universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"How cognitive integration of clinical and methodological expertise enhances data-driven analysis"

verfasst von / submitted by

Caspar Matzhold, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 013
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	MEi:CogSci - Cognitive Science
Betreut von / Supervisor:	Assoc. Prof. Priv.-Doz. Mag. Dr. Peter Klimek

Contents

Chapter 1

Abstract

The amount of medical data generated in healthcare has exponentially increased in recent years. To gain a quantitative understanding of the complex processes involved in healthcare, we need new scientific methods to adequately investigate the heterogeneous and dynamic data they produce. Data-driven analysis, when applied in a clinical context, has the aim to turn complex medical data records into knowledge on how to prevent or treat diseases more effectively. This approach requires the combination of data-related, methodological expertise with in-depth know-how of the involved clinical or medical domain. More precisely, it requires interdisciplinary collaboration, since data analysis competencies needs to be consistent with medical domain expertise. However, successful interdisciplinary research remains a great challenge. In order to transform medical data into knowledge, we need to gain new insights into factors that support interdisciplinary interaction and the emergence of cognitive processes through which disciplinary concepts and methods are integrated. Therefore, this master thesis has two objectives. First, to carry out a data-driven analysis by means of an interdisciplinary research team and, second, to combine our practical experience with theoretical knowledge about cognitive processes to gain knowledge about interdisciplinary research. In terms of methodological data analysis approaches, we applied multiple logistic regression to analyse the dosage-dependent occurrence of osteoporosis in statin patients. This study reveals that there is a highly non-trivial dependence of statin dosage with the odds of osteoporosis. To the best of our knowledge, this is the first study which shows that it is important to consider both potency and dosages when investigating the relationship of osteoporosis and statin therapy. Our results show that the diagnosis of osteoporosis is underrepresented in low-dose and overrepresented in high-dose statin treatment. The second high-level objective was to show how interdisciplinarity promotes data-driven analysis and to gain insights into factors that support the interaction process between researchers from different disciplines and the emergence of a cognitive process among them. We made use of contemporary cognitive science theories and Bourdieu's theory of practice to get an alternative perspective on interdisciplinarity. Based on these theoretical insights combined with our own research experience, we identified body—mind–environment factors that we consider important for a successful research process: personal/habitus-related qualities, environmental/field-related properties, a shared study logic and time. In summary, this work demonstrates how data-driven analysis is

promoted by interdisciplinary research.

Die Anzahl der im Gesundheitswesen generierten medizinischen Daten hat in den letzten Jahren exponentiell zugenommen. Um ein quantitatives Verständnis der komplexen Prozesse im Gesundheitswesen zu erlangen, benötigen wir neue wissenschaftliche Methoden, um die heterogenen und dynamischen Daten, die sie erzeugen, angemessen zu untersuchen. Die datengetriebene Analyse hat im klinischen Kontext das Ziel, komplexe medizinische Datensätze in Wissen umzuwandeln, wie Krankheiten wirksamer verhindert oder behandelt werden können. Dieser Ansatz erfordert die Kombination von datenbezogenem, methodischem Fachwissen und fundiertem Know-how aus dem betroffenen klinischen oder medizinischen Bereich. Genauer gesagt, erfordert dies eine interdisziplinäre Zusammenarbeit, weil die Datenanalysekompetenzen mit der Expertise im medizinischen Bereich übereinstimmen müssen. Erfolgreiche interdisziplinäre Forschung bleibt jedoch eine große Herausforderung. Um medizinische Daten in Wissen umzuwandeln, müssen neue Erkenntnisse über Faktoren gewonnen werden, die die interdisziplinäre Interaktion und die Entstehung kognitiver Prozesse unterstützen, durch die disziplinäre Konzepte und Methoden integriert werden. Daher hat diese Masterarbeit zwei Ziele. Erstens, eine datengetriebene Analyse als ein interdisziplinäres Forschungsteam durchzuführen und zweitens unsere praktische Erfahrung mit theoretischem Wissen über kognitive Prozesse zu kombinieren, um Wissen über interdisziplinäre Forschung zu gewinnen. Bezüglich der methodischen Datenanalyse-Ansätze haben wir multiple logistische Regression angewendet, um das dosisabhängige Auftreten von Osteoporose bei Patienten die mit Satinen behandelt werden zu analysieren. Diese Studie zeigt, dass ein nicht-trivialen Zusammenhang zwischen den Dosierungen von Statinen und Osteoporose besteht. Nach unserem Wissen ist dies die erste Studie, die zeigt, dass bei der Untersuchung der Beziehung zwischen Osteoporose und Statin-Therapien sowohl die Potenz als auch die Dosierung berücksichtigt werden müssen. Unsere Ergebnisse zeigen, dass Osteoporose Diagnosen bei niedriger Dosis unterrepräsentiert und bei hoher Dosis überrepräsentiert sind. Das zweite übergeordnete Ziel bestand darin, aufzuzeigen, wie Interdisziplinarität datengetriebene Analysen fördert, und Einblicke in Faktoren zu gewinnen, die den Interaktionsprozess zwischen Forschern verschiedener Disziplinen und die Entstehung eines kognitiven Prozesses unter ihnen unterstützen. Anhand zeitgenössischen kognitionswissenschaftlicher Theorien sowie Bourdieus Theorie der Praxis, erhielten wir alternative Perspektiven auf Interdisziplinarität. Basierend auf diesen theoretischen Erkenntnissen, kombiniert mit unserer eigenen Forschungserfahrung, haben wir Faktoren identifiziert, die wir für einen erfolgreichen Forschungsprozess als wichtig erachten: persönliche/habitus bezogene Eigenschaften, Umgebungsbezogene/ Feldspezifische Eigenschaften, eine gemeinsame Studienlogik und Zeit. Zusammenfassend zeigt diese Arbeit, wie datengetriebene Analysen durch interdisziplinäre Forschung gefördert werden.

Chapter 2

Introduction

The amount of medical data has increased significantly in recent years. This increase in the available information illustrates the complexity of the healthcare system and the need for new scientific methods to adequately investigate it (Miles, 2009). Data-driven analysis often pursues an exploratory approach to turn complex medical data records into knowledge on how to prevent or treat diseases more effectively. For instance, Kautzky-Willer, Thurner, and Klimek, 2017 could reveal that the use of statins offsets insulin-related cancer risks in type 2 diabetes by using medical claims data of 1.847051 patients.

The challenge of data-driven analysis is not only to make sense out of the large heterogeneous datasets, but also to discover previously unknown patterns which can be meaningfully translated into clinical practice. This is a challenge because research using secondary datasets, such as electronic health records, is confronted with the problem that these heterogeneous are often noisy and incomplete (Holzinger, Dehmer, and Jurisica, 2014). Data-driven analysis therefore requires the combination of data-related, methodological expertise and in-depth know-how from the clinical or medical sector involved. Methodological knowledge for data selection, data transformation and identification of adequate statistical analysis, and medical domain expertise for data interpretation and identification of meaningful results. The analysis of complex structures is a challenging objective that cannot be tackled within one isolated discipline (Beslon et al., 2010).

Scholars have become aware of the need to link disciplinary fields to understand complex phenomena (Aboelela et al., 2007). This recognition has led to an increased interest in interdisciplinary research within universities and laboratories all over the world (Bernini and Woods, 2014). Interdisciplinary studies are a cognitive process by which individuals or groups draw on disciplinary perspectives and integrate their insights to gain a more comprehensive understanding of a complex problem (Repko, 2008).

However, successful implementation of interdisciplinary research remains a great challenge that ranges from practical, methodological to institutional and organisational issues (Fischer, Tobi, and Ronteltap, 2011; Aboelela et al., 2007). This means we need to gain new insights into factors supporting interaction and cognitive integration processes to turn medical data into knowledge.

This master thesis has two objectives. To conduct a medical data-driven analysis as an interdisciplinary research team, and to combine our practical experience with theoretical

knowledge about cognitive processes to gain new insights on interdisciplinary research. For this and to show why interdisciplinary research is beneficial for a discovery-driven-data analysis we provide a report that portrays our study as a cognitive process that proceeds developmentally from problem to understanding (Repko, 2008).

First, we conduct a medical data-driven analysis in cooperation between the Complexity Science Hub Vienna and the Medical University of Vienna. The interdisciplinary research group combines knowledge of medicine, physics, computer science and complexity science. The analysis is based on medical claims data of all Austrians with health claims and uniquely identifiable age and sex who were alive during an observation period from January 2006 to December 20012. We apply a data-driven analysis to explore disease correlations of patients being diagnosed with type 1 diabetes, osteoporosis, depression or cancer (diseases network) and their interdependencies with drug-therapies, including cholesterol lowering and insulin providing drugs (disease–drug network). Besides descriptive statistics, we use advanced statistics like logistic regression to analyse disease correlations and temporal disease progression (trajectories).

Second, we review concepts and definitions of interdisciplinarity to understand what is its goal and when is it appropriate. Then, we broaden our understanding by drawing on current cognitive science concepts and on Bourdieu's sociology theory of practice (P. Bourdieu, 1990; P. Bourdieu, 1998b). From a cognitive science perspective, we consider cognition as a sense-making process that emerges out of the body–mind–environment interaction. Taking this view, we discuss Clark's extended mind approach to explain how environmental externalities like tools support how we process information (Clark and Chalmers, 1998). From a more liberal interpretation of Clark's concept, from a socially extended mind perspective, we describe the highest level of cognitive integration, as the unbounding of disciplinary minds (Bernini and Woods, 2014; Gallagher, 2013). We end our theoretical discussion by taking a cognitive externalist view on Bourdieu's theory of practice including the habitus-field dialectic (Fogle and Theiner, 2017). By making use of Bourdieu's theory we investigate the relationship between the structure and practice to understand the problem of bridging academic habitus of different fields.

Finally, we combine insights gained from our theoretical discussion on interdisciplinary research with those from our study to highlight factors that support the processes of interaction and cognitive integration. We will address the following two questions: Why does an interdisciplinary approach promote data-driven analysis? Which body-mind–environment factors shape interdisciplinary interaction, and cognitive integration processes?

Chapter 3

Interdisciplinary Research

Although several studies have been conducted in health science, involving experts of multiple disciplines, the concept of interdisciplinarity and the associated expectations vary across discipline (Aboelela et al., 2007). This heterogeneity of conceptions may be one of the reasons why successful interdisciplinary cooperation remains a major challenge and points out the need for a clear definition of interdisciplinary research. A definition that informs what is the goal of interdisciplinarity and when it is useful. For this purpose, we first analyse the differences between mono-, multi- and interdisciplinary research. Two definitions follow, one that reflects the common understanding among scientists, and a second that we believe best captures its nature.

3.1 Differences of Scientific Approaches

Disciplinarity (mono-disciplinarity) refers to a system of knowledge specialities called disciplines, such as sociology, medicine or physics. Each discipline has its own epistemological and methodological premise on how to explore the complexity of the world. What they all have in common is the strategy of disciplinary reductionism, which reduces complex things to simpler ones. This approach assumes, that by dividing a complex problem into its constituent parts and studying them separately from various disciplinary perspectives, the gained insights can then be combined into an understanding of the phenomenon as a whole. Although disciplinary reductionism has produced much of what we know about the world, the increased complexity of problems, like food or energy crises, make obvious that we need to develop new strategies on how to approach them. One of them is to conduct a research using insights form two or more disciplines, known as multidisciplinary (Repko, 2008).

Multidisciplinarity approaches problems by placing various valuable insights side by side without trying to integrate them as a whole. The clearly distinguished and sequential studies simply provide different consecutive disciplinary views on the object of interest. Contrary to (mono-) disciplinary research, which aims to study a phenomenon from one specific point of view, multidisciplinary studies provide at least one more perspective on the same phenomena (Repko, 2008). An example of multidisciplinary research is the study of the greenhouse effect, where scientist form two different disciplines, like biology and

chemistry present their perspectives on the particular topic. In order to gain a profound knowledge on a phenomenon, like the greenhouse effect, a process of integration of different perspectives, theoretically as well as methodologically needs to take place. This is the aim of interdisciplinary research, to make connections between the disciplinary perspectives. Interdisciplinary research approaches complex issues, problems or questions from the perspective of two or more disciplines by drawing on their insights and integrating them (Repko, 2008). The process of proactive integration of various insights is used to construct a more comprehensive understanding of the problem. In case of our greenhouse effect example, biologists and chemists would not only provide their particular insights, but rather try to understand how both insights are connected to each other, e.g. how the increase of CO2 affects the flora and how variations of the flora affect the CO2 levels. This example illustrates why interdisciplinary research is fundamental to explore complex problems. Note, one should not conclude that interdisciplinary is in general superior over multidisciplinary. What approach is adequate always depends on the problem and the related research question. Both approaches can be adequate. However, the study of complex interconnected problems always requires an interdisciplinary research design (Repko, 2008).

3.2 Defining Interdisciplinarity

The first definition is intended to provide the common understanding of the goal and how it can be achieved among scientists. Aboelela et al., 2007 conducted expert interviews and a systematic exploration of existing literature to propose a theoretical definition of interdisciplinary research. Their preliminary definition was additionally tested by 12 experienced interdisciplinary researchers. Their final definition is:

> *"Interdisciplinary research is any study or group of studies undertaken by scholars from two or more distinct scientific disciplines. The research is based upon a conceptual model that links or integrates theoretical frameworks from those dis- ciplines, uses study design and methodology that is not limited to any one field, and requires the use of perspectives and skills of the involved disciplines through- out multiple phases of the research process." (Aboelela et al., 2007, p. 341).*

We emphasize this definition because it serves as a first reference point for funding agencies and researchers themselves to identify competences and resources needed for a successful interdisciplinary cooperation. According to this definition, the goal is to design an interdisciplinary study in such a way that different discipline related concepts and methods are integrated, which would otherwise not be the case. And that this process requires the use of perspectives and skills in order to bridge the disciplinary boundaries. In our opinion, however, a definition of interdisciplinary research must take into account that the crucial practice involved, the integration of concepts and methods, is a cognitive process that emerges through the interaction among the individual researchers. To approach interdisciplinarity as cognitive process that proceeds developmentally from problem to understanding is displayed by Repko's definition:

> *"Interdisciplinary studies is a cognitive process by which individuals or groups draw on disciplinary perspectives and integrate their insights and modes of thinking to advance their understanding of a complex problem with the goal of applying the understanding to a real-world problem." (Repko, 2008, p. 102)*

In our opinion, this definition best describes interdisciplinarity because it focuses on the actual practice needed for a successful implementation, namely the emergence of a cognitive integrative process. In line with Aboelela, in his book on interdisciplinary research, Repko also points out that the use of individual skills/qualities as well as a model that guides the process is essential for a successful execution. He refers to different forms of personal qualities which function as a cognitive toolkit. According to Repko, 2008 we can distinguish between 4 types, intellectual capacities, values, traits and skills that affect the cognitive process. Due to the fact that interdisciplinary studies require the integration of multiple concepts of different disciplines, researchers have to acquire certain intellectual capacities, such as perspective taking, critical thinking or meaningful integration. How an individual perceives the world they live in, is partly guided by its own personal values. In the context of interdisciplinarity he mentions empathy, ethical consciousness, appreciation of diversity and humility as important values. As traits that are often advantageous, he lists entrepreneurship and self-reflection. And as skills that promote the interdisciplinary process he mentions communicative skills, abstract thinking and creative thinking. Regarding a model for interdisciplinary research, we would like to mention that a few models and methodological frameworks exist, like Repko's Broad Model (Repko, 2008), the blueprint model (Bernini and Woods, 2014) or the methodology for interdisciplinary research (MIR) framework (Tobi and Kampen, 2018). However, to the best of our knowledge, none of the models has been applied regularly, which is why there are no ratings or recommendations. For this reason and because we believe that most interdisciplinary studies require an individual design, we do not discuss these various models in greater depth here.
Taking together, the goal of interdisciplinary research is to combine both, theoretical and methodological aspects of two or more disciplines in order to gain a new, more comprehensive understanding of a complex phenomenon. Essential for successful research is the emergence of a cognitive integrative process. This requires overcoming the individual disciplinary bias, the disciplinary barriers. As factors that facilitate the research process we identified a model that guides the interaction, and the use of certain skills which function as a cognitive toolkit. However, as reported successful execution still remains a great challenge. One reason could be that interdisciplinary research is cognitively more demanding compared to mono- or multidisciplinary. In order to gain insights on factors that facilitate the cognitive process or reduce the individual cognitive load, we propose a to take a cognitive science perspective on interdisciplinary research.

3.3 Interdisciplinarity from a Cognitive Science Perspective

First of all, in a discussion about cognitive processes, we have to define what we mean by cognition. From a cognitive science perspective, precisely from an enactive perspective we

argue that cognition is an active sense-making process that arises (is enacted) through the body–mind–environment interaction (Stewart et al., 2010) From an embodied perspective we consider disciplines not as disembodied entities, but individuals as the bearers and beholders of disciplinary knowledge and methods. The abstract problem of disciplinary boundaries then turns into a problem of boundaries of the human mind (Bernini and Woods, 2014).

