QUALITY CONTROL

AUTHOR
M. Knigge

INSTITUE
Hanze University
Life Science and
Technology

ORGANIZATION
Eriba
Department of Genetics

DATE
5, May 2018

QUALITY CONTROL

Hanze University of Applied Sciences Groningen

umcg

AUTHOR
M. Knigge

SUPERVISORS
Asst. Prof. Dr. S. Sanna
Asst. Prof. Dr. Y. Li

LECTURER
Dr. M. A. Noback

INSTITUE
Hanze University
Life Science and
Technology

ORGANIZATION
Eriba
Department of Genetics

DATE
5, May 2018

Supervisor's signature

TABLE OF CONTENTS

INTRODUCTION

THE COMPANY

This internship was provided and guided by the immuno-genetics research group in the department of Genetics at the University Medical Center Groningen (UMCG). The hierarchy can been seen as a structure divided in multiple sections (A-F), each with its own director (figure 1). Above these directors is a main director. All sections are divided into departments, and each department has a head. For the Department of Genetics this is prof. Richard Sinke. Under his coordination, several associated and full professors lead different research groups.[1]

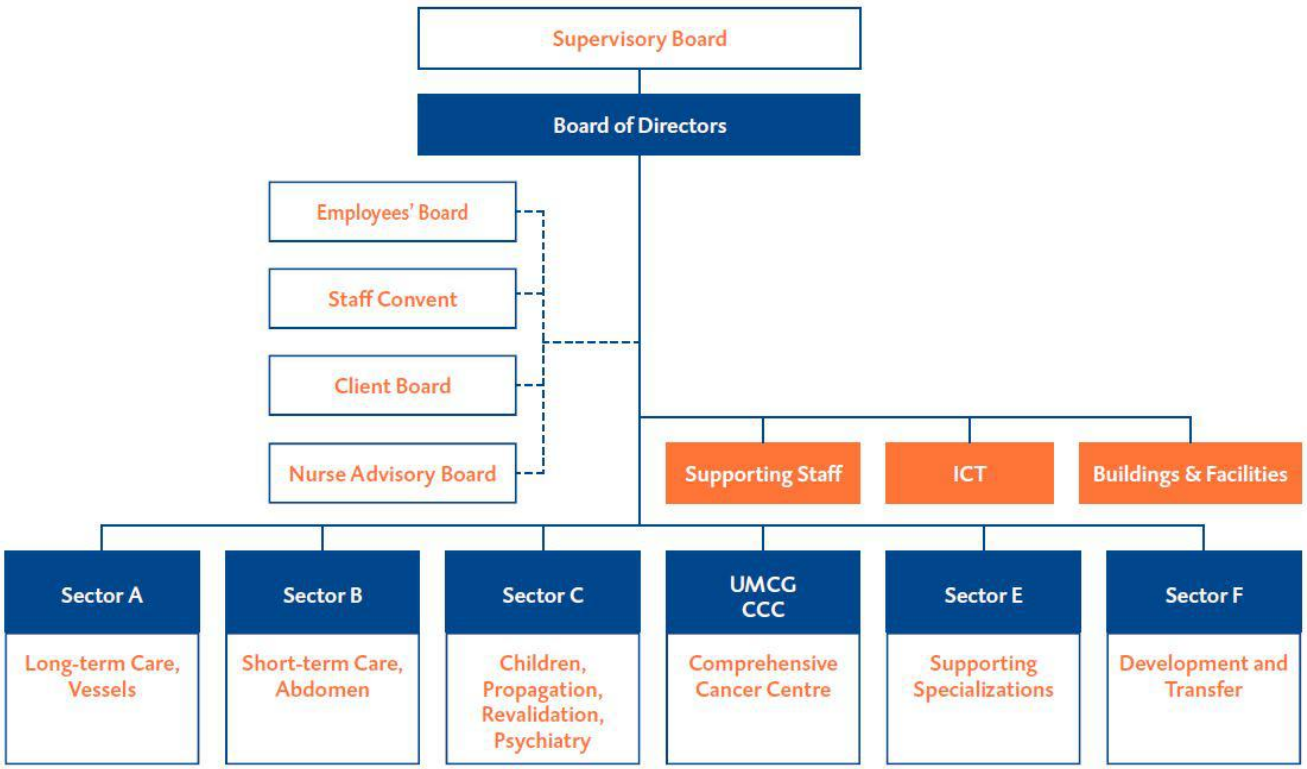

Figure 1: Organogram of the UMCG organizational structure.

