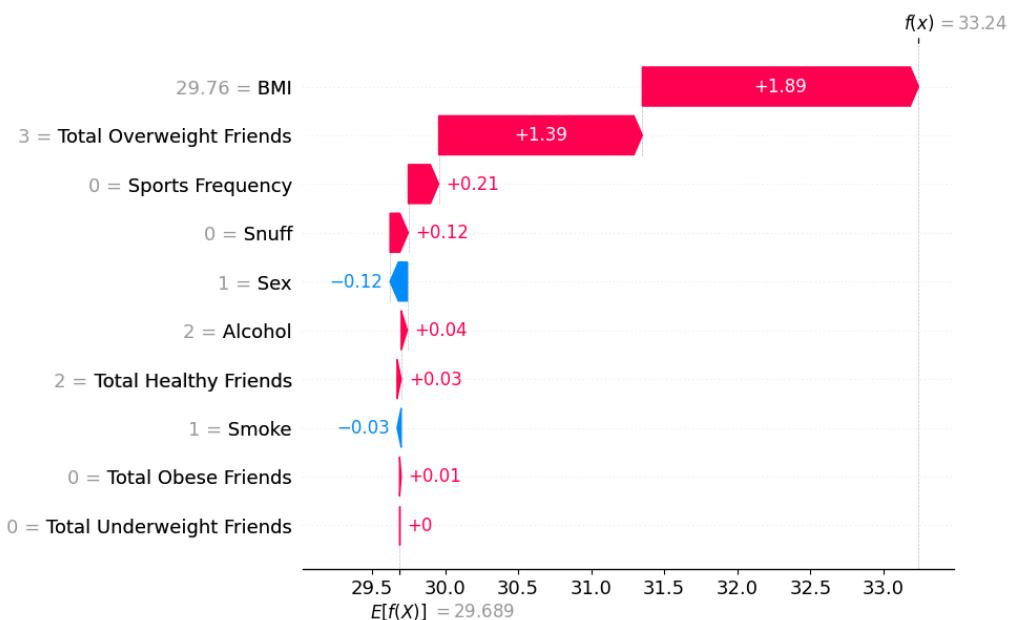


Figure B.6: Waterfall plot with an individual case in the (A) dataset. The initial BMI in FF1 was 29.68, very close to the obese category, which contributed the most to increasing the final FF2 BMI to 33.24, displayed in “ $f(x) = 33.24$ ” at the top of the figure, by quite a lot (+1.89), in comparison with the FF2 average displayed at the bottom of the figure as “ $E[f(x)] = 29.689$ ”. Having 3 overweight friends also contributed to a similar increase of +1.39. Not practicing any sport also contributed to a moderate increase in BMI (+0.21). The rest of the variables' effects' sizes are quite small in comparison to the effects of the first three.

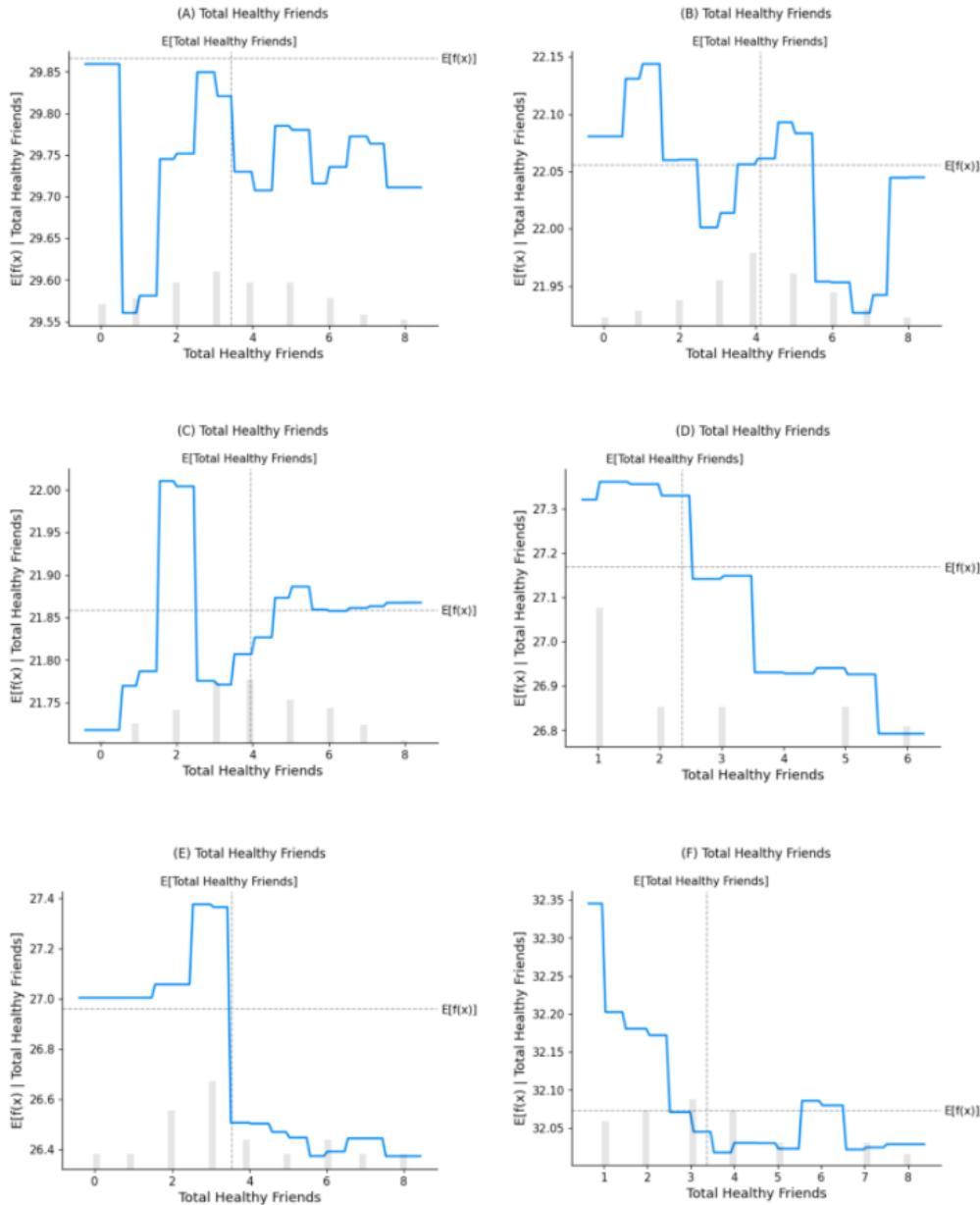


Figure B.7: Partial dependencies plots for datasets A to F in regard to total healthy friends for the RF models. On the X-axis, the total number of healthy friends, with a light grey histogram in the background. On the Y-axis, this variable is expected to modify the model output (BMI) as the variable changes.

