New Statistical Approaches for Modelling Chronic Disease Dynamics

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

Eingereicht von

Dina Voeltz

München, den 15.11.2023

Erstgutachter/in:	Prof. Dr. Annika Hoyer (Universität Bielefeld)
Zweitgutachter/in:	Prof. Dr. sc. hum. Oliver Kuß (Deutsches Diabetes Zentrum, Heinrich-Heine-Universität Düsseldorf)
Drittgutachter/in:	Prof. Dr. rer. nat. HDR Anne-Laure Boulesteix (Ludwig-Maximilians-Universität München)

Tag der Einreichung: 15.11.2023
Tag der Disputation: 06.02.2024

Für Rosi & Stefan

Acknowledgement

First and foremost, I would like to express my greatest possible appreciation and thanks to my supervisor Annika Hoyer. Thank you for all your trust, support, and insightful guidance throughout my academic journey so far. I am deeply indebted to you,…
…for giving me the opportunity to delve into a topic that I am passionate about,
…for your supportive and open-minded nature,
…for your always open door,
…for our "long-distance relationship" to work so well,
…for your valuable advice and encouragement throughout my time as doctoral student,
and for so much more.

Further, I wish to express my sincere thanks to Anne-Laure Boulesteix for joining my examination committee and for supporting me as my former mentor since my very beginning at the LMU. Thank you for sharing your knowledge, experience and advice and that you helped me grow and develop in my academic career. Above all, I am more than grateful that you made me aware of the possibility that I could join Annika to Bielefeld and at the same time maintain my affiliation to the LMU as an external doctoral student.

Furthermore, Oliver Kuß deserves big thanks for the collaboration in joint research projects and for being the second reviewer of this thesis. I very much appreciated your valuable input, helpful suggestions and constructive criticism on manuscripts, reviews, research projects and conference presentations. I would be more than pleased if our first shared publication was not the last.

My special thanks go to David Schindler, who was my former supervisor and who I joyfully worked with as research assistant in my time as Master student. My big thanks for initially planting the idea in my head to pursue a doctorate. Without you, this dissertation might not even exist.

Further, I express my thanks to all my co-authors for the fruitful collaborations, their advice and support during our joint research projects.

I want to thank my colleagues from the working group Biostatistics and Medical Biometry for all the stimulating discussions, as well as for our often cheerful but nonetheless productive working atmosphere. I am more than grateful for us being as heterogeneous as we are and yet our so homogeneously functioning team. It is a pleasure to work with all of you! Evidently, I express my thanks to all the LMU people who have always welcomed me with open arms.

Besides my appreciation of the precious academic support, my personal thanks go to my friends and family. I am more than grateful for their indirect contribution to this work be it through their kind, supportive, critical and motivating words or through any other way.

I express my most enormous thank you to my parents. Thank you, for your unconditional support, love and encouragement at all times, for being patient with me and your understanding when I had a stressful time.

I dedicate this work to Rosi and Stefan. You have played (and will continue to do so in the future!) an immeasurable role for me.

