

The Role of AI

Algorithm to bedside symposia

>>>



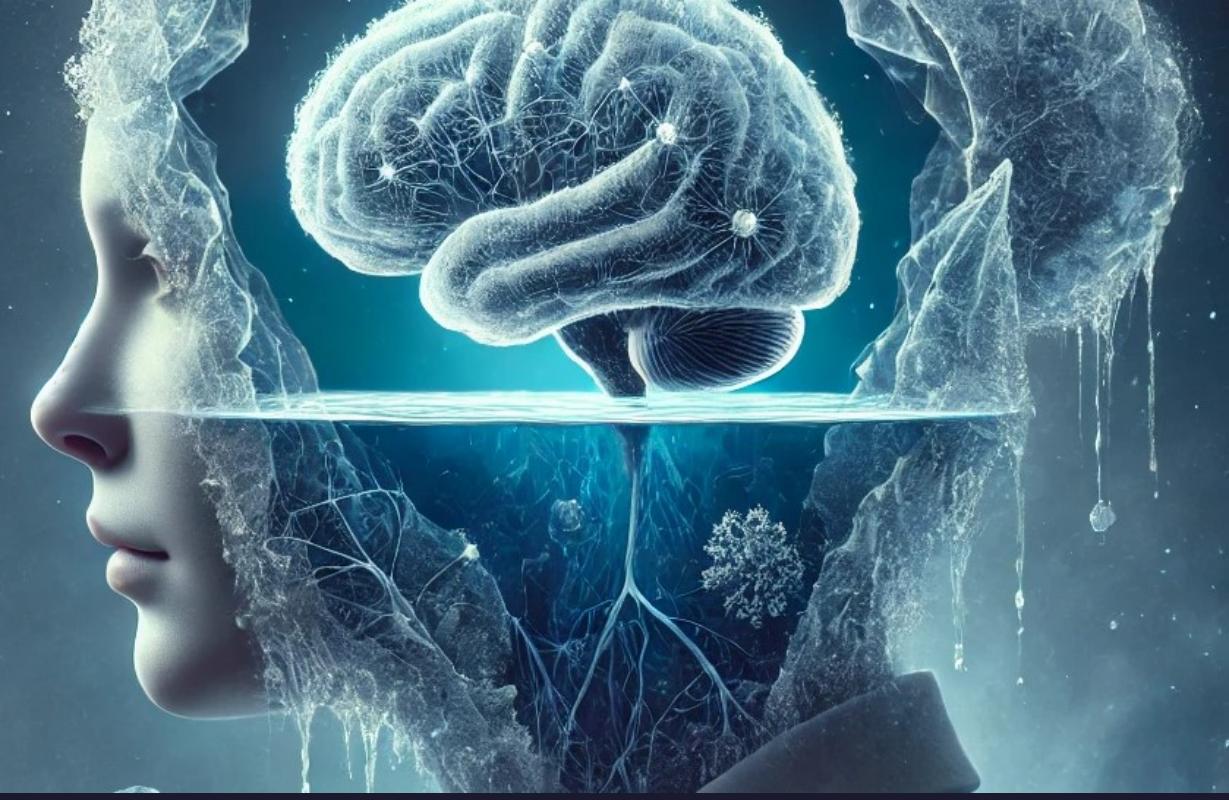
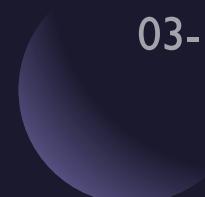


DATA STORIES

From algorithm to bedside



Nermin Ghith, dual B.Sc. in bio/chem.; MPH, PhD
Senior Forsker
Gentofte or Herlev Hospital
03-14 Feb. 2025



Lunch symposia

Omics-Based Clinical and Population Studies (session 1:
Mon. 3 Feb. 2025 - Specific Case Studies)

Contextual Epidemiology of Cardiometabolic
Conditions, Polypharmacy and Multi-morbidity
(session 2: Thurs. 6 Feb. 2025)

Scientific Methods in Biomedical Research (sessions 3
and 4: Mon and Thurs. 10 and 13 Feb. 2025)

- Omics in biomedical research
- Evidence pyramid and analytics
- Global Burden of Disease, Health Analytics, and Access to Therapeutics
- Economic Evaluations





AI AND MIRCO to Macro LEVEL **HEALTH** A NEW FRONTIER

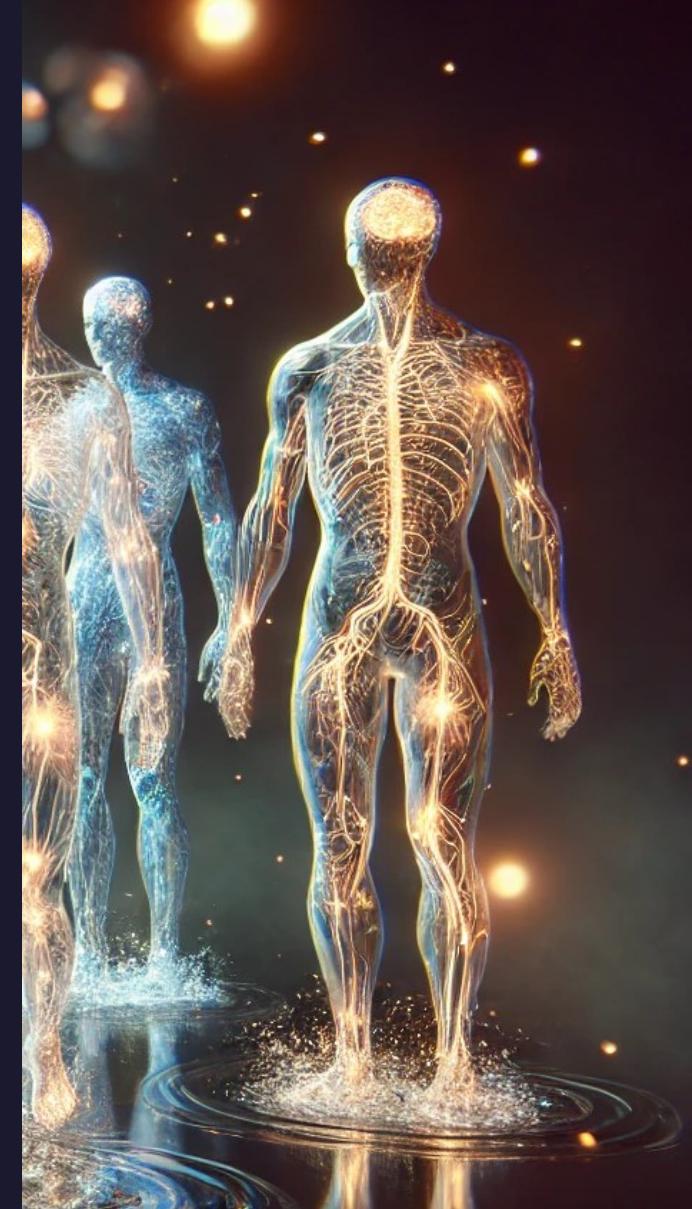


Session 4

- Omics in biomedical research - Clinical trials and risk scores
- Evidence pyramid and analytics - Clinical trials and risk scores
- Global Burden of Disease, Health Analytics, and Access to Therapeutics
- Economic Evaluations

From algorithm to bedside

Risk Scores





RSSs in clinical research

- RCTS
- Clinical practice
- (Metagenomic) population studies
- Quality of patient care and survival
- Polygenic RSSs in prevention

Patient risk scores (RS)

- This is the **RS** for developing critical illness or health outcome (e.g., mortality, readmission, negative prognosis).
- Applied to a **patient** or a population
- Population risk scores:
 - For **screening** purposes.
 - In primary care, with the help of software.
- With any **design**
- Assign the proper level of **severity** and **complexity** of the illness



RS in RCTs

Eligibility for treatments

Estimate the risk of death among critically ill patients

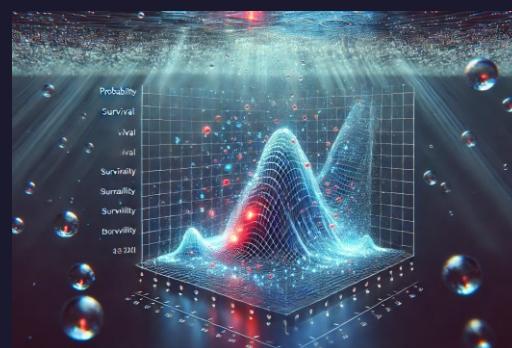
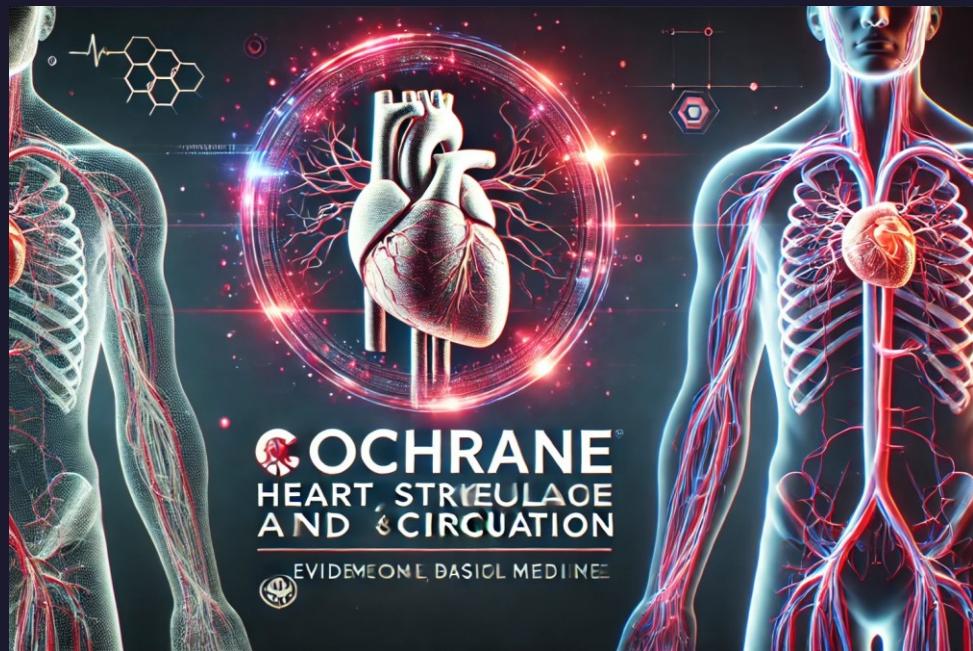
APACHE II scores: double-blind, placebo-controlled, multicenter trial, to assess the efficacy of recombinant human activated protein C, or drotrecogin alfa (activated) (DrotAA), for the treatment of patients with septic shock. (DOI: 10.1056/NEJMoa1202290).

Risk stratification (severity of illness)

- APACHE II scores: Placebo-controlled, multicenter RCT to assess the Efficacy and Safety of Recombinant Human Activated Protein C for Severe Sepsis. (DOI: 10.1056/NEJM200103083441001)
 - APACHE II: the score on the Acute Physiology and Chronic Health Evaluation II.

In subgroup analyses:

- **Framingham risk score:** RCT to assess Rosuvastatin (20 mg daily or placebo) use to Prevent Vascular Events in Men and Women with Elevated C-reactive protein. (<https://doi.org/10.1056/nejmoa0807646>).



Framingham Heart Study



FRAMINGHAM RISK SCORE (FRS) Estimation of 10-year Cardiovascular Disease (CVD) Risk

Step 1¹

In the "points" column enter the appropriate value according to the patient's age, HDL-C, total cholesterol, systolic blood pressure, and if they smoke or have diabetes. Calculate the total points.

Risk Factor	Risk Points		Points	
	Men	Women		
Age				
30-34	0	0		
35-39	2	2		
40-44	5	4		
45-49	7	5		
50-54	8	7		
55-59	10	8		
60-64	11	9		
65-69	12	10		
70-74	14	11		
75+	15	12		
HDL-C (mmol/L)				
>1.6	-2	-2		
1.3-1.6	-1	-1		
1.2-1.29	0	0		
0.9-1.19	1	1		
<0.9	2	2		
Total Cholesterol				
<4.1	0	0		
4.1-5.19	1	1		
5.2-6.19	2	3		
6.2-7.2	3	4		
>7.2	4	5		
Systolic Blood Pressure (mmHg)	Not Treated	Treated		
<120	-2	0	-3	-1
120-129	0	2	0	2
130-139	1	3	1	3
140-149	2	4	2	5
150-159	2	4	4	6
160+	3	5	5	7
Smoker	Yes	4	3	
	No	0	0	
Diabetes	Yes	statin-indicated condition		
	No	0	0	
Total Points				

¹ Adapted from: D'Agostino RB et al. General cardiovascular risk profile for use in primary care. The Framingham Heart Study. Circ 2008;117:743-53.

² Adapted from: Anderson T et al. 2012 Update of the Canadian Cardiovascular Society guidelines for the diagnosis and treatment of dyslipidaemia for the prevention of cardiovascular disease in the adult. Can J Cardiol. 2013;29(2):151-167.

³ apoB: apolipoprotein B stat, CVD: cardiovascular disease, FRS: Framingham Risk Score, HDL-C: high-density lipoprotein cholesterol, LDL-C: low-density lipoprotein cholesterol.

⁴ Statins indicated as lipid therapy.

⁵ Consider LDL-C < 1.8 mmol/L for subjects with acute coronary syndrome (ACS) within past 3 months

Provided courtesy of  Canadian Cardiovascular Society
Leadership. Knowledge. Community.

Date: _____
Patient's Name: _____

Step 2¹
Using the total points from Step 1, determine the 10-year CVD risk* (%).