3.3.1 The Extended Mind

The concept of the extended mind, introduced Clark and Chalmers, 1998 challenges the Cartesian idea that cognition is something that happens just in the head. They argue that when the mind interacts with its environment, under certain conditions, parts of the environment function as parts of the mind. This assumption is based on the parity principle, which is central to the concept of the extended mind:

> *"If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world is (so we claim) part of the cognitive process."* (Clark and Chalmers, 1998, p. 8)

For example, from this view using a notebook to store and recall information about something like the location of a museum, is understood as an extended memory belief. Whereby the notebook functions as a constitutive part of the cognitive processes rather than mere tools or aids on which brain-bound cognition depends, in a straightforward instrumental way (Clark and Chalmers, 1998; Andy Clark, 2008). When the mind interacts, explores and manipulates external tools like the notebook, when the mind couples with these externalities, feedback, feedforward and feed-around loops emerge that promiscuously criss-cross the boundaries of brain, body, and world. These coupled systems, they claim, can be seen as complete cognitive system of its own right. The basic idea behind the extended mind concept is:

> *"In general, evolved creatures will neither store nor process information in costly ways when they can use the structure of their environment and their operations upon it as a convenient stand-in for the information processing systems concerned. That is, know only as much as you need to know to get the job done." (Andy Clark, 1989, p. 64)*

In short, we use tools to reduce our cognitive load, to process information in cost-effective manner. Through repeated interaction with externalities like tools, certain conditions arise under which they may become part of our individual sense-making process, and new ways of thinking emerge which otherwise would not exist. This means in the context of interdisciplinary research we should include the use of tools as a factor that reduces the individual cognitive load, which in return enhances the cognitive integration process.
Inspired by the extended mind concept, Bernini and Woods, 2014 propose to approach interdisciplinarity as the interaction of what they call individuals' disciplinary minds. They

suggest that the boundaries of the human mind, of the individual disciplinary minds, may under certain conditions be unbounded, allowing them to extend and interact with the environment, including other disciplinary minds. If the boundaries of the disciplinary mind become porous, enabling the mind to partially extend into the world, new forms of thought emerge which were previously difficult or impossible. In this sense, disciplinary minds discover in the interaction and extension with other minds what is possible to do and what not (Bernini and Woods, 2014; Menary, 2007). This concept of strong interdisciplinarity describes the bridging of disciplinary boundaries that enables a high-level cognitive integration process. At the highest level, when the mind partially extends into the world, a full cognitive integration takes place. This is of course desirable, but it requires an intensive and time-consuming collaboration among scientists. In this sense, we argue that there are different levels of integration, although a full integration should be the goal, most important is that it occurs to some extent at all (Klein, 2010). Whether a full or partial integration takes place depends on time and the proactive processes of interaction and integration. Bernini and Woods offer a cognitive science perspective on the body–mind–environment interaction and on how it is related to the problem of disciplinary boundaries. Accordingly, the individual disciplinary minds can also be regarded as externalities which get under the right conditions unbounded. In the same way, as external objects play a significant role in aiding cognitive processes, so can disciplinary minds.

Although the extended mind approach offers a useful alternative perspective on cognition, pointing out that externalities within the environment can have a significant impact on cognitive processes, its explanation of the unbounding of disciplinary minds is due to the criteria of the parity principle not quite appropriate. Clark lists, reliably, automatically, easy accessible as the three conditions that need to be met by external physical processes if they are to be included as part of an agent's cognitive process (Andy Clark, 2008). These criteria narrow down cognitive processes to those who are activated through reliable, automatic loops between the agent and the easily accessible tools. This means in the context of interdisciplinary research, the extended mind approach enables us to understand the functional role of tools, but it does not explain the unbounding of disciplinary minds. That's why we believe that the 'socially extended mind' thesis, introduced by Gallagher, 2013 provides a better account to explore the dynamical cognitive processes of interaction and integration.

3.3.2 The Socially Extended Mind

Gallagher, 2013 introduced the socially extended mind, as a liberal and specifically social extension of the extended mind hypothesis. In part, it was meant to move beyond the conservative parity principle criteria. Gallagher argues that these criteria do not apply to all cognition, especially when it comes to cognitive processes and activities such as problem solving, judging or interpreting. He argues that some externalities that allow you to think through a problem, without which you would not be able to think through the problem are neither reliably available nor easily accessible, as a notebook. But these externalities, although they violate Clarks' rules are still part of the cognitive process. Furthermore, they

can have a significant impact on how we accomplish certain cognitive processes. Gallagher points out that certain institutional or collective practice that has a direct impact on cognitive processes and behaviour may even introduce greater stability than is available in a single biological system. Based on these considerations, he introduced his concept of mental institutions. A mental institution is not only an institution with which we accomplish certain cognitive processes, but also without which such cognitive possesses would no longer exists. He states:

> *"In each case a mental institution*
> *1. includes cognitive practices that are produced in specific times and places, and*
> *2. is activated in ways that extend our cognitive processes when we engage with them (that is, when we interact with, or are coupled to them in the right way)."*
> *(Gallagher, 2013, p. 4)*

For example, the legal system can be described as an expression of several minds externalized and extended into the world, which then function as an external memory to accomplish certain aims, to reinforce certain behaviours and to solve certain problems. This also applies to research practices, they are not only the product of certain lived experiences or cognitive exercises, but also used as a tool to enable the emergence of the experience of the specific research practice, including the specific research thinking-style that shapes our cognitive process.

In this sense, we argue mental institution describe what Bernini and Woods' call disciplinary mind. From an enactivism perspective, a mental institution forms a coupled system between the living agent and the institutional lived experience in a way that enacts cognitive processes, that would otherwise not be possible. The basic modus operandi of this engagement, and of cognition in general is sense-making (Stewart et al., 2010). Mental institutions promote institutional specific sense-making processes such as certain research practices, and in this sense, describe what we call disciplinary bias. Although mental institution or disciplinary mind is beneficial for monodisciplinary research, in the context of interdisciplinary research we consider them as a barrier, as a cognitive bias which needs to be overcome. However, knowing that cognitive practice is produced and activated in spatial dependency can also be used to design a study in way that it promotes the interactions among disciplinary minds. Based on their own research experience Bernini and Woods proposed a conceptual model, the blueprint model that takes into account the importance of structural, temporal factors for interdisciplinary research (Bernini and Woods, 2014). The blueprint model includes a project space, in which subgroups of the interdisciplinary team can interact throughout the project and a central meeting place, the we-space. Diachronic or longer-term entanglements emerge when agents move together to explore the complexity of phenomena and thereby they slowly adjust to other disciplines-thinking styles. The we-space, is the place where all disciplinary minds interact intensively with each other and form synchronic entanglements which may activate so-called 'enhancing loops'(Bernini and Woods, 2014).

We believe Bernini and Woods suggest the following study design for an interdisciplinary research. It should be designed in way that it allows an adaptation of the individual disciplinary minds which in return promotes the unbounding of them and the emergence of a cognitive integration process among them. The two related but different forms of practice should take place in separate spaces, so that the space specific properties activate certain cognitive practice.

3.3.3 Cognitive Science Conclusion

Taking together, we argue from an embodied/ enactive view that the brain is not the place where all mental processing happens, but as part of a larger embodied and enactive system (Stewart et al., 2010; Gallagher, 2013). Through the body–mind–environment interactions different forms of system couplings emerge. These system couplings are distinguishable in terms of their exclusivity. Coupling systems that have been created out of individual tool-related experiences are more exclusive to the agent. Coupling systems that are activated, for example, by the legal systems are less exclusive, since they also depend on the lived experiences of others. What counts, in both cases is the reliability of the interaction. As Gallagher phrased it:

> *"In the context of extended cognition, where we can speak of interaction with institutions as well as with tools, instruments, technologies, etc., the point is that cognition just is any interaction or engagement that produces meaning for the agent, where the production of meaning is not just an individual enterprise."(Gallagher, 2013, p. 7)*

Clark's extended mind approach as well as the more enactive influenced socially extended mind approach increased our understanding of interdisciplinary research as a process among disciplinary minds which under the right conditions form a coupled system in a way that allows new cognitive processes to emerge, processes that would otherwise not be possible. From this point of view, we argue that interdisciplinary research is one large cognitive sense-making process between disciplinary minds that depends on time and space specific structures and practises, as well on external tools within the space. Besides skills and a model, we consider the intelligent use of externalities including tools and lived-experiences as factors that may reduce cognitive load of an individual or of individuals. To gain a deeper understanding of the relationship between the structure and practice, how disciplinary minds are produced, and why it is such a challenge to bridge them, we propose to make use of Bourdieu's theory of practice, including the habitus–field dialectic. Additionally, we aim to combine insights gained from our cognitive science discussion with those provided by Bourdieu. In this sense, we take a cognitive externalist perspective on Bourdieu's theories (Fogle and Theiner, 2017).

3.4 Bourdieu's Theory of Practice

Central to Bourdieu's considerations about functionality of the world is his theory of practice, more precisely the theory of social practice. Through practice the structure of the

world comes alive (Fogle and Theiner, 2017; P. Bourdieu, 1998b). According to Bourdieu, the social practice acts as natural unremarkable tool to effectively manipulate and reproduce the social order. In accordance with Clark, Bourdieu's concept of practice is characterised as the reliable (re-) production of acceptable and advantageous responses, which function up to a large extend automatically (Fogle and Theiner, 2017). But practice is still an active process involving extension and adaptation of operations, and therefore it cannot be reduced to an automatic or mechanic reflex. It relies upon a form of knowledge that stays close to its environmental conditions (material conditions of existence characteristics of a class condition), ready to produce certain kinds of action appropriate to the current situation (Bourdieu, 1977). According to Bourdieu, repeated practice produces a hierarchically structured space in which a variety of social fields relationally position themselves (Kneer, 2004).

3.4.1 The University Field

Bourdieu describes the world as a multi-dimensional social space (a macrocosm) in which several social fields (microcosms) are located, which function autonomous but simultaneously depend on the macrocosm (Kneer, 2004) The social space continuously differentiate itself into diverse distinctive fields of practice, such as the economic field or the university field. From an enactive perspective we claim, that each field can be understood as a self-organizing system which is operationally closed but still environment-dependent (Stewart et al., 2010). The differentiation process is primarily based on the competition for the accumulation of rare resources (forms of capital). Based on the old class model, Bourdieu lists four types of capital: the economic capital, the cultural capital, the social capital and symbolic capital. The economic capital is the most valuable one, because in contrast to the others it can be transformed into other forms of capital (Schwingel, 2003). In the university field, the embedded agents are in a permanent battle for academic capital and prestige. Through this competitive practice the hierarchically organised university field is produced and reproduced (P. Bourdieu, 1988). Bourdieu describes the academic capital as a product of the combined effects of cultural transmission by the family and the cultural transmission by the academic career. (P. Bourdieu, 1984; P. Bourdieu, 1988). Whereby, the academic capital is strongly tied to the economic capital, to earning potential which allow further academic experiences, further accumulation of academic capital (P. Bourdieu, 1988). The individual fields differ according to their field-specific logic, which justifies the practical sense of the field specific practice (Schwingel, 2003).

From a cognitive externalist perspective, the field can be described as a distributed network of social knowledge that serves as an external template for individual actions appropriate to the field, to the current situation (Fogle and Theiner, 2017). In terms of mental institution, the distributed network of social practical knowledge can form structures that support and extend our cognitive abilities. By repeated application of field-specific practice, reliable structures emerge which then affects the embodied knowledge, the habitus of the embedded agent. Different homogenous environmental conditions form different forms of field-specific practical knowledge. The habitus is the vehicle that carries the diverse forms of specific

practical knowledge on how to reproduce field-specific conditions. And the embedded agents create the field-specific practice through their repeated actions forming the field-specific structure (P. Bourdieu, 1998a). Therefore, the habitus and the field influence each other mutually. Bourdieu calls this proactive process as the habitus–field dialectic (Schwingel, 2003).

3.4.2 The Academic Habitus–Field Dialectic

The habitus is a systematic collective assembly of dispositions, which are constitutive for the agent's practical actions and thinking.

> *"The structures constitutive of a particular type of environment (e.g. the material conditions of existence characteristic of a class condition) produce habitus, systems of durable, transposable dispositions, structured structures predisposed to function as structuring structures, that is, as principles of the generation and structuring of practices and representations which can be objectively "regulated" and "regular" without in any way being the product of obedience to rules, objectively adapted to their goals without presupposing a conscious aiming at ends or an express mastery of the operations necessary to attain them and, being all this, collectively orchestrated without being the product of the orchestrating action of a conductor. " (Bourdieu, 1977, p. 72)*

It is as a form of intelligence, primarily body based that functions to a high degree automatically, nonetheless it allows a sophisticated and flexible adaptation of the world we live in, linking higher and lower functions faculties. It serves as an economical principle, by answering urgent demands of everyday life with a fluidity for which other faculties are poorly suited (Fogle and Theiner, 2017). The habitus functions as both, as modus operandi of certain practice and as the result of it, reflecting the field specific habitus. It is the link between structure and practice of a field (P. Bourdieu, 1990). And through the repeated interaction between habitus and the field (habitus–field dialectic), the power structures of the social space, including the diverse sub-fields is reproduced.
From a cognitive externalist view, the habitus is a form of knowledge that is not only embodied but also fundamentally environment-dependent. As a modus operandi of reliable practice patterns, it functions as a cognitive enhancement by reducing cognitive load and thereby enabling other higher cognitive functions to operate. And the dialectic between the habitus and the field, can be described as an interactive process between the agent's embodied knowledge (habitus) and the distributed network serving as an external template for individual schemes of perception and action (field) (Fogle and Theiner, 2017). The academic habitus encompasses the embodied knowledge of certain academic practice, like research practice including the usage of tools like a computer for computer scientist or laboratory equipment for chemist as well as epistemological forms of thinking that produce certain structures, which, over time are incorporated by the agent forming its academic habitus. In this sense, we argue that Bourdieu's concept of the academic habitus describes what we previously called the disciplinary bias.

The academic habitus is a system which students and professors maintain by conforming to the established habitus which encompasses university field-specific schemes and strategies of accumulation of more academic capital and prestige, which in turn reinforces the structural power dynamics of the university field (P. Bourdieu, 1988). The dialectic between the academic habitus and a specific university field reproduce certain barriers that challenge new ways of thinking and new approaches such as interdisciplinarity. This problem has also been recognized by Bourdieu. He criticizes that the power structure and the adoption of the established habitus of the academic order, which has produced and legitimized a conforming scholar, can negatively affect the generation of knowledge. Negative in terms of reducing the potential for the development of alternative research methods and perspectives, as the agents conform to the established standards (P. Bourdieu, 1988).
The academic habitus is the modus operandi of certain university-specific practice and simultaneously the product of it, reflecting field-specific attitudes, behaviour and thinking. In this sense, the habitus is the product of past experiences, and at the same time it ensures their presence, which is reflected in form of perception, thinking and actions schemes:

> *"The habitus, a product of history, produces individual and collective practices - more history - in accordance with the schemes generated by history. It ensures the active presence of past experiences, which, deposited in each organism in the form of schemes of perception, thought and action, tend to guarantee the "correctness" of practices and their constancy over time, more reliably than all formal rules and explicit norms." (P. Bourdieu, 1990, p. 54)*

The vehicle of these schemes is the habitus which serves the agent as an orientation within the social space, encompassing the practical sense and the feel for the game, a template for field specific accurate thinking and acting, as well as the social sense, including sense for morality for fairness etc. (Schwingel, 2003). It is composed of lived experiences and those lived by others. Whereby the habitus must always be seen relational to the agent's position within the social space. Since the position defines the environmental conditions and therefore the habitus. This means a habitus can be understood not only as a durable system of disciplinary research practice and thinking, but also as a system of personal views on politics, faith, norms, values etc., which are termed personal bias. A personal bias is often even more subtle, since it has been developed one's whole live (Repko, 2008). Therefore we argue, that the problem of overcoming the individual disciplinary bias should also take into account personal bias.
Taking together, Bourdieu's theory of practice, including the habitus–field dialectic offers a deeper understanding of why the unbounding of the disciplinary minds, of the individual academic habitus is such a challenge. The main reason is, that academic habitus is a durable system of dispositions that shape our perception, thought and action. This embodied knowledge has been produced in relation to a field-specific position, to specific forms of distributed knowledge, starting from early childhood on over the academic carrier to date. The habitus is the product of a complex interconnected system that depends on capital (position within the field), lived experiences and the choice of the academic carrier. Although any academic habitus has agreed to act according to the logic of the university

field, to accumulate more academic capital and prestige, each sub-field has its own logic on how to achieve that. Individuals who share a similar position within a field, have a similar logic, a similar sense-making process. This similarity reproduce itself through the habitus–field dialectic and forms the individual academic habitus, a system of durable. On the one hand this process fosters field-specific actions, like interaction and integration, but on the other hand it reduces the potential to develop alternative practice like new research methods or ways of thinking. This reproductive process of similarity challenges the interdisciplinarity field that aims to make use of the power of heterogeneity. To break through this reproductive process, through the disciplinary barriers, we believe it is necessary to create a common shared logic. One that corresponds to the logic of the individual sense-making processes. However, to achieve a high level of cognitive integration process, as proposed by Bernini and Woods, we think the creation of a new university sub-field, which we call the interdisciplinary field is required. Like any other university sub-field, the interdisciplinary field would follow the logic of accumulating academic capital, prestige as well as money. This means interdisciplinary research must be attractive in terms of academic capital and prestige in order to be attractive to agents embedded within the field. As long as the interdisciplinary research is somehow outside the university field because it does not belong to one of the established disciplines, not enough agents will participate. Without repeated practice the necessary structures cannot arise.

3.5 Conclusion of The Cognitive Externalist Discussion

The aim of this section was to define the nature of interdisciplinarity and to identify factors that influence the interdisciplinary research process. We first discussed the differences between the various research approaches, followed by two definitions to understand what is the goal of interdisciplinary research and when it is appropriate. Here we want to present our definition of interdisciplinarity which reflects our insights:

> *Interdisciplinarity describes a process of proactive interaction between at least two scientists with a different expertise in such a way that a meaningful cognitive integration emerges to advance their understanding of a complex problem.*

As factors that promote the process we first identified a conceptual model and the use of specific skills. By drawing on current cognitive science perspectives we gained new insights on the factors promoting the integrative process. From an extend mind perspective we consider the intelligent use of tools as a way to reduce the individual cognitive load, which in return fosters the cognitive integration. And a high-level integration process can be achieved through the unbounding of the individual disciplinary minds. By using Bourdieu's theoretical perspective we understood that the problem of disciplinary barriers, the barrier of the academic habitus, is a problem of shared similar dispositions (embodied knowledge) that reproduces field specific practice and thinking styles. This similarity reproduction gives raise to the power structure of the university field, reducing the potential to develop alternative research methods and perspectives, as the agents conform to the established standards. As a possible way to break through this practice–structure process, we identified

the creation of a common shared logic that corresponds to the different sense-making processes. For strong interdisciplinarity however we think it has to become a field itself. To achieve this, the interdisciplinary field needs to be more attractive in terms of academic capital and prestige.

Chapter 4

Report of an Interdisciplinary Data-Driven Analysis

Complex diseases like diabetes or cancer are increasing rapidly in prevalence worldwide. The International Diabetes Federation has estimated that the number of diabetes patients will increase up to 592 million by 2035 (Shi and Hu, 2014). This identifies the need to gain new knowledge that helps to recognize, prevent and treat diseases.

In medicine, clinical trials are constantly generating more knowledge, which helps us to better understand the complexity of diseases and to develop new drugs for their treatment. However, due to the nature of the clinical trials these findings are limited. The relatively small clinical trials and the homogenous sample sizes do not allow to detect all adverse drug events (Tatonetti et al., 2012).