Zusammenfassung

Chronische Krankheiten, wie die koronare Herzkrankheit, Krebs, Demenz und Diabetes, gehören weltweit zu den häufigsten, folgenreichsten und wirtschaftlich bedeutendsten Gesundheitsproblemen. Auf globaler Ebene zählen chronische Erkrankungen, auch bekannt als nicht übertragbare Krankheiten (non-communicable diseases, NCD), zu 7 von 10 der häufigsten Todesursachen. Sie sind verantwortlich für 74 % aller Todesfälle pro Jahr (ca. 41 Millionen Todesfälle) und betreffen Menschen aller Altersgruppen, Regionen und Länder. Der zunehmende Anteil der an NCDs leidenden Menschen ist besorgniserregend, da diese Erkrankungen irreversibel sind und eine lebenslange ärztliche Betreuung erfordern, die Aktivitäten des täglichen Lebens einschränken und mit erheblich höheren Gesundheitskosten verbunden sind. Folglich sind die epidemiologische Forschung und die statistische Modellierung chronischer Krankheitsdynamiken, beispielsweise im Hinblick auf deren Prävalenz und Inzidenz, persistente Themen, die immer wichtiger und dringlicher werden. Die Literatur bietet einen umfangreichen statistischen „Werkzeugkasten“, der verschiedene Modelle chronischer Krankheiten umfasst. Diese Vielfalt bietet Chancen, kann jedoch die Wahl eines geeigneten Ansatzes erschweren. Alle Methoden haben ihre Vor- und Nachteile und sind daher für bestimmte Forschungsfragen mehr oder weniger geeignet. Unterschiedliche Modelle benötigen unterschiedliche Eingabefaktoren, basieren auf parametrischen oder nicht-parametrischen Annahmen und erfordern möglicherweise Individualdaten, während in anderen Fällen aggregierte Daten ausreichen. Die Implementierung kann einfach oder anspruchsvoll sein, ein Modell und seine Ergebnisse können leicht verständlich oder komplex und kompliziert zu interpretieren sein, ebenso können sich die Methoden hinsichtlich der Recheneffizienz unterscheiden. Das Ziel dieser kumulativen Dissertation besteht darin, das statistische Instrumentarium der Modellierung chronischer Krankheiten zu ergänzen, es zugänglicher und verständlicher zu machen und eine Anleitung für die Auswahl geeigneter statistischer Ansätze zu geben. Die Entwicklung, Anwendung und Bewertung statistischer Ansätze zur Quantifizierung und Prognose der Belastung durch chronische Krankheiten spielen in den fünf Beiträgen eine entscheidende Rolle. Um Verfügbarkeit und Transparenz zu maximieren, sind die Quellcodes aller beitragenden Artikel öffentlich abrufbar.

Der erste Teil dieser Dissertation befasst sich mit epidemiologischen Vorhersagemodellen. Der erste Beitrag vergleicht dazu drei Projektionsmethoden. Dabei liegt ein besonderer methodischer Schwerpunkt auf der Anwendung partieller Differentialgleichungen (PDE), die Prävalenz, Inzidenz und Mortalität in Beziehung setzen und das Illness-Death Modell (IDM) beschreiben. Der Vergleich zeigt, dass sich die Ergebnisse der Methoden erheblich unterscheiden. Die Verwendung sehr simpler Methoden, die zeitliche Trends bei Inzidenz und Mortalität außer Acht lassen, führt zu einer erheblichen Unterschätzung der künftigen Zahl von Menschen mit chronischen Krankheiten. Der zweite und dritte Beitrag schätzen die Inzidenz und prognostizieren künftige Fallzahlen von Diabetes in Deutschland, sowie damit verbundene wirtschaftliche Folgen. Beide Anwendungen unterstreichen das Potenzial des IDMs und der zugehörigen PDE für die epidemiologische Modellierung.

Der zweite Teil dieser Arbeit befasst sich mit der Ereigniszeitanalyse. In den letzten Jahren wurden nicht- und semiparametrische Modelle und Hazard Ratios als ihr primäres Ergebnis immer wieder hinsichtlich Interpretation, technischer Umsetzung und Flexibilität kritisiert. Als Abhilfe wird im vierten Artikel ein neues parametrisches additives Modell vorgestellt und dessen Einsatz und Leistungsfähigkeit in einer beispielhaften Anwendung und einer Simulationsstudie demonstriert. Per Definition überwindet das vorgeschlagene Modell die oben genannten Einschränkungen und kann als leistungsstarkes Werkzeug zur Ereigniszeitanalyse dienen.

Der letzte Teil ist dem Bereich von Simulationsstudien zum Vergleich verschiedener Modelle gewidmet, um dem Mangel an Leitlinien für geeignete methodische Entscheidungen in der medizinischen Forschung entgegenzuwirken. Ungeeignete statistische Methoden können zu ungenauen Ergebnissen und irreführenden Schlussfolgerungen führen, die die Qualität der Wissenschaft, der Evidenz und letztendlich die Qualität der Patientenversorgung gefährden können. Im Hinblick darauf können vergleichende Simulationsstudien Abhilfe schaffen. Solche Studien ermöglichen es, die Leistung und Eigenschaften statistischer Methoden in einem Umfeld zu verstehen, zu bewerten und zu vergleichen, in dem die „Grundwahrheit“ bekannt ist. Der fünfte Beitrag zielt darauf ab, den Status neutraler Vergleichsstudien zu stärken. Dabei wird zudem ein Beitrag zur medizinischen Forschung geleistet, indem drei frequentistische Ansätze zur Meta-Analyse von Diagnosestudien in verschiedenen Szenarien systematisch verglichen werden.