Total Points	Men	Women
-3 or less	<1	<1
-2	1.1	1.0
-1	1.4	1.2
0	1.6	1.5
1	1.9	1.7
2	2.3	2.0
3	2.8	2.4
4	3.3	2.8
5	3.9	3.3
6	4.7	3.9
7	5.6	4.5
8	6.7	5.3
9	7.9	6.3
10	9.4	7.3
11	11.2	8.6
12	13.3	10.0
13	15.6	11.7
14	18.4	13.7
15	21.6	15.9
16	25.3	18.5
17	29.4	21.5
18	>30	24.8
19	>30	27.5
20	>30	>30
21+	>30	>30

* Double cardiovascular disease risk percentage for individuals between the ages of 30 and 59 without diabetes if the presence of a positive history of premature cardiovascular disease is present in a first-degree relative before 55 years of age for men and before 65 years of age for women. This is known as the modified Framingham Risk Score.

Step 3^{2,3}
Using 10-year CVD risk from Step 2, determine if patient is Low, Moderate or High risk.* Indicate Lipid and/or Apo B targets

Risk Level*	Initiate Treatment If:	Primary Target (LDL-C)	Alternate Target
High FRS ≥20%	• Consider treatment in all (Strong, High)	• ≤2 mmol/L or ≥50% decrease in LDL-C (Strong, Moderate)	• Apo B ≤0.8 g/L or Non-HDL-C ≤2.6 mmol/L (Strong, High)
Intermediate FRS 10-19%	• LDL-C ≥3.5 mmol/L (Strong, Moderate) • For LDL-C <3.5 mmol/L consider it: • Apo B ≥1.2 g/L • OR Non-HDL-C ≥4.3 mmol/L (Strong, Moderate) • Men ≥50 and women ≥60 with 1 risk factor: low HDL-C, impaired fasting glucose, high waist circumference, smoker, hypertension	• ≤2 mmol/L or ≥50% decrease in LDL-C (Strong, Moderate)	• Apo B ≤0.8 g/L or Non-HDL-C ≤2.6 mmol/L (Strong, Moderate)
Low FRS <10%	• statins generally not indicated	• statins generally not indicated	• statins generally not indicated
Statin-indicated conditions**	• Clinical atherosclerosis* • Abdominal aortic aneurysm • Diabetes mellitus • Age ≥ 40 years • 15-year duration for age ≥ 30 years (DM1) Microvascular disease • Chronic kidney disease (age ≥ 50 years) • eGFR <60 mL/min/1.73 m ² or ACR > 3 mg/mmol	• statins generally not indicated	or Apo B: _____

Framingham Heart Study

Risk Functions



Risk Functions →

Atrial Fibrillation 10 year Risk

Heart Failure in Atrial Fibrillation (10-year risk)

Cardiovascular Disease (10-year risk)

Cardiovascular Disease (30-year risk)

Congestive Heart Failure

Hard Coronary Heart Disease (10-year risk)

Coronary Heart Disease (10-year risk)

Riesgo de Enfermedades Coronarias Cardiacas en 10 Años

Recurrent Coronary Heart Disease

Coronary Heart Disease (2-year risk) – First Event

Diabetes

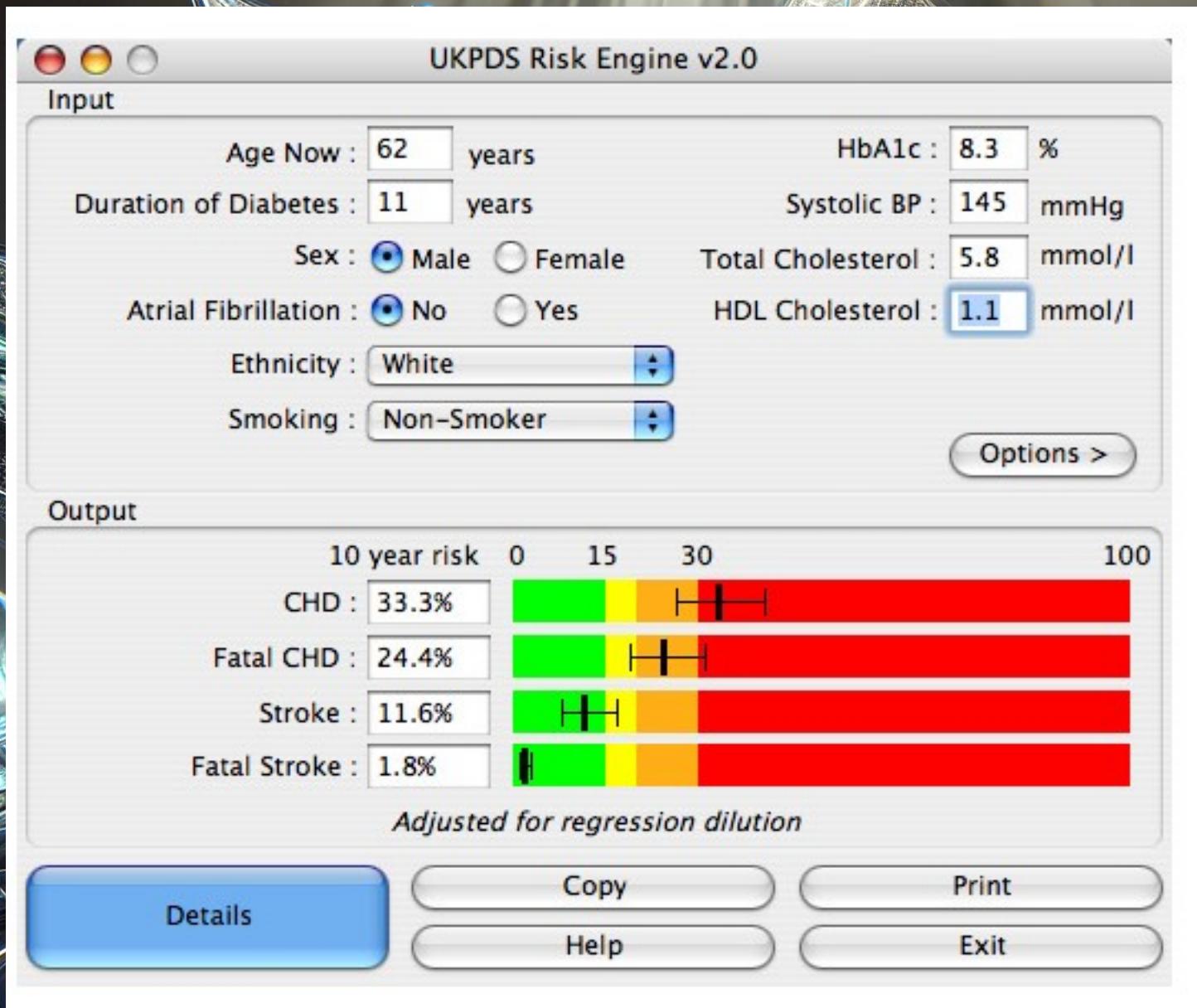
Fatty Liver Disease Risk Function

Hypertension

Intermittent Claudication

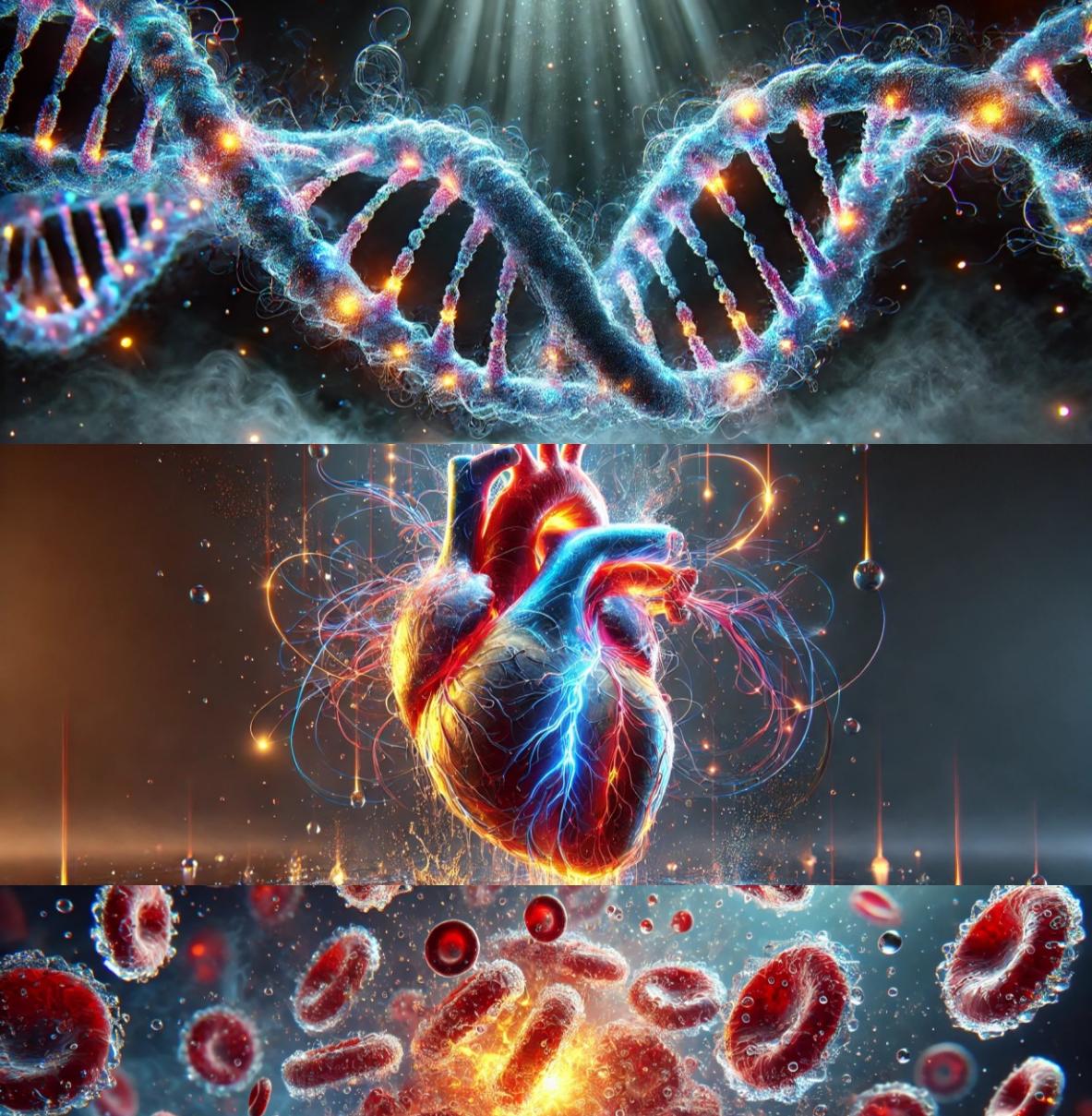
Stroke

Type 2 diabetes-specific risk calculator

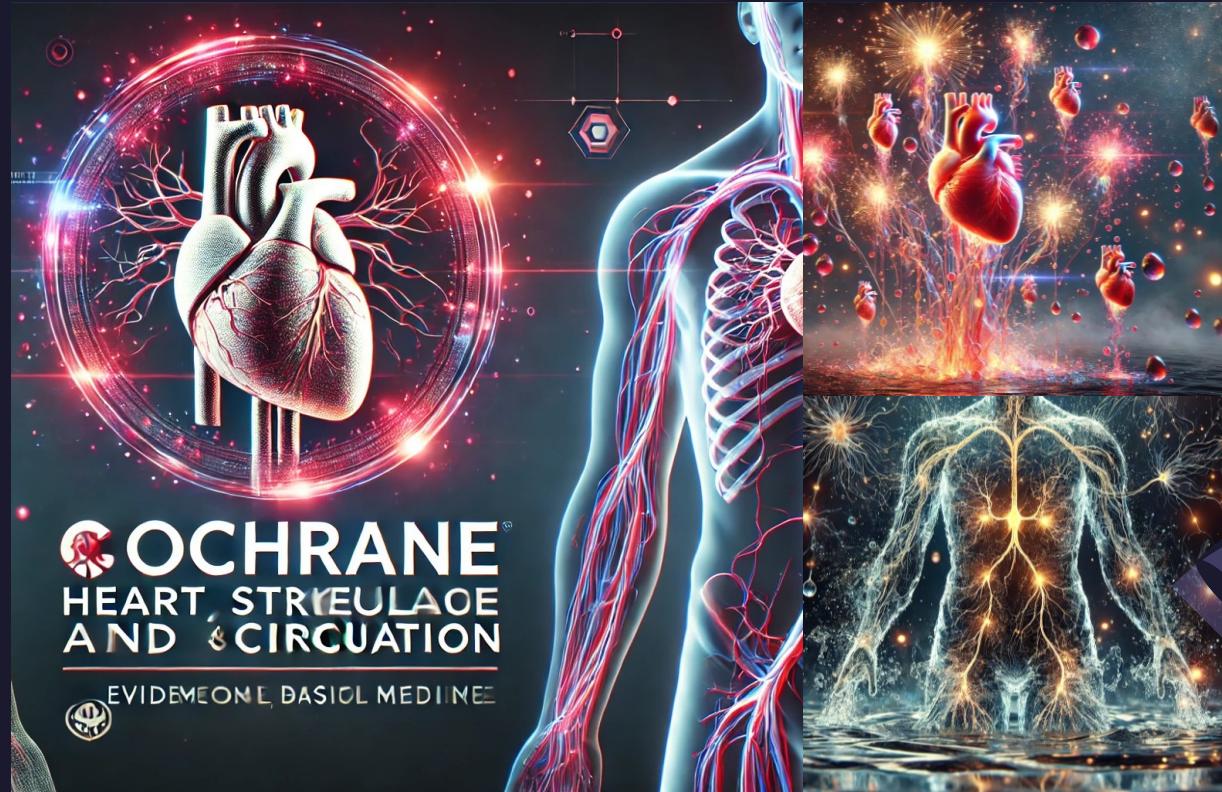


RS in clinical practice/studies

- Charlson Comorbidity Index
- **Johns Hopkins Aggregated Diagnosis Groups (ADGs)** - Predict Mortality in a General Adult **Population**
- **Elixhauser Comorbidity Score** - to predict 30-day, **in-hospital**, and 1-year mortality in older adults
- QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease
- The **HAS-BLED** score – risk of bleeding in patients with atrial fibrillation.