THE PROJECT

Evidence suggests that the human gut microbiome plays an important role in metabolic and immune function.[2] Empirical evidence of Microbiome-Wide Association Studies (MWAS) in population cohorts show significant associations identified systematically between the gut microbiome and multiple complex traits and intrinsic factors, like Inflammatory Bowel syndrome (IBD), body mass index (BMI), food allergies and blood cells composition.[3] Discovery of these links is key to understanding disease, and potentially

disease treatment.

The body of a human is a complex interconnected ecosystem, and the gut is where the body acts as a first line of defense. Where it interacts with the "outside world", functioning as a frontline of the immune system, which is constantly exposed to new microbes and molecules.[4] The whole collection of microbes and molecules that are present in and on the human body is known as the microbiota.[5] The microbiome refers to the whole set of genes within these microbes. The role of the microbiome composition/function is considered as an acting organ in the body's operation. It is presumed that it has an effect on aging, digestion, mood, cognitive function, and immune system.[6] The immune is a defensive system from the host entailing many biological structures and processes within an organism to protect against diseases, and infection. The function of the immune system relies on the ability to detect and distinguish a wide arrange of agents known as pathogens, viruses, and parasites from self and non-self.[7]

The aim of this research is the identification of causality links between the microbiome composition/function and immune system: does the microbiome influence the immune system (cell counts, cytokines, globulin levels), or/and does the immune system influence the microbiome. We have access to the largest population cohort with gut metagenomic sequencing (LL-DEEP), but this cohort has not been characterized in depth for immune traits.And we have access to another cohort, the human function genome project 500FG[8] which was specifically designed to assess the immune and metabolic system (500 phenotypes available), and also microbiome, but it is very small. This cohort has also genetic and gene expression available. By using gene expression data, genetic data, and transcriptomic data from the 500 Functional Genomic (500FG) cohort a linear model is constructed that explains immune traits/functions between gene expression data, genetic data, genetic data combined gene expression data and the 500FG cohort with Elastic Net Regularization.[9] This constructed model, that is based on genetic data and gene expression data is used to predict immune traits/functions in LL-Deep data which contains genetic, and gene expression data from a large number of individuals which lacks immunogenic information. These predicted immune phenotypes will be used to forecast a causal link with microbial composition/function, with one-sample Mendelian Randomization (MR), see figure 2.

We expect to find causal links between microbiome composition/function and the immune system that can help us understand the interplay between the gut microbiome and the immune system, and ultimately help understand disease, and develop disease treatment; restoring microbiome composition and/or function through personalized nutrition or treatments. However, to efficiently translate findings into clinical practice, it is essential to discriminate between microbiome features that are on the causal pathway to disease from those that are a consequence.