Serious side effects and drug interactions remain a significant source of mortality and morbidity around the world with costs estimated at several billion dollars each year (Bates et al., 1995; Classen et al., 1997). The detection of adverse drug events requires extensive medical data that provide information on long-term effects and interactions with other drugs and diseases.

Fortunately, the amount of medical data generated in healthcare has exponentially increased in recent years and these databases present the opportunity to study drug effects from patient population data (Tatonetti et al., 2012). This increase in medical data allows new research methods, such as data-driven analysis to discover hidden patterns of interaction that have not been detected in clinical trials. Data-driven analysis, when applied in a clinical context, has the aim to turn complex medical data records into knowledge to predict adverse drug events so that future patients can be protected from the sometimes serious consequences (Tatonetti et al., 2012).

For instance, Kautzky-Willer, Thurner, and Klimek, 2017 used medical claims data of 1,847,051 patients to investigate interconnections between type 2 diabetes and cancer. They explored the impact of different type 2 diabetes therapies on cancer risks, revealing that the use of statins offsets insulin-related cancer risks. This insight allows them to make recommendations on how treatment can become more efficient. In another data-driven study, Tatonetti et al., 2012 investigated clinical databases to predict drug effects and interactions. They discovered adverse drug events in patients treated with combined

therapy of pravastatin and paroxetine. Both drugs are among the most widely prescribed medications in the world, but the synergistic side effect, that interaction between both drugs increases blood-glucose levels was previously unknown. These examples illustrate the high value of data-driven analysis in medical research.
The challenge of data-driven analysis is not only to make sense out of the large heterogeneous datasets, but also to discover previously unknown patterns which can be meaningfully translated into clinical practice. This is a challenge because research using secondary datasets, such as electronic health records, is confronted with the problem that these heterogeneous and distributed records are often noisy and incomplete. According to Holzinger, Dehmer, and Jurisica, 2014 this problem can be divided into three categories:

- Heterogeneous data sources (data integration)
- Complexity of the data
- Noisy, uncertain data (data cleaning)

These categories demonstrate why data-driven analysis requires the combination of data-related, methodological expertise with in-depth know-how of the involved clinical or medical domain. Methodological knowledge for data selection, data transformation and identification of adequate statistical analysis, and medical domain expertise for data interpretation and identification of meaningful results. Therefore, the analysis of complex structures cannot be tackled within one isolated discipline (Beslon et al., 2010).
More precisely, it requires interdisciplinary collaboration, since data analysis competencies needs to be consistent with medical domain expertise. Interdisciplinary research creates a feedback loop between individual carriers of methodological and medical expertise who support the research processes of data cleansing, data understanding and knowledge discovery. Using the knowledge of two disciplines increases the chances to discover new patterns of drug-drug and diseases-drug interactions, enabling better prediction of adverse drug events and improved disease treatments.
A study analysis of 17,9 million papers covering all scientific fields revealed that teams are 37.7% more likely to insert novel combinations of knowledge into familiar knowledge domains than solo authors (Uzzi et al., 2013). This illustrates that the power of interdisciplinary research fosters the emergence of new knowledge that benefits the individual disciplines involved.

4.1 Objectives

This report portrays how we approached a discovery driven data analysis as an interdisciplinary research team. We pursue two goals: first, to generate new knowledge on how to prevent or treat diseases more effectively by applying a data-driven analysis, and second, to show why such a study requires the cooperation of different disciplines, why interdisciplinarity is beneficial. We not only present the methods we used and the obtained results, including preliminary and main results, but also a review of the underlying

research process itself, how we proceeded developmentally from problem to understanding. As pointed out by Holzinger, Dehmer, and Jurisica, 2014 and colleagues, the main challenge of discovery-driven data analysis is not only to make sense out of the data, but extract hidden patterns that are meaningful. This report should make it clear why an exploratory data-driven analysis is a challenging objective that cannot be tackled within a single isolated discipline.

The report consists of 4 parts. It contains a description of our group and the meetings, of the data, of the individual research phases and a discussion of the main results. In addition, we describe in detail in the appendix how we extracted our final results. By the descriptions of the individual meetings in the report, whereby one description portrays the process of several meetings, we show how we discussed as a group what can be studied based on the data, why it could be relevant and how best to do it. This requires the combined use of methodological and medical expertise. Former expertise is needed to identify the appropriate methods, interpret the statistical results and present them properly. Medical knowledge for the interpretation of the quality of the data, and for the identification of meaningful results. Therefore, this study was carried out in cooperation between the Complexity Science Hub Vienna and the Medical University of Vienna. To simplify the identification of the two different groups in the report, we refer to members of the Complexity Science Hub Vienna as, the hub-group and to the doctors from the Medical University as the medical-group.

4.2 Group and Meeting Description

The number of participants of both groups varied over the whole research process from 3-7 for each group. The medical-group included experts in various medical specialities such as gender medicine, diabetes or mental illness. Together with the hub-group, which consisted of experts in physics, computer science and complexity science, the interdisciplinary research group had the necessary knowledge to carry out this study.

At regular held meeting, at least once a month, both groups explored the medical datasets aiming to interpret and discover meaningful patterns of complex diseases. The meetings took place at medical university of Vienna in a room which includes a table with enough space for any group member, a beamer for power-point presentations and a blackboard used to externalize urgent issues in diverse forms. Besides these meetings, in which all members participated, we also had meetings of our subgroup, which was formed during our research process. In total, we had around 27 meetings, of which 12 were subgroup meetings.

The practise of these meetings was structured as follows: First each group member gave a brief overview of what he/she had been working on over the past week, pointing out relevant findings. Then the other members added information that they considered being useful, and ideas on how the presented project might be related to their own research. In addition, these meetings were used to discuss research-based specific problems, mainly methodological issues but also fundamental issues were addressed.

Furthermore, in weekly held meetings at the Complexity Science Hub all members of the Hub, who worked on various projects under the umbrella of complexity science, shared results and findings which might be relevant to others. These meetings offer the great opportunity to identify possible collaborations among the various projects and to gain insights on how other researcher approach complex phenomena.

4.3 Data

The study is based on medical claims data of patients with hospital stays during 2006–2012. All patients were anonymized and described by a fictitious identification number (ID). In total, we had 5 different datasets which can be divided into 3 parts:

1 Datasets P1 and P2 contain information on the year of birth and sex of patients. P1 encompasses all Austrian who were alive from January 2006 to December 2007. The sample size consisted of 7,945,775 (male= 3,712,939, female= 4,232,836). P2 consists of 1,674,266 patients (male=873,332, female= 953,577) from Lower Austria who were alive from December 2006 to January 2012

2 Datasets D1 and D2 provide information on 1,642 different diagnoses, including secondary diagnosis, which patients have received during a hospital stay. For each diagnose we extracted the corresponding International Classification of Diseases, 10th revision (ICD10) codes. D1 contains a list of diagnoses for patients of dataset P1 and D2 for the corresponding patients of dataset P2.

3 Dataset M contains information on prescriptions of 59 different drugs that allows us to determine which anonymised patient has received on what day how much of a certain medication. For each drug, we extracted the corresponding Anatomical Therapeutic Chemical (ATC) Classification System codes.

4.4 Review of the Research Process

The following review aims to portray how we approached a discovery-driven data analysis as an interdisciplinary research team. Whereas we only report on steps which had a direct impact on the research process. This means we do not provide a report of all meetings neither on all conducted analyses. Retroperspective the research process can be divided into 3 phases. In phase 1, data-cleaning, phase 2, understanding the data, and phase 3, identification of meaningful patterns of interaction.
Note, all results in phase 1 and 2 are preliminary, so are the graphics which display them. This means we do not guarantee for their correctness. Since the presentation of our interaction as an interdisciplinary research group is part of the objective of this review, we provide the original graphics and slides used to make sense out of the data. Therefore, the figures may not always be aesthetically pleasing. In phase 3 we present the main findings of our study. The methods and results in phase 3 as well as their interpretation are to be

regarded as valid. Since we intend to publish our key findings, phase 3, and in particular the discussion of the results, may already include excerpts of our paper.

4.4.1 Phase 1: Data-Cleaning

Meeting: "Kick-off Meeting"

The purpose of the first meeting of our interdisciplinary investigation was to get to know each other and to identify a common goal. Although some members of both groups already knew each other from previous studies, we started with a short introduction round. Every member presented his/her background and stated their research interest. To identify a common goal, we then discussed the characteristics of the different datasets to understand what can be studied. Concluding that dataset P2 offers the opportunity to investigate medical-treatment-trajectories in relation to certain disease. Due to the fact that diabetes mellitus is one of the most prevalent chronic diseases in Western society (Shi and Hu, 2014) and that some members of our research group recently have conducted a study on associations between different type 2 diabetes therapies and their impact on cancer risks, we decided to further investigate the complexity of the diabetes mellitus (Kautzky-Willer, Thurner, and Klimek, 2017). As our first research interest, we agreed to analyse disease risks of patients with type 1 diabetes mellitus (DT1) in relation to their medical treatments. Note, in the following analysis we refer to a disease risk as the relationship between a diagnosis of a disease and a patient-level variable such as prescription of a drug or sex. We then discussed comorbid diseases and drugs that are relevant to the study. The medical-group identified cancer, depression and osteoporosis as the most interesting comorbidities for DT1 patients. The hub-group presented the individual drugs that are contained in the medication prescription dataset M. For the first part of the drug–diseases analysis, we identified 12 insulins that can be qualitatively distinguished in terms of their medical effect into short-acting, intermediate and long-acting insulins. Our first research question was, RQ1:

> *Is there a correlation between the three qualitatively different types of insulin therapies and the risks of being diagnosed with cancer, osteoporosis and depression in patients with diabetes mellitus type 1?*

Phase1: Exploring diabetes mellitus type 1 treatment-dependent comorbidities

The analysis began by examining the research question agreed upon in the previous interdisciplinary meeting. We agreed to identify diabetes mellitus type 1 patients and possible co-occurring diseases (comorbidities) based on diagnoses made during a hospitalization. The individual diagnosis of a patient, regardless if main or side diagnoses, has been identified according to the ICD codes:

- E11: diabetes mellitus type1 patients
- F32: depression

- C00–C88: different types of cancer
- M80– M82: different types of osteoporosis

Out of the 1,826,909 patients in the P2 dataset, we identified 4,716 patients with uniquely identifiable age and sex who received the diagnoses diabetes mellitus type1 (DT1). These 4,716 DT1 patients (male=2,498; female= 2,218) with an average age of 52.93, represented our group DT1. From this group, we identified 549 individuals diagnosed with cancer, 403 diagnosed with depression and 262 patients diagnosed with osteoporosis. The individual treatment histories for group DT1 were extracted based on dataset M. Out of the 59 possible variables, 12 were extracted, including 5 different fast-acting insulins, 4 intermediate-acting insulins and 3 long-acting insulins. According to the different types of insulin therapies group DT1 was divided into 3 sub-groups:

- Group 1 encompasses all DT1 patients treated with fast-acting insulins and analogues for injection (ATC code starting with A10AB) N= 1,856
- Group 2 contained DT1 patients treated with intermediate- or long-acting insulin analogues for injection (ATC code starting with A10AD) N= 1,507
- Group 3 included DT1 patients treated with long-acting insulins and analogues for injection (ATC code starting with A10AE) N= 1,078

All 3 sub-groups encompassed together 4,441 DT1 patients, including multiple entries. This means that a patient can be a member of group 1 and at the same time listed as member in group 2. As uniquely identifiable patients treated with at least one of three insulin therapies we obtained a total number of 2,992. The different case numbers of patients with diagnosis and those who have received medical treatment indicated that we are missing information on insulin therapies from 1,724 DT1 patients. Due to the difference between the expected and observed case numbers, we decided to test the consistency of the individual insulin prescription histories. Based on the prescription dates, we extracted the time series of each insulin and visualised them as bar charts. In order to interpret the results, we consulted the medical-group.

Meeting: Making sense of data

The hub-group presented the obtained results of the analysis, the number of identified DT1 patients and the insulin based grouped case numbers. When we referred to the different case numbers, that we are missing from 1,724 patient information regarding their insulin treatment, the medical- group pointed out that this cannot be the case. They explained that diabetes mellitus type 1 requires constant insulin therapy. It is impossible for a DT1 patient to not receive any kind of insulin therapy. To find out what could have caused this difference, we discussed the individual prescription time series using bar charts. Fig. 1 displays the time series of all intermediate- or long-acting insulin analogues for injection. Apart from the obvious difference in prescription frequency before and after 2008, the time series seemed to be consistent. The medical-group explained that this difference was due

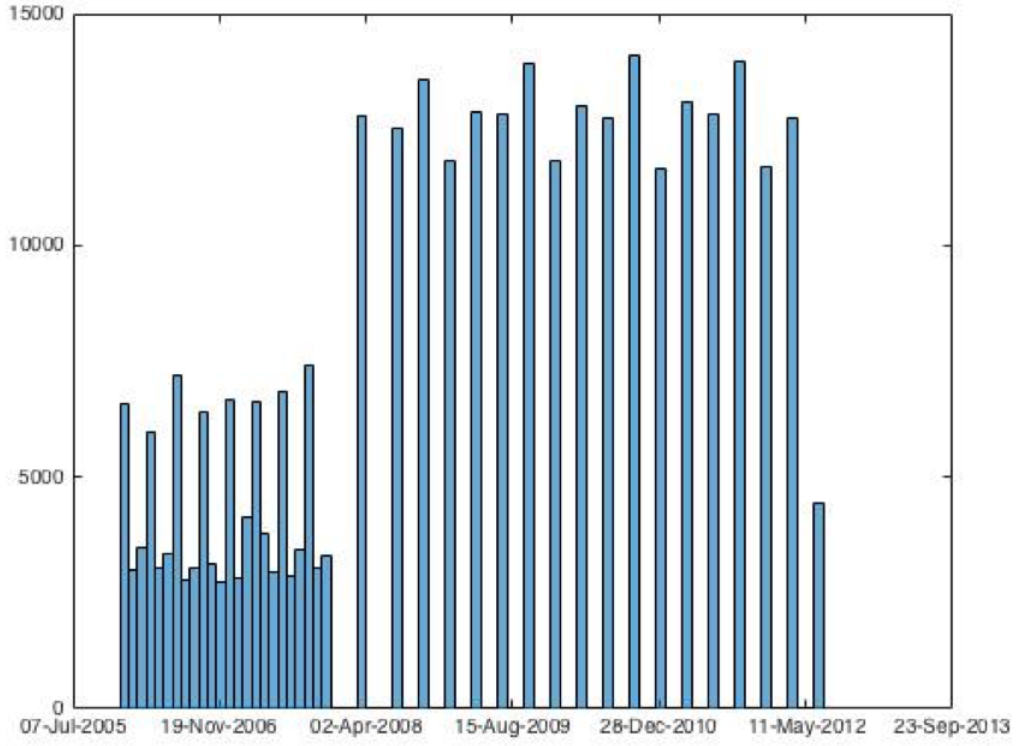

Figure 1: Prescription Time Series of Intermediate- or Long Acting Insulins

to the change of the medical prescription guidelines. The fact that we were able to observe this change reinforced our assumption that the dataset M provides reliable information. As an alternative explanation for the noise of the data, the doctors assumed, based on their own practical experience, that DT1 patients were accidentally entered as diabetes type 2 patients. Anyways we decided as a group to change our methodological strategy for the extraction of diabetes type 1 patients. Instead of diagnoses, we use DT1-related insulin therapies to identify the group. To get a meaningful DT1 group based on medical treatments, we set 4 inclusion-criteria that needed to be met:

C1 Valid Treatment

Only patients who were treated with intermediate-acting insulins (ATC code starting with A10AD), fast-acting insulins (ATC code starting with A10AB) or with the drug combination of long-acting and fast-acting insulins (ATC code starting with A10AB and A10AE) were classified as valid diabetes mellitus type 1 patients.

C2 Regularly applied treatment

Only if a patient had a minimum of 8 different prescription entries in terms of dates of a given drug, we assigned him/her as being a valid drug user of the corresponding medical treatment. This is equivalent to a medication treatment, regularly applied over two years.

C3 Observation minimum

Medications with at least 35 prescriptions were included, otherwise excluded for further analysis.

C4 Daily-defined-dosage (DDD) maximum

Patients who had a daily-dosage average of a particular drug lower than the maximum dosage recommended by Vienna General Hospital were included, otherwise they were excluded as a valid drug user of the specific medication. The recommended maximum dosage was for all insulins the same, 200 units per day (information is available on request: *(AKH) Allgemeines Krankenhaus der Stadt Wien* n.d.).

Though we changed our methodological approach to identify DT1 patients, our research interest remained unchanged (RQ1).

Phase 1: Treatment-based identification of DT1 patients

Based on criterion 1, valid treatment, we identified 18,057 patients treated with intermediate-, fast-acting insulins and/or treated with either one of the 15 possible (5 fast-acting x 3 long-acting) valid drug combinations. According to criterion 2 out of 18,057 we extracted those patients with at least 8 different dates of prescriptions, reducing the number of valid patients to 11,466. Each of the 12 medications was then reviewed for the number of observations. Out of the 12 different insulins, 3 medications with less than 35 observations were excluded (C3). This was the case for 1 fast-acting insulin, 1 intermediate-acting insulins and 1 long-acting insulin. In order to identify patients with a daily dosage average lower than the maximum recommended by the Vienna General Hospital, we had to calculate the daily average for each insulin for each patient individually. To obtain the individual averages, we extracted the individual drug histories, including information on dates of received prescriptions and the corresponding amount of the given insulin. The average was calculated by dividing the sum of all amounts of the administered drugs by the sum of treatment days, minus the days a patient spent in a hospital. The hospital days were deducted on the assumption that patients were provided with insulin for the time spent in the hospital. Patients with a higher daily dosage average than recommended were then excluded (C4), reducing the number of valid patients to 11,426 (male= 5,375; female= 6,051) with an average age of 61.75. The extraction process is displayed by Tab1 1.

Medication	Insulin	Filter C2	Filter C4
A10AB	8,769	4,501	4,491
A10AD	11,956	6,814	6,781
A10AE	4,249	2,199	2,194
Valid-Combination	B-E = 3,695	B- E= 1,984	B-E = 1,978
Total	18,057	11,466	11,426

Table 1: Extraction Summary

In order to understand the quality of our new cohort, we additionally identified patients being diagnosed with diabetes mellitus type 1 (ICD code E14) and those being diagnosed with diabetes mellitus type 2 (ICD code E11). Out of the 11,426 patients, 22% had a DT1 and 41.6% a DT2 diagnosis. These results confirmed the assumption of the medical-group, that the diabetes mellitus data may be noisy probably due to incorrectly entered but not incorrectly generated diagnoses. To determine how to continue with the analysis, we consulted the medical-group.