Charlson Comorbidity Index



- **Charlson Comorbidity Index (CCI), 1987**, is an assessment tool used to predict **long-term mortality** across various **clinical** populations (e.g., medical, surgical, ICU, trauma, and cancer patients).
- It can also predict **in-hospital mortality**, though it may not perform as well for certain groups like **ICU** or trauma patients when compared to tools designed specifically for those situations.
- The CCI is useful in providing a **valid assessment** of a patient's unique clinical condition.
- It helps **differentiate** major diagnostic and prognostic differences among patients with the same medical diagnosis.

- **weighted scoring system** - predict the **risk of death within one year** after hospitalization for patients with certain **comorbid/multi conditions**.
- The index includes **19 conditions**, each assigned a **weight from 1 to 6** based on its impact on **1-year mortality**



Conditions	Assigned weights for each condition
MI	1
CHF	1
Peripheral vascular disease	1
Cerebrovascular disease	1
Dementia	1
Chronic pulmonary disease	1
Connective tissue disease	1
Ulcer disease	1
Mild liver disease	1
Diabetes	1
Hemiplegia	2
Moderate or severe renal disease	2
Diabetes with end organ damage	2
Any tumor without metastasis	2
Leukemia	2
Lymphoma	2
Moderate or severe liver disease	3
Metastatic solid tumor	6
AIDS	6

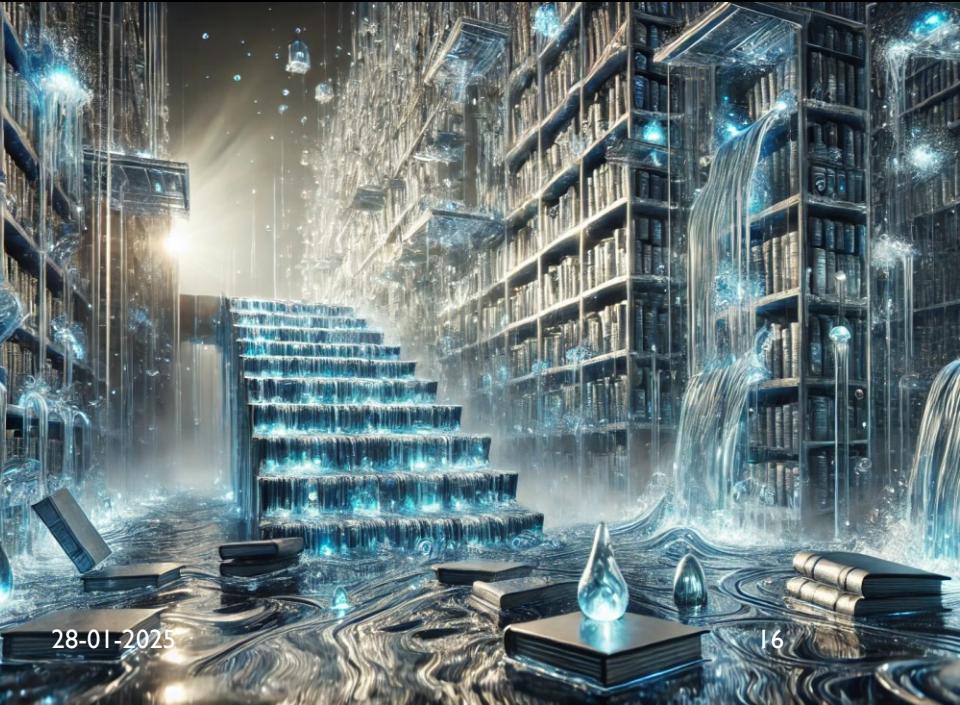
MI, myocardial infarction; CHF, congestive heart failure.



Re-weighted versions of the CCI

Conditions	Original CCI	Quan's "Updated Charlson"	Ghali CABG	Klabunde prostate	Klabunde breast	Chaundhury inpatients	Schneeweiss elderly	Elixhauser hospitalized patients
MI	1		1	1	3		1	
Peripheral vascular disease	1		2	1	1		1	1
Cerebrovascular disease	1		1	1			1	
CHF	1	2	4	2	2		1	1
Peptic ulcer disease	1							1
Diabetes	1			1	2		1	1
Chronic pulmonary disease	1	1		2	2		2	1
Connective tissue disease	1	1		3	3			1
Dementia	1	2					3	
Mild liver disease	1	2		2	3		2	1
Diabetes with end organ damage	2	1		2	2		2	1
Moderate or severe renal disease	2	4	3	2	4	2	3	1
Hemiplegia	2	2		1	3		1	
Any tumor without metastasis	2	2				2	2	1
Leukemia	2	2					2	
Lymphoma	2	2					2	1
Moderate or severe liver disease	3	4				2	3	
AIDS	6	4				3	6	1
Metastatic solid tumor	6	6					6	1
<i>Additional Elixhauser conditions</i>								
Arrhythmia								1
Coagulation								1
Fluid electrolyte								1
Alcohol dependence								1
Pulmonary circulation disorder								1
Hypertension								1
Hypothyroidism								1
Obesity								1
Weight loss								1
Anemia (blood loss)								1
Deficiency anemia								1
Drug abuse								1
Psychosis								1
Depression								1
Neurologic or degenerative disorder								1
Paralysis								1
Citations	39,225	2,849	296	1,644	1,644	309	1,569	7,361

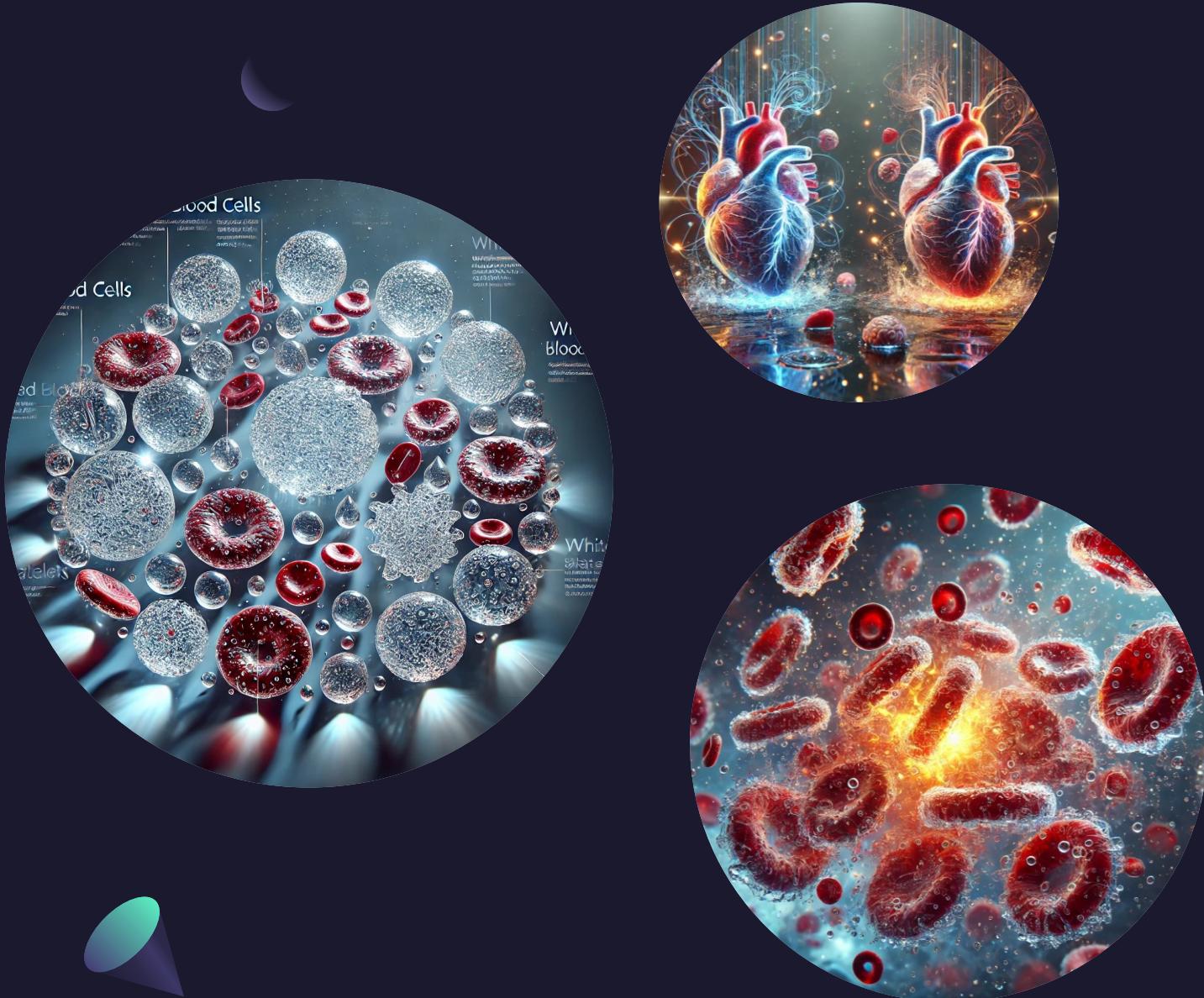
MI, myocardial infarction; CHF, congestive heart failure.



HAS-BLED score

Bleeding Risk in Anticoagulated Patients With Atrial Fibrillation

- The HAS-BLED score, 2010, is a real-world cohort of 3450 anticoagulated patients with AF.
- (hypertension, abnormal renal/liver function, stroke, bleeding history or predisposition, labile international normalized ratio [INR], elderly [age ≥ 65 years], drugs/alcohol concomitantly)
- 3 risk stratifications: a score of 0 indicates low risk, 1–2 indicates moderate risk, and ≥ 3 indicates high risk. Recently.
- Overlapping risk scores: HEMORR₂HAGES, ATRIA, CHADS₂, or CHA₂DS₂-VASc scores



RESEARCH ARTICLE

Short Term Survival after Admission for Heart Failure in Sweden: Applying Multilevel Analyses of Discriminatory Accuracy to Evaluate Institutional Performance

Nermin Ghith^{1,2*}, Philippe Wagner^{1,3}, Anne Frølich², Juan Merlo¹

1 Unit for Social Epidemiology, Faculty of Medicine, Lund University, Malmö, Sweden, **2** Research Unit of Chronic Conditions, Bispebjerg University Hospital, Copenhagen, Denmark, **3** Centre for Clinical Research, Västmanland, Uppsala University, Västerås, Sweden

* nermin_m.ghith@med.lu.se



OPEN ACCESS

Citation: Ghith N, Wagner P, Frølich A, Merlo J (2016) Short Term Survival after Admission for Heart Failure in Sweden: Applying Multilevel Analyses of Discriminatory Accuracy to Evaluate Institutional Performance. PLoS ONE 11(2): e0148187. doi:10.1371/journal.pone.0148187

Editor: Pablo Garcia de Frutos, IIBB-CSIC-IDIBAPS, SPAIN

Received: September 10, 2015

Accepted: January 14, 2016

Published: February 3, 2016

Copyright: © 2016 Ghith et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The database we analyzed is not publicly available for ethical and data safety reasons according to the Swedish National Board of Health and Welfare. However, the same dataset can be constructed by request to the Swedish National Board of Health and Welfare after approval of the research project by an Ethical Committee and by the data safety committee at the Swedish National Board of Health and Welfare.

Funding: This work was supported by the Swedish Research Council (VR) [#2013-2484, Juan Merlo] and by grants from the University of Copenhagen and

Abstract

Background

Hospital performance is frequently evaluated by analyzing differences between hospital averages in some quality indicators. The results are often expressed as quality charts of hospital variance (e.g., league tables, funnel plots). However, those analyses seldom consider patients heterogeneity around averages, which is of fundamental relevance for a correct evaluation. Therefore, we apply an innovative methodology based on measures of components of variance and discriminatory accuracy to analyze 30-day mortality after hospital discharge with a diagnosis of Heart Failure (HF) in Sweden.

Methods

We analyzed 36,943 patients aged 45–80 treated in 565 wards at 71 hospitals during 2007–2009. We applied single and multilevel logistic regression analyses to calculate the odds ratios and the area under the receiver-operating characteristic (AUC). We evaluated general hospital and ward effects by quantifying the intra-class correlation coefficient (ICC) and the increment in the AUC obtained by adding random effects in a multilevel regression analysis (MLRA). Finally, the Odds Ratios (ORs) for specific ward and hospital characteristics were interpreted jointly with the proportional change in variance (PCV) and the proportion of ORs in the opposite direction (POOR).

Findings

Overall, the average 30-day mortality was 9%. Using only patient information on age and previous hospitalizations for different diseases we obtained an AUC = 0.727. This value was almost unchanged when adding sex, country of birth as well as hospitals and wards levels. Average mortality was higher in small wards and municipal hospitals but the POOR values were 15% and 16% respectively.