Abstract

Chronic diseases, such as coronary heart disease, cancer, dementia, and diabetes, are among the most common, consequential and economically important health problems worldwide. At a global level, 7 out of 10 leading causes of death are chronic conditions, also known as non-communicable diseases (NCD). They account for 74% of all deaths each year (approx. 41 million deaths), affecting people of all ages, regions and countries. The increasing proportion of people suffering from NCDs is worrying as these conditions are irreversible and necessitate ongoing, lifelong medical attention, constrain activities of daily living and are associated with considerably higher healthcare expenses. Consequently, epidemiological research that addresses NCDs and statistical modelling of disease dynamics in terms of, for example, prevalence and incidence, are persistent topics that continue to become greater in importance and urgency. Literature provides a large statistical "toolkit" that encompasses various chronic disease models. This variety provides opportunity, but may complicate the choice of an appropriate approach. All methods come with their advantages and disadvantages, and therefore are more or less suitable for certain research question. Different models rely on different input factors, they may require individual data while in other cases aggregated data is sufficient, approaches may be based on parametric or nonparametric assumptions, their implementation may be rather simple or challenging, a model and its results might be easily understandable or more complex and complicated to interpret, and the methods will presumably differ in terms of computational efficiency. The aim of this cumulative dissertation is to complement the statistical toolkit of chronic disease modelling, to make it more accessible and comprehensible, and to provide guidance for selecting suitable statistical approaches. The development, application and evaluation of statistical approaches to quantify and project chronic disease burden play a crucial role in the five contributing articles. To make the methodology accessible and to maximise transparency, the source codes of all contributing articles are publicly available.

The first part of this dissertation deals with epidemiological projections. The first contribution compares three projection methods, with a particular methodological focus on the application of partial differential equations (PDE) that relate prevalence, incidence and mortality and describe the illness-death model (IDM). The comparison shows that the methods' results differ substantially. It appears that using too simplistic methods which ignore temporal trends in incidence and mortality leads to severe underestimation of the future number of people with chronic diseases. The second and third contribution estimate the incidence and project the future number of people with diabetes in Germany and its related economic consequences. The two applications underline the potential of the IDM and its associated PDE for epidemiological modelling.

The second part of this thesis addresses the analysis of time-to-event outcomes. In recent years, non- and semi-parametric models and hazard ratios as their primary outcome have continuously been criticized in terms of interpretation, technical implementation, and flexibility. As a remedy, the fourth article proposes a new parametric additive hazard model and demonstrates its use and performance in an exemplary application and a simulation study. By definition, the proposed model overcomes the above-mentioned limitations and may serve as a powerful tool for analysing time-to-event outcomes.

The last part is devoted to the field of comparison studies to counteract the lack of guidance on suitable methodological choices in medical research. Inappropriate statistical methods may lead to inaccurate results and misleading conclusions which could jeopardise the quality of science, evidence, and ultimately, the quality of care given to patients. This is addressed by comparative simulation studies. Such studies allow to understand, evaluate and compare the performance and properties of statistical methods in a setting where the "ground truth" is known. The fifth article aims to reinforce the status

of neutral comparison studies and contributes to medical research by systematically comparing three frequentist approaches for the meta-analysis of diagnostic test accuracy studies in various scenarios.

List of Abbreviations

ADEMP	Aims, Data-Generating Mechanisms, Estimands, Methods, Defining Aims
AFT	Accelerated Failure Time
AKI	Acute Kidney Injury
ARI	Absolute Risk Increase
ARR	Absolute Risk Reduction
AUC	Area under the Curve
CDF	Cumulative Distribution Function
CONSORT	Consolidated Standards of Reporting Trials
DTA	Diagnostic Test Accuracy
FSO	Federal Statistical Office
HALLUCA	Halle Lung Cancer
HARKing	Hypothesizing After the Results are Known
HR	Hazard Ratio
ICD	International Classification of Diseases
IDM	Illness-Death Model
LADA	Latent/ Late Onset Autoimmune Diabetes in Adults
MODY	Maturity-Onset Diabetes of the Young
MRR	Mortality Rate Ratio
NCD	Non-Communicable Disease
NGAL	Neutrophil Gelatinase associated Lipocalin
NNH	Number Needed to Harm
NNT	Number Needed to Treat
NSCLC	Non-Small Cell Lung Cancer
ODE	Ordinary Differential Equation
OR	Odds Ratio
PDE	Partial Differential Equation
PDF	Probability Density Function
PH	Proportional Hazard
RD	Risk Difference