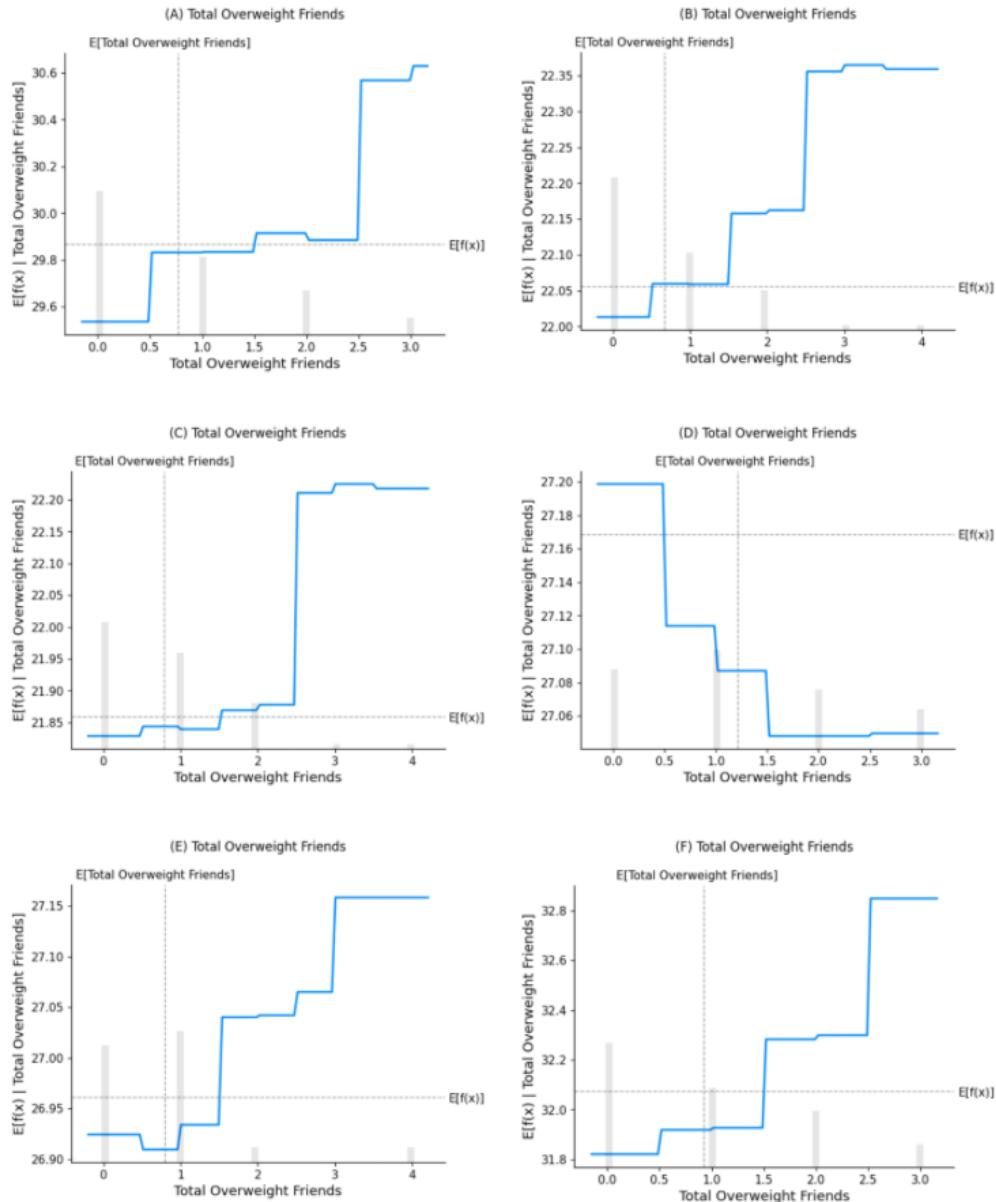


Figure B.8: Partial dependencies plots for datasets A to F in regard to total overweight friends for the RF models. On the X-axis, the total number of healthy friends, with a light grey histogram in the background. On the Y-axis, how this variable is expected to modify the model output (BMI) as the variable changes.

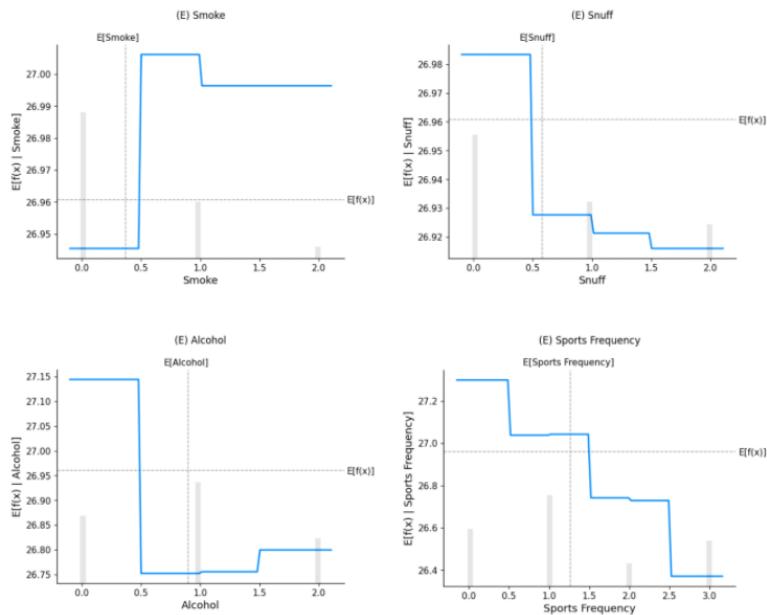


Figure B.9: Partial dependencies plots for datasets E regarding recreational drugs and sport frequency for the RF models. On the X-axis, the total number of healthy friends, with a light grey histogram in the background. On the Y-axis, how this variable is expected to modify the model output (BMI) as the variable changes.

B.4 Result IV

Frequency consumption of medication and social network influence in a general youth population.

In previous studies there has been a concerning trend in self-medication and in particular the overuse of painkillers for non-therapeutic purposes; mostly as a recreational drug within Norway [341, 468, 469] which coincide with worldwide trends and usage [333, 337–339]. With this study, we aim for two objectives. First, to update the data on self-reporting medication with the FF1's 2010 data, and if possible, include up to Fit Futures 3 data done throughout the year 2022. And second to investigate if there's a social influence component as to whether students tend to self-medicate.

We analyzed the frequency of self-reported consumption of medicines and diseases in our population (figure B.10). We can observe that there are plenty of dermatological diseases, however, the usage of dermatological medicines is quite low. In contrast, the amount of pain-related diseases is low, but the consumption of painkillers, or anti-inflammatory medicaments is extremely high in comparison.