Meeting: Making sense of the group

The hub-group reported first the identified problem, that the medical treatment based defined DT1 group is still noisy in terms of wrong or missing diagnoses. The medical-group explained although our valid treatment criterion is mostly related to DT1 patients, we cannot rule out that this group contains patients diagnosed with diabetes mellitus type 2, as they are also being treated with these types of insulin. However, we agreed that our methodological approach remains the best way to identify DT1 patients, but also that we need additional information. Information that allows a meaningful investigation of RQ1 and interpretation of obtained results. For this purpose, we first discussed, how the quality of our dataset can generally be assessed and through which information we can validate the results of group DT1 and secondly, what are possible confounding variables. In order to assess the quality of our data and of our DT1 analysis, we have decided to design our analysis similar to that used by Kautzky-Willer, Thurner, and Klimek, 2017 who investigate the association between individual diabetes mellitus type 2 therapies, statin use and site-specific cancer risks. We wanted to find out, based on our dataset, if we can identify the same trend as revealed by Kautzky-Willer et al., that the use of statin offsets insulin-related cancer risks. However, we were interested in investigating the statin effect for patients treated with insulin therapies for DT1, not DT2 patients. The research interest was, RQ2:

> *Is there a correlation between DT1 insulin therapies, statin use and site-specific cancer risks?*

In order to analyses RQ2 and RQ1, possible association between individual DT1 insulin therapies and the risks of being diagnosed with cancer, osteoporosis and depression, we discussed common side medications that may affect insulin-related risks. Based on considerations of the medical-group and in relation to the available information about drug treatments contained in M we identified 31 possible confounding drugs (variables) which can be divided into 3 groups according to their medical effect: antidiabetics, cholesterol-lowering drugs and glucose-lowering drugs excluding insulins. Whereas, we were interested in investigating the first 2 types of drug in combination with insulins, group 3 was regarded solely as a confounding variable, since it acts like insulins by lowering the glucose level. The adapted research interest was formulated as follows, RQ3:

Is there a relationship between the individual DT1 insulin therapies, the use of additional medications and the risks of being diagnosed with cancer, osteoporosis and depression?

To investigate insulin–diseases relations as well as insulin–drug interactions, we decided to extract different types of DT1 groups, namely one group with DT1 patients who received additional medication and one group that excludes all patients treated with a specific side medication.

4.4.2 Phase 2: Understanding the Data

To investigate the new, adapted research questions RQ2 and RQ3 we first had to identify patients treated with the confounding drugs/variables. We extracted 31 different drugs and grouped them according to their medical effect into group statin (cholesterol-lowering drugs), group fibrate (lipid-lowering drugs), group metformin (antidiabetics) and group A10BX (glucose-lowering drugs excluding insulins). All groups were extracted according to the corresponding ATC codes:

- Group Statin consisted of patients treated with either one of the 7 different statins (ATC code starting with C10AA).
- Group Metformin comprised patients treated with metformin (ATC code A10BA02).
- Group BX encompasses all patients treated with least one of the 23 different glucose-lowering drugs that are not insulins (ATC code starting with A10BB, A10BD, A10BF, A10BG, A10BH and A10BX

As decided by the interdisciplinary research group, all members of group BX were excluded for the further analysis to eliminate the confounding effect of glucose-lowering drugs that are not insulins. The 12 remaining drugs were extracted according to the defined criterion 2: regular applied treatment and criterion 3: observation minimum. As a result of these two criteria, 2 out of the 7 statins were excluded, reducing the number of drugs to 6 (5 statins and metformin). Based on the 2 groups, treated as variables we have then formed different DT1 groups, first to understand the effect of insulins by eliminating all possible confounding variables, and second, to investigate the effect of insulin therapies in combination with one of the other variables. Patients were grouped according to whether they were treated with insulins, metformin or statins. As shown in Tab. 2, the 4 new treatment-based generated groups were named after whether they included or excluded DT1 patients or whether they consisted of a combination of DT1 patients with other patients. Group Diabetes Type1 consisted of all DT1 patients who can additionally be treated with metformin, but who had not been treated with statins. Group Diabetes Type1 and Statin included all DT1 patients, who can be additionally be treated with statins and/or metformin. Group Diabetes Type1 and Statin-2 comprised all DT1 patients who may also be treated with statins, but not with metformin. Group Statin or Diabetes Type1 consisted of statin patients and according to the 4 criteria for valid insulin patients. Patient of this group can additionally be treated with metformin. Group Statin included all patients who received

	Diabetes Type1	Diabetes Type1 & Statin	Diabetes Type1 & Statin 2	Statin OR Diabetes Type1	Statin OR Insulin	Statin	Control
Medication	- comb. Insulin Medication	- comb. Insulin Medication and Statin	- comb. Insulin Medication and Statin	-Statin or comb. Insulin Medication	-Statin or Insulin	-Statin	
Additional Medication	Metformin	Metformin		Metformin	Metformin		
Excluded Medication	Statin		Metformin			-Metformin -Insulin	-Metformin -Insulin -Statin
Excluded for All Groups	Medication A10BX	Medication A10BX	Medication A10BX	Medication A10BX	Medication A10BX	Medication A10BX	Medication A10BX

Table 2: Extraction Summary

statin treatment, excluding those who were treated with metformin or either with one of the 12 different insulins. The control group consisted of all patients of dataset P2 who were not treated with any of these medications, metformin, insulin, statin or glucose-lowering drugs (group BX). Tab. 3 provides an overview of the quantitative differences between the 6 groups. Each of these 5 groups were compared to the control group to analyse sex- and age- dependent disease risks, and odds-ratios (ORs) of being diagnosed with depression, cancer or osteoporosis. The individual results were then compared to examine possible differences between the groups, whether including or excluding confounding variables affect the overall risk of a given group (RQ1, RQ3). To investigate associations between DT1 insulin therapies, statin use and site-specific cancer risks, we computed age- and sex-dependent odds-ratios of 88 different cancer types (RQ2). The individual effect of each single medication was calculated by logistic regression. Due to the different age averages of the individual groups compared to the control group (see Tab. 3), we decided for the regression analysis to include only patients above 50 years to rule out age as a confounding variable when comparing patients with and without treatment. Each patient was described according to the number of medications in a particular group. For example, patients of Group Diabetes Type1 and Statin, were described by 18 binary variables, each one corresponding to a drug (12 insulins, 5 statins, metformin). The variable was 1 if the drug was prescribed to a patient, and 0 if not. Among the 2^{18} possible drug combinations those actually occurring were identified, whereby only drug combinations with at least 32 observations (patients) were included. Each medication group was weighted by their number of patients. Based on the remaining medication groups we calculated the risk of being diagnosed with one of the comorbidities and the site-specific cancer risks. The obtained comorbidity risks for the individual drugs was computed as the relative number of patients who had been diagnosed with the particular comorbidity. To control for age dependencies, we used the average age of all patients of a given medication group. The obtained results were discussed in meetings with the medical-group.

	Diabetes Type1	Diabetes Type1 & Statin	Diabetes Type1 & Statin No Metformin	Statin OR Diabetes Type1	Statin	Control
N:	3,399	8,391	5,255	71,023	54,715	1,485,191
Male	50%	49%	53%	45%	44%	48%
Female	50%	51%	47%	55%	56%	52%
Mean Age	61.6	70	67	76	77	43
DT1	26.5%	21%	25%	2.6%	0.04%	0%
DT2	30%	42%	37%	9.1%	2%	0%

Table 3: Extraction Summary

Meeting: Investigating insulin–disease and insulin–drug interactions

At this stage of the research process, the primary goal was still to make sense out of the data and to detect relevant differences and trends by which hypotheses can be generated. Note, as mentioned earlier, the following results are preliminary and the graphs and tables are original versions used in the meetings. Furthermore, I would like to point out once again that in this review we only present results which had an impact on the further research process. Due to the large number of discussed results, for instance alone in one meeting we discussed more than 50 slides, we cannot report all events. The individual research questions were discussed based on the corresponding visualised results. Fig. 2, 3 and Fig. 5 show the relationship between the individual drugs of a particular group and the risk of being diagnosed with depression, cancer and osteoporosis for female (red) and male (blue) patients. The composite disease risk is displayed by group both (green). The results shown are logarithmic odds (log-odds) obtained from the individual logistic regression analyses. For each log-odd, we have assigned the standard errors to identify significant results.

> *RQ1: Is there a correlation between the three qualitatively different types of insulin therapies and the risks of being diagnosed with cancer, osteoporosis and depression in patients with type 1 diabetes mellitus?*

RQ1 was discussed based on the results of group Diabetes Type1, because the group-properties allowed an assessment of the different insulin effects by excluding possible statin effects. As shown by Fig. 2, there is a trend towards increased risk of being diagnosed with osteoporosis and depression for female (red) DT1 patients treated with fast-acting insulins. However, the hub-group explained that interpretation of the results is limited, due to the

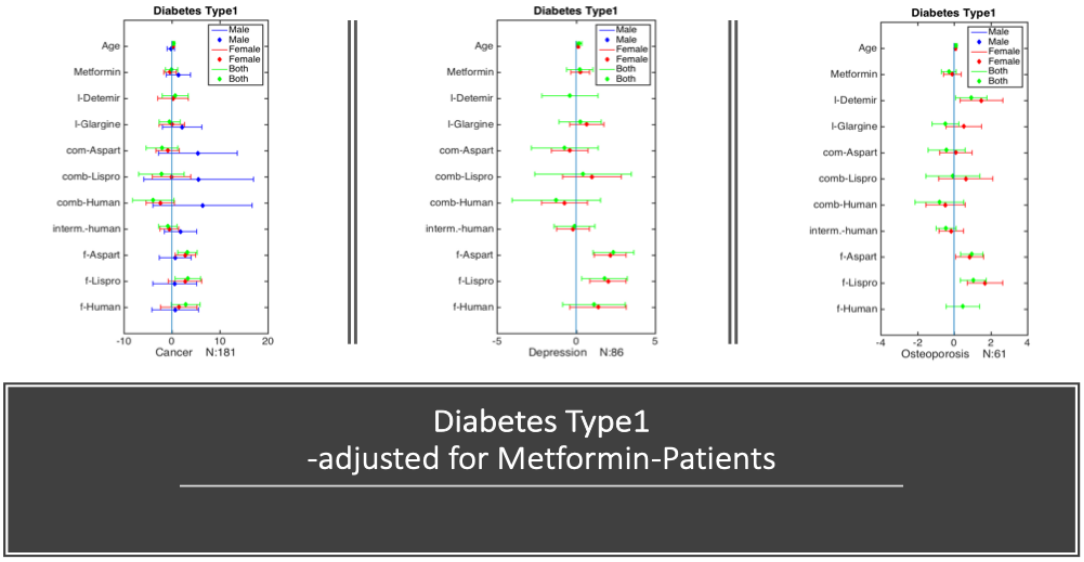

Figure 2

small case numbers (depression: 86; osteoporosis: 61). This problem was illustrated by the fact that no results were found for men (blue) as the number of male-observations was below 32 and therefore excluded from the analysis.

> *RQ3: Is there a correlation between the individual DT1 insulin therapies, the use of additional medications and the risks of being diagnosed with cancer, osteoporosis and depression?*

RQ3, was discussed by group Diabetes Type1 and Statin, which allowed the interpretation of possible effects between the individual insulins and the confounding variables, statin and metformin. Comparing the results of Fig. 2 and Fig. 3, we concluded that metformin does not affect insulin-related risks of being diagnosed with depression or osteoporosis. For patients treated with statins we could observe a decreased cancer risks. This trend was further investigated by RQ2 in Fig. 4 and Fig. 5

> *RQ2: Is there a correlation between insulin therapies, statin use and the risks of being diagnosed with cancer?*

Significant ORs for male and female patients between site-specific cancer risks and the different groups of medications compared to the control group were displayed by Fig 4. We observed the trend that DT1 insulin-related therapies are associated with an increased cancer risk (red), whereas statin-treated patients had decreased odds (blue). In this sense, we replicated the reported trend by Kautzky-Willer, Thurner, and Klimek, 2017 that the use of statin offsets insulin related cancer risks. The relationship between the individual statin therapies was then further discussed by Fig. 5 which verifies partly the previous observed trend, that some statin therapies are associated with a decreased cancer risk (left plot). This was the case for Lovastatin and Pravastatin. As noted by the medical-group there seems to be another trend, that statins with higher potency (Simvastatin,

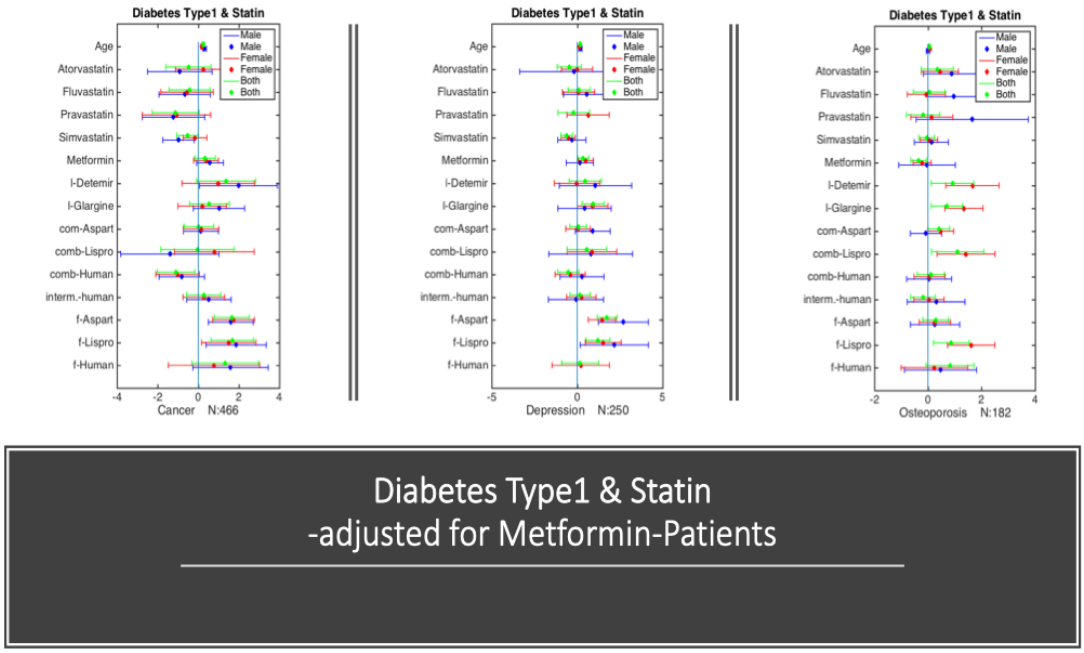

Figure 3

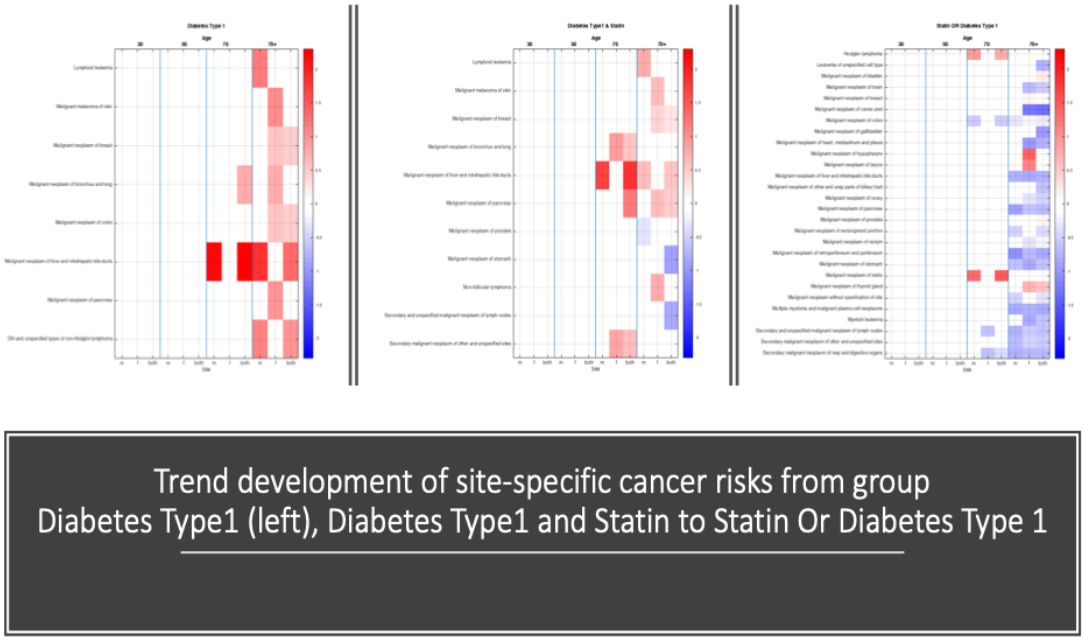

Figure 4

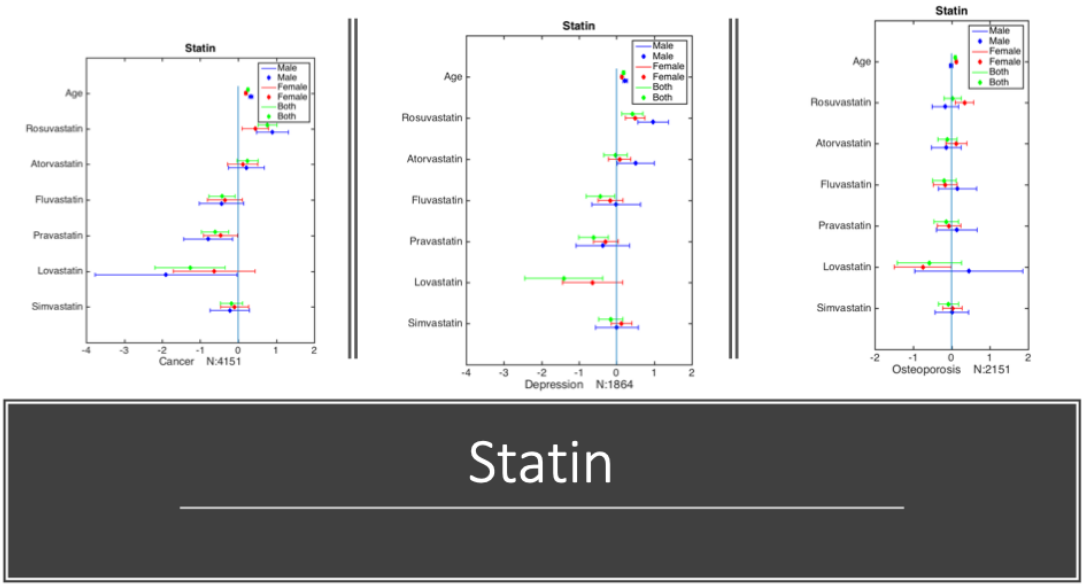

Figure 5

Atorvastatin and Rosuvastatin) correlate with an increased cancer and depression risk (middle plot) in both sexes and with an increased osteoporosis risk for females (right plot). This was surprising since statins should not increase the risk of diseases like cancer or osteoporosis (Kautzky-Willer, Thurner, and Klimek, 2017; Lin, Chou, et al., 2018). The medical-group pointed out that we might have discovered a meaningful trend that is worth being further explored.