ROC	Receiver-Operating-Characteristics
RR	Relative Risk/ Risk Ratio
RRR	Relative Risk Reduction
SHI	Statutory Health Insurance
SIR	Susceptible-Infected-Removed
SOTA	State of the Art
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TNM	Tumor-Node-Metastasis
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
WHO	World Health Organisation
YLL	Years of Life Lost

List of Tables

List of Figures

Contents

1 Introduction

"The major factors that brought health to mankind were epidemiology, sanitation, vaccination, refrigeration, and screen windows." [1]
(Richard Lamm)

While infectious diseases have an essential impact on the health of people worldwide, chronic conditions have evolved to a major global health threat. Chronic diseases, also called non-communicable diseases (NCD), such as coronary heart disease, cancer, dementia, and diabetes, impose great burden on humanity [2]. NCDs are among the most common, consequential and economically important health problems worldwide [3,4]. They are particularly worrying as these diseases are irreversible, require lifelong medical care, restrict activities of everyday living, and are related to remarkably higher healthcare costs. Globally, NCDs belong to 7 of the top 10 causes of death [5]. They affect people of all ages, regions and countries and account for 74% of all deaths (about 41 million deaths) per year [2,3,5]. By 2030, the number of deaths related to NCDs are estimated to reach 52 million per year. Further, according to the World Health Organization (WHO), costs for treating the worldwide NCD epidemic are projected to increase to $30 trillion by 2030 [4]. However, epidemiologic studies provide evidence that up to 80% of all cases associated with NCDs such as diabetes, heart diseases and strokes, as well as 40% of all cancer diagnoses are attributable to a number of known and, above all, preventable risk factors [4]. Thus, further research and the promotion to modify major risk factors such as smoking, high cholesterol alcohol use, poor diet, high blood pressure, obesity and physical inactivity, would likely result in reducing a considerable proportion of the burden caused by NCDs.

Overall, the rising proportion of people suffering from NCDs and the associated, ever-increasing healthcare costs require more knowledge about diseases and forecasting of future chronic disease burden in order to initiate preventive measures and estimate expected requirements. Consequently, epidemiological research that inherently focuses on statistical modelling of chronic disease dynamics is a persistent topic of great importance and urgency. The main purpose of this cumulative dissertation is to advance statistical modelling in the context of chronic diseases, as well as to make it more accessible and comprehensible. To achieve this, the thesis is made up of three distinct parts, each with a different focus, which underline the relevance and usefulness of statistical approaches for chronic disease modelling. The five contributing articles that accompany the three parts focus on the development, application and evaluation of statistical methods to quantify and project chronic disease burden.

Part I - Epidemiological Projection Models

Epidemiology is the study of the distribution and determinants of diseases in human populations and focuses on the description of health-related states, rates, events and trends. Compared to clinical research, which primarily works at the level of individual patient data, epidemiology concentrates on research questions at the population level[6]. The essential aim is to inform health professionals and the public at large, ultimately improving general health situations. For instance, epidemiologists study chronic and long-duration diseases (e.g., asthma) as well as infectious diseases (e.g., cholera) which may be derived from the idea of an "epidemic" (i.e., the appearance of a particular disease in a large proportion of people in a particular region, community, or population at the same time)[6]. Since even in the developing world, chronic diseases are among the most common causes of death, the importance and amount of epidemiological research that focuses on the context of chronic conditions