RS - quality of patient care and survival



<https://doi.org/10.1371/journal.pone.0148187>





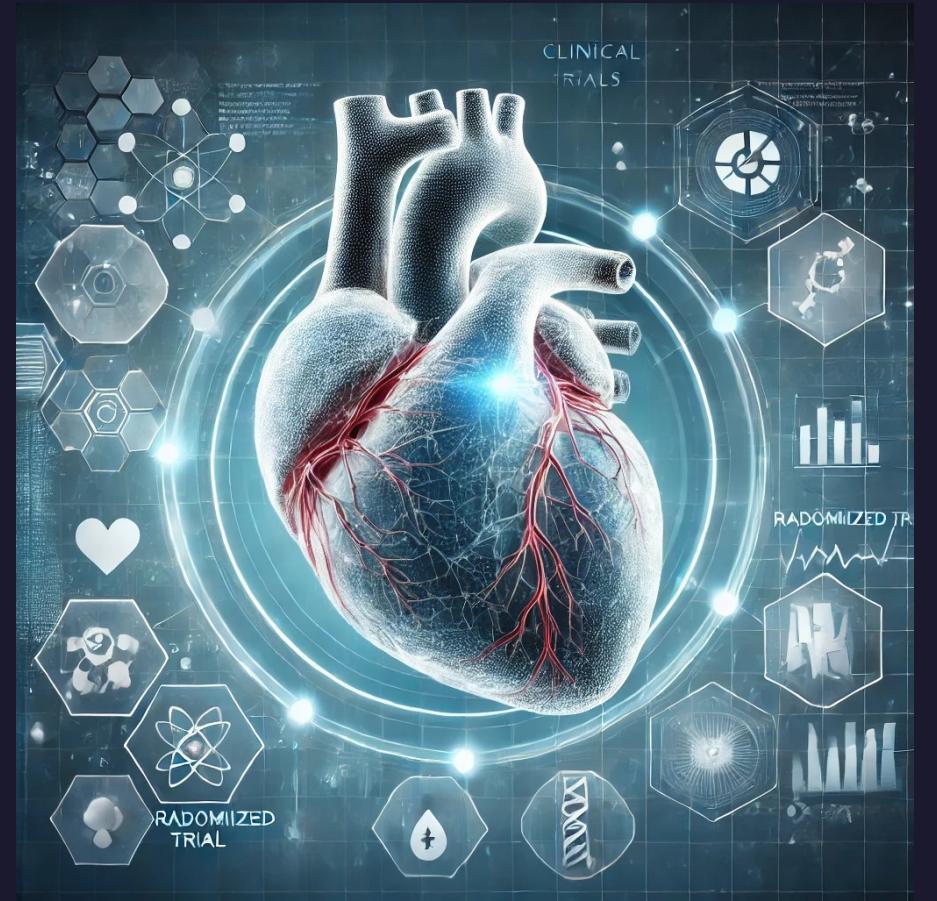
RS - quality of patient care and survival

- Study (patient population) specific risk score – Ghith et al. (2016):
<https://doi.org/10.1371/journal.pone.0148187>
- knowing the patient age and previous diseases (i.e., **the RS**) was enough to achieve an $AUC = 0.727$ which was only slightly better when knowing the sex and ethnic status of the patient (+0.002 units).
- However, this patient-level information was only marginally improved by knowing the **hospital or the ward** where the patient was treated.

Study I – Methods and findings

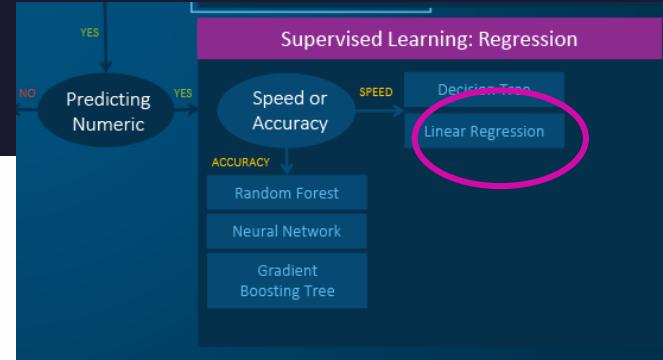
Data: using a unique personal identification number, the research database included linked data from

- the Swedish patient register with
 - the Cause of Death Register and
 - the Longitudinal Integrated Database for Health Insurance and Labour Studies (LISA).
-
- **Population:** covered 36,943 patients aged 45-80 who were treated in 565 departments at 71 hospitals.
 - with a discharge diagnosis of heart failure (International Classification of Diseases code I50) ; (2007-2009)

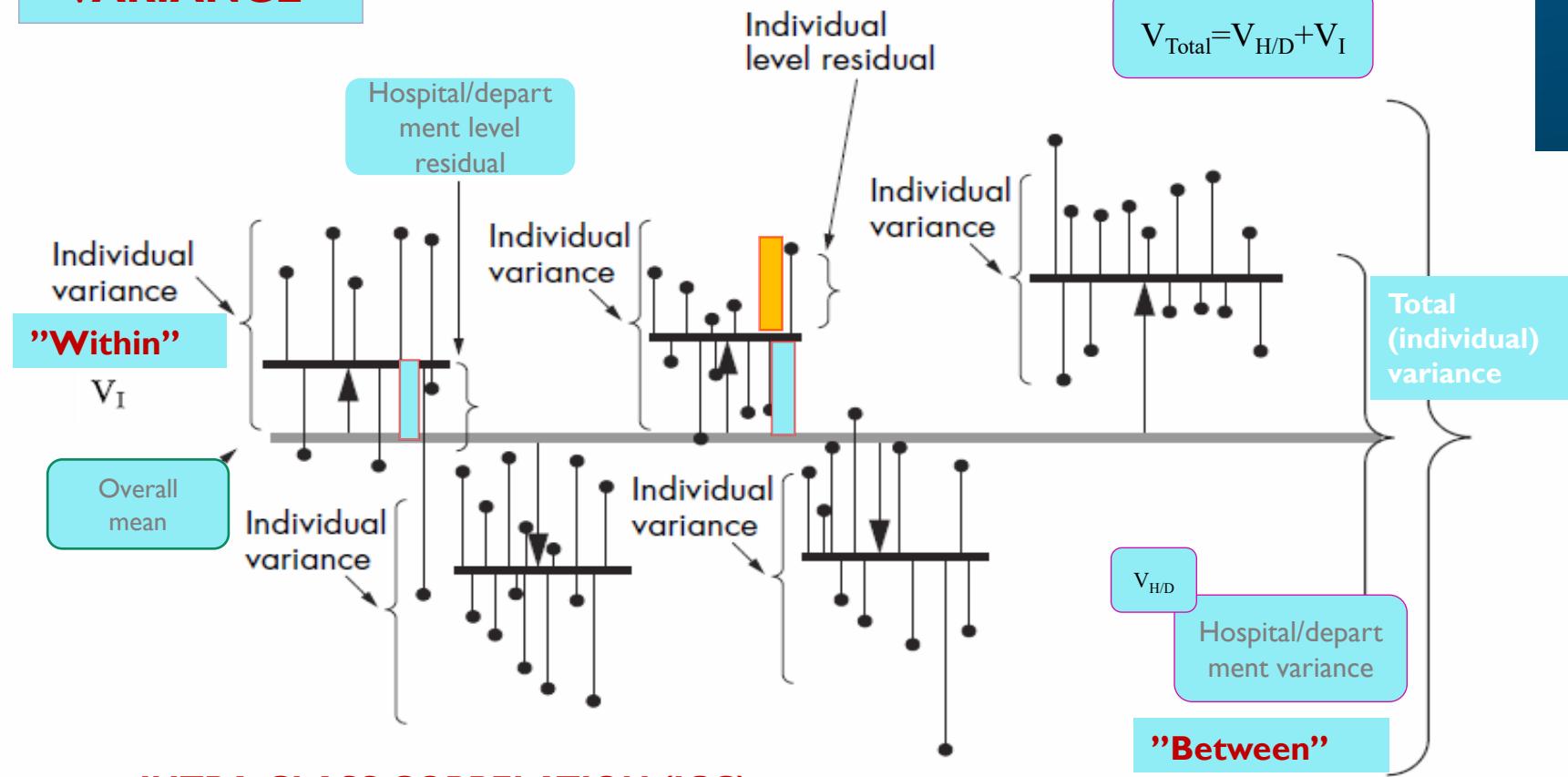


VISUAL STATISTICS...

Multilevel analysis



“VARIANCE”



INTR-CLASS CORRELATION (ICC)

$$VPC/ICC = \frac{\text{Between } V_{\text{Hospital/department}}}{\text{between } V_{\text{Hospital/department}} + \text{within } V_I}$$

Total individual variance = within + between

Ghith and Merlo et al (2016): <https://doi.org/10.1371/journal.pone.0148187>



From my PhD Thesis: Variance measures are expressed as median values and 95% credible intervals (CIs)

	Single level Logistic Regression models		Three Level Multilevel Logistic Regression Models	
Measures of the GCE	Model 1 (RS) <small>36 variables on patient case-mix</small>	Model 2 (RS+gender/age)	Model 3 (RS+gender/age)	Model 4 (RS+ gender/age+ patient volume, hospital classification)
AUC	0.727 (0.719-0.736)	0.729 (0.721-0.738)	0.753 (0.745-0.762)	0.752 (0.743-0.760)
ICC _H			0.04% 0.004 (0.000-0.014)	0.002 (0.000-0.015)
ICC _D			5.3% 0.053 (0.035-0.081)	0.036 (0.020-0.064)

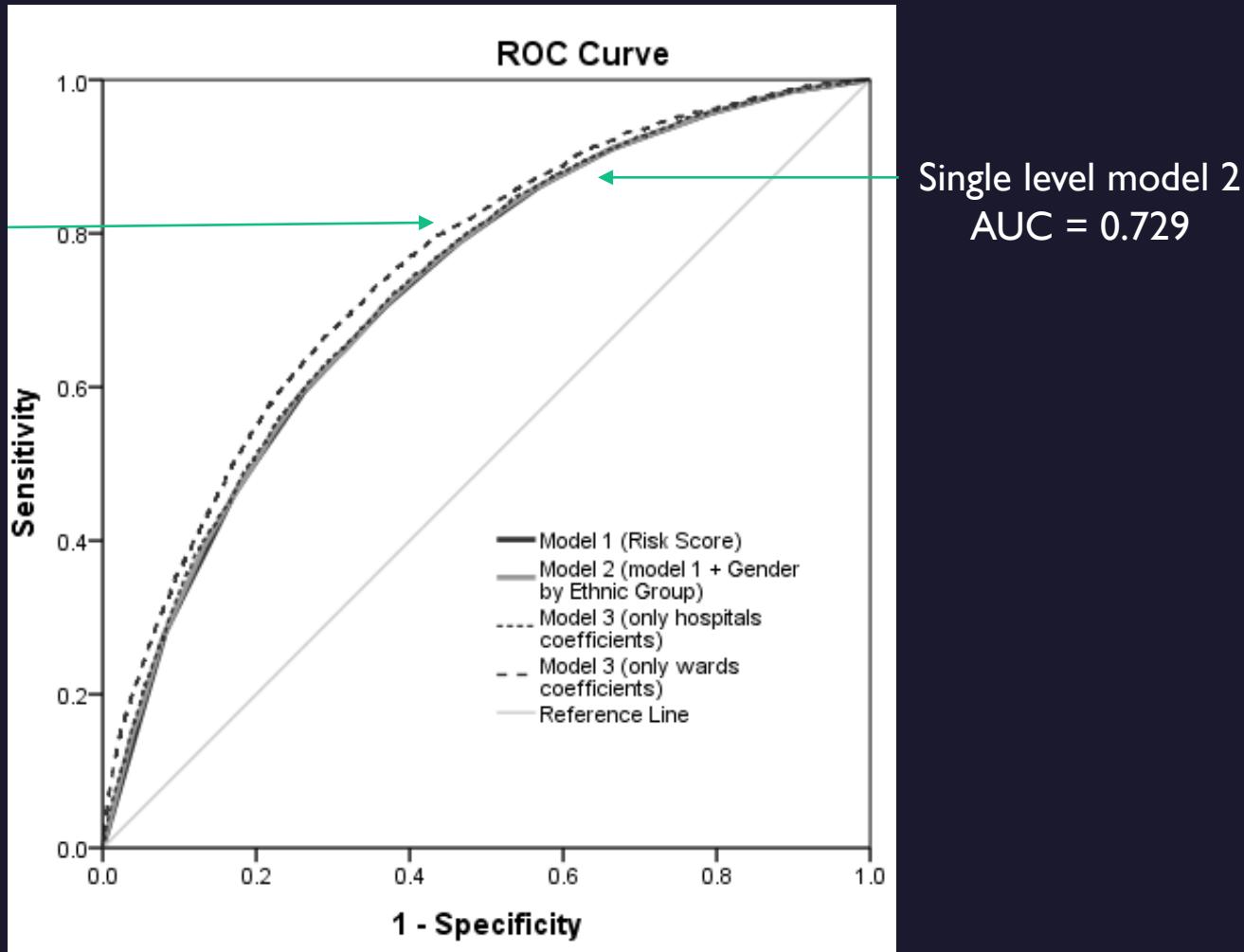
Average 30-day mortality was 9 %



From my PhD Thesis: Variance measures are expressed as median values and 95% credible intervals (CIs)

Multilevel model 3
AUC ≈ 0.753

the increase in the AUC in model 3 compared to model 2 represents the *ceiling* of the hospital's general contextual effects.



Since measures on the GCE were so small, information on the specific contextual effects (patient volume and hospital classification) are not relevant.



RS - quality of patient care and survival

Using the **ICC** in combination with the average **30-day mortality**.

- If the overall survival is **good** (i.e., **high**) and the **ICC low**, all hospitals were doing homogeneously well.
- If the overall survival is **low** and the **ICC low**, all hospitals were doing homogeneously **bad**.
- focus on intervention should be targeting all hospitals in the country.
- However, the higher the **ICC**, the more appropriate to act on the hospitals with the **worst** survival.