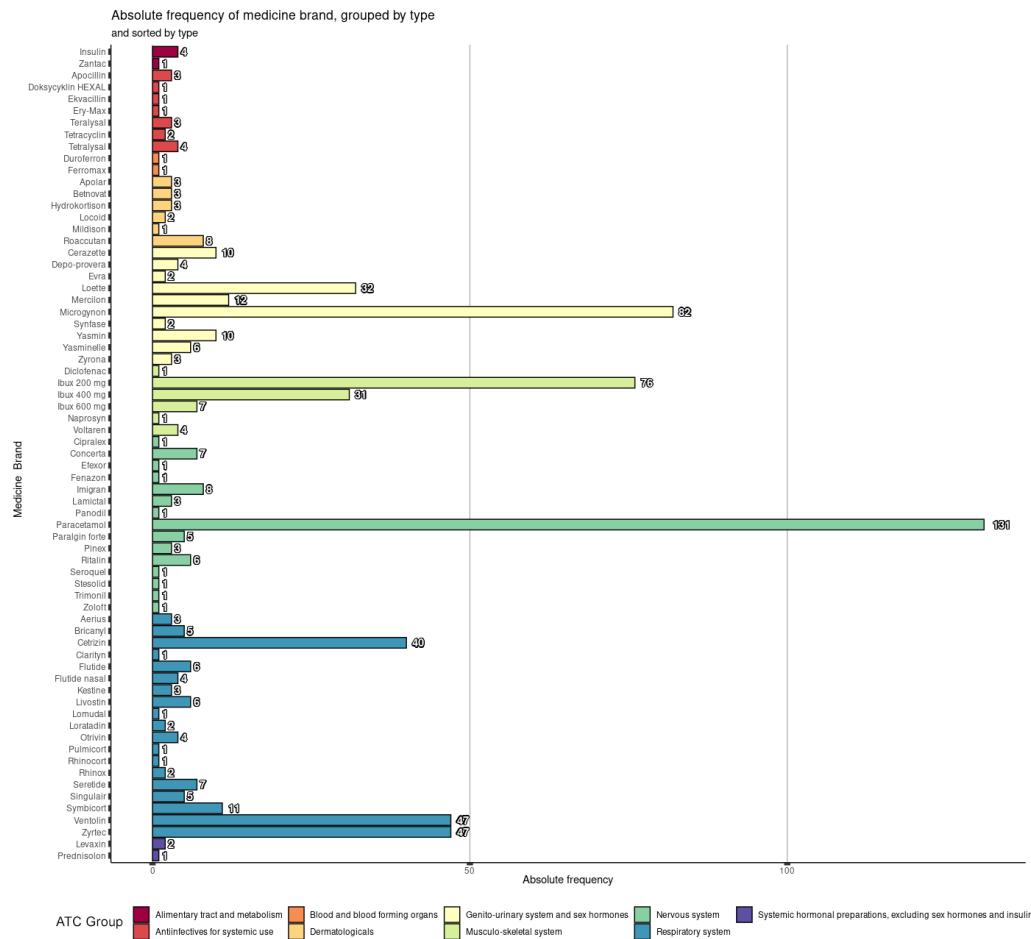


Figure B.10: Absolute frequency of medicine consumption sorted by ATC group. Hormonal contraceptives, anti-inflammatories, painkillers, antiasthmatic, and antihistaminics dominate the frequency of use.

Females seem to have a higher consumption of medicines in general (figure B.12), and also higher disease prevalence (figure B.11). However, the disease types and medicine types don't match; meaning that self-medication is higher among the female population.

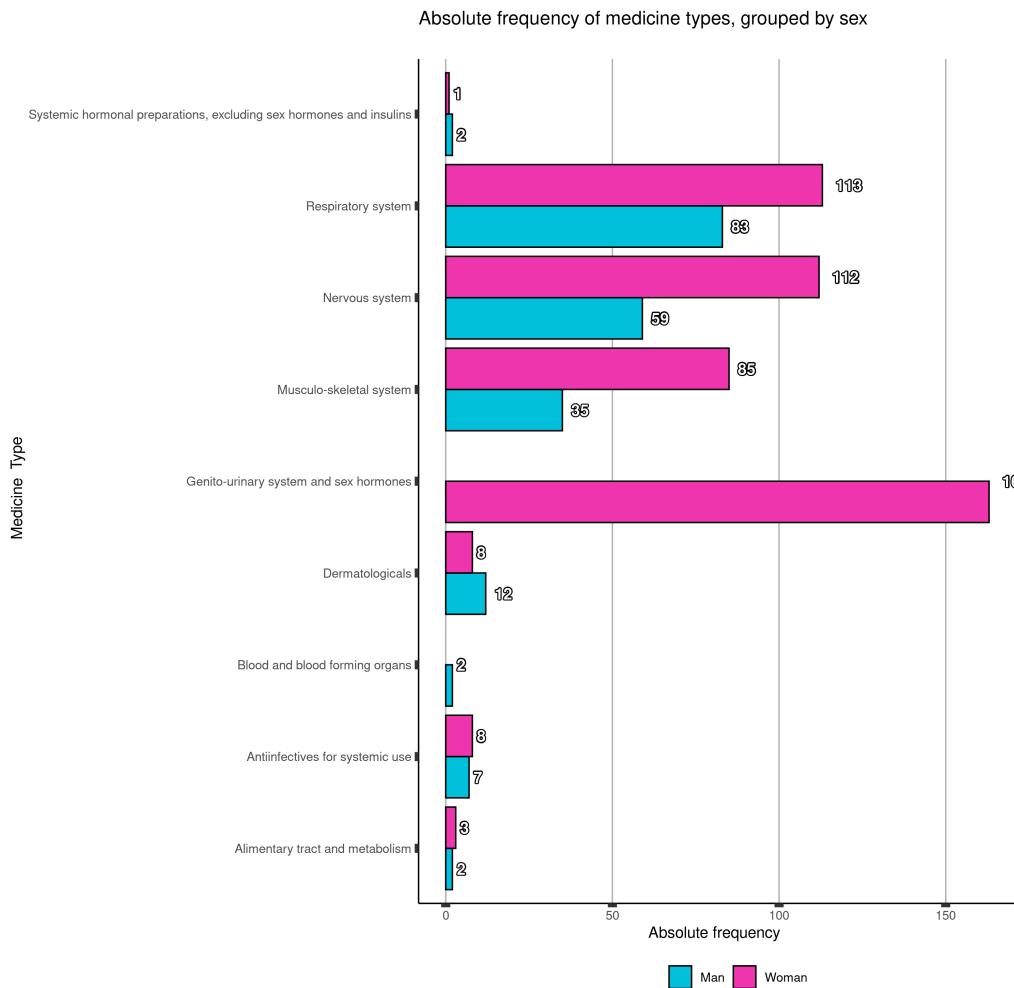


Figure B.11: Absolute frequency of medicine consumption group by ATC group and divided by sex, which seems to indicate a higher consumption among females

Table B.12: X^2 significant bias of each medicine type.

Bias Type	Medicine	Significance	
Highschool	Hormonal (women only)	Brand	< 0.0001 ****
	Type	< 0.0001 ****	
	Yes / No	0.0098 **	
Sex	OTC	Antinflammatory	0.21 ns
		Antihistamine	0.16 ns
		Painkiller	0.0008 ***
	OTC	Antiinflammatory	< 0.0001 ****
		Antihistamine	0.33 ns
		Painkiller	0.0004 ***

Summary of the X^2 tables showing bias for high school and sex. The left column indicates the bias type, either high school or sex. The medicine column indicates which type of medicine is being analyzed, with hormonal contraceptives done only with the female population. The significance columns show the p-value in both number and GP Prism 5.04/d format (asterisks). The table shows bias in high school for hormonal contraceptives and painkillers, and in sex for both anti-inflammatories and painkillers.