increases. With the constant rise of chronic disease burden over the past decades, epidemiological measures continue to become more relevant and popular[6]. Being two of the fundamental measures, the incidence and prevalence of diseases are commonly reported[7]. The incidence rate refers to the number of new cases with a condition of interest within a specified period of time, while the prevalence summarises the proportion of existing cases versus the considered population number at a particular point in time[6,8]. Further, projections of measures that quantify disease frequency and anticipating future demands play a central role in disease surveillance and management as well as in the guidance of future healthcare resources and policies. For this aim, statistical projection models are powerful tools but obviously, different models may vary in their outcomes and closeness to reality. In the context of chronic disease projections, several methodological approaches have been advocated[9]. Among these statistical techniques, a common approach simply combines the previously estimated prevalence of the disease of interest and future population numbers to project case numbers[10]. This method requires little data and statistical knowledge. However, it might be too simplistic to fully capture the complex nature of chronic diseases. As an alternative, multistate models, also called compartment models, are able to incorporate underlying disease-specific transition rates and have proven to more accurately mirror reality[10,11]. The illness-death model (IDM) is a typical compartment model in the epidemiology of chronic diseases that considers the three states healthy, diseased and dead and that represents the transition rates from one state to another as continuous-time stochastic processes[9]. The theoretical background of the IDM allows to derive differential equations that relate prevalence, incidence and mortality[12,13]. As there remains considerable debate about the appropriate methodological approach for chronic disease projections, this thesis derives, employs and critically discusses selected methodologies. By assessing the strengths and limitations of the different techniques, this dissertation aims to help researchers to refine estimations of future chronic disease burden. As one of the PDE approaches shows particular potential in the context of chronic disease epidemiology, this thesis illustrates the method using aggregated and routinely collected data to estimate and project epidemiological and economic measures of diabetes in Germany.

Part II - Time-to-Event Models

Survival analysis, or more generally time-to-event analysis, refers to a set of statistical procedures for analysing the time which elapses until some event occurs. In health care research, the time period and the well-defined endpoint of interest could be the diagnosis of a disease, relapse from remission, or recovery, as for instance, the time period between a confirmed response versus the first relapse and recurrence of cancer, the cessation of breast feeding, the time between fertilization and conception or the discharge from hospital after surgical treatment of a distal radius fracture [14,15]. Often in medical research and as indicated by the term "survival analysis", the endpoint considers the death of a person. Such survival or time-to-event outcomes are mostly analysed by the Cox Proportional Hazards (PH) model, which is an essentially non-parametric method [16]. As a result of the Cox regression model, it is common practice to report the hazard ratio (HR), i.e. the ratio of hazard rates in e.g. an exposed versus non-exposed group. Conveniently, the estimated HR summarizes the treatment effect over the entire length of the trial in a single number whereas other measures such as median survival only compare survival time at one point. Recently however, the Cox model and the HR have been criticized for a number of reasons. Most relevant points of concern were (i) that the HR must be interpreted judiciously and that it is often mistaken as relative risk [17], (ii) its inherent selection bias or left truncation bias even in randomized trials [18] and (iii) its non-collapsibility, i.e. conditioning on a covariate that is related to the event generally changes the HR, even if this covariate is unrelated to the exposure [19]. Despite these disadvantages, the Cox model and the HR remain one of the most used statistical models and conventional effect measure in time-to-event analysis. This gives rise to the question why parametric survival models are not preferred over the Cox model, particularly when knowing that in

regression models for continuous, binary, nominal, or ordinal outcomes typically rely on parametric modelling. Parametric methods have been developed long before the Cox model, they are easier to estimate, interpret and communicate [20,21]. This thesis addresses this controversy and, as a remedy, proposes a new parametric additive hazard model that does not suffer from the disadvantages mentioned. The exemplary application using data from a study investigating medical care to lung cancer patients shows that the approach works well in practice.

Part III - Comparison and Benchmark Studies

In many scientific fields, there are strong incentives to devote to the development of new methods. New methods are expected to "improve our world", be it by simplifying processes or bringing results of statistical analyses closer to the truth. For authors, it seems appealing to entail new analytical approaches in their research as it is one of the most straightforward ways for a work to be considered as novel and innovative, a prerequisite for publication[22]. Consequently, the development of methods is an active area of research in statistics. This situation is a double-edged sword: it provides opportunity but at the same time it is a serious challenge as the choice of a method may considerably affect results[23]. With the constant introduction of new algorithms and models, it is becoming increasingly difficult to remain informed and to select the most appropriate approach. This problem is addressed by comparison and benchmark studies which aim to systematically analyse and compare several methods in an unbiased manner[23]. Using simulated or real data sets and by assessing the methods in different conditions, they provide guidance for suitable method choices depending on the respective study setting, hypothesis or research question. This dissertation aims to contribute to the improvement of research practices and methods by (i) reinforcing the status of neutral comparison studies and (ii) contributing to medical research by conducting a comparison of three frequentist approaches for the meta-analysis of diagnostic test accuracy (DTA) studies.