Since we still struggled to identify a meaningful DT1 group and therefore also to interpret the results, we have decided as a group to change our research interest from insulin to statin therapies. The new research question was, RQ4:

Is there a relationship between disease risks and the different statin therapies?

Due to the new research interest, we decided to redefine group Statin. For this purpose, we discussed statin-therapy-related diseases and possible confounding variables/ drugs. As disease risks of interest, we identified the same three as for the previous analysis, depression, cancer and osteoporosis, and as possible confounding variables we identified 4 different drugs. Three out of the 4 medications were already identified for group DT1, namely metformin, insulins and other glucose-lowering drugs that are not insulins (group BX). We decided to additionally control for lipid lowering drugs (fibrates). To obtain group Statin, we have agreed to apply the same inclusion criteria as in group DT1, except criterion 1, valid treatment combination, which became irrelevant.

After redefining our research goal, one doctor has expressed his interest in a closer cooperation to explore possible associations between the individual statins and disease risks. We therefore agreed to meet more frequently as a small group, independent of the regular group meetings, to discuss the current results of the analysis. This group consisted of one doctor with medical expertise and two members of the hub-group, including myself, who have a deep expertise in statistics and methodology.

Phase 2: Exploring disease occurrence in statin patients

We first extracted group Statin which included all patients that received regular statin treatment over 2 years, (C2), except those who were treated with glucose-lowering drugs (group BX). Out of the 7 statins, one with less than 35 observations was removed (C3). Patients who had a higher daily dosage average of a particular statin as recommended by Vienna General Hospital were also excluded (C4). The recommended maximal dosage per day for each statin is as follows: simvastatin (80mg), lovastatin (80mg), pravastatin (40mg), fluvastatin (80mg), atorvastatin (40mg) rosuvastatin (80mg);(information is available on request: *(AKH) Allgemeines Krankenhaus der Stadt Wien* n.d.). Group Statin consisted of 63,778 patients (male: 44.5%, female: 55.5%) with an average age of 77.1. For these patients, we controlled for possible drug interactions with insulin, metformin and fibrates. Insulin and metformin patients were extracted according to the corresponding ATC codes as described earlier. To control the effects of fibrates, we extracted all patients treated with a lipid-lowering drug (ATC code, starting with C10AB). Due to inclusion criteria 2 and 3 we had to exclude one out of the 4 fibrates. The control group encompassed all patients of dataset P2 who were not members of group Statin or group BX. This group consisted of 1,610,542 patients (male: 48.5%, female: 51.5%) with an average age of 45.6. For group Statin, we calculated the age- and sex-dependent disease risks compared to the control group. The individual statin-dependent disease risks were calculated by logistic regression as described before for group DT1 (see analysis phase 2). However, for a better interpretation of the results, we additionally grouped the individual medications according to their medical effect. For instance, the 6 individual insulins then appeared as one variable, in this case as the variable insulin.
We discussed the preliminary results of group Statin within our small sub-group.

Subgroup Meeting1: Exploring group statin

To understand the characteristics of the new group we first interpreted the disease risks of group Statin. The sex-dependent disease risks of patients treated with statins compared to patients with no statins (control group) were shown in Tab. 4. Patients of group Statin show in all three cases an increased disease risk (cancer-ORs: 4.75; osteoporosis-ORs: 6.05; depression-ORs: 4.38) when compared to the control group. Note that these results, as explained by the doctor, must be interpreted carefully because disease risk is a multifactorial phenomenon, meaning that other factors like comorbid diseases or other medications may affect the observed overall disease risk. Nonetheless, Tab. 4 served as a first reference point to understand group properties. Of particular interest, we identified the risk of being diagnosed with osteoporosis (ORs: 6.05), because it had the highest sex-independent risk factor of all three diseases. Based on the results of the logistic regression we then discussed drug–drug and disease–drug interactions. Out of the three diseases, the results of patients being diagnosed with osteoporosis were the most interesting. For the individual confounding variables, we could not observe significant effects, but when grouped together to one variable, we observed an increased osteoporosis risk in statin patients who also received insulin treatment, see Fig. 6.

Disease-Risk-Analyses
Statin Patients vs. Control Group
All Ages

		chi2	p-value	Odds
Female	'Cancer'	3.695,40	<0.05	3,75
	'Osteoporosis'	6.926,60	<0.05	5,90
	'Depression'	3.279,22	<0.05	4,26
Male	'Cancer'	8.384,86	<0.05	6,10
	'Osteoporosis'	1.278,75	<0.05	5,50
	'Depression'	1.320,48	<0.05	4,28
All	'Cancer'	11.300,00	<0.05	4,75
	'Osteoporosis'	8.761,16	<0.05	6,05
	'Depression'	4.815,70	<0.05	4,38

Table 4

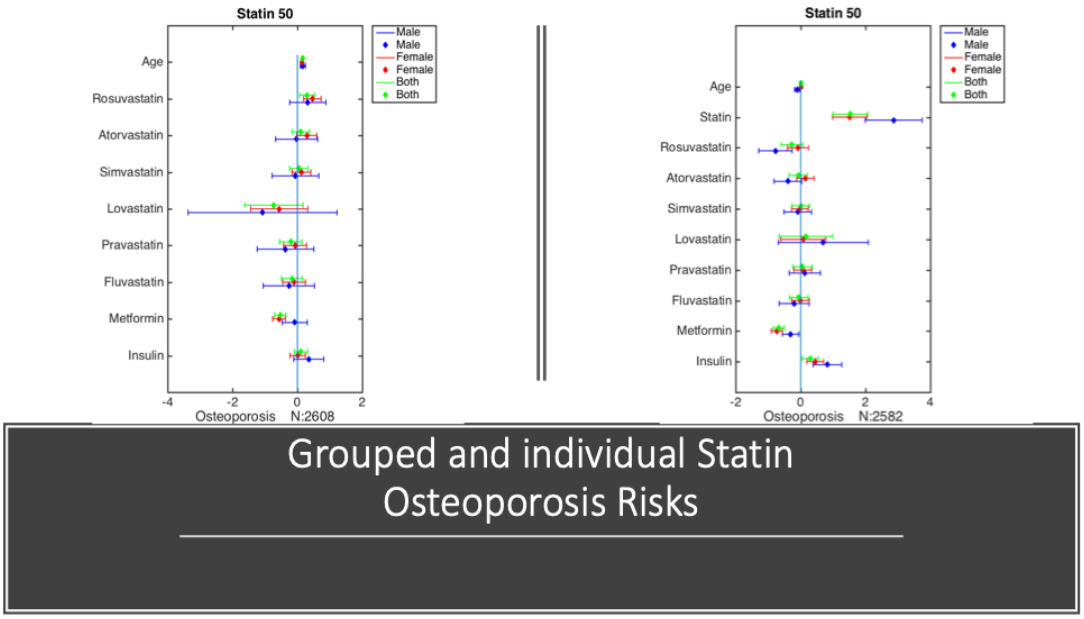

Figure 6

The graphic on the right of Fig. 6 shows the combined effect of all statins. It shows the previously observed trend that both sexes in group Statin have an increased risk of being diagnosed with osteoporosis. The left graphic shows the disease–drug interactions of the individual statins. We could observe again the trend that statins with a higher potency are related with an increased osteoporosis risk. However, significant results were only observed for female patients treated with atorvastatin (ORs: 1.33) and rosuvastatin (ORs: 1.64). The doctor explained that this is not surprising, because it is known that females compared to males have an increased risk of being diagnosed with osteoporosis (Black and Rosen,

2016). Contrary to that, we could not observe significant sex differences between male and female statin patients when compared to the respective sexes in the control group.
However, we agreed that the common goal of our analysis is the investigation of the relationship between statin-treated patients and their risk of being diagnosed with osteoporosis. First of all, we decided to investigate why we could not observe the expected sex difference. The first research interest was formulated as follows:

RQ5: Is there a sex difference in statin patients diagnosed with osteoporosis?

The two hub members, the statistical and methodological experts of this group, suggested to extract the group Statin based on dataset P1 to test whether we obtain the same results as for group Statin from dataset P2. The idea was based on our previous grouping experiences of DT1 patients, where we discovered that dataset P2 seems to be noisy in terms of missing or incorrectly entered diagnoses. We agreed to first compare results of group Statin of dataset P2 with those of dataset P1 before proceeding with the analysis.

Phase 2: Exploring the relationship between statin treatment and osteoporosis diagnosis

Group Statin was extracted from dataset P1 according to the same criteria used for the extraction in P2. Except criteria 2, regular applied treatment, we had to adapt to the observation duration of dataset P1, which includes information of patients who were alive from January 2006 to December 2007. We therefore reduced the duration of regular treatment to 1 year (C2). The obtained group consisted of 300,411 patients (female: 50.9%, male: 53.37%) of whom 10,179 had been diagnosed with osteoporosis. For this group, we extracted the corresponding control group, which included 7,645,364 patients (female: 53.37%, male: 39.19%). The sex-dependent osteoporosis risks were calculated for both sexes compared to the control group. The obtained results were presented in the next meeting.

Subgroup Meeting: Comparing datasets P1 with P2

We first discussed the overall case numbers of the two Statin groups. In dataset P2 we identified 18,083 patients with either one of the 3 ICD codes M80–M82, of which 2,682 were treated with statins. In dataset P1, we could identify 84,025 osteoporosis patients of which 10,179 received a statin treatment. Although the different case numbers can be explained by the different sample sizes of both data-sets (P1= 7,945,775, P2= 1,826,909), we cannot exclude the possibility that missing entries in P2 are the cause. From a statistical point of view, the larger sample size of P1 also has the benefit that it allows for a more accurate statistical analysis. Furthermore, we could observe in group statin based on dataset P1 the expected osteoporosis risk difference in sex. Based on these 2 findings, the sex and size difference, we decided to carry out the further analysis based on dataset P1. We then discussed in detail what our research questions are and how to approach them methodologically. In addition to the already formulated research interest RQ5, we decided to analyse the relationship between the individual statins, including possible potency-

and dosage-dependent differences, and the risk of being diagnosed with osteoporosis. The corresponding research questions were:

> *RQ6: Is there a relationship between osteoporosis diagnoses and the different statin potencies?*
>
> *RQ7: Is there an association between osteoporosis diagnoses and the daily defined average dosage of the individual statins? However, before conducting the analysis, we decided to consult the interdisciplinary research team.*

Meeting: Defining the final analysis

We informed the interdisciplinary group about our research process and the obtained insights. First, based on the results of the data-set P2 extracted group statin, we decided to investigate the relationship between sex-dependent osteoporosis risk and the individual statin medications. Second, we could not detect the medically expected sex-dependent risk difference and we therefore decided to extract group statin based on dataset P1. It was found that P1 includes 65,942 more patients with an osteoporosis diagnoses than P2, and that female patients of data-set P1 showed an increased risk compared to males. Indicating that P1 is better suited for the further analysis. We then presented our research interest and the corresponding 3 research questions.
The interdisciplinary research group has agreed to both, our research questions and our methodological decision to switch to dataset P1. Once more we discussed as a group the properties of our final group statin, every single research question and how to approach them methodologically. The group considered all extraction criteria as appropriate, except the exclusion of patients treated with glucose-lowering medications (BX group). The medical-group explained that the exclusion of this group was only relevant for the analysis of insulin related effects but not for statins. Due to the large sample size of P1 (N=7,945,775) the group further suggested to exclude patients who were older than 90 years to gain a more homogenous group. Concerning the research question the medical-group pointed out that RQ7, the analysis of dosage-dependent risks, is particularly interesting because there is not much knowledge available yet.
We decided that the further analysis will be carried out by the sub-group, which will continue to report on its current state of research.

4.4.3 Phase 3: Identification of Meaningful Patterns of Interaction

Note, so far, we discussed preliminary results, needed to understand the data. The results of the following analysis phase reflect the most important (meaningful) findings of our interdisciplinary study. Therefore, the methods and results described as well as their interpretation are to be regarded as valid.
The final cohort encompassed all Austrians with health claims and uniquely identifiable age and sex who were alive during the entire observation period from January 2006 to December 2007 (N= 7,945,775). Patients who were born in these years or who were older than 90 years were excluded to gain a more homogenous group of patients. The obtained

cohort consisted of 7,897,449 patients (male: 3,702,572; female: 4,194,877). Based on this cohort we identified all patients who have been treated at least once with one out of the 7 currently available statins on the market. We additionally controlled for possible effects of other prescribed medications including 39 different kinds of insulin-sparing or -providing medications and for 3 fibrates. In total, we tested for 49 different medications, 7 statins, 39 insulins and 3 fibrates, whereas statins were the main inclusion criteria for this group. For each medication, we tested if the patient was a regular drug user or not (C2). Only if a patient had a minimum of four different prescription entries for a given drug, we assigned him/her as being a valid drug user, meaning that the treatment was regularly applied over one year. Medications with less than 35 valid patients were excluded (C3). We excluded 1 statin (cerivastatin) and 18 different insulins from the 49 different medications, which reduces the number of drugs due to C3 to 31. Patients who had a daily dosage average of a particular drug lower than the recommended maximum dosage by Vienna General Hospital were included (see analysis phase 2), otherwise they were excluded as valid drug users of the specific medication (C4). In total, group Statin consisted of 353,591 patients (male: 175,557 female: 178,034). For group Statin, we identified 11,707 patients (male=1,766, female= 9,941) having one of the three osteoporosis diagnoses (M80, M81, M82). The corresponding control group was represented by all patients of our cohort who were not members of group statin. The control group consisted 7,543,858 patients (male= 3,527,015, female= 4,016,843), including 68,693 patients (male= 10,409, female= 58,284) diagnosed with osteoporosis.

The time series of the redeemed prescription of each statin was analysed (see analysis phase 1) to ensure the completeness of our data. To understand the characteristics of the groups in relation to our research interest, we first calculated the sex-dependent osteoporosis risks and the corresponding confidence interval (CI) for the new cohort and for group statin compared to the control group.

Subgroup Meeting: Understanding the final cohort

First of all, we looked at the results of the individual statin time analysis, concluding that all prescription histories are consistent. We then interpreted the result of our cohort and of group Statin. In our cohort, we observed the expected sex-dependent osteoporosis risk differences. Females are at higher risk of being diagnosed with osteoporosis when compared to males (OR: 5.01, 95% CI: 4.92–5.10, $p<0.01$). This trend was also observed in the comparison of patients of group statin compared to the control group. Both sexes of group Statin show an increased risk of being diagnosed with osteoporosis, females [OR (f): 3.94, 95% CI: 3.85–4.03, $p<0.01$] and males [OR (m): 3.33, 95% CI: 3.15–3.51, $p<0.01$], whereby the OR for females are significantly increased with respect to males ($p<0.01$).

Although these results suggest that there is a statin-dependent difference of being diagnosed with osteoporosis in statin patients (RQ5), the medical expert of the subgroup pointed out that additional information is required to validate a statin-dependent risk hypothesis. The two hub members of the subgroup proposed to perform an additional sex- and age-dependent risk analysis of patients of the same group. By comparing female to male

patients in the same group, we gain an alternative view on the sex-specific risk, namely how it changes as a function of statin use. To distinguish between the two analyses, we refer to the comparison of patients in the same group as intragroup and to the comparison of patients between the two groups as intergroup analysis.

Phase 3: Exploring age- and sex-dependent risks

The age- and sex-dependent relative risks were calculated in 10 year intervals for patients of the same group (intragroup) and for patients treated with statin compared to those without (intergroup). In order to detect significant sex differences, we visualized the ORs with the corresponding confidence intervals (CI). The results of the intragroup analysis were presented as a table allowing a better comparison of the risk development of both groups.

Subgroup Meeting: Inter- and Intragroup analysis

The intergroup risk analysis of group Statin compared to the control group displays that osteoporosis diagnoses are overrepresented in statin-treated individuals, with significant stronger effects in females than males, see Fig. 7.