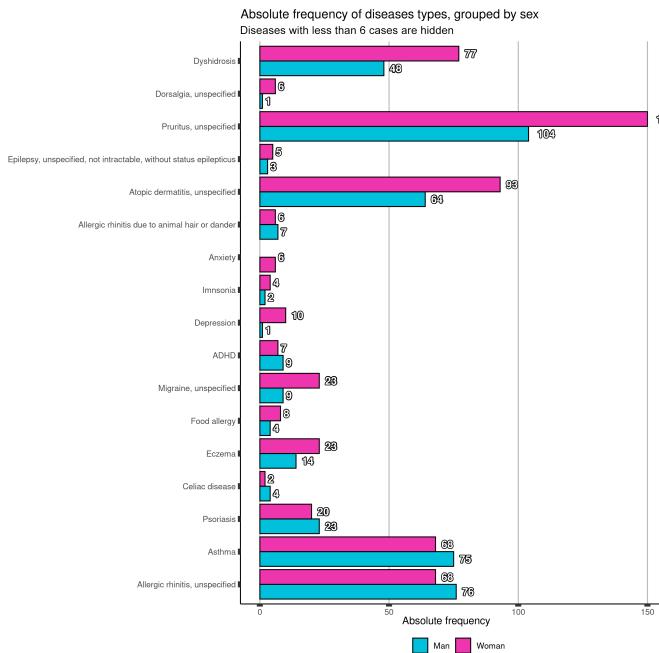


Figure B.12: Absolute frequency of self-reported diseases divided by sex. To avoid visual cluttering, diseases with 5 or fewer instances are not included in this figure. Females also lead in dermatological and psychological conditions, plus migraines. Males seem to have a short lead in respiratory conditions.

We analyzed if there was any bias concerning sex and high school using over-the-counter medicines and also concerning hormonal contraceptives in women. Our results (table B.12) indicate that women sharing the same high school are biased towards hormonal contraceptive usage. Our simulations also show that women who are friends are biased to share the same hormonal contraceptive brand. One of the possible reasons is due to direct recommendations among them and later on asking their family doctor for the same brand. Another one is that women going to the same school tend to live close by and thus share the same family doctor who tends to prescribe the same medication.

The results coincide with teenagers' trend of misusing over-the-counter medication.

Appendix C: Olink panel

C.1 Introduction

The "Olink Inflammatory Panel 96" is a tool for measuring inflammation in biological samples. It analyzes 92 protein biomarkers that play a role in inflammation. The name "96" means that it measures 92 biomarkers, plus another 4 quality control samples. The analysis is conducted using Olink's proprietary Proximity Extension Assay (PEA) technology, which allows for measuring several markers simultaneously. However, the results don't include the absolute concentrations for proteins measured in plasma or serum samples. Instead, you get a Normalized Protein eXpression (NPX) which is Olink's arbitrary unit in the Log2 scale that can't be converted back to absolute concentrations. For practical purposes, it means that 2 or more biomarker levels can't be combined.

The results are given in batches, with each batch specifying a different Limit of Detection (LOD) value for each biomarker level. A given sample that is under LOD means that the machine cannot guarantee that the given value is correct. Usually, these values are censored to the left with the LOD being the minimum.

From here on, we present the list of biomarkers listed here sorted by their acronym in alphabetical order. Each biomarker is presented by its acronym and name in the header. Inside the body of each, the acronym links to Wikipedia if an entry exists, the protein ID links to the Uniprot website <https://www.uniprot.org/>, and "technical" links to the Olink site where further literature references can be found alongside the technical data details regarding sampling. Within this thesis, there's a comprehensive summary of how they affect inflammation processes described in this document.

Since the inflammatory panel for FF1 was performed, some of the proteins have changed their names over the years. The names that are expressed in the data may differ from the names found on the Olink page but ultimately are the same protein. If clarification is needed, both names are included in the heading.

C.2 Biomarkers

C.2.1 ADA - Adenosine Deaminase

[ADA P00813 Technical](#)

ADA is an enzyme that plays an essential role in the metabolism of the nucleotide adenosine. This enzyme is involved in the development and function of the immune system. Adenosine Deaminase is required for the normal function of T-cells, which play a crucial role in regulating the immune response and preventing inflammation. Deficiencies in Adenosine Deaminase levels have been associated with immune system dysfunction, resulting in increased susceptibility to infections and inflammatory disorders.

C.2.2 ART -Artemin

[ARTN Q5T4W7 Technical](#)

Artemin is a protein that belongs to the family of neurotrophic factors. It plays an important role in the development and maintenance of sensory neurons. Studies have shown that artemin can also modulate inflammation by reducing the production of pro-inflammatory cytokines and increasing the secretion of anti-inflammatory cytokines. It has been found to play a crucial role in the regulation of pain and inflammation.

C.2.3 AXIN1 - Axin-1

[AXIN1 O15169 Technical](#)

Axin-1 plays a critical role in regulating the Wnt signaling pathway; this pathway is a complex cell-to-cell communication pathway that plays a crucial role in various biological processes, including embryonic development, tissue regeneration, and stem cell proliferation and is associated with insulin sensitivity. Studies have shown that defects in the Axin-1 protein can lead to chronic inflammation, autoimmune diseases, and cancer.

C.2.4 BDNF - Brain-derived neurotrophic factor

[BDNF P23560 Technical](#)

This protein is no longer in the Olink 96 inflammation panel and has now been

moved into the "Cardiometabolic II" panel instead.

BDNF is a protein that promotes the growth and survival of neurons in the brain. It is a key factor in the development and plasticity of the nervous system. Several studies have reported that BDNF is involved in the regulation of immune function and inflammation. BDNF has been found to suppress inflammation by reducing the expression of pro-inflammatory cytokines and promoting the production of anti-inflammatory cytokines. It has been suggested that BDNF may have therapeutic potential in the treatment of inflammatory and autoimmune disorders.

Is related to IL-6, as both of them are myokines.

C.2.5 BNGF - Beta-nerve growth factor

[BNGF P01138 Technical](#)

This protein is involved in regulating the activation and proliferation of immune cells, including T-cells, B-cells, and natural killer cells. Beta-grown factor also regulates the production of cytokines. Studies have shown that deficiencies in Beta-grown factors can lead to immune system dysfunction, resulting in increased susceptibility to infections and inflammatory disorders.

C.2.6 CASP8 - Caspase-8

[CASP8 Q14790 Technical](#)

There are 12 known caspase proteins in humans and all of them play a role in programmed cell deaths (apoptosis). Its name derives from cysteine-aspartic proteases. Proteases are proteins that break other proteins.

Caspase-8 is one of the initiators of apoptosis and is involved in the activation of the pro-inflammatory cytokines IL-1 β , IL-6, and TNF- α . Is related to many other markers in this list which also promote apoptosis by activating CASP8.

C.2.7 CCL2 / MCP1 - Monocyte chemotactic protein 1

[MCP1 P13500 Technical](#)

MCP-1 is a chemokine involved in the recruitment of macrophages to the site of inflammation. It is produced by injured cells, endothelial cells, and activated immune

cells. CCL2 is related to pathogens of diseases characterized by monocytic infiltrates.