Outline

Section 2.1 is devoted to statistical projections methods for epidemiological research. It introduces and compares different approaches for estimating and projecting common epidemiological measures, i.e., the incidence, prevalence and case number of people with a specific disease, as well as disease-related healthcare expenses. Section 2.2 deals with statistical methods used for time-to-event analysis. It critically discusses commonly used approaches as well as the most popular effect measure in that field, the HR. Building on this background, the main focus of section 2.2 is to introduce a new parametric additive hazard model. Section 2.3 gives a brief introduction to the background and relevance of comparison and benchmark studies. Further, the section provides information on three recently published frequentist methods for the meta-analysis of DTA studies that are then systematically compared in a simulation study. Chapter 3 provides a short overview of the software used to implement the statistical models. Further, it specifies whether and how the data and code that support the results of the studies can be accessed. Finally, Chapter 4 concludes with a general discussion of the main findings and implications.

Contributing articles

Each of the sections 2.1, 2.2 and 2.3 encompasses at least one contributing article which has been either published in a scientific journal, has been accepted or is currently under review for publication. In the latter case, the latest version of the submitted article is included. Alongside the articles, a specification of the contributions of all authors involved in creating each of the manuscripts is provided.

2 Methodological and General Background

Part I

2.1 Epidemiological Projection Methods

"It is far better to foresee even without certainty than not to foresee at all." [24]
(Henri Poincare)

One of the main purposes of epidemiology is to study the patterns, causes and effects of health and disease on population level, as well as to describe the frequency and distribution of a disease in a specified population [8]. In order to understand the epidemiology of a disease, it is essential to continuously and systematically collect, analyse and interpret health data. This ongoing collection of information and monitoring of a disease is known as surveillance [8]. It is among the main goals of disease surveillance systems to determine the need for public health actions, support information-based decision making in healthcare, and contribute to planning, implementing and evaluating public health practice, policies, prevention activities and disease management programs. Further, epidemiological research and disease surveillance aim at informing about temporal changes of a health problem and potential future burden caused by the surveyed disease [8].

In order to reach the goals of epidemiology and disease surveillance, various quantitative measures, e.g., incidence, prevalence, disease-related complications and mortality, are estimated and projected. Particularly the incidence and prevalence are among the central concepts and are commonly encountered in epidemiology [8,25]. The prevalence represents the proportion of people in a population having a disease [8]. By that, this measure of disease status may also function as indicator on the burden of a disease in a certain population in terms of monetary costs, quality of life, life expectancy or morbidity [25]. Such knowledge is often crucial for decision makers to determine where investments in health care should be targeted. While the prevalence reflects on the existing cases of a disease, the incidence refers to the number of new disease cases within a certain period [25]. The incidence rate is an important measure for assessing the risk and burden of a disease on the population, for the economy and the health sector. Insights on the incidence rate are key for instance for decision making about future public health interventions and policies, as well as for planning and evaluating prevention activities [8,25].

Knowing that the proportion of people suffering from chronic diseases is continuously increasing and that disease management activities have shown the potential to reduce the risk of acute complications, premature mortality and the general suffering caused by chronic diseases, epidemiological research and disease surveillance are of utmost importance in the context of chronic diseases [26,27]. Though, effective responses to any chronic disease require accurate estimates of current and future chronic disease burden to tailor healthcare activities. Consequently, projections of epidemiological measures are central for planning and anticipating future need of health care resources, for evaluating prevention activities as well as for identifying potential risk factors or high-risk groups [11,28]. However, in order to be of use, information from disease surveillance must be representative, timely and efficient. This can be achieved by using mathematical models and the best available, i.e., most representative, comprehensive and up-to-date, data allowing to compile reasonable projections under certain assumptions [11,28]. Thereby, statistical projection models can be valuable tools to speculate on future distribution of diseased and non-diseased.