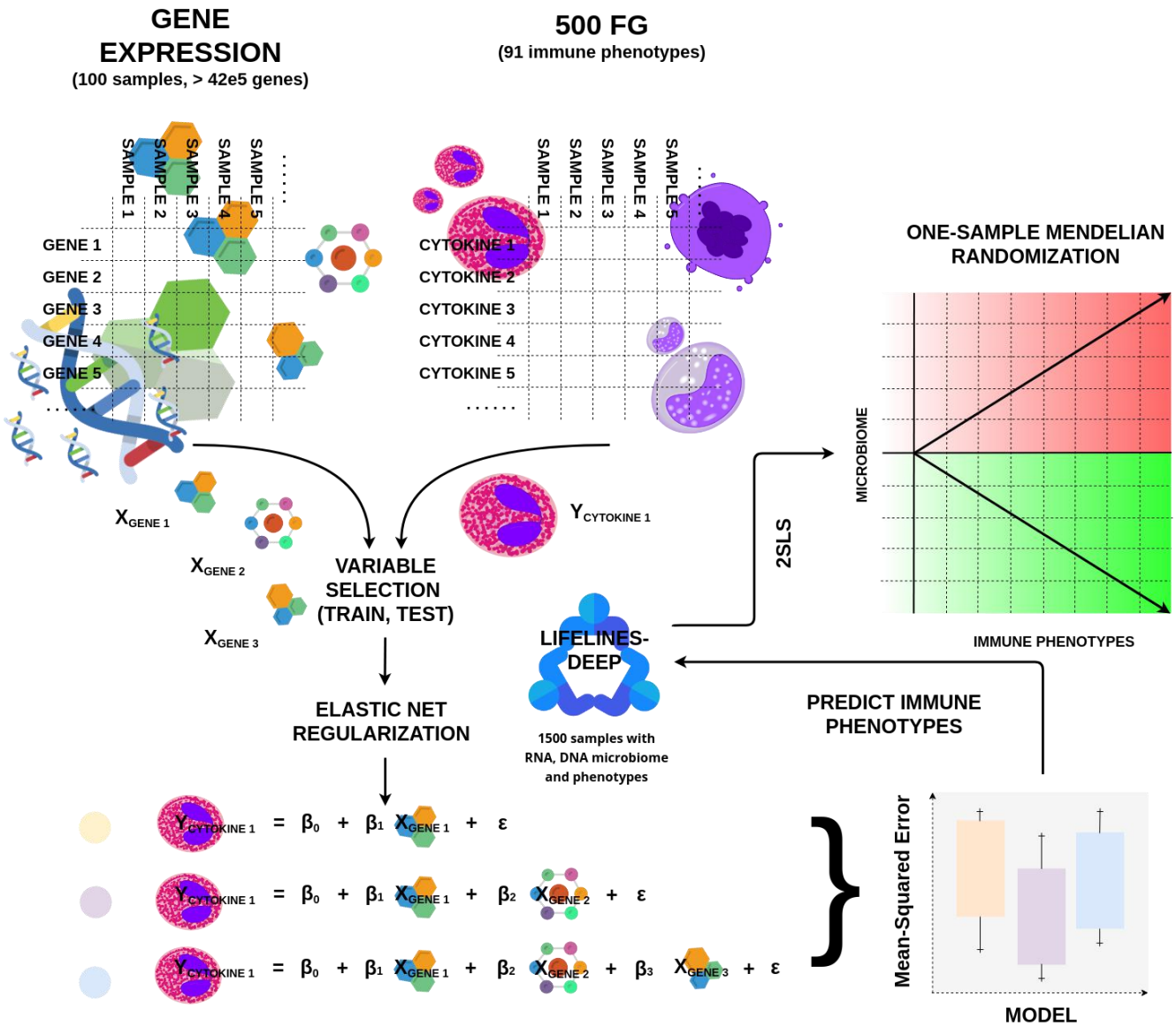

Figure 2: Overview of the computational infrastructure. For every immune phenotype in the 500FG cohort, multiple models will be build with a different set of genes that tries to explain a certain immune phenotype with the Elastic Net approach. After the model creation, these will be evaluated and the best one will be used to predict immune phenotypes in LL-Deep, which in turn, are going to be used to access causality links between the gut microbiome and immune system.

INTERNATIONAL ORGANIZATION FOR STANDARIZATION

The International Organization for Standardization (ISO) is an international standard-setting embodiment build out of representatives from a variety of national standard organizations. The ISO has been founded on 23 February 1947, this organization promotes worldwide industrial and commercial standards. It is headquartered in Geneva, Switzerland, and is enforced in 162 countries.[10, 11]

The ISO is an independent organization, that completely uninfluenced by governmental organizations, and the members of this organization are part of standardized sub organizations in 162 member countries. ISO is the world's greatest developer and employer of voluntary international standards and facilitates world trade by providing uniform standards between nations. Over twenty thousand standards have been developed: ranging from products manufacture, and technology to warrant food safety, and health-care. The use of these standards help the development of products and services that are safe, reliable and of good quality, national wide.