CCL3 - C-C motif chemokine 3

[CCL3 P10147 Technical](#)

Also known as macrophage inflammatory protein-1 α (MIP-1 α). CCL3 plays a role in the recruitment of macrophages and T cells to inflamed tissues and is considered a pro-inflammatory chemokine. It has been implicated in the pathogenesis of rheumatoid arthritis.

C.2.8 CCL4 - C-C motif chemokine 4

[CCL4 P13236 Technical](#)

This protein is involved in the recruitment of T cells and monocytes to inflamed tissues. It is considered a pro-inflammatory chemokine and has been implicated in rheumatoid arthritis and multiple sclerosis. CCL4 has also been shown to attract immune cells to tumor sites.

C.2.9 CCL7 / MCP3 - Monocyte chemotactic protein 3

[MCP3 P80098 Technical](#)

Same as CCL2. This protein can bind to heparin. Heparin is produced by basophils and mast cells and it is an anticoagulant agent. It also helps leukocytes move from the inside of blood vessels to the outside (diapedesis and extravasation).

C.2.10 CCL8 / MCP2 - Monocyte chemotactic protein 2

[MCP2 P80075 Technical](#)

Same as CCL2, but CCL8 is a potent inhibitor of the most common strain of HIV. This protein can also bind to heparin.

C.2.11 CCL11 - Eotaxin

[CCL11 P51671 Technical](#)

Is a chemokine involved in the recruitment of eosinophils. It is overexpressed in

several inflammatory and allergic disorders including asthma and rhinitis. Another two eotaxins exist, CCL24 and CCL26 not included within the analyzed biomarkers.

C.2.12 CCL13 / MCP4 - Monocyte chemotactic protein 4

[MCP4 Q99616 Technical](#)

Same as CCL7. It might be related to monocyte activities in atherosclerosis.

C.2.13 CCL19 - C-C motif chemokine 19

[CCL19 Q99731 Technical](#)

CCL19 is a chemokine that plays a critical role in immune system homeostasis. Is essential for the proper function of dendritic cells and T-cells. Deficiencies in CCL19 expression have been associated with immune system dysfunction, resulting in increased susceptibility to infections and inflammatory disorders.

C.2.14 CCL20 - C-C motif chemokine 20

[CCL20 P78556 Technical](#)

It acts for the chemotaxis of dendritic cells, T-cells, and B-cells. Is a chemotactic factor that attracts lymphocytes and, slightly, neutrophils, but not monocytes. It mainly acts on skin and mucosal surfaces for both homeostatic and inflammatory conditions.

It has been suggested that CCL20 may be involved in the pathogenesis of several inflammatory disorders, including psoriasis and inflammatory bowel disease, and also positively regulates sperm motility

C.2.15 CCL23 - C-C motif chemokine 23

[CCL23 P55773 Technical](#)

CCL23 has been shown to play a role in allergic asthma. It may also have a role in cancer immunity, as it has been found to be elevated in some cancer patients.

C.2.16 CCL25 - C-C motif chemokine 25

[CCL25 O15444 Technical](#)

This protein is involved in the recruitment and activation of T cells in the gut mucosa. It has been shown to play a role in intestinal inflammation and is important for the maintenance of gut immune homeostasis. Potentially involved in T-cell development.

C.2.17 CCL28 - C-C motif chemokine 28

[CCL28 Q9NRJ3 Technical](#)

This chemokine is produced by mucosal tissues to attract CD4+ and CD8+ T-cells, and eosinophils. It is believed to play a role in host defense against infectious agents and maintaining mucosal homeostasis. CCL28 has been implicated in the pathogenesis of inflammatory bowel disease.

C.2.18 CD244 - Natural killer cell receptor 2B4

[CD244 Q9BZW8 Technical](#)

This protein is expressed on the surface of natural killer cells and interacts with ligands on target cells to activate NK cell cytotoxicity. Activation of this receptor can also modulate NK cell cytokine production (IFN γ and TNF α). Studies suggest that altered expression or function of 2B4 may contribute to the development of autoimmune diseases and cancer.

C.2.19 CD5 - T-cell surface glycoprotein CD5

[CD5 P06127 Technical](#)

Is a CD protein that acts as a co-stimulatory or co-inhibitory receptor depending on the context, and its expression is often upregulated in conditions of chronic inflammation. Studies suggest that CD5 may modulate immune responses by regulating signaling through other receptors, such as the T-cell receptor and CD40.

C.2.20 CD6 - T-cell differentiation antigen CD6

[CD6 Q8WWJ7 Technical](#)

Another CD protein is involved in T-cell activation, proliferation, and migration. There are several similar types of CD6, each for different purposes. The expression of CD6 is often upregulated in multiple sclerosis and rheumatoid arthritis.

C.2.21 CD40 / TNFSF5R - Tumor necrosis factor receptor superfamily member 5

[CD40 P25942 Technical](#)

Is a CD protein that activates APCs to promote T-cell activation, survival, and cytokine production, leading to inflammation. Dysregulation of the CD40 pathway has been implicated in Hyper IgM syndrome, which is an immunodeficiency disorder characterized by low IgM, and high IgG, IgA, and IgE. Effectively T cells cannot receive the class switching signaling. Is linked with Alzheimer's, systemic lupus erythematosus, and rheumatoid arthritis.

C.2.22 CDCP1 - CUB domain-containing protein 1

[CDCP1 Q9H5V8 Technical](#)

It has puzzling significance in humans and it might be involved in T-cell chemotaxis, cell adhesion, cell-matrix association, anchorage versus migration, proliferation versus differentiation, leukemia, hematopoietic stem cell subsets, tumor progression, and metastasis.

C.2.23 CSF1 - Macrophage colony-stimulating factor 1

[CSF1 P09603 Technical](#)

This cytokine is produced by various cell types, including macrophages, stem cells, endothelial cells, and T cells. It also regulates the recruitment and activation of these immune cells and promotes the release of proinflammatory chemokines.

Related to the vitamin D chapter, it regulates osteoclast proliferation and differentiation and the regulation of bone resorption.

C.2.24 CST5 - Cystatin D

[CST5 P28325 Technical](#)

Proteases break down proteins into smaller peptide fragments or amino acids. They play a crucial role in many biological processes, including digestion, blood clotting, and protein turnover and recycling.

Cystatin D is an inhibitor of cysteine proteases and is expressed in immune cells. Cystatin D can regulate inflammation by inhibiting the activities of proteases released by immune cells, thereby preventing tissue damage caused by inflammation. Its dysregulation is found in asthma and rheumatoid arthritis.

C.2.25 CX3CL1 - Fractalkine

[CX3CL1 P78423 Technical](#)

Fractalkine is unique among chemokines, as it exists in both a soluble form and a membrane-bound form. The membrane-bound form acts as an adhesion molecule, promoting the adhesion and migration of immune cells to sites of inflammation. The soluble form can also be chemotactic for immune cells, playing a role in the recruitment of immune cells to sites of inflammation.

C.2.26 CXCL1 - C-X-C motif chemokine 1

[CXCL1 P09341 Technical](#)

CXCL1 acts as a potent chemoattractant for neutrophils, which is critical in the defense against pathogens. Dysregulation of CXCL1 expression has been implicated in asthma and chronic obstructive pulmonary disease.