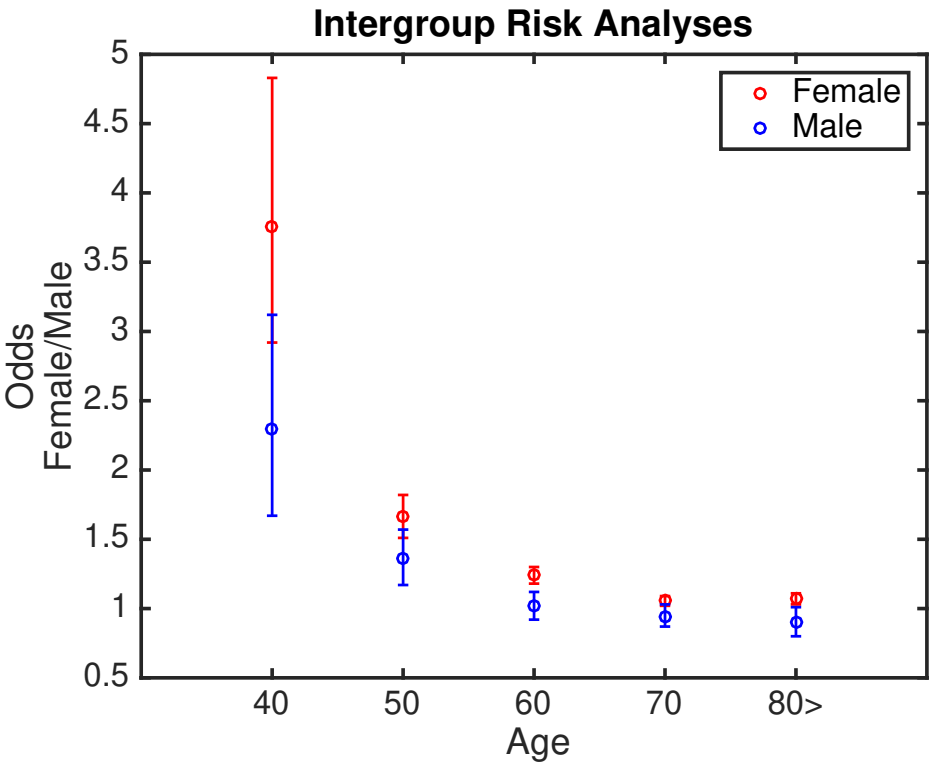

Figure 7

In addition, Fig. 7 shows that in the age class 40–50 years, the relationship between statin treatment and increased ORs of osteoporosis is stronger than in all other age groups. This sex-dependent difference fades with an increase of age. In contrast, the intergroup analysis shows that the risk of being diagnosed with osteoporosis increases with age for

	Statin	Control
Age	OR (95%-CI)	OR (95%-CI)
40-50	3.27 ** (2.22–4.84)	1.99** (1.82–2.18)
50-60	3.43** (2.91–4.04)	2.81 (2.66–2.97)
60-70	4.67** (4.23–5.16)	3.84** (3.67–4.01)
70-80	5.11** (4.71–5.54)	4.57** (4.40–4.75)
80-90	4.67** (4.16–5.25)	3.94** (3.77–4,12)

** p<0.01; *p<0.05

Table 5: Intragroup Analysis

females when compared to males of the same group, as shown by Tab. 5. It presents the results of the intragroup analysis, the odds of being diagnosed with osteoporosis with the odds of being female in the statin-treated and the control group, compared to males respectively. The intragroup analysis confirms that female sex is significantly associated with increased odds of osteoporosis across all ages in both groups. By comparing the effect sizes between the statin-treated and the control group, we see that statin treatment coincides with significant increased odds for females particularly in younger age groups (below 70 years), when compared to males. For instance, in the age group 40–50 years we find that females with statins have a 3.27fold increased risk to be diagnosed with osteoporosis when compared to men, whereas without statin treatment females have only a 1.99fold increased risk for osteoporosis compared to males.
Nonetheless, we observed that the ORs increase as a function of age in both groups, Tab. 5. With increased age, there is an increased osteoporosis risk in the control group which gives an age-dependent decrease in the OR between osteoporosis and statin treatment. Based on these results, we have agreed that we need additional information to make a valid statement about the relationship between the use of statin and the occurrence of osteoporosis diagnosis. We decided to further investigate the osteoporosis risk in relation to the individual statins (RQ7) and their dosages dependencies (RQ8).

Phase 3: Exploring dosage-dependent risks

The relationship between the individual statins and the risk of being diagnosed with osteoporosis was computed for each sex by logistic regression as described in analysis phase 2. Each patient was described by 31 variables corresponding to the individual drugs (5 statins,

17 insulins, metformin, 3 fibrates). To investigate the individual statin-dosage-depended risk, if differences in dosage have an effect on the risk of being diagnosed with osteoporosis, we applied a multiple logistic regression analysis. For a better comparison between individual statins and their dosage-dependent risks we first converted the Daily Defined Dose units (DDD) into milligram (milligram scale). This was done for each statin according to its conversion factor. The individual conversion factors were calculated on the basis of information provided by WHOCC on statins (*WHOCC Guidelines for Daily Defined Dosage for Statins* n.d.). Each statin was then further divided into 5 different milligram–categories (0-10, 10-20, 20-40, 40-60, 60-80 [mg]). Patients were assigned to one of the five categories according to their daily average dosage in milligrams. Each patient was described by 31 binary variables, one variable corresponding to the individual statin dosage category, thus 30 variables corresponding to the other 30 medications (5 statins, 17 insulins, metformin, 3 fibrates). A 1 indicated that a patient had been treated with the medication or that the patient was in a particular dosage category, a 0 if neither was the case. Depending on the number of patients per dosage category, we observed a range of 2,546–3,342 actual cases out of the 5×2^{30} unique possible drug combinations. Drug combinations with fewer than 32 observations were excluded, reducing the range to 237- 296 medication combinations for which we calculated the individual risks. The composite osteoporosis risk for each drug combination was computed as the relative number of patients who had been diagnosed with at least one of three osteoporosis diagnoses (ICD codes: M80- M82). To control for age dependencies, we used the average age of all patients of a given medication group. To assess the goodness of fit of the regression models we controlled for adjusted R-squared obtained from the logistic regression and for the variance inflation factor (VIF) to test for multicollinearity (see Appendix). The obtained logarithmic odds were again transformed into odds-ratio (OR) to facilitate the interpretation of the obtained results. To identify (significant) trends we visualised the individual ORs with the corresponding CI.

Subgroup Meeting: Exploring dosage-dependent risks

The dosage-dependent osteoporosis risks of the individual statins (RQ7) were discussed in relation to their potency (RQ6). The statin potency ranking from high to low is:$rosuvastatin > atorvastatin > simvastatin > pravastatin > fluvastatin > lovastatin$ (information is available on request: *(AKH) Allgemeines Krankenhaus der Stadt Wien* n.d.). Consistent dosage-dependent trends were observed in higher potency statins as shown by Fig. 8. The figure presents the development of the dosage-dependent ORs in higher potency statins. Black markers display the ORs of all individuals treated with the specific statin, blue markers denote male and red female ORs. To each OR we assigned the corresponding confidence interval. Fig. 8 shows the trend that a higher dose is associated with increased occurrence of osteoporosis diagnosis. Tab. 6 presents the dosage-dependent ORs for 6 statins. Fig. 8 and Tab. 6 reveal the following picture: osteoporosis is significantly underrepresented for all individuals who received a low dose statin treatment (0-10mg) of pravastatin (OR:0.69, CI:0.53–0.90), lovastatin (OR:0.39, CI:0.18–0.84), simvastatin (OR:0.70, CI:0.57–0.87) or rosuvastatin (OR:0.69, CI:0.55–0.89). In the sex-specific analysis only low dose simvastatin

All	Fluvastatin	Pravastatin	Lovastatin	Simvastatin	Atorvastatin	Rosuvastatin
0-10 mg	1	**0.69****	**0.39***	**0.70****	1.04	**0.69****
CI	(1-1)	(0.53-0.90)	(0.18-0.84)	(0.57-0.87)	(0.87-1.25)	(0.55-0.88)
10-20 mg	**0.59****	0.87	1.06	0.83	**1.35****	0.90
CI	(0.42-0.82)	(0.7-1.08)	(0.68-1.64)	(0.68-1.02)	(1.11-1.65)	(0.71-1.14)
20-40 mg	0.85	1.01	1.06	1.07	**1.78****	**2.04****
CI	(0.69-1.05)	(0.81-1.26)	(0.83-3.08)	(0.87-1.32)	(1.41-2.24)	(1.30-3.20)
40-60 mg	0.91			**1.65****	**2.12****	
CI	(0.75-1.12)			(1.31-2.07)	(1.47-3.07)	
60-80 mg	1.09			**3.31****	**3.15****	
CI	(0.88-1.35)			(2.37-4.63)	(1.77-5.58)	

Female	Fluvastatin	Pravastatin	Lovastatin	Simvastatin	Atorvastatin	Rosuvastatin
0-10 mg	1	0.74	**0.41***	**0.75***	1.12	0.85
CI	(1-1)	(0.54-1.01)	(0.18-0.9)	(0.59-0.95)	(0.9-1.4)	(0.64-1.11)
10-20 mg	**0.59****	0.88	0.95	0.88	**1.57****	1.2
CI	(0.4-0.87)	(0.68-1.14)	(0.61-1.49)	(0.7-1.11)	(1.23-2)	(0.9-1.6)
20-40 mg	0.93	1.05	1.80	1.14	**2.48****	**3.01****
CI	(0.73-1.19)	(0.81-1.37)	(0.93-3.5)	(0.9-1.45)	(1.84-3.34)	(1.74-5.25)
40-60 mg	0.98			**1,85****	**2.77****	
CI	(0.77-1.25)			(1.42-2.41)	(1.69-4.52	
60-80 mg	1.1			**3.72****	**4.81****	
CI	(0.85-1.42)			(2.57-5.4)	(2.19-10.54)	

Male	Fluvastatin	Pravastatin	Lovastatin	Simvastatin	Atorvastatin	Rosuvastatin
0-10 mg	1.53	0.71	1.05	**0.59***	1.46	0.64
CI	(1-1)	(0.4-1.24)	(1-1)	(0.37-0.93)	(0.95-2.26)	(0.38-1.09)
10-20 mg	0.65	0.94	1.05	0.69	**1.80***	1.10
CI	(0.3-1.38)	(0.6-1.48)	(0.35-3.12)	(0.44-1.06)	(1.14-2.85)	(0.64-1.86)
20-40 mg	0.68	1.08	1.05	1.05	**2.23****	1.83
CI	(0.43-1.09)	(0.68-1.72)	(1-1)	(0.68-1.63)	(1.32-3.76)	(0.74-4.56)
40-60 mg	0.9			**1.63***	**3****	
CI	(0.57-1.44)			(1-2.65)	(1.42-6.36)	
60-80 mg	1.53			3.57	**3.14***	
CI	(0.93-2.5)			(1.78-7.18)	(1.01-9.77)	

** p<0.01; *p<0.05

Table 6: Individual statin-ORs ($95\%CI_s$) obtained from the logistic regression model of the dosage-dependent osteoporosis risks.

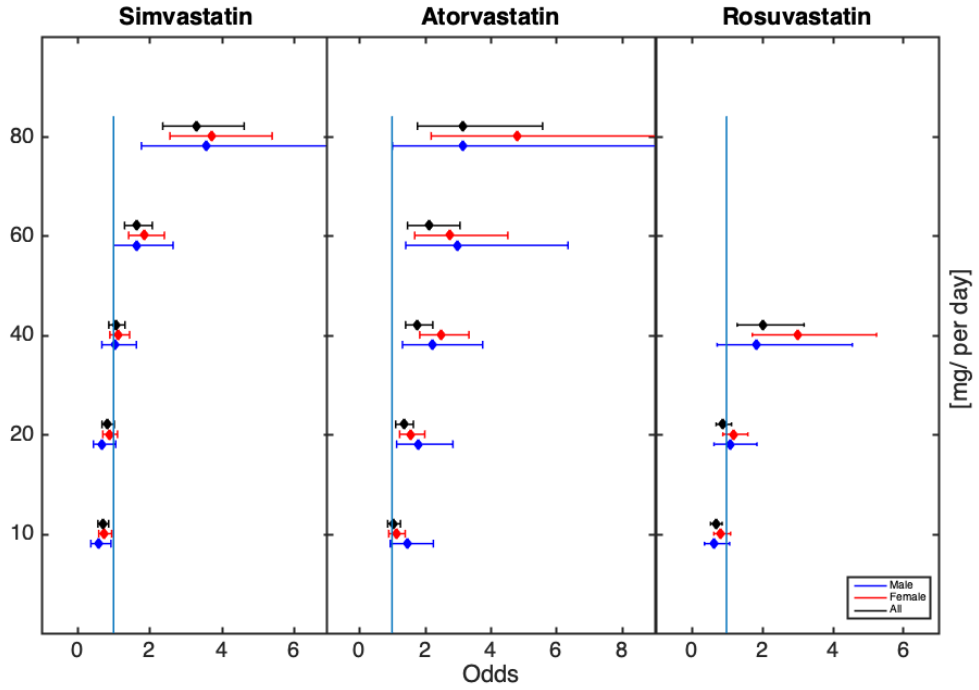

Figure 8: Dosage dependency of three most potent statins

treatment is related to an underrepresentation of diagnosed osteoporosis in males and a low dose lovastatin and simvastatin treatment in females. With increased dosage, however the relationship between statin treatment and osteoporosis reverses. In particular, the ORs of simvastatin and atorvastatin show the risk trends between statin dosage and osteoporosis. Osteoporosis diagnoses are overrepresented in the group treated with 40mg of atorvastatin per day (OR:1.35, CI:1.11–1.65) and increased with the increase of the daily dosage (60mg atorvastatin: OR: 1.78, CI:1.41–2.24; 80mg atorvastatin: OR: 3.15, CI:1.77–5.58). The fact that osteoporosis is significantly overrepresented in patients treated with high doses is also observed for simvastatin (60mg: 1.65 OR, CI:1.31–2.07; 80mg: 3.31 OR, CI:2.37–4.63). In both females and males, the increase in dose of simvastatin is related to a higher occurrence of a disease diagnoses, whereas for atorvastatin this relationship only remains significant in female patients (see Tab 6).

Taken together, we observe a dosage-dependent trend in statin-treated patients, which shows that higher dosages are associated with higher incidence of osteoporosis diagnosis, with females at increased risk of being diagnosed with osteoporosis compared to males. In order to validate this hypothesis, we discussed possible confounding comorbidities which may affect the observed trend. As possible confounding diseases, we identified 6 different groups:

1 Rheumatoid arthritis (ICD code: M06)

2 Ischaemic heart diseases (ICD codes: I20–I25)

3 Diseases of arteries including arterioles and capillaries (ICD codes: I70–I79)

4 Stroke (ICD codes: I63, I64)

5 Diabetes mellitus type 1 and 2 (ICD codes: E10, E11)

6 Chronic renal insufficiency (ICD codes: N17–N19)

Phase 3: Testing for confounding diseases

From group Statin, we identified all osteoporosis patients who also had one of the 6 comorbidities. Patients were extracted and grouped according to the ICD codes. To test whether or not the observed dosage-dependent trend is independent of the identified comorbidity groups, we performed a multiple logistic regression analysis of the dosage-dependent osteoporosis risks, as described in analysis phase 6, with the difference that the patients of the respective confounding disease group were excluded. The obtained results of each regression were visualized as one graphic, to compare the individual risk trajectories.

Subgroup Meeting: Final interpretation of the results

The relationship between the dosage-dependent osteoporosis risk and the confounding disease groups was discussed for each statin, as shown in Fig. 9 for simvastatin and in Fig. 10 for atorvastatin.

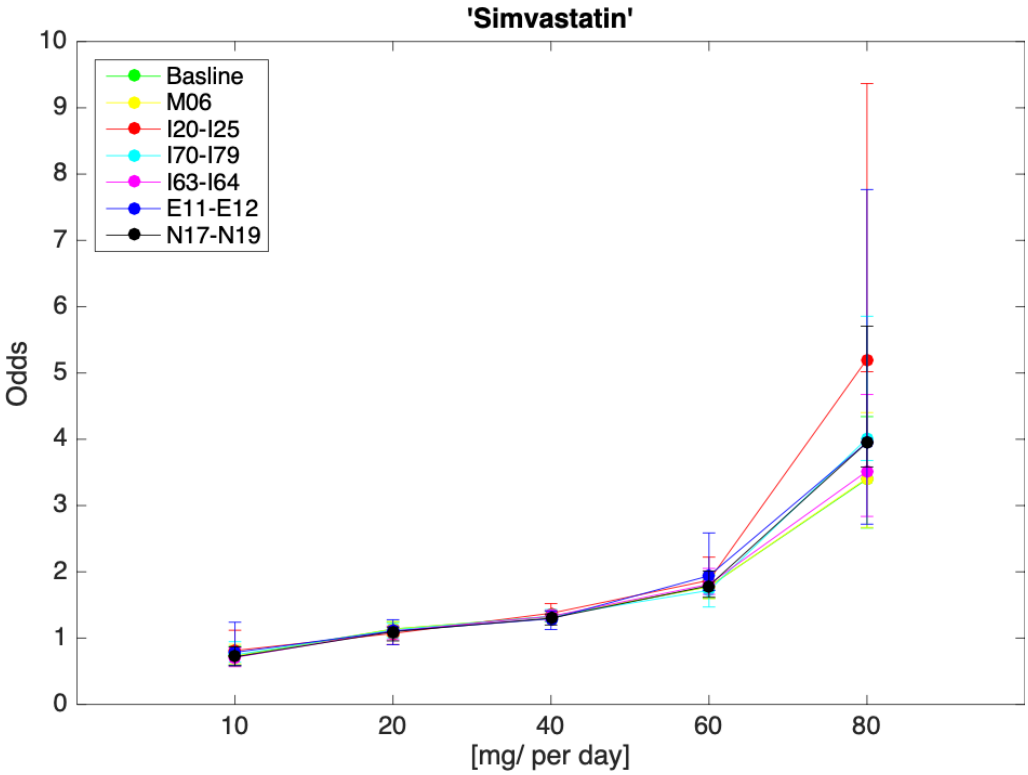

Figure 9: Relationship between the dosage-dependent osteoporosis risk and the confounding disease groups for simvastatin

Fig. 9 presents the dosage-dependent osteoporosis risk trajectories for the individual disease groups of simvastatin. The baseline (green) denotes all individuals treated with simvastatin and their dosage-dependent osteoporosis risk development. The other 6 trajectories display

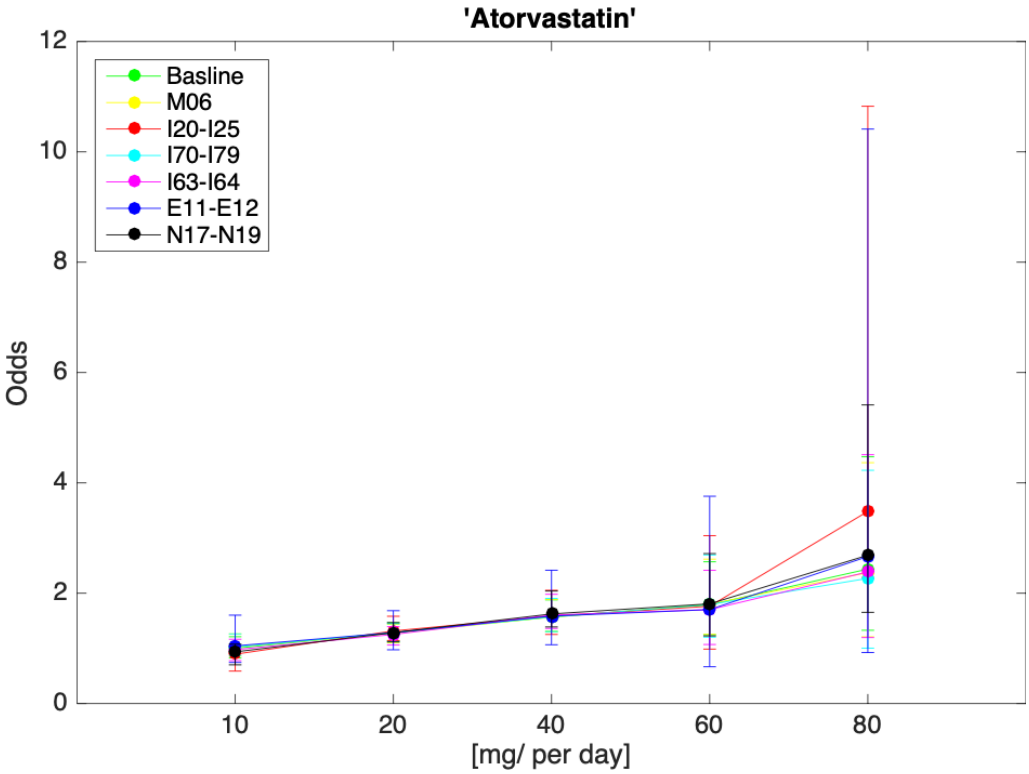

Figure 10: Relationship between the dosage-dependent osteoporosis risk and the confounding disease groups for atorvastatin

how the risk develops when patients of a particular disease group are excluded. Fig. 9 shows that all trajectories follow the same trend that a higher dose is associated with higher incidence of osteoporosis diagnosis. Although we observe a higher variance of ORs in the highest dosage category (60–80 mg), the differences can be neglected as they are likely to result from the small case numbers of this dosage category. This is indicated by the large CI. The interpretation of the results is further confirmed by the results of other statins as shown for atorvastatin in Fig. 10. The dosage-dependent osteoporosis risk analysis of atorvastatin in relation to the 6 disease groups reveals a similar picture as shown in Fig. 9 for simvastatin. All trajectories follow the same trend, except in the dosage category (60–80 mg) we observe again a higher variance. Nonetheless, we concluded that Fig. 9 and Fig. 10 indicate that the observed dosage-dependent risk development is independent of comorbidity groups.