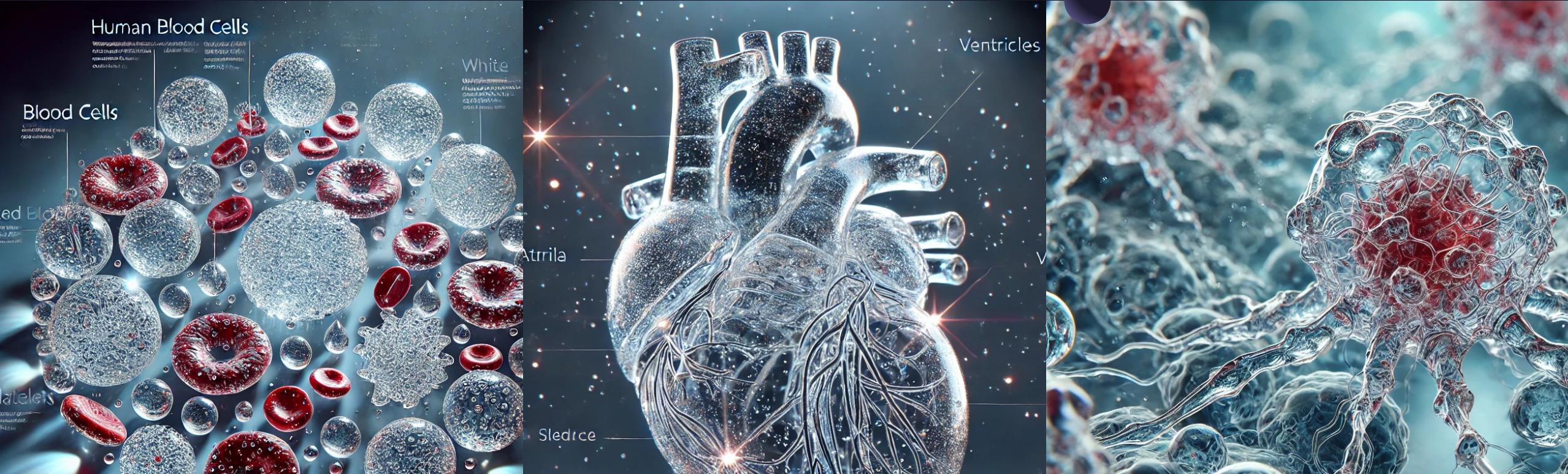
Using the **AUC** in combination with the average **30-day mortality**.

- If the hospitals and wards do **not add discriminatory accuracy**, it would not be effective to focus on certain hospitals.
- If we aim to **decrease mortality**, we should focus on improving care in the patients with **higher mortality** (using **RS**) while it does not matter which is the hospital of treatment.



RSs in Metagenomic population studies

Ghith et al (2022).





Sustainability Development Goals and biomarkers of Antimicrobial Resistance: differential abundance analyses

Nermin Ghith, Hannah-Marie Martiny, Patrick Munk, Amalia Bogri, and Frank Møller Aarestrup

Research Group of Genomic Epidemiology, Technical University of Denmark (DTU)



Research group of Genomic Epidemiology
Technical University of Denmark
Mobile: +4528190249
research@dtu.dk
<https://www.ncbi.nlm.nih.gov/nuccore/50059311/>

Presentation date: 6/12/2022
PRESENTATION (Publication) NUMBER 2581

Revised abstract

Background:

Inequality in global health outcomes such as antimicrobial resistance (AMR) is shaped by macro determinants of health within and between countries.

Therefore, this study aims to assess the potential relationship between the country's development level - progress towards achieving the sustainability development goals (SDG) and the burden of AMR.

The study identifies resistance genes, phenotypes (antibiotics), classes of antimicrobial resistance and bacterial taxa that are differentially abundant between the sampling sites within countries, and compliance with SDGs goals that might explain variation in the global burden of AMR between countries.



Methods:

The Global Sewage Surveillance project consortium collects untreated sewage samples to assess the global burden of AMR. Samples were sequenced to obtain data on the microbial compositions within site-country covering information on resistance genes and bacterial taxa. Samples were collected between 2016 and 2018 and came from 233 cities (102 countries).

A combination of models is developed to describe and characterize the data on AMR in association with SDG indices and goals. A number of SDG indices were added to the analyses to evaluate the study aim. Data on SDGs indices were extracted from the World Bank and the Global Burden of Disease data portals.

Results:

The microbial communities across sites differ from one another, which are attributed to site-country development levels (e.g., measured with SDG index and goals). The study identified a number of differentially abundant taxa at different levels of the bacterial taxonomy as well as differentially abundant genes, phenotypes and classes of AMR.

Conclusions:

Further interventions focusing on improving adherence to SDG goals have the potential for elevating inequality in the global burden of AMR. Yet, a complementary assessment for additional sampling locations is required to get data that are more granular on the global and local burden of AMR and to determine the most appropriate course of action.

Methods

Metagenomics samples were sequenced to obtain information on 19 classes of antimicrobial resistance, 65 phenotypes, 562 resistance genes and bacterial taxa.

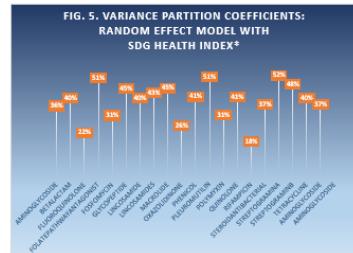
To assess the study aim, the findings were triangulated using the Wilcoxon-Mann-Whitney test (WMW), negative binomial models (NB), random effect models, compositional data analysis, and analysis of compositions of microbiomes with bias correction (ANCOM-BC or ABC).

Applying random effect models, the contextual effect of the city was quantified using the variance partition coefficient (VPC) by adding the random effects of the city in the analyses.

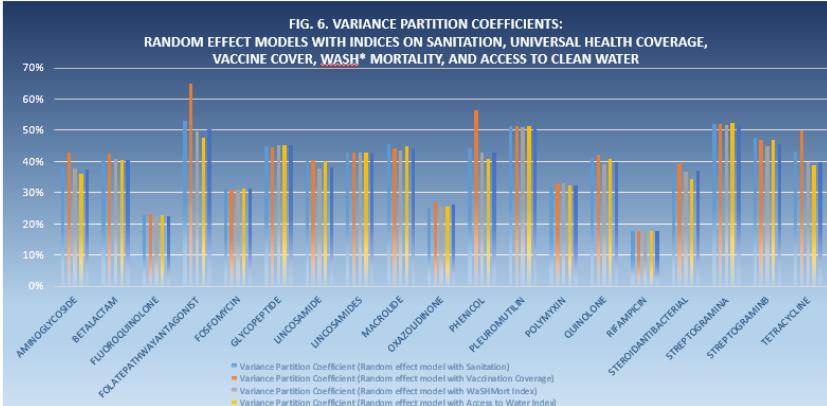
A number of SDG indices were added into the models to evaluate the study aim. Data on SDGs indices were extracted from the World Bank and the Global Burden of Disease data portals. The model variance was estimated as the median of the posterior distribution obtained by the Markov Chain Monte Carlo (MCMC) method.



*The index covers all the 17 goals

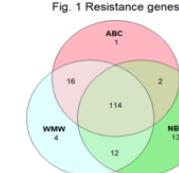


* The index covers 41 health indicators



* Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (exposure to unsafe Water, Sanitation and Hygiene for All (WASH))

Fig. 1 Resistance genes



Results

The microbial communities across sites differ from one another, which is attributed to site-country development levels (e.g., measured with SDG index and goals). The study identified 114 resistance genes, 35 phenotypes and 11 classes (Figs. 1-3) that are differentially abundant between countries.

Classes of AMR that have the smallest p-values across the random effect model analyses are: Aminoglycoside, Folate pathway antagonist, Lincosamide, Macrolide, Oxazolidinone, Phenicol, Pleuromutilin, Polymyxin, Streptogramin-A, Streptogramin-B, and Tetracycline.

Adding the SDG global index and SDG health scores into the random intercept models explained around 18%-52% of the between city variation in total AMR (Figs 4 and 5). Similar findings obtained by adding the indices on Sanitation, Universal Health Coverage (UHC), Vaccine Cover, WaSH Mortality, and Access to water individually into the models. The variance partition coefficients (VPC) values are between 18%-65% (Fig 6).

In random effect models (Figs 4-6), considering the SDG indices, the largest reduction in between city variation in the burden of AMR occurred for folate pathway antagonist, pleuromutilin and streptogramin-A classes, and the least reduction (still quantitatively large) was for fluoroquinolone, oxazolidinone and rifampicin classes.

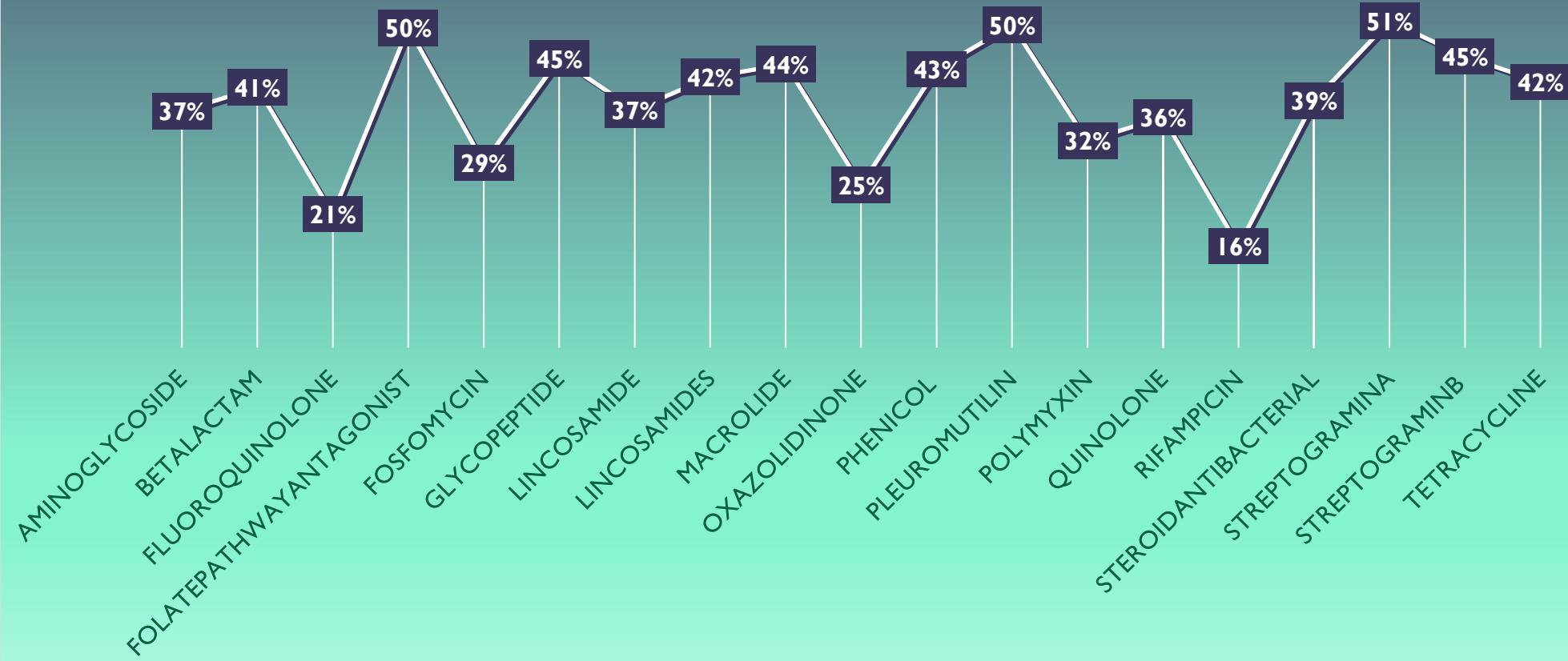
Conclusions

Actions focusing on sustained compliance with the SDG goals have the potential for elevating inequality in the global burden of AMR. There is a need for additional data that cover more sampling locations to develop more comprehensive global monitoring of AMR, and to determine the most appropriate course of action.