C.2.27 CXCL5 - C-X-C motif chemokine 5

[CXCL5 P42830 Technical](#)

Attracts and activates neutrophils to sites of inflammation. Excessive recruitment of neutrophils in response to CXCL5 may lead to tissue damage and chronic inflammation.

C.2.28 CXCL6 - C-X-C motif chemokine 6

[CXCL6 P80162 Technical](#)

Similar to CXCL5, studies have suggested that CXCL6 may also be involved in the pathogenesis of certain autoimmune diseases due to its effects on immune cell activation and trafficking. Is highly expressed in the presence of bacteria.

C.2.29 CXCL9 - C-X-C motif chemokine 9

[CXCL9 Q07325 Technical](#)

Similar to CXCL6, but it recruits T cells instead.

C.2.30 CXCL10 - C-X-C motif chemokine 10

[CXCL10 P02778 Technical](#)

Pro-inflammatory. Similar to CXCL9. It has anti-angiogenesis properties which translate into anti-tumor effects. IL-12 also increases the production of IFN γ which increases the production of CXCL10.

C.2.31 CSCL11 - C-X-C motif chemokine 11

[CXCL11 O14625 Technical](#)

Same as CXCL10, plus it is also chemotactic for interleukin-activated T-cells, neutrophils, and monocytes. May play a role in skin immune responses.

C.2.32 DNER - Delta and Notch-like epidermal growth factor-related receptor

[DNER Q8NFT8 Technical](#)

Is a transmembrane protein that interacts with the immune system by modulating the differentiation and activation of T cells. Research suggests that DNER is involved in the pathogenesis of multiple sclerosis, an inflammatory disease of the central nervous system, and has the potential to be a therapeutic target for MS.

C.2.33 EIF4EBP1 - Eukaryotic translation initiation factor 4E-binding protein 1

[EIF4EBP1 Q13541 Technical](#)

Is a regulatory protein that modulates protein synthesis in cells. Research shows that 4E-BP1 has anti-inflammatory effects by suppressing the production of pro-inflammatory cytokines. Additionally, 4E-BP1 has been found to play a role in regulating immune cell function and immune responses.

C.2.34 FGF5 - Fibroblast growth factor 5

[FGF5 Q8NF90 Technical](#)

FGF is introduced in section 3.4.2.

FGF5 has been found to inhibit the production of pro-inflammatory cytokines TNF- α and IL-6. Increase the secretion of anti-inflammatory cytokine IL-10. FGF5 is also involved in the regulation of immune cell differentiation, migration, and proliferation.

FGF5 is also required for a normal hair growth cycle. Functions as an inhibitor of hair elongation, by promoting progression from anagen (hair growing) into catagen (hair dying).

C.2.35 FGF19 - Fibroblast growth factor 19

[FGF19 O95750 Technical](#)

Is a hormone-like protein that plays a role in suppressing bile acid synthesis and lipid metabolism. Recent studies suggest that FGF19 may have anti-inflammatory effects by modulating the activity of immune cells and suppressing the production of pro-inflammatory cytokines.

C.2.36 FGF21 - Fibroblast growth factor 21

[FGF21 Q9NSA1 Technical](#)

Has anti-inflammatory properties and regulates immune function. Animal studies suggest that FGF21 can reduce macrophage activation and cytokine production while increasing the production of regulatory T-cells. FGF21 has also been shown to protect against inflammation-induced damage in the liver and heart. It also stimulates glucose uptake in adipocytes.

C.2.37 FGF23 - Fibroblast growth factor 23

[FGF23 Q9GZV9 Technical](#)

Is a hormone-like protein that regulates phosphate, PTH, and vitamin D metabolism. Research suggests that FGF23 may play a role in modulating immune responses by

regulating the differentiation and activation of immune cells. Additionally, FGF23 has been found to be involved in the pathogenesis of rheumatoid arthritis.

C.2.38 FLT3LG - Fms-related tyrosine kinase 3 ligand

[FLT3L P49771 Technical](#)

This protein plays a crucial role in the development of immune cells, including dendritic cells, which are responsible for initiating immune responses. It is known to enhance the production and survival of dendritic cells, promoting their migration to lymph nodes where they can activate T-cells to fight off infections. Additionally, FLT3LG has been shown to reduce inflammation by inhibiting the migration of neutrophils and monocytes, to sites of infection or tissue damage.

C.2.39 GDNF - Glial cell line-derived neurotrophic factor

[GDNF P39905 Technical](#)

While traditionally known for its role in promoting the survival and function of neurons, glial cell line-derived neurotrophic factor has also been found to have anti-inflammatory effects on the immune system. It has been shown to reduce the production of pro-inflammatory cytokines, such as TNF- α and IL-1 β while increasing the production of anti-inflammatory cytokines like IL-10. This can help to decrease the severity of inflammatory responses in chronic diseases.

C.2.40 HGF - Hepatocyte growth factor

[HGF P14210 Technical](#)

It has been shown to promote the expansion and differentiation of regulatory T-cells, particularly in the prevention of autoimmune diseases. It has also been found to have anti-inflammatory effects by reducing the production of pro-inflammatory cytokines and inhibiting the activation of NF- κ B.

C.2.41 IFNG - Interferon gamma

[IFNG P01579 Technical](#)

Described in section 3.4.3.

C.2.42 IL1A - Interleukin-1 alpha[IL1A P01583 Technical](#)

Described in section 3.4.3.

C.2.43 IL2 - Interleukin-2[IL2 P60568 Technical](#)

Described in section 3.4.3.

C.2.44 IL2RB - Interleukin-2 receptor subunit beta[IL2RB P14784 Technical](#)

Described in section 3.4.3.

C.2.45 IL4 - Interleukin-4[IL4 P05112 Technical](#)

Described in section 3.4.3.

C.2.46 IL5 - Interleukin-5[IL5 P05113 Technical](#)

Described in section 3.4.3.

C.2.47 IL6 - Interleukin-6[IL6 P05231 Technical](#)

Described in section 3.4.3.

C.2.48 IL7 - Interleukin-7[IL7 P13232 Technical](#)

Described in section 3.4.3.

C.2.49 IL8 / CXCL8 - Interleukin-8[IL8 P10145 Technical](#)

Described in section 3.4.3.

C.2.50 IL10 - Interleukin-10[IL10 P22301 Technical](#)

Described in section 3.4.3.

C.2.51 IL10RA - Interleukin-10 receptor subunit alpha[IL10RA Q13651 Technical](#)

Described in section 3.4.3.

C.2.52 IL10RB - Interleukin-10 receptor subunit beta[IL10RB Q08334 Technical](#)

Described in section 3.4.3.

C.2.53 IL12B - Interleukin-12 subunit beta[IL12B P29460 Technical](#)

Described in section 3.4.3.

C.2.54 IL13 - Interleukin-13[IL13 P35225 Technical](#)

Described in section 3.4.3.

C.2.55 IL15RA - Interleukin-15 receptor subunit alpha[IL15RA Q13261 Technical](#)

Described in section 3.4.3.

C.2.56 IL17A - Interleukin-17A

[IL17A Q16552 Technical](#)

Described in section 3.4.3.

C.2.57 IL17C - Interleukin-17C

[IL17C Q9P0M4 Technical](#)

Described in section 3.4.3.