4.5 Results and Discussion

Osteoporosis is a chronic disease characterised by a reduced bone mineral density induced by an imbalance in osteoblastic and osteoclastic bone formation and resorption (Stone and Hosking, 1991). Due to the elevated fracture risk, osteoporosis can have detrimental effects on a patient's quality of life and is associated with a higher mortality and morbidity as well as being an economic burden (Zethraeus et al., 2007). By now, numerous studies about osteoporosis and its treatment have been conducted – many of which revolve around the

question whether statins affect bone metabolism (An et al., 2017; Chan et al., 2000; Lin, Chou, et al., 2018; Lin, Liou, et al., 2018; Meier et al., 2000; Van Staa et al., 2000).
Statins play a crucial role in the management of hypercholesterolemia, which makes them a commonly used drug (Johansen et al., 2014). Actual guidelines for the treatment of hypercholesterolemia in high-risk patients suffering from cardiovascular disease or diabetes have been issued, recommending cholesterol levels to be as low as possible (Catapano et al., 2016). Due to the sheer number of patients under statin therapy, research on the connection between statin usage and osteoporosis risk is of great importance. However, data about the impact of different kinds of statins and their dosages is sparse so far. Therefore, the aim of this study is to investigate the effect of statin treatment on the diagnosis of osteoporosis in males and females.
This study reveals that diagnosis of osteoporosis in statin treated patients was dosage-dependent. Fig. 7 and Tab. 6 display that there is a highly non-trivial dependence of statin dosage with the odds of osteoporosis. For all individuals who received a low dose statin treatment osteoporosis is significantly underrepresented. With increased dosage, however the relationship between statin treatment and osteoporosis reverses. In both, females and males the increase of the dosage of simvastatin was related to an overrepresentation of a diagnosed osteoporosis, whereas for atorvastatin and pravastatin this relationship only remained significant in female patients.
Taken together, our data suggest that osteoporosis is overrepresented in high-dosage statin treatment but underrepresented in low-dosage statin treatment. New guidelines for cholesterol lowering therapies for prevention of cardiovascular complications advise to reduce plasma LDL-C levels as low as 70mg/dl (Catapano et al., 2016). Especially the current recommendations show a trend towards increasing doses in statin therapy regiments (Rodriguez et al., 2017). Due to conflicting results of previous studies on statins and bone mineral density, we propose that monitoring high-risk patients, i.e. post-menopausal female patients under high-dosage statin therapy, might be useful in order to offer an individual therapy to prevent or treat osteoporosis.

Chapter 5

Study based Insights

In this section, we combine insights gained form our theoretical discussion on interdisciplinary research with those from our study. The goal is to show how our data-driven analysis has benefited from interdisciplinary collaboration. And to highlight aspects/factors which we experienced as important in the course of our study.

5.1 Why does an interdisciplinary approach promote data-driven analysis?

In our research portray we have shown why the interdisciplinary approach is beneficial for an exploratory data-driven analysis. By integrating medical with methodological expertise, we gained new insights on the occurrence of osteoporosis in relation to statin dosages. Regarding the level of cognitive integration, we do not believe that we achieved the highest level, namely the unbounding of disciplinary minds, but it partially has taken place, which was sufficient to advance our understanding of the complex relationship between osteoporosis and applied medical treatments. Without the interdisciplinary cooperation between the Complexity Science Hub Vienna and the Medical University of Vienna the statin-dosage-dependent occurrence of osteoporosis would not have been discovered. Since none of the individual groups, neither the hub-group nor the medical-group, has the necessary knowledge to carry out this analysis themselves, our study exemplifies why interdisciplinary research promotes data-driven analysis. At every phase of our research process, this became clear.

In phase 1 of our research process, medical expertise was important to understand the quality of our data. Noisy, uncertain data requires domain knowledge to make sense out of the data (Holzinger, Dehmer, and Jurisica, 2014). As described in phase 1, the medical-group provided insights on the insulin treatment for diabetes type 1 patients which made us aware that our data is noisy. And the hub-group could offer an alternative methodological approach (through medical treatment) on how to extract the group of interest.
In phase 2, the hub-group prepared the various disease–drug relations in such a way that the medical group was able to identify the statin-dosage-dependent trend. However, as

reported, we could not observe the expected sex difference in osteoporosis. Which in turn called into question the quality of our data and our observed trend. The hub-group then had the key idea of changing from data-set P2 to P1.
In phase 3, both expertises were needed again to validate our results. Medical knowledge was required to eliminate possible confounding variables which affect the identified disease-drug interaction. And statistical knowledge was required to select and perform the required analyses.
In short, our study illustrates why the use of the different disciplinary expertise, why interdisciplinary collaboration promotes discovery-driven data analysis. Interdisciplinarity is beneficial because all phases of an data-driven analyses, data-cleaning, understanding the data and the identification of meaningful results, require the combination of different expertise. Methodological knowledge to carry out adequate analysis and domain expertise to understand the results achieved (Holzinger, Dehmer, and Jurisica, 2014).

5.2 Which body–mind–environment factors shape interdisciplinary practice?

In the course of our study, we have experienced different aspects / factors that influence interdisciplinary interaction, and cognitive integration processes. Now we would like to list those factors that we believe enhance the interactions between scientists from different disciplines in such a way that a cognitive process emerges between them. 4 distinguishable factors came to the fore:

1 Habitus-related qualities
    [a] Integration-related personal qualities
    [b] Interaction-related personal qualities
2 Environmental/ field-related properties
    [a] Group size
    [b] Tools
3 A shared study logic
    [a] The what, why and how of a study
4 Time

5.2.1 Habitus-related qualities

In this discussion we take a more liberal perspective on Bourdieu's theory of practice. We extend the concept of habitus to personal qualities that may require conscious effort. Furthermore, in the following discussion, we refer to personal qualities that are important for interdisciplinary research, to those that function as cognitive toolkit, as suggested by Repko (2008, pp. 128- 142).

a Integration-related personal qualities

The essential criterion of integration is that it is meaningful, in this sense it involves critical evaluation of the exchanged disciplinary knowledge to create a common ground among them and to construct a more comprehensive understanding of the phenomenon of interest. We experienced that meaningful integration is fostered by the intellectual capacity, critical thinking and by the two skills abstract thinking and creative thinking. Although the capacity to analyse, critique and assess information is fundamental to any mono-discipline, for humanities as well as for natural science, the required analytical skills differ. From an interdisciplinary view, critical thinking requires the evaluation and integration of multiple assumptions and insights of different disciplines concerning one complex issue. This process of sense-making of various concepts aims to shift one's focus from a narrow disciplinary context to a broader interdisciplinary one. Another important skill involved in the process of cognitive integration of different disciplinary views is abstract thinking. Contrary to concrete thinking, which limits thought to what is right in front of you, abstract thinking takes into account that a concept's meaning may differ from discipline to discipline. In this sense, abstract thinking is characterized as the mental adaptability and flexibility needed to construct a more comprehensive understanding of a multi-dimensional complex phenomenon. In some cases, it even might be required to think 'outside the box', also described as creative thinking. This thinking style aims to combine previously unrelated ideas through identifying new relationships among concepts or theories. This was not required of our data-analysis but we experienced it in the context of another study which was also discussed during our meetings.

b Interaction-related personal qualities

The proactive process of integration of different concepts and theories depends on how researchers interact with each other, and is therefore influenced by personal qualities like communicative skills, empathy and perspective taking. In general, communicative skills refer to your capability to apply the personal qualities discussed above.
Communication skills have proven to be important to us in finding the common research goal and in determining what the individual disciplines can contribute to its exploration. This includes an open discussion of one perspective strengths and the limitations of other perspectives. This proactive process can cause conflicts, which is why we consider empathy as an important value for successful collaboration. It refers to the process of identification with another's thoughts, pain or situation in a way that it can be communicated to others. Empathy, as experienced, is particularly important when your work is negatively influenced due to personal circumstances of another member, for example when he/she feels too stressed to finish the job in time. Showing understanding of the other's situation will have a positive effect on further cooperation as the person in difficulty probably recognises your behaviour.
On a less emotional and more general level, the ability to view a problem from different perspectives has a positive impact on interactive and integrative processes. To take an alternative perspective into account is achieved by not only understanding how others

view something in a particular way but also why they do so. It improves efficiency of integrating information and therefore facilitates assembling new sets of potential solutions to a complex problem.

5.2.2 Environmental/ field-related properties

From an enactive perspective, we defined cognition as a sense-making process that is enacted through a mind–body–environment interaction. Therefore, the environmental conditions play a crucial role in interactions. We experienced two environment-related factors that had a direct impact on the interdisciplinary research process, group size and tools.

a Group size

During our research process, we experienced that the size of the group has a significant impact on the research process. This effect became particularly apparent in the interdisciplinary meetings.

As mentioned in the study description, the number of participants varied over the whole research process. That is why we have seen how different group sizes affect the interdisciplinary collaboration. At the time, as at least 4 researchers wanted to present their current results, the quality of the interdisciplinary meetings and therefore of the whole study has suffered. As a reminder, in addition to the 4 members of the hub-group (those who presented their results), there were always at least 3 medical-group researchers present, which is a total of at least 7 disciplinary minds. Due to this increase in participants, there was not enough time for all researchers to present their studies nor to discuss the presented ones in sufficient depth. From a cognitive science perspective, the main problem was that the group size did not allow for intense interaction between the researchers needed for the cognitive process to emerge.

We tried to compensate for this by adapting our meeting practice. We decided to limit the number of slides of each power-point presentations held in the meetings, and to additionally create a shared online folder for the detailed presentations. This did not lead to the desired result, because it became obvious that a dialogue between the individual groups was necessary for the interpretation of the results. As a third attempt to increase the effectiveness of meetings, we decided to print the main results as charts or spreadsheets instead of Power Point presentations. This change in presentation style has had a positive impact on the research process as we have been able to discuss several studies.

However, the interaction has improved significantly only through the formation of subgroups, consisting out of 3-4 researchers. Due to the more frequent subgroup meetings, the interaction has intensified, improving the cognitive process of integration among the disciplinary minds.

That the formation of subgroups is beneficial for the whole process, has also been reported by Bernini and Woods, 2014. They described in their research based conceptual model, that the close co-operation of some participants in addition to the joint group meetings in the we-space was beneficial. When researcher move together within the project space to explore the complexity of a phenomenon they slowly adjust to other discipline thinking styles, which in return promotes the cognitive integration process.

In short, the size of the group must allow close and intense interaction between the researches to develop a mutual understanding of each others thinking style which promotes the emergence of a cognitive integration process. The challenge is to design the study in a way that precisely this is possible. Such a study design should take into account the size of the group and the use of tools, e.g. different presentations.

b Tools

We experienced that tools, in our case mainly visualisation have a significant impact on how we process information. Depending on the individual academic habitus, on the individual practice history, visualisations aid the cognitive process of integrating new information. On the highest level, as proposed by Andy Clark, 2008, tools enable the mind to extend partially into the world allowing it to act and interact with the environment .
However, we argue in the context of interdisciplinarity that the identification of tools has to be the aim that aid the cognitive processing of all, or at least as many as possible, academic habitus. This identification, as described in the previous section, depends on the group size and the time available. That this can be a challenge is reflected in our research report. As shown, we have used different forms of visualisations, to make sense out of the data. Besides histograms, bar charts, tables we also used pie charts and network visualisations. The latter two were for our purpose insufficient, which is why we did not include them in the research report.
Although choosing the right tools is a time-consuming challenge, we believe it is highly important. Tools that support the individual sense-making processes of all members can significantly improve the interdisciplinary research process. It aids the process of bridging the individual academic habitus.
In the case of data-driven analysis we consider visualisation in its various forms as an essential tool to turn data into knowledge. It not only supports the individual sense-making processes, but also the transformation of knowledge between persons or, at best, in the whole group (Burkhard, 2004).
Although the choice of the right, appropriate tool depends on the type of research and the associated academic habitus of the individuals, we believe that visualisation in all its varieties covers a very broad spectrum. In general, and in particular for interdisciplinary research, we recommend considering, that cognition arises out of the mind–body–environment interaction and externalities within the environment have a significant impact on how we process information. Therefore, the methodological discussion at the beginning of the study should not be limited to tools for the analysis, but should include those tools which help the transformation of knowledge between persons. Identifying the right tools will help to overcome the disciplinary barriers and thus support interaction and cognitive integration processes.

5.2.3 A shared study logic (what,why,how)

In the course of our research process we experienced that a commonly shared goal and a strategy to achieve it, positively influence both the interaction among researchers and

the cognitive integration process. From a cognitive externalist perspective, a shared study logic that determines what, why and how to study something improves the interdisciplinary practice.

We argue that a commonly shared study logic promotes the interdisciplinary research process in the same way that the logic of a field promotes certain practices. A common understanding of what should be achieved serves as the basis for a common cognitive process on how to achieve it.

The importance of a clearly defined goal has also been acknowledged by Tobi and Kampen, 2018. In their model for a methodology for interdisciplinary research (MIR) they promote that the research question, the hypotheses should be leading for all decisions made during the research process. The idea is to put the research's common goal at the center rather than the diversity of the different disciplinary backgrounds. By focusing on the common goal, it should be easier to overcome the disciplinary barriers, thereby improving interdisciplinary collaboration.

Tobi and Kampen suggest that in any phase of the research process, the whole research group should explicitly discuss the what, why and how of the study. Although we did not use their model, we followed this strategy. In every meeting, we discussed what are the results, why they might be important and how to proceed. We think by addressing the what and the why of a study, the chance of identifying a research question which is relevant to more than one discipline increases and the how helps to develop a common shared strategy to achieve the goal. This may sound trivia since this is, or should be part of any study process. However, we argue that it makes a difference whether the what, why and how is explicitly discussed by the entire group or not.

5.2.4 Time

Although time is obviously an important factor for any study, we list it here because we want to create the awareness that interdisciplinary research is very time consuming. We think it is important for any researcher who aims to conduct an interdisciplinary study to take this into account. Time is the variable on which almost any other factor depends. It takes time to identify a shared goal which draws on the individual disciplines, to adapt to the different research styles including disciplinary thinking styles, to identify tools that support all individual sense-making processes and it takes time for a meaningful cognitive process to emerge. Time is also the critical factor to reach the highest level of cognitive integration, the unbounding of the disciplinary minds. Bernini and Woods, 2014 proposed, based on their own interdisciplinary research experience, that under certain conditions the individual disciplinary minds may be unbounded, allowing them extended and interact with disciplinary minds. They further reported that had intensive fortnight meetings. This shows that their concept of the unbounding of disciplinary minds is based on time consuming, intensive interaction process. As pointed out, we experienced that a shared goal, a precise research question as well as a small group size and the intelligent use of tools, which enhance the cognitive process among researchers, have a positive effect on the efficient use of time.

Chapter 6

Conclusion

In this thesis, we showed that interdisciplinary research fosters the cognitive process of turning data into knowledge. We successfully conducted a data-driven analysis in a clinical context by combining methodological with medical expertise. We identified highly non-trivial associations between statin-dosage-dependent treatment and the occurrence of osteoporosis diagnoses. Osteoporosis was overrepresented in high-dose and underrepresented in low-dose statin treatment when compared to control subjects without statins. The underrepresentation of osteoporosis in low-dose statin treated patients could also be reproduced in a sex-specific analysis for simvastatin in males and for lovastatin and simvastatin in females. In both, females and males the increase of the dosage of simvastatin and atorvastatin was related to a significant overrepresentation of a diagnosed osteoporosis, whereas for rosuvastatin, this relationship only remained significant in females. As mentioned we aim to publish our study, because our results add some new insight into the relationship between statin treatment and the risk of osteoporosis.
In addition we created a research portray to show why the use of the different disciplinary expertise, why interdisciplinary collaboration promotes discovery-driven data analysis. Interdisciplinarity is beneficial because all phases of a data-driven analyses, data-cleaning, understanding the data and the identification of meaningful results, require the combination of different expertise. Methodological knowledge to carry out adequate analysis and domain expertise to understand the results achieved.