Confidential

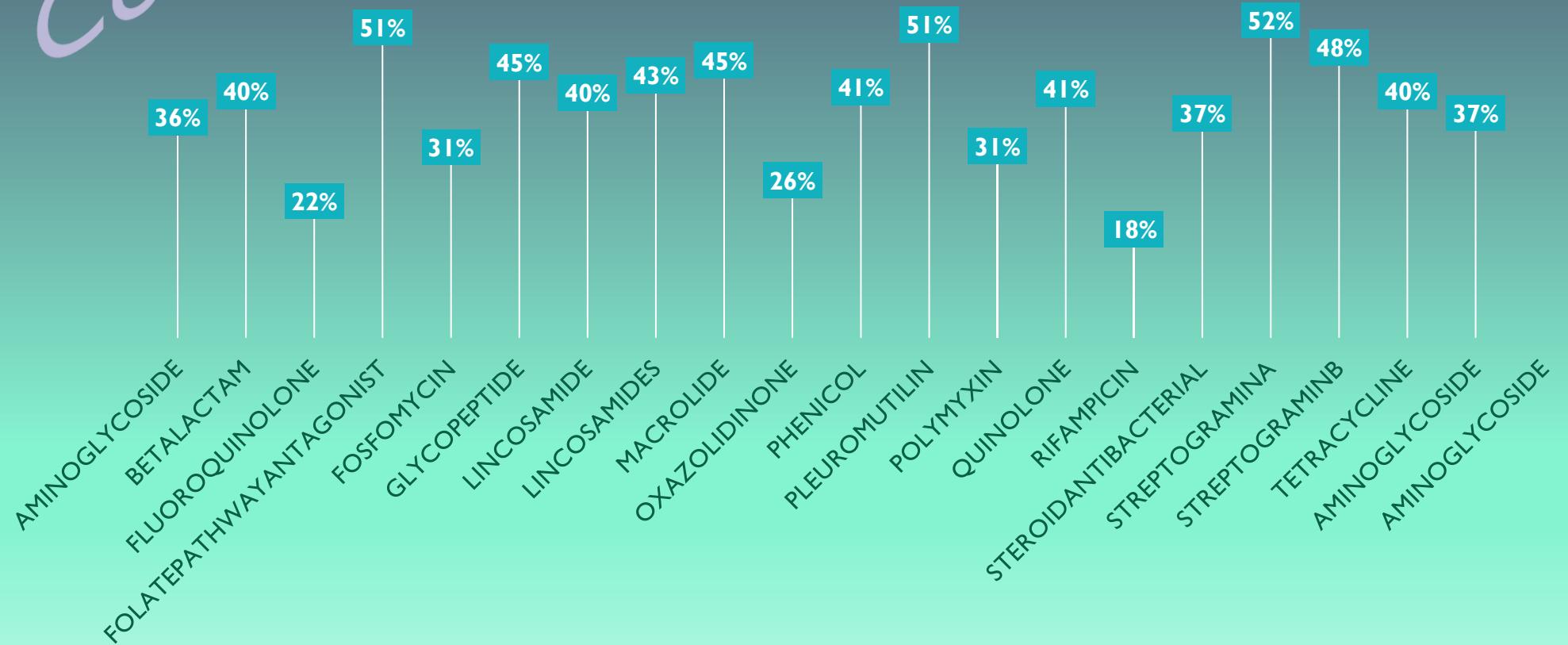
FIG. I. EXPLAINED VARIANCE: RANDOM EFFECT MODEL WITH SDG GLOBAL INDEX SCORE*



*The index covers all the SDGs 17 goals



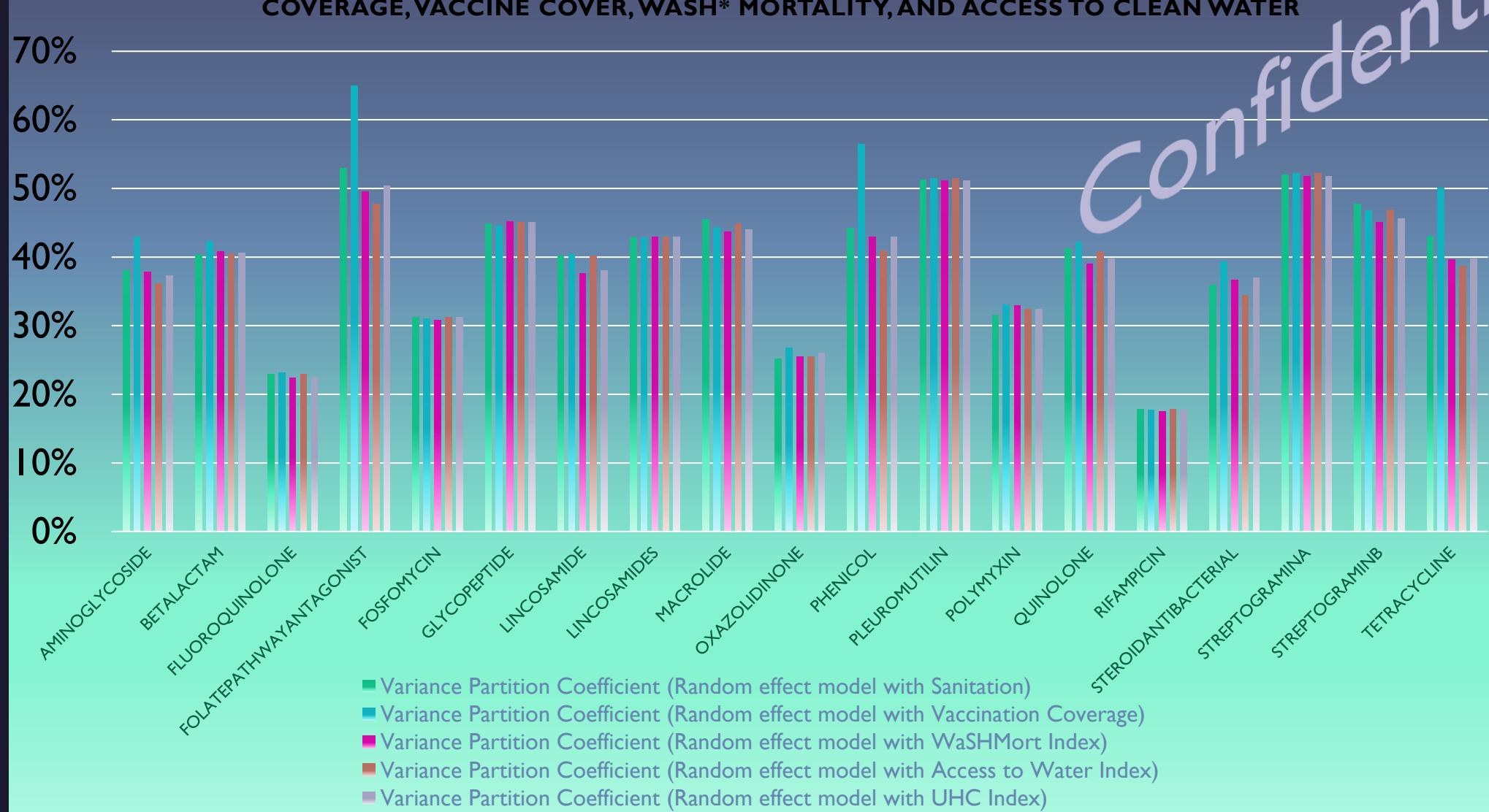
Confidential

**FIG. 2. EXPLAINED VARIANCE:
RANDOM EFFECT MODEL WITH
SDG HEALTH INDEX***

* The index covers 41 health indicators (health index)

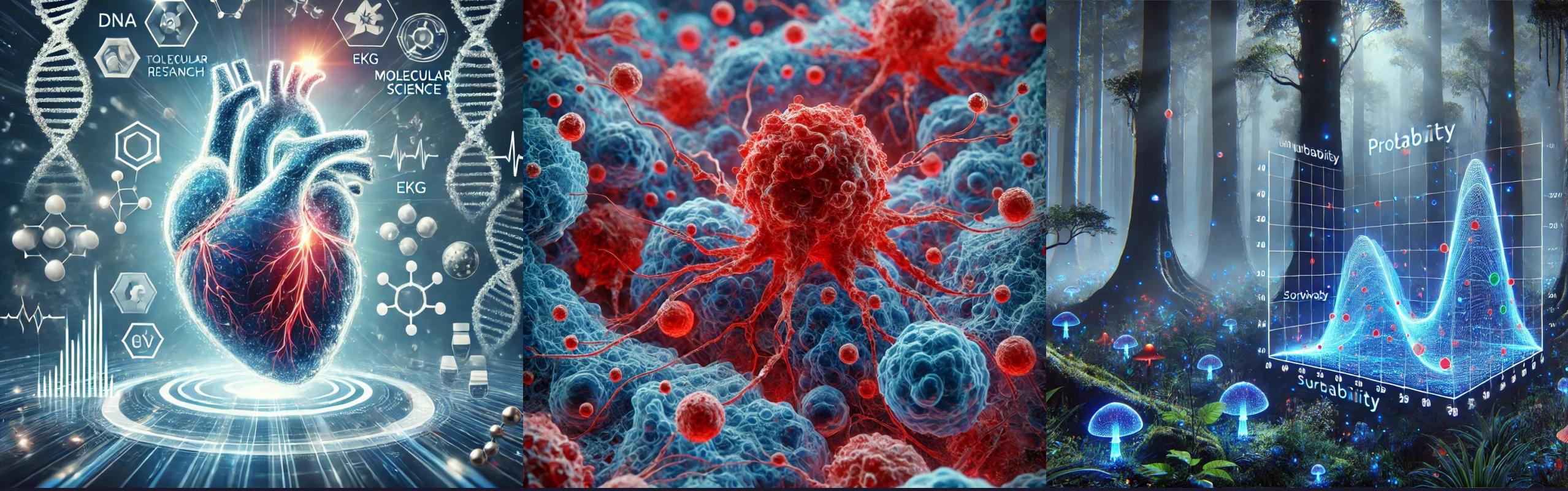


**FIG. 3. EXPLAINED VARIANCE:
RANDOM EFFECT MODELS WITH INDICES ON SANITATION, UNIVERSAL HEALTH
COVERAGE, VACCINE COVER, WASH* MORTALITY, AND ACCESS TO CLEAN WATER**



* Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (exposure to unsafe Water, Sanitation and Hygiene for All (WASH))





Polygenic RSs in prevention

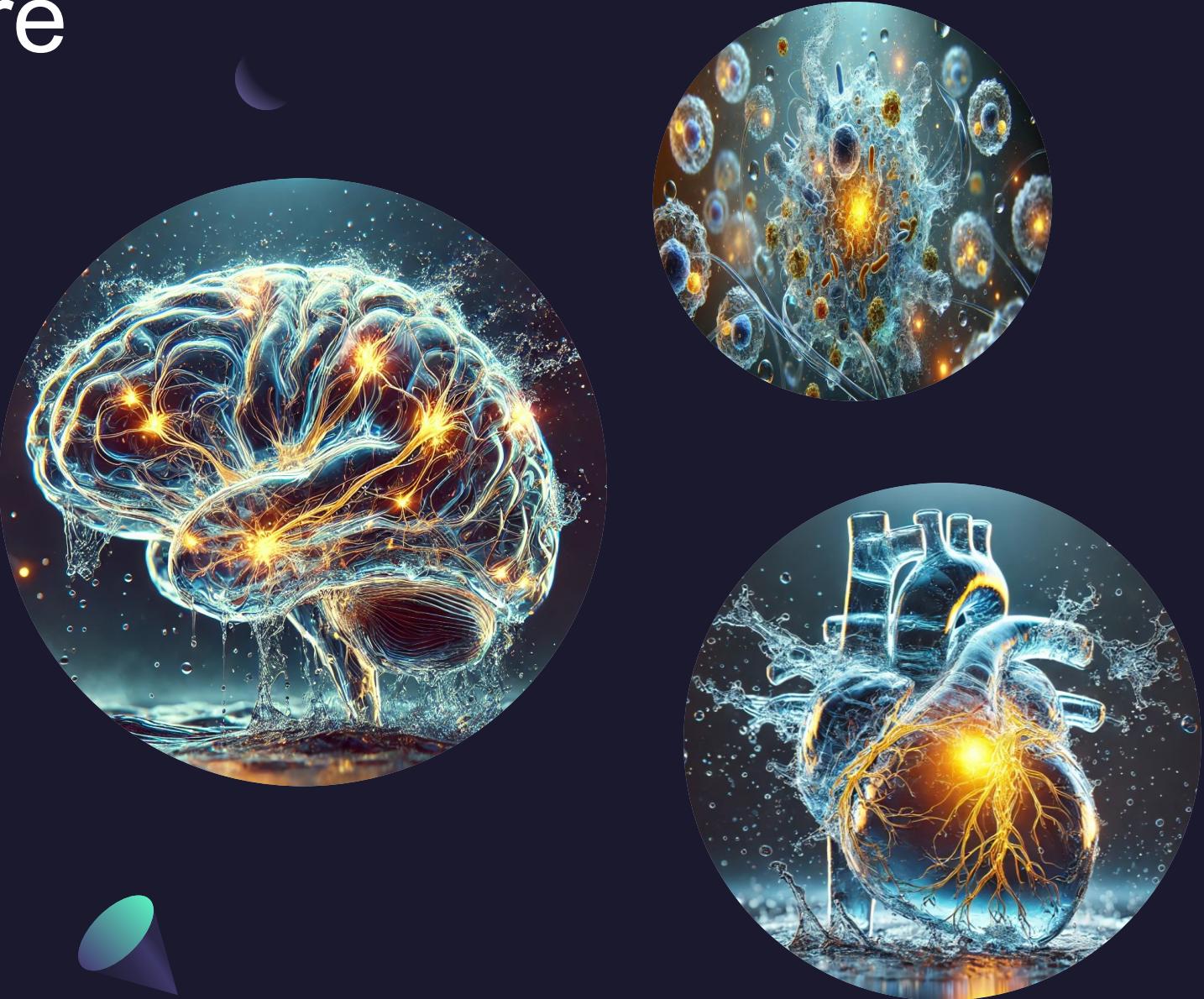
- Incorporating non-genetic risk factors into polygenic risk score models - Source: <https://doi.org/10.1038/s41598-023-27637-w>
- Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes – Source :
[DOI: 10.1056/NEJMoa0804742](https://doi.org/10.1056/NEJMoa0804742)

Polygenic Risk Score

Many people have an illness, or several illnesses, that are affected by **changes** in either one or many of their genes, frequently coupled with **environmental** factors.

Researchers are studying these **changes** to understand the role that genetics plays in **diseases** across different populations.

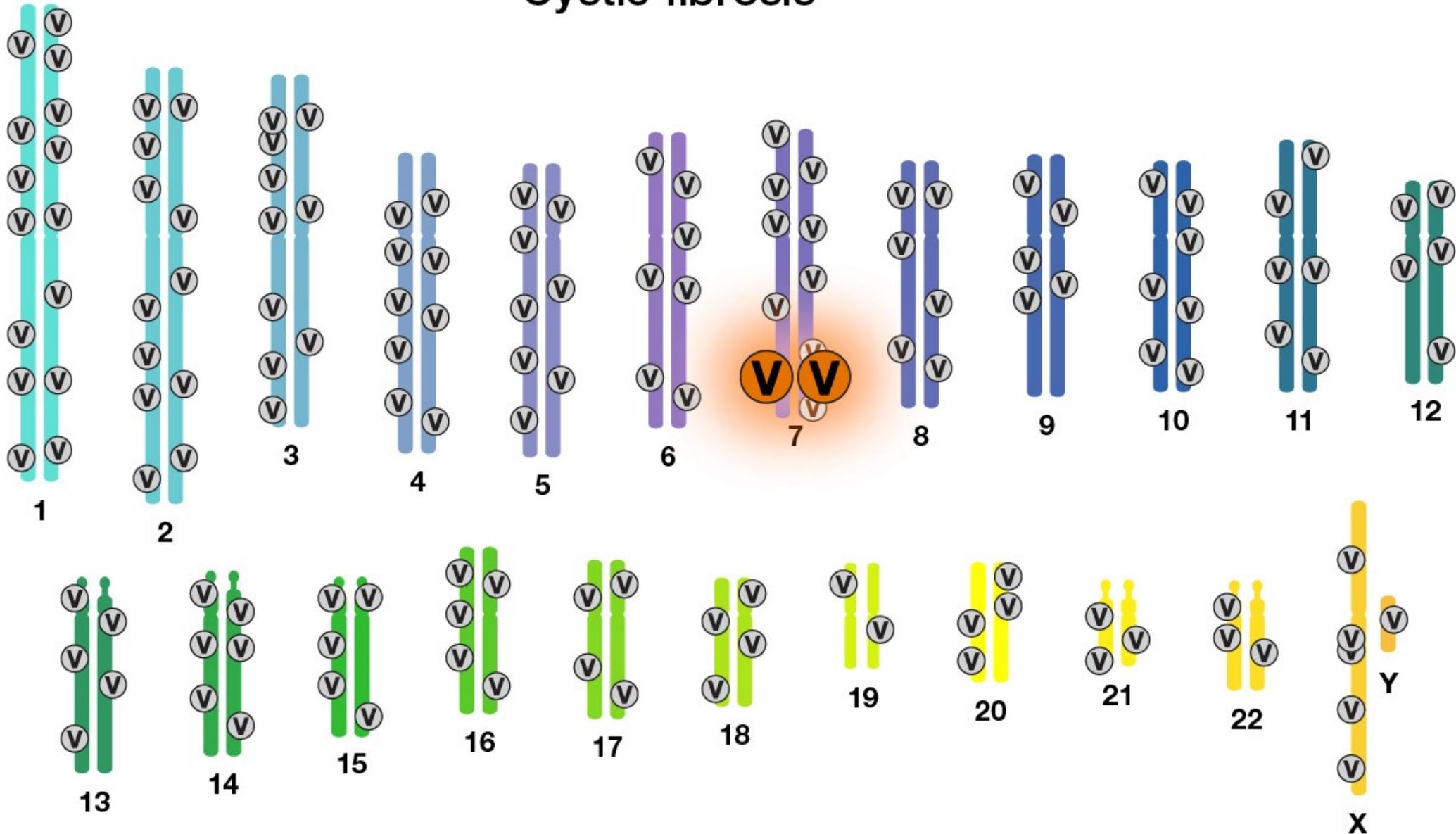
A “**polygenic risk score**” is one way by which people can learn about their risk of developing a disease, based on the total number of **changes** related to the disease.