C.2.58 IL18 - Interleukin-18

[IL18 Q14116 Technical](#)

Described in section 3.4.3.

C.2.59 IL18R1 - Interleukin-18 receptor 1

[IL18R1 Q13478 Technical](#)

Described in section 3.4.3.

C.2.60 IL20 - Interleukin-20

[IL20 Q9NYY1 Technical](#)

Described in section 3.4.3.

C.2.61 IL20RA - Interleukin-20 receptor subunit alpha

[IL20RA Q9UHF4 Technical](#)

Described in section 3.4.3.

C.2.62 IL22RA1 - Interleukin-22 receptor subunit alpha-1

[IL22RA1 Q8N6P7 Technical](#)

Described in section 3.4.3.

C.2.63 IL24 - Interleukin-24

[IL24 Q13007 Technical](#)

Described in section 3.4.3.

C.2.64 IL33 - Interleukin-33

[IL33 O95760 Technical](#)

Described in section 3.4.3.

C.2.65 KITLG / SCF - Stem cell factor

[SCF P21583 Technical](#)

This cytokine plays a very important role in the formation of blood cells during embryonic development, the formation of spermatozoids, and the formation of melanocytes (cells producing pigmentation).

Besides that, is involved in several signaling pathways, and is believed to act synergistically with interleukins.

C.2.66 LIF - Leukemia inhibitory factor

[LIF P15018 Technical](#)

This cytokine has been found to have both pro-inflammatory and anti-inflammatory effects depending on the context. It has been shown to promote the differentiation and survival of immune cells, including T-cells and B-cells, and to upregulate the production of pro-inflammatory cytokines like IL-6 and TNF- α . However, leukemia inhibitory factor has also been found to have anti-inflammatory effects by reducing the production of reactive oxygen species and inhibiting the NF- κ B pathway.

C.2.67 LIFR - Leukemia inhibitory factor receptor

[LIFR P42702 Technical](#)

This receptor is expressed on various cell types, including immune cells, and is activated by leukemia inhibitory factors. Activation has been found to promote the survival,

proliferation, and differentiation of T-cells and B-cells. Additionally, it has been shown to have anti-inflammatory effects by reducing the production of pro-inflammatory cytokines and inhibiting the NF-κB pathway.

C.2.68 MMP1 - Matrix metalloproteinase-1

[MMP1 P03956 Technical](#)

This is an enzyme that is involved in breaking down extracellular matrix proteins such as collagen. MMP-1 is produced during the process of inflammation as a response to tissue injury or infection. Its activity helps to deconstruct damaged extracellular matrix and promote tissue remodeling.

MMPs are associated with rapid tumor growth and metastasis because of the destruction of the extracellular matrix. This characteristic is shared by other markers in this list.

C.2.69 MMP10 - Matrix metalloproteinase-10

[MMP10 P09238 Technical](#)

Very similar to MMP1, but instead it breaks down another type of structure in the extracellular matrix. It is also linked with cancer development.

C.2.70 NRTN - Neurturin

[NRTN Q99748 Technical](#)

Neurturin belongs to the GDNF family of ligands. Neurturin has been shown to play a role in the development of the nervous system and the maintenance of adult neurons. While its role in inflammation is not fully understood, studies have suggested that Neurturin may have anti-inflammatory properties. It has been shown to modulate the activity of immune cells, reduce inflammation, and promote tissue repair.

C.2.71 NT3 - Neurotrophin-3

[NT3 P20783 Technical](#)

Neurotrophin-3 is a protein that belongs to the neurotrophin family and has mainly been studied for its role in the development and maintenance of the nervous system.

Studies have now shown that it is an anti-inflammatory protein, by inhibiting the production of inflammatory cytokines and chemokines. It also involves the immune system by modulating the activation of T-cells and B-cells. It also has been shown to enhance the survival of immune cells.

C.2.72 OSM - Oncostatin-M

[OSM P13725 Technical](#)

Is a cytokine that belongs to the IL6 superfamily and is similar to LIF. It is not clear whether this cytokine is anti or pro-inflammatory due to limitations in lab technology when the first experiments were conducted. Nowadays is known that it maintains P-selectin activated for a longer time promoting inflammation. But it also has been shown to reduce joint destruction in arthritis.

C.2.73 PD-L1 - Programmed cell death 1 ligand 1

[PD-L1 Q9NZQ7 Technical](#)

PD-L1 suppresses the adaptive immune system during pregnancy and inhibits autoimmune diseases. However, this allows for better proliferation and evasion of cancer cells.

C.2.74 S100A12 / ENRAGE - Protein S100-A12

[ENRAGE P80511 Technical](#)

Is a calcium-binding protein that is involved in inflammation and immune responses. Calcium signaling is involved in several pathways that initiate the activation and proliferation of immune cells. One example is the activation of T cells. When a T cell receptor recognizes a specific antigen, it initiates a signaling cascade that leads to an increase in intracellular calcium levels. This increase in calcium activates the protein calmodulin, which in turn activates the protein phosphatase calcineurin. Calcineurin then triggers the nuclear translocation of the transcription factor NFAT, which regulates the expression of several genes involved in T cell activation and proliferation.

S100s proteins are considered DAMPs. Research suggests that S100A12 is a potential biomarker for chronic inflammatory diseases, such as inflammatory bowel disease,

and may contribute to the pathogenesis of these conditions by activating immune cells and promoting inflammation.

C.2.75 SIRT2 - SIR2-like protein 2

[SIRT2 Q8IXJ6 Technical](#)

Protein is primarily found in the cytosol and plays a crucial role in regulating various cellular functions such as cell proliferation, metabolism, and aging.

It has anti-inflammatory properties. It promotes survival in T-cells. Additionally, SIRT2 can modulate the function of macrophages by regulating their polarization towards an anti-inflammatory phenotype, thereby reducing chronic inflammation.

C.2.76 SLAM / SLAMF1 - Signaling lymphocytic activation molecule

[SLAMF1 Q13291 Technical](#)

SLAM is a protein that is involved in the activation and proliferation of T cells, differentiation of B cells, and the production of Ig. It also promotes the production of anti-inflammatory cytokines.

C.2.77 SULT1A1 - Sulfotransferase 1A1

[ST1A1 P50225 Technical](#)

SULT1A1 is an enzyme that aids the metabolism of both internal and external compounds that are inside the body, such as hormones (internal) or drugs (external). It has anti-inflammatory effects by regulating the production of cytokines and chemokines. Deficiencies in SULT1A1 have been linked with chronic inflammation and autoimmune diseases.

C.2.78 STAMB P - STAM-binding protein

[STAMB P O95630 Technical](#)

STAM-binding protein is an anti-inflammatory protein that regulates cytokine signaling. It binds to the interleukin-2 receptor and inhibits T-cell activation and proliferation. STAM-binding protein also regulates the production of nitric oxide (section 3.4.3).

C.2.79 TGFA - Transforming growth factor alpha

[TGFA P01135 Technical](#)

TGF- α is a type of growth factor (section 3.4.2). It inhibits pro-inflammatory cytokines and promotes anti-inflammatory cytokines. TGF- α has been shown to be protective against autoimmune diseases and chronic inflammation.