Second, we made use of cognitive externalist insights to gain a deeper understanding of the problem of bridging different disciplines. From a cognitive science perspective, we defined cognition as a sense-making process that emerges out of the body–mind–environment interaction, and individuals as the bearers and beholders of disciplinary knowledge and methods. Taking a cognitive externalist perspective on Bourdieu's theory of practice, the vehicle that carries the embodied knowledge, is the academic habitus. It is a durable system of dispositions that shape our perception and action, which acts according to the logic of the university field. Bourdieu describes the academic habitus as a system which students and professors maintain by conforming to the established habitus. This reproductive process of similarity challenges the bridging of disciplinary fields as the academic habitus diminishes the potential to develop alternative practice like new research methods or ways of thinking.

Finally, we combined insights gained from our theoretical discussion on interdisciplinary research with those from our study to identify body–mind–environment factors that support interdisciplinary interaction, and cognitive integration processes. Where 4 distinguishable factors came to the fore: personal/habitus-related qualities, environmental/ field-related properties, a shared study logic and time.

In the course of our study, we have learned that certain personal/ habitus-related qualities support the emergence of a cognitive integrative process between researcher of different disciplines. Skills like abstract or creative thinking promote the process of cognitive integration. The essential criterion of integration is that it is meaningful. The intellectual capacity to think critically helps in the evaluation of new disciplinary knowledge and therefore fosters a meaningful integration of it. We have also experienced that communication skills, perspective taking and values such as empathy have a positive impact on interdisciplinary collaboration. They support the interaction process between scientists with different expertise, different academic habitus and therefore the emergence of a cognitive process between them.

As important environmental/ field-related properties we identified the size of the group and the intelligent use of tools. The size of the group must allow close and intense interaction between the individual researchers to develop a mutual understanding of each other's thinking style, so that a cognitive integration process can emerge. If the number of participants within the field, if the number of academic habitus no longer permits this, we recommend the formation of subgroups. Inspired by Clark's extended mind approach, we learned that the intelligent use of tools in our case mainly visualisation, can have a significant impact on the cognitive integration process. In line with Clark we argue that tools, under certain conditions support how we process information. In case of data-driven analysis we believe that visualisation not only helps to interpret the result and to discover meaningful trends but also to transform knowledge between persons or, at best, between the whole group. In the context of interdisciplinarity we argue that goal has to be the identification of tools that aid the cognitive processing of all, or at least as many as possible, academic habitus. Identifying the right tools can help to overcome disciplinary barriers.

As another important factor that supports the bridging of disciplinary minds, we identified the creation of a common study logic. We argue that a shared study logic that determines what, why and how to study something improves interdisciplinary collaboration. A logic that supports the individual sense-making processes and therefore enhances the processes of interaction and meaningful integration of new information. The idea is to put the common research goal at the center rather than the diversity of the different disciplinary backgrounds.

We list time as the last important factor because we want to create the awareness that interdisciplinary research is very time-consuming. It takes time to identify a shared goal which draws on the individual disciplines, to adapt to the different research styles, to identify tools that support all individual sense-making processes and it takes time for a meaningful cognitive process to emerge. We believe that it is important for any researcher who aims to conduct an interdisciplinary study to take this into account.

As a concluding commentary on data-driven analysis we highly recommend an interdisciplinary study design. For a successful execution we recommend putting a shared goal at the center of the research process. At each phase of the research, we advise that the whole group discusses the what, why and how of the study in order to generate a common understanding, a common study logic. This discussion should include the identification of tools which support the sense-making process of as many individuals as possible. Both the identification of suitable tools and a shared study logic support the interaction and cognitive integration processes, which in return allows a more effective use of the rare resource time.

Chapter 7

Methods

To investigate the individual statin-dosage-dependent risks that is, if differences in dosage have an effect on the risk of being diagnosed with osteoporosis, we applied a multiple logistic regression analysis. First, we had to extract for each patient the average daily defined unit of each medication.

7.1 Daily Defined Dose Average in Units

Based on the prescription information, we extracted each patient's individual intake time and amount for each of his/her prescribed medications. Time vectors from the first to the last date entry of a prescription were extracted for each patient according to his/her medical treatments. Assuming that patients receive medication during hospitalization, days spent in a hospital (Th_i) were excluded if they fall into the prescription observation time series ($Th_i \in Tm_{i,j}$). The intake time of a patient i for a particular drug j was then calculated as the sum of the time vector of a particular drug ($Tm_{i,j}$), and the corresponding amount of the particular drug as the sum of all amounts ($Vol_{i,j}$). A patient's daily unit average of a medication ($Av_{i,j}$) was computed by dividing the sum of all amounts of the administered drugs by the sum of treatment days, minus the days a patient spent in a hospital:

$$Av_{i,j} = \frac{Vol_{i,j}}{Tm_{i,j} - Th_i}$$

7.2 Dosage-dependent Osteoporosis Occurrence

For a better comparison between individual dosage-dependent risks we first converted the Daily Defined Dose units (DDD) into milligram (milligram scale). This was done for each statin according to its conversion factor. The individual conversion factors were calculated on the basis of information provided by WHOCC on statins (*WHOCC Guidelines for Daily Defined Dosage for Statins* n.d.). The dosage-dependent occurrence of osteoporosis was calculated for each statin individually by a logistic regression.
Statins were divided into 5 different milligram categories ($0-10, 10-20, 20-40, 40-60, 60-80[mg]$). Patients were assigned to one of the five categories according to their

daily-average-dosage in milligrams ($P_{i,d}$). For instance, if a patient received an average statin treatment of $10 - 20[mg]$ he/she was assigned a 2 ($P_{i,2}$). In total each patient was described by 31 variables, 1 statin-dosage categorical variable $(1 - 5)$ and 30 binary variables corresponding to the other 30 medications (5 statins, 17 insulins, metformin, 3 fibrates). For the binary variables, a 1 indicates that a patient has been treated with the medication, and a 0 if not. For each statin, out of the 5×2^{30} unique possible medical treatment combinations we identified those who actual occurred. Medical treatment groups with less than 32 patients were excluded. As a consequence, drugs that have occurred only in groups with less than 32 observations, dropped out of analysis. With the remaining medical groups we created the predictor matrix (X) to compute the predicted osteoporosis occurrence. Each row of (X) represents a uniquely identifiable observed drug combination, each row has 31 columns, one for each medication.
In order to compare the individual groups with patients of the control group, with individuals who did not receive any of the drugs tested, we added them as reference group to predictor matrix (X). Each group was weighted by their number of patients $(w_{1)}$, and by relative number of patients who had been diagnosed with at least one of three osteoporosis diagnoses (ICD code: M80–M82) (w_2). To control for age dependencies, we used the average age of all patients of a given group (w_3). The group specific properties $w1, w2, w3$ were computed for all and for each sex.
In the logistic regression model (Urban and Mayerl, 2006; Peter and John, 1989) the linear predicted osteoporosis occurrence (Z) was a function of all included medications $(x_1...x_n)$ of a given medical treatment group, the group's number of patients (w_1), the relative number of osteoporosis patients (w_2) and the group's average age (w_3). This gives a log. regression model of type:
$Z \sim X + w1 + w2 + w3$
The obtained logistic regression describes the disease risk as a function of the log-odds (logarithmic odds):

$$\text{Logit(Z)} = \beta_0 + \beta_1 x_1 + \cdots + \beta_{31} x_{31} + \gamma_1 w_1 + \gamma_2 w_2 + \gamma_3 w_3$$

Where $Logit(Z)$ and the log odds for a patient medication x_i are given by the coefficients. The obtained log-odds were transformed into odds-ratio ORs to facilitate the interpretation of the obtained results. The individual medications coefficients were used to interpret the effect of a given medications on the risk of being diagnosed with osteoporosis. Of particular interest were the coefficients which displayed the dosage-dependent odds evolution.

7.3 Goodness of the model

To assess the goodness of our model, we controlled for:

1 Coefficient of determination (R^2)

2 Variance inflation factor (VIF)

3 False discovery rate (Benjamini– Hochberg procedure)

Ad1) The R^2, the coefficient of determination, is the proportion of the explained variance of the predicted variable through the predictors. The higher the R^2 the closer is the data to the fitted regression line. The coefficient of determination ranges from 0 to 1 (Urban and Mayerl, 2006; Peter and John, 1989). We tested the individual R^2 of each statin-dosage-dependent regression for each sex.
In the regression model for composite disease risk, we received the following R^2 values for: simvastatin: 0.936, lovastatin: 0.943, pravastatin: 0.934, fluvastatin: 0.949, atorvastatin: 0.955, and for rosuvastatin: 0.950.
For the individual regression analysis in females we received the following R^2 values for: simvastatin: 0.952, lovastatin: 0.963, pravastatin: 0.955:, fluvastatin: 0.960, atorvastatin: 0.965, and for rosuvastatin : 0.964.
Due to the small number of male subjects we received slightly lower R^2 values for: simvastatin: 0.812, lovastatin: 0.782, pravastatin: 0.796:, fluvastatin: 0.782, atorvastatin: 0.787, and for rosuvastatin: 0.777.

Ad2) To control for multicollinearity among the individual medications, whether they are independent of each other, we calculated the variance inflation factor (VIF) for each regression model (VIF). The VIF measure corresponds to the tolerance (1/tolerance) reciprocal transformation. The tolerance is the $1 - R^2$ difference. The R2 is derived from the regression of the corresponding predictor variables for all other model predictor variables (Urban and Mayerl, 2006).
Values greater than <= 10 indicate strong correlations among the predictor variables (O'brien, 2007; Urban and Mayerl, 2006). For the individual statin regression models we find the max. abs. VIF is for: simvastatin: 2.736, lovastatin: 4.197, pravastatin: 3.041:, fluvastatin: 3.177, atorvastatin: 2.861, and for rosuvastatin: 3.287.

Ad3) When applying multiple statistical test like logistic regressions, some p-values will be by chance less than the p-value of your choice. To control for these false- positives we have used the Benjamini–Hochberg procedure (McDonald, 2009; Benjamini and Hochberg, 1995).
First, all p-values are ranked from smallest to largest. Where the smallest has a rank of $i = 1$. Second, each individual p-value is compared to its Benjamini– Hochberg crtitical value$(i/m)Q$, where i is the rank, m is the total number of tests, and Q is the false discovery rate of your choice. Third, the largest p-value that has $P < (i/m)Q$ is significant. All smaller p-values are also significant, even if they are not less than their Benjamini—Hochberg crtitical value (McDonald, 2009).

Appendices

Bibliography

[1] Sally W Aboelela et al. "Defining interdisciplinary research: Conclusions from a critical review of the literature". In: *Health services research* 42.1p1 (2007), pp. 329–346.

[2] *(AKH) Allgemeines Krankenhaus der Stadt Wien.* https://https://www.akhwien.at/default.aspx?pid=575. Accessed: 2018-07-12.

[3] T An et al. "Efficacy of statins for osteoporosis: a systematic review and meta-analysis". In: *Osteoporosis International* 28.1 (2017), pp. 47–57.

[4] David W Bates et al. "Incidence of adverse drug events and potential adverse drug events: implications for prevention". In: *Jama* 274.1 (1995), pp. 29–34.

[5] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995), pp. 289–300.

[6] Marco Bernini and Angela Woods. "Interdisciplinarity as cognitive integration: auditory verbal hallucinations as a case study". In: *Wiley Interdisciplinary Reviews: Cognitive Science* 5.5 (2014), pp. 603–612.

[7] Guillaume Beslon et al. "From digital genetics to knowledge discovery: Perspectives in genetic network understanding". In: *Intelligent Data Analysis* 14.2 (2010), pp. 173–191.

[8] Dennis M Black and Clifford J Rosen. "Postmenopausal osteoporosis". In: *New England Journal of Medicine* 374.3 (2016), pp. 254–262.

[9] Bourdieu. *Outline of a Theory of Practice. (R. Nice, Trans.)* Vol. 16. Cambridge: Cambridge University Press Cambridge, 1977.

[10] Pierre Bourdieu. *Der Einzige und sein Eigenheim.* Hamburg: Vsa-Verlag, 1998.

[11] Pierre Bourdieu. *Distinction: A social critique of the judgment of taste. (R. Nice, Trans.)* Cambridge, MA: Harvard University Press, 1984.

[12] Pierre Bourdieu. *Homo academicus. (P. Collier, Trans.)* Stanford: Stanford University Press, 1988.

[13] Pierre Bourdieu. *Practical reason: On the theory of action.* Stanford, CA: Stanford University Press, 1998.

[14] Pierre Bourdieu. *The logic of practice. (R. Nice, Trans.)* Stanford, CA: Stanford University Press, 1990.

[15] Remo A Burkhard. "Learning from architects: the difference between knowledge visualization and information visualization". In: *null*. IEEE. 2004, pp. 519–524.

[16] Alberico L Catapano et al. "2016 ESC/EAS guidelines for the management of dyslipidaemias". In: *European heart journal* 37.39 (2016), pp. 2999–3058.

[17] K Arnold Chan et al. "Inhibitors of hydroxymethylglutaryl-coenzyme A reductase and risk of fracture among older women". In: *The Lancet* 355.9222 (2000), pp. 2185–2188.

[18] A Clark and D Chalmers. "The extended mind, analysis 58, 7–19". In: *Reprinted in the philosopher's annual (2000), Ridgeview* (1998), pp. 59–74.

[19] Andy Clark. *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge: MIT Press, 1989.

[20] Andy Clark. *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford, England: Oxford University Press, 2008.

[21] David C Classen et al. "Adverse drug events in hospitalized patients: excess length of stay, extra costs, and attributable mortality". In: *Jama* 277.4 (1997), pp. 301–306.

[22] Arnout RH Fischer, Hilde Tobi, and Amber Ronteltap. "When natural met social: a review of collaboration between the natural and social sciences". In: *Interdisciplinary Science Reviews* 36.4 (2011), pp. 341–358.

[23] Nikolaus Fogle and Georg Theiner. *The "Ontological Complicity" of Habitus and Field: Bourdieu as an Externalist (RE-REVISED)*. https://www.researchgate.net/publication/305725525_The_Ontological_Complicity_of_Habitus_and_Field_Bourdieu_as_an_Externalist_RE-REVISED. Accessed: 2018-01-4. Jan. 2017.

[24] Shaun Gallagher. "The socially extended mind". In: *Cognitive Systems Research* 25 (2013), pp. 4–12.

[25] Andreas Holzinger, Matthias Dehmer, and Igor Jurisica. "Knowledge discovery and interactive data mining in bioinformatics-state-of-the-art, future challenges and research directions". In: *BMC bioinformatics* 15.6 (2014), p. I1.

[26] Michael Edward Johansen et al. "Cardiovascular risk and statin use in the United States". In: *The Annals of Family Medicine* 12.3 (2014), pp. 215–223.

[27] Alexandra Kautzky-Willer, S Thurner, and P Klimek. "Use of statins offsets insulin-related cancer risk". In: *Journal of internal medicine* 281.2 (2017), pp. 206–216.

[28] Julie Thompson Klein. "A taxonomy of interdisciplinarity". In: *The Oxford handbook of interdisciplinarity* 15 (2010), pp. 15–30.

[29] Georg Kneer. *Differenzierung bei Luhmann und Bourdieu. Ein Theorienvergleich*. Frankfurt am Main: Suhrkamp, 2004.

[30] Tsung-Kun Lin, Pesus Chou, et al. "Long-term effect of statins on the risk of new-onset osteoporosis: A nationwide population-based cohort study". In: *PloS one* 13.5 (2018), e0196713.

[31] Tsung-Kun Lin, Yi-Sheng Liou, et al. "High-potency statins but not all statins decrease the risk of new-onset osteoporotic fractures: a nationwide population-based longitudinal cohort study". In: *Clinical epidemiology* 10 (2018), p. 159.

[32] John H McDonald. *Handbook of biological statistics*. Vol. 2. sparky house publishing Baltimore, MD, 2009.

[33] Christoph R Meier et al. "HMG-CoA reductase inhibitors and the risk of fractures". In: *JAMA* 283.24 (2000), pp. 3205–3210.

[34] Richard Menary. *Cognitive integration: Mind and cognition unbounded*. New York: Springer, 2007.

[35] Andrew Miles. "Complexity in medicine and healthcare: people and systems, theory and practice". In: *Journal of Evaluation in Clinical Practice* 15.3 (2009), pp. 409–410.

[36] Robert M O'brien. "A caution regarding rules of thumb for variance inflation factors". In: *Quality & quantity* 41.5 (2007), pp. 673–690.

[37] McCullagh Peter and Nelder John. *Generalized linear models*. London: Chapman and Hall, 1989.

[38] Allen F Repko. *Interdisciplinary research: Process and theory*. London: Sage, 2008.

[39] Fatima Rodriguez et al. "Association between intensity of statin therapy and mortality in patients with atherosclerotic cardiovascular disease". In: *JAMA cardiology* 2.1 (2017), pp. 47–54.

[40] Markus Schwingel. *Pierre Bourdieu zur Einführung*. Hamburg: Junius, 2003.

[41] Yuankai Shi and Frank B Hu. "The global implications of diabetes and cancer". In: *The lancet* 383.9933 (2014), pp. 1947–1948.

[42] John Stewart et al. *Enaction: Toward a new paradigm for cognitive science*. Cambridge, Massachusetts: MIT Press, 2010.

[43] MD Stone and DJ Hosking. "Treatment of osteoporosis: current and future." In: *Annals of the rheumatic diseases* 50.10 (1991), p. 663.

[44] Nicholas P Tatonetti et al. "Data-driven prediction of drug effects and interactions". In: *Science translational medicine* 4.125 (2012), 125ra31–125ra31.

[45] Hilde Tobi and Jarl K Kampen. "Research design: the methodology for interdisciplinary research framework". In: *Quality & quantity* 52.3 (2018), pp. 1209–1225.

[46] Dieter Urban and Jochen Mayerl. *Regressionsanalyse: theorie, technik und anwendung*. Vol. 2. Springer, 2006.

[47] Brian Uzzi et al. "Atypical combinations and scientific impact". In: *Science* 342.6157 (2013), pp. 468–472.

[48] TP Van Staa et al. "Use of statins and risk of fractures." In: *Journal of Bone and Mineral Research* 15 (2000), S155–S155.

[49] *WHOCC Guidelines for Daily Defined Dosage for Statins*. https://www.whocc.no/atc_ddd_index/?code=C10AA&showdescription=yes. Accessed: 2019-03-06.

[50] N Zethraeus et al. “Cost-effectiveness of the treatment and prevention of osteoporosis—a review of the literature and a reference model”. In: *Osteoporosis international* 18.1 (2007), pp. 9–23.