Single-gene diseases

Cystic fibrosis

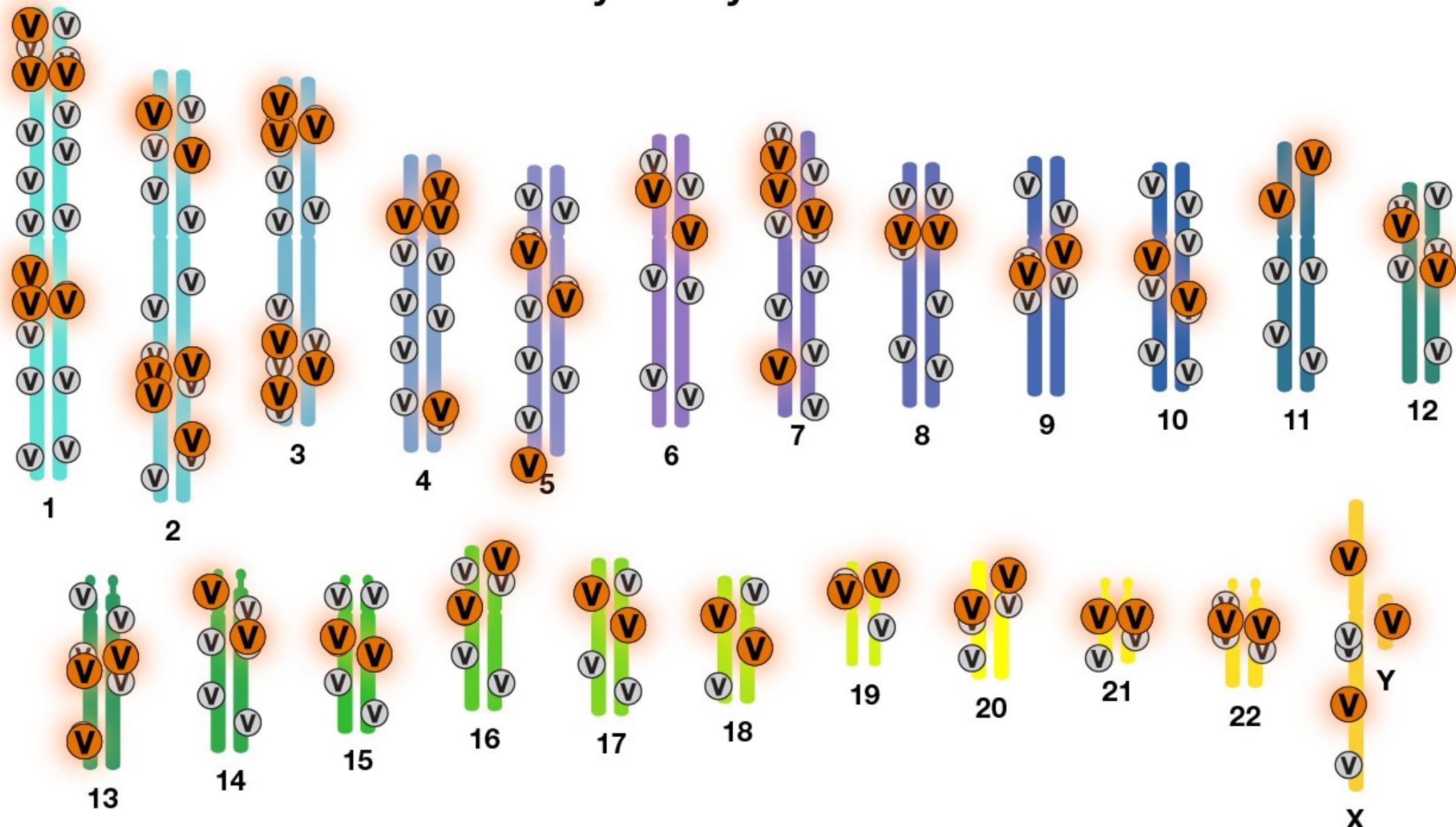
Person
one



Complex diseases - polygenic

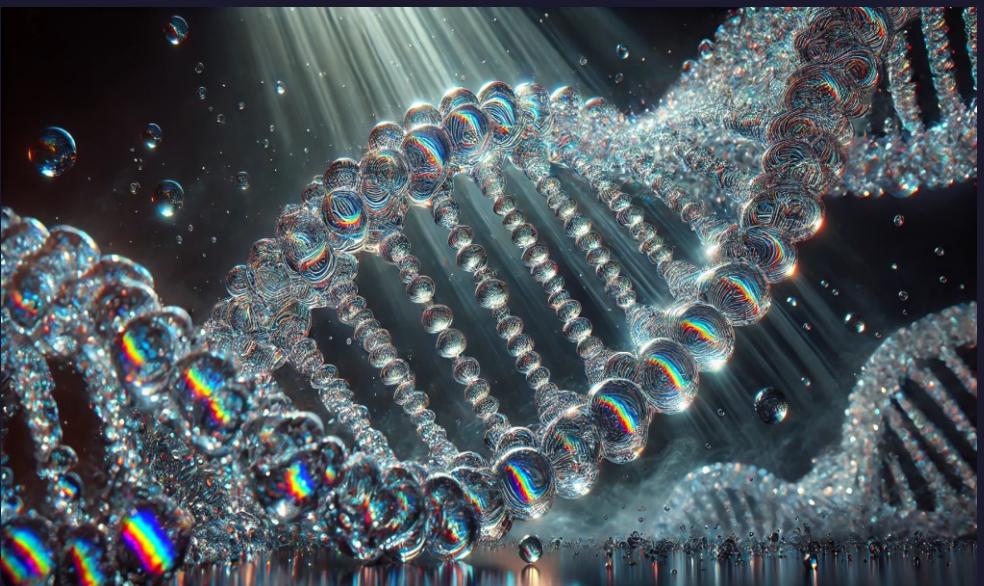
Coronary artery disease

Person
two



The necessity of incorporating non-genetic risk factors into polygenic risk score models

- **Commercial genetic risk scores** - misleading as they overlook easily accessible risk factors like **sex, BMI, age, smoking habits, family history, and physical activity**.
- models using **UK Biobank data** and externally testing them in the **Lifelines cohort**.
- Incorporating **common risk factors** improves the ability to identify the **top 10% of high-risk individuals** for **type 2 diabetes (T2D)** and **coronary artery disease (CAD)**.
 - Age, BMI, physical activity, sex, parental disease and smoking status
- Comparing different models (**genetics-based, risk factor-based, and combined**),
 - the **incidence of T2D in the highest risk group** increases from **3.0-fold and 4.0-fold to 5.8-fold**.
 - **CAD**, risk increases from **2.4-fold and 3.0-fold to 4.7-fold** when using the combined model.
- **Genetic Risk assessment should incorporate additional variables** instead of relying solely on commercial genetic tests.



Genotype Score in Addition to Common Risk Factors for Prediction of Type 2 Diabetes

Background:

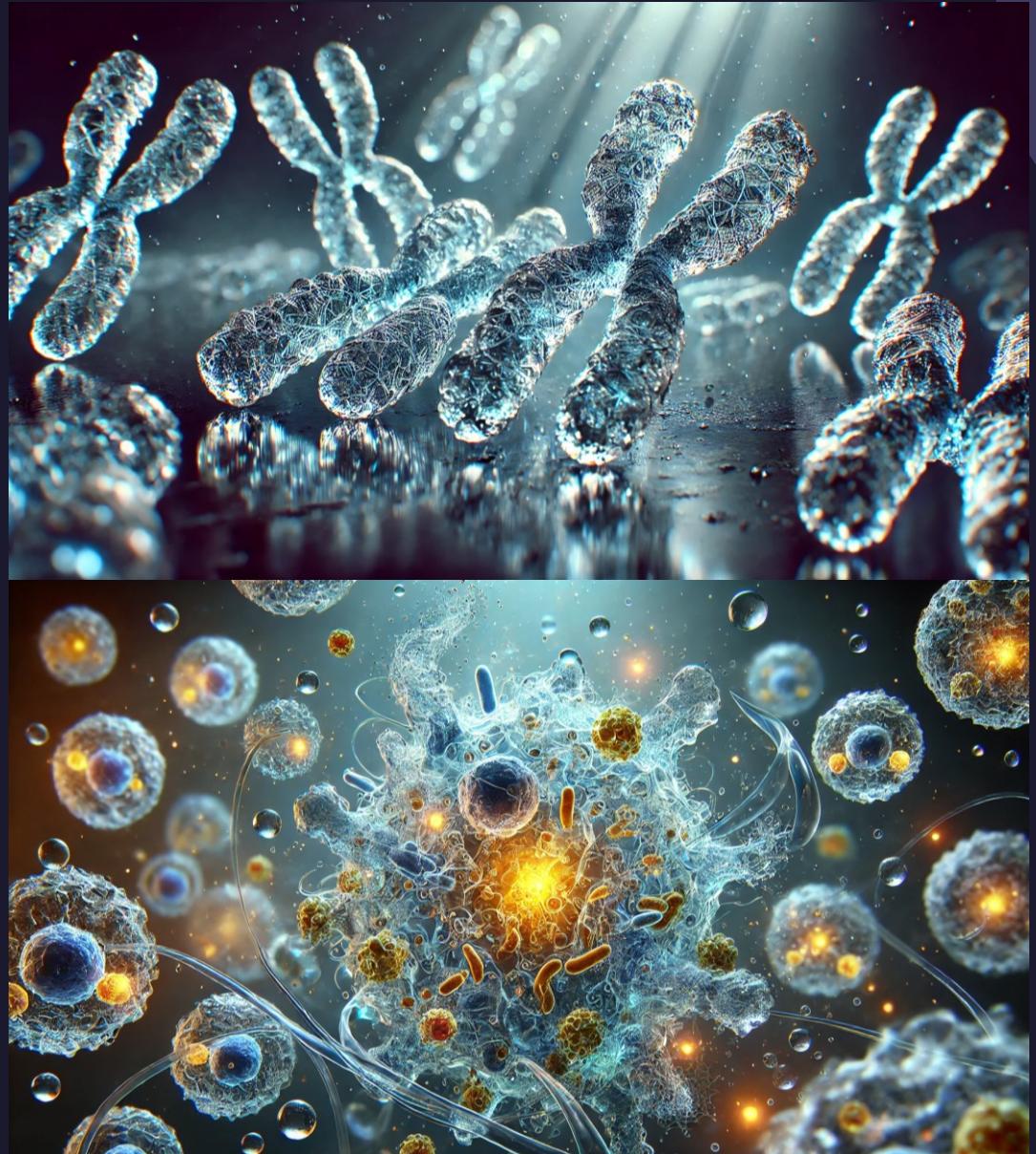
- The hypothesis suggests that understanding **genetic loci linked to type 2 diabetes risk** offers better risk prediction than relying solely on **common phenotypic risk factors**.

•age, sex, family history, body-mass index, fasting glucose level, systolic blood pressure, high-density lipoprotein cholesterol level, and triglyceride level

- The genotype score led to **risk reclassification** for, at most, **4% of participants**.

Conclusions:

- A genotype score incorporating **18 risk alleles** successfully identified new diabetes cases in the community.
- However, it only **slightly improved risk prediction** compared to using **common risk factors alone**.





From algorithm to bedside

Global Burden of Disease, Health Analytics, and Access to Therapeutics

- Pandemics influence RCTs
- Market size influences the design and scope of RCTs, as well as access to therapeutics





From algorithm to bedside

Economic evaluations

- Traditional RCS vs omics-based RCTs ??