The standards that ISO provides aid companies in increasing productivity while minimizing errors and inefficacy. They enable products from different markets to become objectively comparable. Additionally, they help companies entering new markets by aiding in the production and development of national trade, on an equally transparent basis. Furthermore, these standards also provide a safeguard for consumers. One important method through which they achieve this is through obligatory certificates which conform to the minimum standards set internationally. This gives consumers a verifiable measure of the quality of their purchases.[10, 11, 12]

WHY ISO NORMS

The ISO norms are developed with the purpose of getting a grip on country specific norms, that can be used to create norms that are the same across the entire globe. The use of these standards help the development of products and services that are safe, reliable and of good quality through the trust that is created about these qualities in association with ISO norms national wide. The ISO norms aid companies in important aspect that benefits profits and company image, such as by increased productivity, safe guarding of customers, and bringing about transparency. Companies that are ISO certified must satisfy all requirements defined by that norm, which of course, are defined within that ISO norm. The responsibility to meet those requirements lies in the hands of the company, and how they want to achieve is, is also upon themselves. This might require some effort and resources at first but will benefit any party involved on the long term. The ISO norms can be seen as a tool, which helps optimize a company, find weak spots, and

improve any organization of a company. Its effect does not only take place within the company, but also it helps the company to enter market in which a certain standard is required. This will improve sales for example, because when a company adheres to these norms, there product can be guaranteed to meet a certain quality standard that will encourage customers to choose this product instead of a product from a company that lacks this norm. Before a company gets a ISO certificate, the company needs be tested, also known as audited. When the company passes this accreditation, they will acquire this certificate.[13]

UMCG AND ISO NORMS

The first medical center in the Netherlands that is certified according the ISO quality management system is the UMCG. Back in January 2016 the UMCG received the ISO 9001 certificate for healthcare for patients, research, training and education.

It is important for the UMCG that this was rewarded by an independent institution such as ISO. The ISO 9001 certificate shows that the quality of the organization and professional services from the UMCG are of a high quality, which is especially important for maintaining a reputation of adequate health care and objective high-quality research, and the effort that the UMCG is willing to put into achieving this. The certificate is internationally recognized, which is important for the international relations and collaborations that the UMCG wants to maintain. External auditing is repeated every year to maintain value of the certificate and high level performance in the organization. The examination returns every year because the certificate must be updated to current policies and practices which patients must be able to count on for their safety and care. Independent external parties also examine the quality of the care from the UMCG. The UMCG has to report periodically to other independent parties such as (1) the Healthcare Inspectorate and (2) the Ministry of Health, Welfare and Sports.[14] To make matters more specific: the UMCG is certified for the NEN-EN-ISO 9001:2008 norms for research and education, training, and healthcare since January 2016. The ISO 9001 norm is based on seven core quality management principles. (the QMPs):

- **Customer focus**: The primary focus of quality management is to meet customer requirements and to strive to exceed customer expectations.

- **Leadership**: Leaders at all levels establish unity of purpose and direction and create conditions in which people are engaged in achieving the organization's quality objectives.

- **Engagement of people**: Competent, empowered and engaged people at all levels throughout the organization are essential to enhance its capability to create and deliver value.

- **Process approach**: Consistent and predictable results are achieved more effectively and efficiently when activities are understood and managed as interrelated processes that function as a coherent system.

- **Improvement**: Successful organizations have an ongoing focus on improvement

- **Evidence-based decision making**: Decisions based on the analysis and evaluation of data and information are more likely to produce desired results.

- **Relationship management**: For sustained success, an organization manages its relationships with interested parties, such as suppliers.

The genetics department at the UMCG uses every core management tool to improve and maintain the standards mentioned before. The core principles that are most notable in terms of result-oriented management are process approach, improvement, and evidence based decision making. The tactics through which these principles are maintained consists of meetings, presentations, discussions and reviewing. At the department of Genetics there are meetings every day, these meetings are all organized with approximately the same goal but the setup differs slightly every time. Researchers have to present their work every now and then to give updates to the entire department, the research group, and their supervisors. The items that people have to present at meetings are the goal of the research, the schedule, the results, and why certain decisions are made. The presentations are also setup in the form of a collaborative work meeting, in which everybody can ask questions, give feedback, and make suggestions for the presenter. The setup of the meetings helps researches to think deeper about their research and motivates them to make evidence-based decisions which helps to improve individualized research. Besides the fact that researchers have to warrant their results and research, the department also wants to create and maintain a certain level of knowledge by keeping everyone informed about high impact research or research that is important for projects similar to tasks and research performed by the Genetics Department. Informing employees and researchers is done through a journal club where three people have to study six recently published papers and present them to the entire department every week.