With regards to statistical projection methods of chronic diseases, there remains considerable debate about methodological approaches. Obviously, different models may vary in their assumptions, outcomes and closeness to reality [28]. Further, projection methods are inherently limited by the availability of epidemiological and demographic data. In the context of projecting chronic disease case

numbers, one 'status quo approach' is most often used which relies on a simple application of the observed prevalence from a base year to population projections [29-31]. Precisely, the age- and sex-specific prevalence $p(t,a)$ at time t can be obtained from

$$p(t,a) = \frac{I(t,a)}{H(t,a) + I(t,a)} \quad (1)$$

Where $I(t,a)$ and $H(t,a)$ denote the number of people with and without a disease, respectively. To project the number of age- and sex-specific future cases $I(t,a)$, the estimated prevalence $\hat{p}(t,a)$ is then applied to the age-, sex- and time-dependent future population number $N(t,a)$ as in

$$I(t,a) = \hat{p}(t,a) \times N(t,a) \quad (2)$$

Other epidemiological factors, such as the incidence, are only incorporated implicitly in the prevalence. However, this method disregards the fact that prevalence is a consequence of incidence and mortality and ignores any other epidemiological determinant. Hence, it is questionable whether this method is able to accurately mirror reality. Vice versa, statistical methods that reflect on underlying disease-specific transition rates, i.e., incidence and mortality, may better mirror the complex nature of chronic diseases. Therefore, some (more advanced) studies rely on multistate models that relate disease-specific transition rates. These include for instance time-discrete Markov models [11,28], Poisson regression to model disease-specific transition rates [32,33] or differential equations [12,13,34].

Overall, further developing and spreading the knowledge about accurate projection methods is essential to counteract the ever-worsening chronic disease situation. For that aim, the part on epidemiological projection models is structured as follows: Section 2.1.1 introduces the theoretical background of multistate models and specifically, the illness-death model (IDM) as one popular example of multistate models in epidemiology. Section 2.1.2 focuses on the projection of case numbers via a set of two partial differential equations (PDE). Section 2.1.3 relates the prevalence, incidence and mortality of a chronic disease in a single PDE. The purpose of Section 2.1.4 is to give a practical example of how to model the future epidemiological and economic burden of diabetes mellitus in Germany. The overall section 2.1 serves as basis for the first three contributing articles which are all devoted to projections of the current and future burden of type 1 and type 2 diabetes (T1D, T2D) in terms of case numbers, prevalence, incidence and a variety of healthcare cost measures.

2.1.1 The Illness-Death Model

Most generally, a multistate model represents a stochastic process $X(t)$ with a set of two or more discrete and disjunct states which determines the state space S. Depending on the context, multistate models are also referred to as compartment or state models. It is assumed that all individuals of a population are in one of the states. As implied by the function of t, multistate models are dynamic in the way that the number of individuals in each compartment may fluctuate in the course of time. The progression from one state to another is described by probabilities to occupy one of the different states and the intensities for the transitions.

Multistate models are most often built around deterministic differential equations [9]. Such deterministic models produce consistent outcomes for a given set of inputs, i.e., each problem belongs to one set of specified values and only one solution. Its mathematical characteristics are known, none of them is random and from deterministic perspective, there is no allowance of error. Deterministic models have the advantage of being both, relatively easy in terms of implementation and conceptually simple. However, they lack stochasticity which may be a relevant feature in the context of modelling of chronic conditions. In contrast to deterministic models, stochastic methods are able to incorporate this random element at some level by integrating unknown components into the model. Essentially,

compartment models can also be used with a stochastic (random) framework [9,35,36]. In some cases, representing disease onset or disease-related mortality in a human population may be more realistically represented by a stochastic process that comprises a certain level of randomness [35]. In that view, a stochastic framework may sometimes be more suitable and more realistic for epidemiological modelling, although it may be more complex to analyse.

Compartmental models are a general modelling technique and nowadays, a common tool for mathematical modelling of infectious and chronic diseases. For example, they are valuable statistical tools for predicting how a disease spreads within a certain population, for estimating or projecting the total number of diseased, to model the duration of an epidemic or pandemic, or to computing various other epidemiological measures such as the incidence, reproduction number or years of life lost (YLL) as a measure of premature mortality. The origin of such models dates back a long time to the early 20th century where they have first appeared in the context of infectious diseases in pioneering works of Ross (1916) [37], Ross and Hudson (1917) [38], Kermack and McKendrick (1927) [39] and Kendall (1956) [40]. In chronic disease epidemiology, compartment models have appeared later and gained more attention since the 1950th [41]. In epidemiology, a well-known compartment model is the competing risks model (Figure 1)[42]. Most generally, competing risks are events which preclude the occurrence of a primary event or outcome of interest or modify the risk of the considered outcome[43,44]. In medical context, one common example of competing risks may refer to disease relapse and death in remission[43,44]. For instance, a competing risks model may consider different causes of death, and the final state "death" is divided into two or more states (see Figure 1)[42]. Two special cases of the competing risk model that are commonly applied in epidemiology, are the susceptible-infected-removed model (SIR model) (Figure 2) and the IDM (Figure 3). The SIR model originates in infectious disease modelling. In this work, the focus is on the IDM (Figure 3) as it has been proven valuable in reflecting on the complex interplay of three basic epidemiological parameters, namely, the incidence, prevalence and mortality of a chronic disease [13,35].