Evidence pyramid and analytics

Machine Learning vs. Statistics



Feature	Statistics	Machine Learning
Goal	Explain relationships, hypothesis testing, and inference	Predict outcomes and optimize models for accuracy
Focus	Understanding causality , estimating parameters, deriving meaning from data	Pattern recognition, generalization , and automation
Causality vs. Correlation	causal inference (e.g., experimental design)	focuses on correlations – difficult to identify ‘why’
Data Needs	Works well with small , structured datasets	Handles massive, unstructured datasets (images, text, etc.)
Data Abundance	Structured manageable size data	Big data : video, audio, text, images,
Methods	Regression models, hypothesis testing, confidence intervals	Neural networks, decision trees, support vector machines
Complexity	Relatively simple models, interpretable results	Deep learning models with millions of parameters
Algorithmic Advances and complex data structures	struggle with high-dimensional data	ML techniques (e.g., random forests, deep learning)
Feature Engineering	Manual	Automated with e.g., Deep learning methods, convolutional networks in image recognition
Black Box Models vs. Interpretability	Interpretable	Sometimes like "black boxes"
Non-parametric vs. Parametric Models	Usually assume a distribution	often uses non-parametric
Computational power	Classical computation	GPUs , TPUs, and distributed computing

Evidence pyramid and analytics

Common statistical tests in RCTs

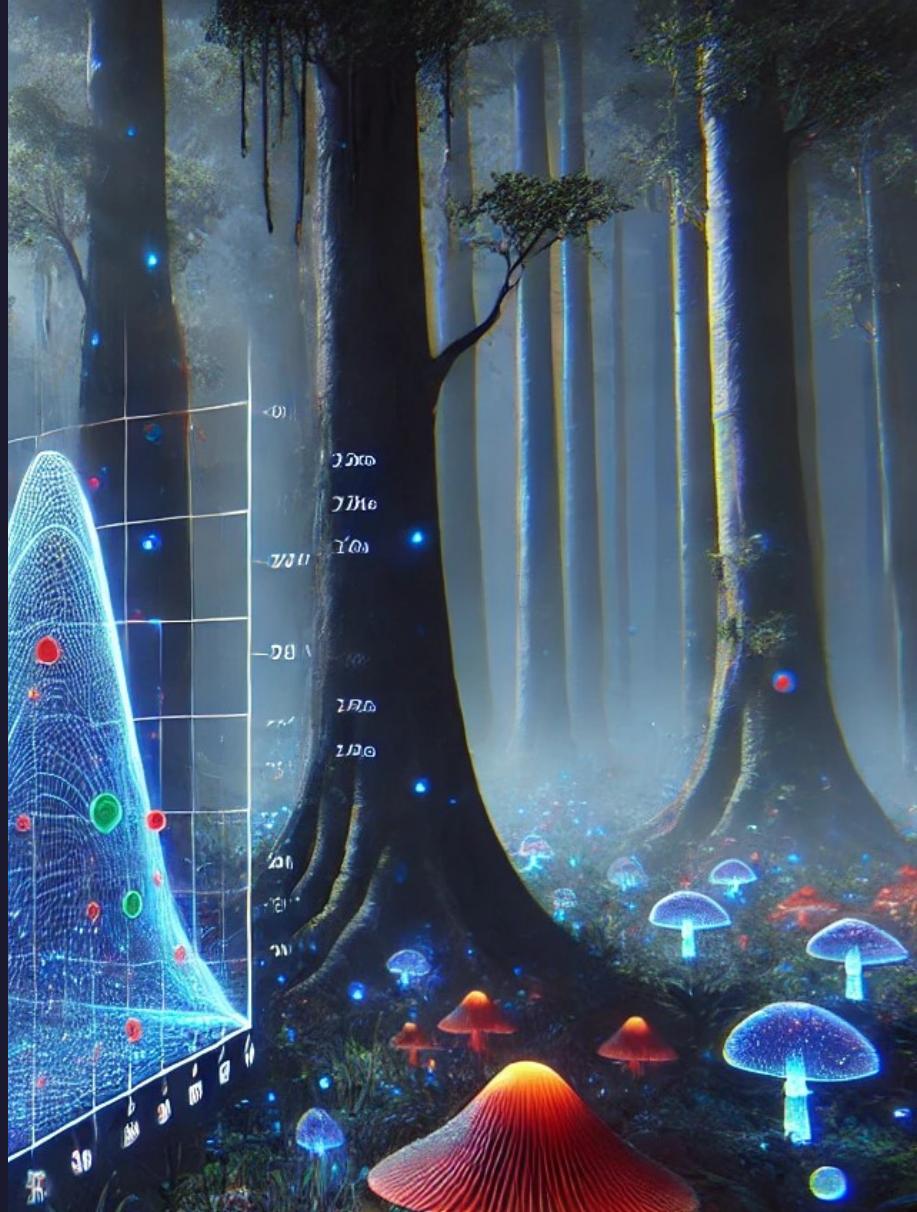


Figure 1: Statistical methods for comparing two independent groups or samples

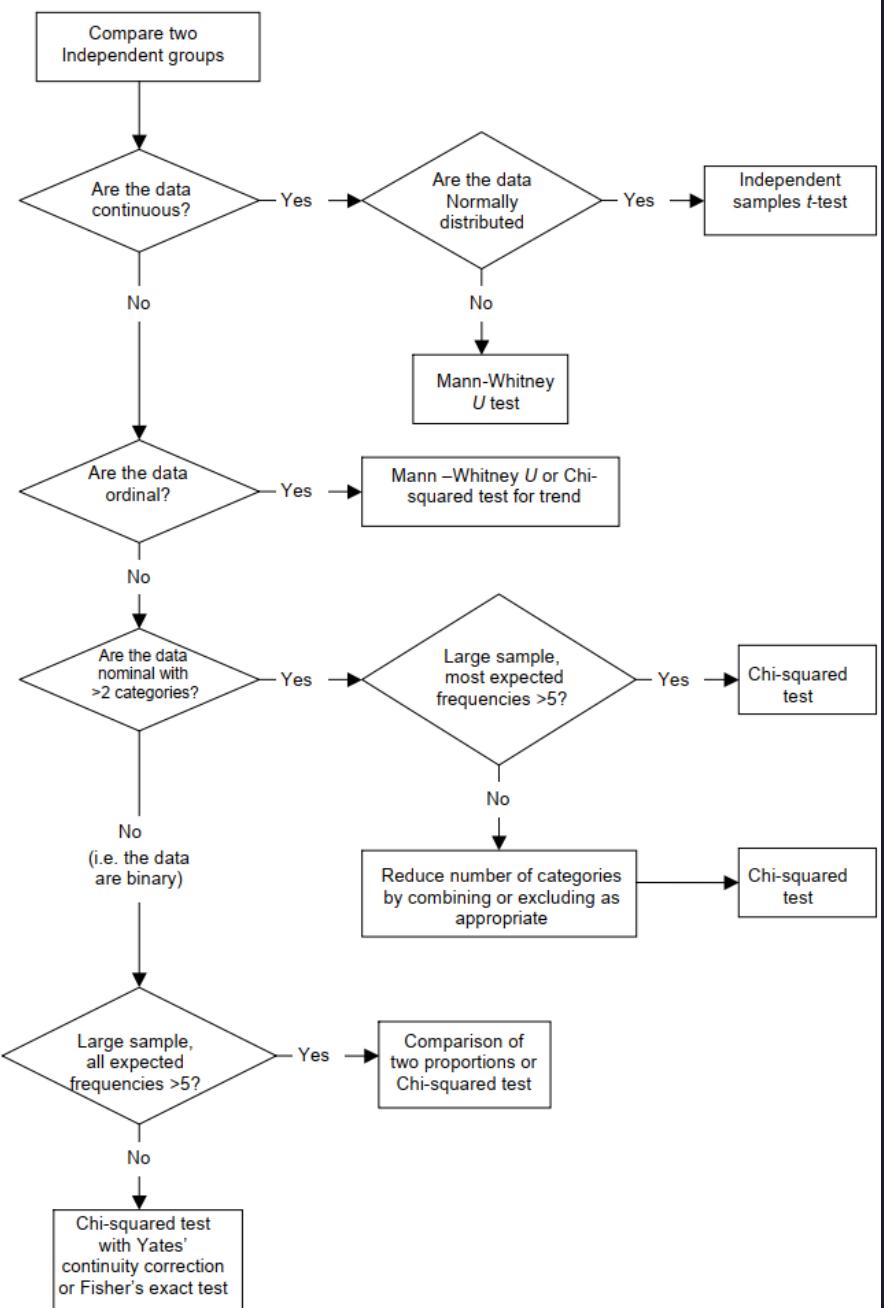


Figure 2: Statistical methods for differences or paired samples

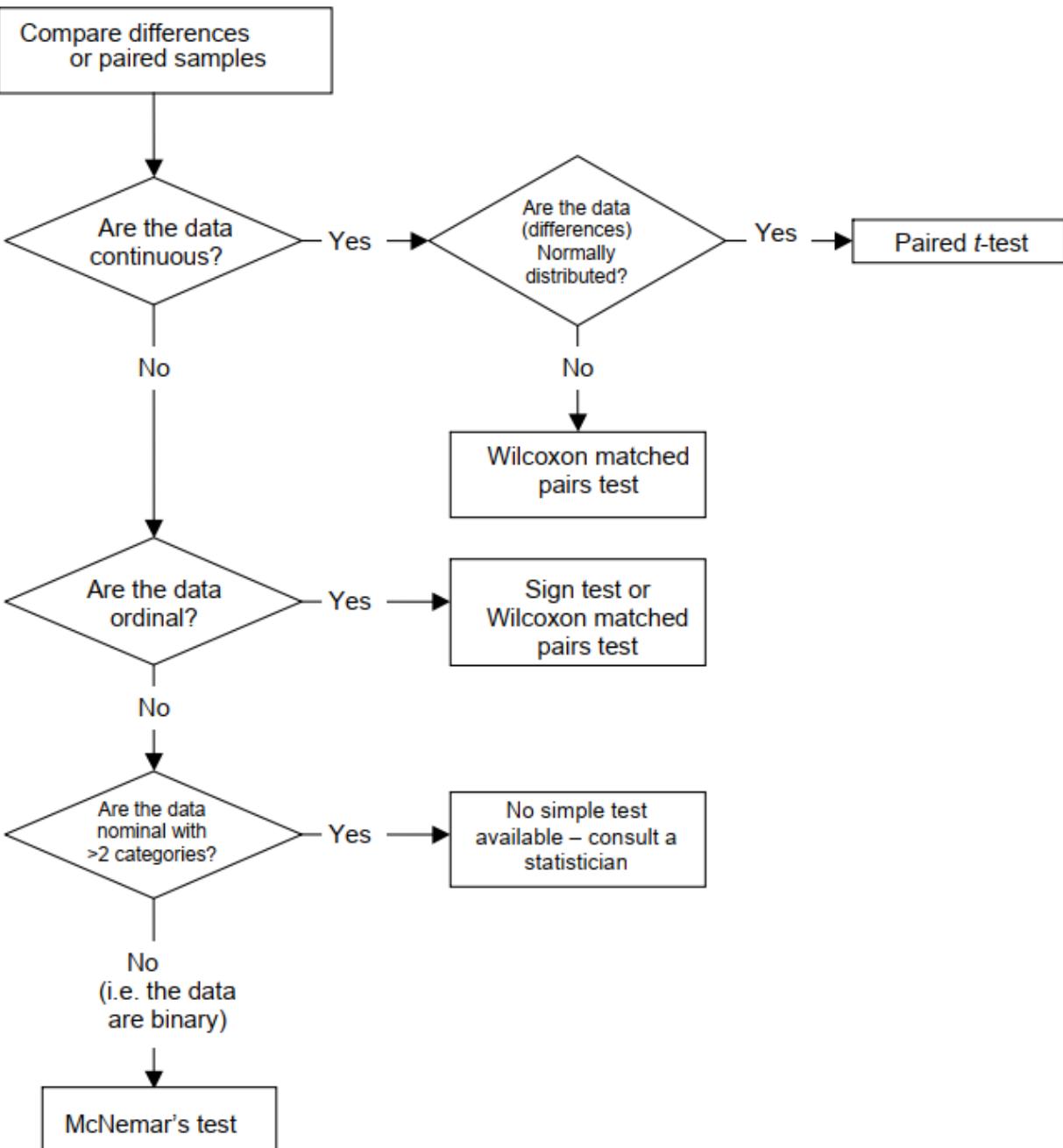
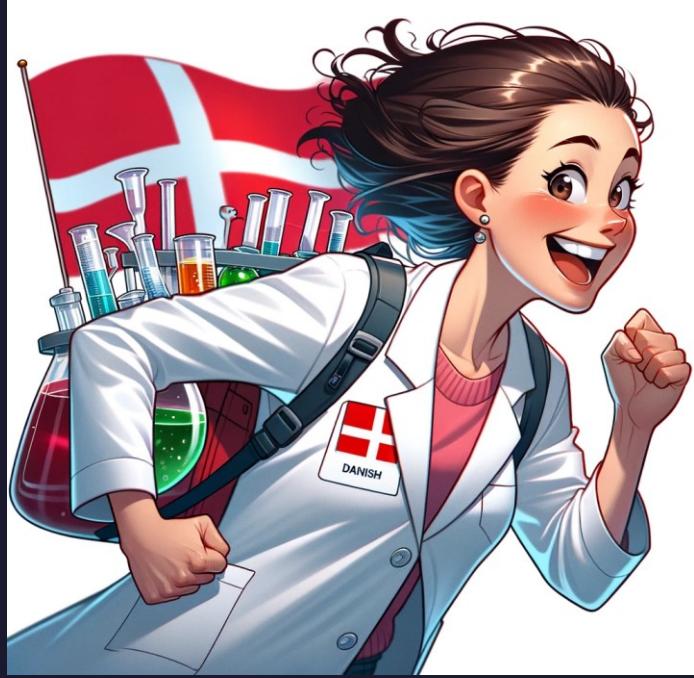


Table 1: Statistical methods for relationships between two variables measured on the same sample of subjects

	<i>Continuous, Normal</i>	<i>Continuous, non-Normal</i>	<i>Ordinal</i>	<i>Nominal</i>	<i>Binary</i>
Continuous	Regression Correlation: (Pearson's r)	Regression Rank Correlation: (Spearman's r_s)	Rank Correlation: (Spearman's r_s)	One-way Analysis of Variance	Independent samples t -test
Continuous, non-Normal		Regression Rank Correlation: (Spearman's r_s)	Rank Correlation: (Spearman's r_s)	Kruskall-Wallis test	Mann-Whitney U test
Ordinal			Rank Correlation (Spearman's r_s)	Kruskall-Wallis test	Mann-Whitney U test Chi-squared test for trend
Nominal				Chi-squared test	Chi-squared test
Binary					Chi-squared test Fisher's exact test



**(omics) experimentation will be part of
routine personalized care**



Thank You

Linkedin connect!