ISO 27001

The ISO 27001 is security standard for information security standard. This is published by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). These organizations have specified a management system that is a tool to help bring information security under management control and gives specific requirements.

Organizations can have a number of information security controls. But, without an information security management system (ISMS), it is very easily to have a disorganized security system. Security controls in companies typically address aspects of IT or data security specifically. This ISO specifies how implement a good security system on storing data that is sensitive, that it may not be shared. But also on how to act

when there is for example a data breach. The UMCG is very strict in these rules and when violated it can mean the end of the relation between employer and employee. As employee of the department of Genetics people mostly work with patient data, and when working with this kind of data employees have to adhere to the rules of the UMCG. It is not allowed to store patient data on private laptops or external storage devices, and everyone has to sign a form of conventionality.

QUALITY ASSURANCE AND CONTROL

MEETINGS

To warrant the high-level research, multiple meetings are organized per week for researchers to come together and free up some time for discussions. Meetings are organized department-wide, research specific, and with attendance of supervisors. These meetings help the researchers to stay on track and gives a moment to discuss the results or methods with people from another field who might bring in some fresh perspective. In a way, they force attending researchers to gain in depth knowledge about their projects, because, as an example, at the department wide meetings, researchers from different disciplines are present to which they also have to justify their choices. This requires that the presenter breaks his project down to understandable information for everyone to comprehend.

REPLICATION

The replication of a study means being able to repeat a study while using the exact same methodology, but with different subjects, experimenters, data, or patients. The theory discovered in a research project will be applied to new situations, performed by different researchers, in order to determine generalizability to different situations, data, subjects, or patients. The main purpose of replication is:

- assurance that the results are reliable and valid
- determination of confounding variables
- application of previous results to new situations
- inspire new research into combining previous findings from other related studies.

Take this situation for example: a team is researching the use and effect of a certain painkiller medication in patients before surgery. Before conducting the trial, the literature will be consulted for finding similar or problems or the same problem. The main question here is: how does the research group know that the results from that study are applicable and transferable into the research of their team? The research team will perform a replication of previous research procedures in their clinical setting and thus will be able to strengthen the evidence of previous research finding.

Replication is one of the most important parts if research, it is needed to warrant results found by some

study. At the department this is done by replicating some analysis in another data cohort, or by letting multiple researchers conduct the same analysis. As an Bioinformatician this is of great importance because code can vary a lot, and human error is frequent if research only involves the perspective of one person.

STATISTICAL TESTS

As a Bioinformatician, the performance of statistical tests is an important part of the research, and is the main part of the research. This entails statistics to compare distributions with chi-square tests, finding associations, determining causality links between genotype-phenotype, pruning data, and quality control. For example, with to genotype datasets, can they just be compared, and is it needed to remove outliers, this can be done with a Principle Component Analysis (PCA). What amount of variance explained is needed, and how many components are needed to determine genetic outliers. These are just a couple steps and questions that are important for this research. These kind of methods are discussed every week to make sure no mistakes are made , and all analysis are replicated in other cohorts to validate methods and findings. For any analysis, when possible in this research everything is written by us, because we want to gain as much knowledge as possible, so we do not go for the easy option, which is the use of existing R-packages, we write our own libraries. At the department of genetics we work with large datasets: genetic, expression, phenotypic and other kinds of data. When analyzing these large amounts of data the overview is easily lost, and a statistical test always gives an answer, and the most tricky part in this research is that we use a machine learning approach which makes it difficult to interpret errors since it is sometimes not clear what is going on, this is due to the large amount of data, there will always be a p-value to calculate. So it is important to setup the correct hypotheses. As I would like to quote Arne Poortinga - garbage in, garbage out-, there is always an answer, but is it honest?