C.2.80 LAP TGF- β 1 / TGFB1 - Latency-associated peptide transforming growth factor beta-1

[TGFB1 P01137 Technical](#)

LAP TGF- β 1 binds to TGF- β and keeps it inactive until it is needed. TGF- β is a type of growth factor (section 3.4.2) with anti-inflammation properties which is related to IL-4, IL-10, and NO. TGF- β acts synergistically with TGF- α .

LAP TGF- β 1 has also been shown to be protective against autoimmune diseases and chronic inflammation.

C.2.81 TNFSF1 / LTA / TNFB - Tumor Necrosis Factor Beta

[TNFB P01374 Technical](#)

Explained in section 3.4.3.

C.2.82 TNFSF2 / TNF / TNFA - Tumor necrosis factor

[TNF P01375 Technical](#)

Explained in section 3.4.3

C.2.83 TNFSF9 - Tumor necrosis factor receptor superfamily member 9

[TNFRSF9 Q07011 Technical](#)

TNFSF9 is a protein, member 9 of the TNF superfamily (section 3.4.3). It is also classified as CD137L.

It is expressed on activated T-cells and helps to activate them by binding to their ligand. On T-cells, it promotes the survival of a particular sub-type of memory T cells not explained in this document, and suppresses regulatory T cells; as such it has a pro-inflammatory effect.

Furthermore, TNFRSF9 can also enhance the activity of NK cells.

C.2.84 TNFSF10 / TRAIL - TNF-related apoptosis-inducing lig-and

[TRAIL P50591 Technical](#)

TRAIL is a protein, member 10 of the TNF superfamily (section 3.4.3), and is also known as TNFSF10, and is produced by most tissue cells and causes apoptosis in cells that have become tumorous. It also happens to be CD253.

Caspase-8 (section C.2.6) activates downstream effector caspases, which in time activates TRAIL. Later on, it activates NF- κ B (section 3.4.2) which translates the needed proteins in cell apoptosis and inflammation.

TRAIL is expressed on activated T-cells and NK cells and can bind to its receptor, TRAIL-R, on various cell types; mainly cancer cells. TRAIL-R is also present in T-cells which signal the destruction of activated T-cells and inhibit memory T-cells proliferation.

C.2.85 TNFSF11 / RANKL / TRANCE - TNF-related activation-induced cytokine

[TRANCE O14788 Technical](#)

TNFSF11 is a cytokine, member 11 of the TNF superfamily (section 3.4.3). It is also known as RANKL. It is also known as TRANCE (**T**NF-**R**elated **A**ctivatio**N**-induced **C**ytokin**E**).

The main function of RANKL is to destroy bone tissue and dump calcium into the blood when the levels of calcium are too low. This is further expanded in the Vitamin D chapter (section 3.6.4). RANKL binds to RANK (also known as TNFRSF11A), by doing so it promotes the activation of osteoclasts. But RANKL is secreted by osteoblasts which promote the opposite; calcium fixation into the bones. Let's explain briefly this mechanism before further confusion arises.

Bones typically suffer micro-ruptures by daily activities, but more heavily while doing hard physical activity. Osteoblasts detect these tiny cracks and release RANKL which binds to monocytes. When this happens, several monocytes can fuse together forming an osteoclast. Then, the osteoclast will start drilling the fractured bone tissue forming holes known as Howship's lacunae, and by doing so it will release calcium and phosphate ions into the bloodstream. Eventually, osteoclasts will reach an osteocyte which is randomly scattered inside the bones, which will promote osteoclast apoptosis, and bone tissue destruction will be stopped. Later on, osteoblasts will repair the lacunae, and by doing so some will be trapped inside, which will turn into osteocytes.

RANKL-driven osteoclastogenesis also leads to the release of growth factors, such as previously described FGF23, HGF, and LAP TGF- β 1, which subsequently promote the differentiation and activity of immune cells like T-cells. The bones are also filled with bone marrow, which is filled with hematopoietic stem cells (figure 3.31), which in turn differentiate into, T cells, B cells, basophils, neutrophils, eosinophils, and monocytes. RANKL-mediated activation of immune cells can lead to the production of pro-inflammatory cytokines which may hinder the proper functioning of bone marrow and prevent the differentiation of new lymphocytes.

C.2.86 TNFRSF11A / RANK / TSLP - Thymic stromal lymphopoietin

[TSLP Q969D9 Technical](#)

The RANK function is explained in the previous section. Is the first receptor of RANKL By itself it also enhances the stimulation of T-cells by promoting dendritic cells to do so.

C.2.87 TNFRSF11B / OPG - Osteoprotegerin

[OPG O00300 Technical](#)

TNFRSF11B is the second receptor of RANKL. Its name means "protein (-in) which protects (-protege-) the bones (osteo-)". Osteoprotegerin is primarily produced by both osteoblasts and osteoclasts and plays a vital role in regulating bone resorption which will also be discussed in the vitamin D chapter. Osteoclasts break down bone tissue, and by doing so, several pro-inflammatory cytokines and chemokines are released. Osteoprotegerin inhibits the activity of osteoclasts and acts as an anti-inflammatory protein by

extension

Furthermore, osteoprotegerin reduces the production of macrophages and modulates the differentiation of T-cells and B-cells.

C.2.88 TNFSF12 / TWEAK - TNF-related weak inducer of apoptosis

TWEAK O43508 Technical

TNFSF12 is a cytokine, member 12 of the TNF superfamily (section 3.4.3). It works very similar to TNF (section 3.4.3), but it is found in a higher variety of tissues.

Is also a powerful inducer of NF- κ B related inflammatory responses. TWEAK can also activate T-cells, promote differentiation of regulatory T-cells, and promote differentiation of dendritic cells.

C.2.89 TNFSF14 / LIGHT - Tumor necrosis factor ligand superfamily member 14

TNFSF14 O43557 Technical

TNFSF14 is a cytokine, member 14 of the TNF superfamily (section 3.4.3). It also known as LIGHT (homologous to Lymphotoxin, exhibits Inducible expression and competes with HSV Glycoprotein D for binding to Herpesvirus entry mediator, a receptor expressed on T lymphocytes)

TNFSF14 Is a cytokine predominantly expressed by activated T-cells and can activate TNF- α receptor signaling pathways, leading to inflammation. Is related especially to the TNFRSF9 receptor.

As such it shares functions related to T-cells with TNFRSF9. It also enhances the production of IL-6 and TNF- α . It also activates APCs to further stimulate T-cell responses.

C.2.90 UPA - Urokinase-type plasminogen activator

UPA P00749 Technical

uPA is a serine protease that dissolves blood clots and it is used clinically to break down thrombosis and embolisms. Elevated levels of uPA are highly correlated with tumor malignancy.

Additionally, it can activate macrophages and neutrophils, and stimulate the production of TNF- α and IL-6.

C.2.91 VEGFA - Vascular endothelial growth factor A

[VEGFA P15692 Technical](#)

VEGF-A is a growth factor that is involved in the regulation of angiogenesis. It plays an important role in wound healing and tissue repair. Sadly, similarly to uPA, is associated with tumor growth.

It also activates macrophages and neutrophils and stimulates the production of TNF- α and IL-6. Plus it can promote the differentiation of pro-inflammatory Th17 cells, and inhibit the function of regulatory T cells.