QUALITY OF CODE

The quality of the is very important, it needs to be readable and reproducible for other studies, the code written needs to be publication ready. For my own research I write my own packages that are used the version control system GIT, which helps to keep track of the code, makes it easy to document the code, and when something is lost, it is easily resolved. For R I use the basic template that all package depositories use, it needs a description of the functionality, description of the parameters, description of the output, and many more things. This publication ready programming and having the repository publicly available forces the programmer to write well documented code. And having a GIT system has the advantage of also being a logbook. For an example I want to refer to my Mendelian Randomization package, which can be easily installed into R, and has a tutorial: https://bitbucket.org/MatthijsKnigge/mendelianRandomization/

CODE PIPELINES

The building of pipelines for our analyses is very easy, since the data that we use are large matrices, and the

system that we use to deploy our analysis is SLURM. This kind of research is something completely different than for example text-mining based studies, there are not a lot of errors that can be made, there are no weird variables that can be missed. It is namely the application of some statistical test on a very large dataset.

DOCUMENTATION LAYOUT & LOGBOOK

It is important to keep track of what has taken place in the research. For the computational part I take advantage of the GIT system, which helps me keep track of the code that I have written, what I have changes and why, this is all extremely well documented. See previous link to Mendelian Randomization package.

For logging results, and the output of the analysis, I use R markdown, which is perfect because when I want to discuss preliminary results in a meeting I always use this document because it shows code, results, and I can write my findings and conclusions about it and then discuss it.

REFERENCES

[1] Organisatie . (2018). Umcg.nl. Retrieved 1 May 2018, from https://www.umcg.nl/NL/UMCG/overhetumcg/organisatie/Paginas/default.aspx

[2] Zhernakova, A., Kurilshikov, A., Bonder, M., Tigchelaar, E., Schirmer, M., & Vatanen, T. et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science, 352(6285), 565-569. doi:10.1126/science.aad3369

[3] Schirmer, M., Smeekens, S., Vlamakis, H., Jaeger, M., Oosting, M., & Franzosa, E. et al. (2016). Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity. Cell, 167(4), 1125-1136.e8. doi:10.1016/j.cell.2016.10.020

[4] Andrew L. Kau, Jeffrey I. Gordon. 2018. "Human Nutrition, The Gut Microbiome, And Immune System: Envisioning The Future". Nature 474 (7351): 327. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3298082/.

[5] Ursell, L., Metcalf, J., Parfrey, L., & Knight, R. (2012). Defining the human microbiome. Nutrition Reviews, 70(suppl_1), S38-S44. Retrieved from https://academic.oup.com/nutritionreviews/article-abstract/70/suppl_1/S38/1921538?redirectedFrom=fulltext

[6] Josef Neu, J. (2011). Cesarean versus Vaginal Delivery: Long term infant outcomes and the Hygiene Hypothesis. Clinics In Perinatology, 38(2), 321. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3110651/

[7] Massimo Mangino, T. (2017). Innate and adaptive immune traits are differentially affected by genetic and environmental factors. Nature Communications, 8. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5227062/

[8] 500 Functional Genomics Project. (2018). Human Functional Genomics Project. Retrieved 19 April 2018, from http://www.humanfunctionalgenomics.org/site/?page_id=82

[9] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal Of The Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320. Retrieved from https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x

[10] Members. (2018). Iso.org. Retrieved 3 May 2018, from https://www.iso.org/members.html

[11] About us. (2018). Iso.org. Retrieved 3 May 2018, from https://www.iso.org/about-us.html

[12] Developing standards. (2018). Iso.org. Retrieved 3 May 2018, from https://www.iso.org/developing-standards.html

[13] ISO Register.nl | Het nationaal register van ISO gecertificeerde bedrijven. (2018). Isoregister.nl. Retrieved 4 May 2018, from https://www.isoregister.nl/certifcering/hoe-krijg-ik-een-iso-certifcaat.html

[14] Kwaliteitskeurmerken . (2018). Umcg.nl. Retrieved 4 May 2018, from https://www.umcg.nl/NL/UMCG/Kwaliteit/kwaliteitskeurmerken/Paginas/default.aspx