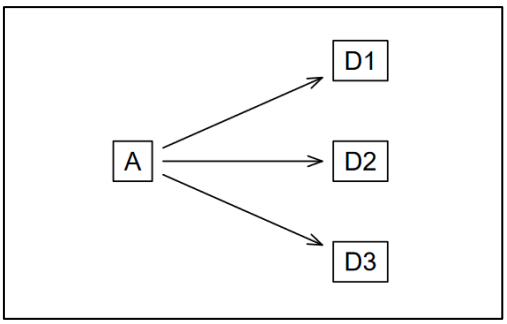

Figure 1: Exemplary competing risk model. The model consists one state "Alive" (A) and three states "Dead" that represent different causes of death (D1, D2, D3).

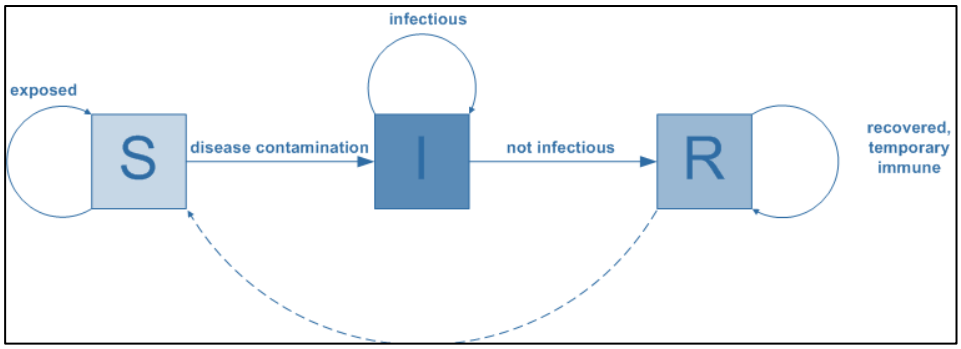

Figure 2: SIR model. The state S denotes the number of susceptibles, i.e., individuals that may infect with a specific disease. State I represents the number of infectives which may transmit the disease to other susceptibles. State R refers to the number of recovered people who return back to state S, who remain in state R with immunity or who died.

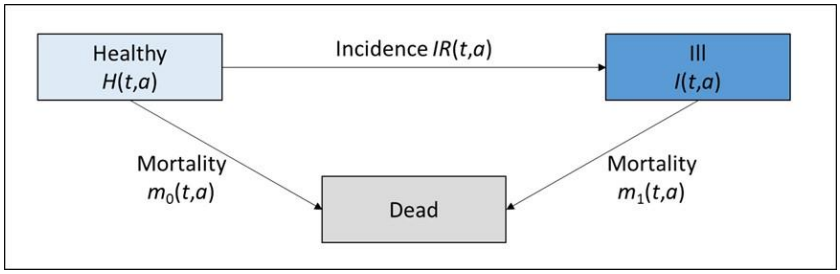

Figure 3: Illness-Death-Model of a chronic disease. All people in a population are assumed to be in one of the three states: Healthy, Diseased, or Dead. H(t,a) and I(t,a) denote the number of people without and with the disease of interest, respectively. It is assumed that at birth, all people start in the healthy state. Depending on time t and age a of each respective person, they will then transition to another state which is described by the incidence IR, the mortality of the non-diseased m_0, and the mortality of the diseased m_1.

The IDM is a multistate model that represents a Markov process in continuous time $\{X_t; 0 \leq t < \infty\}$. In the context of chronic diseases, the IDM can be restricted to modelling three states, i.e., a state space $S = \{1, 2, 3\}$, as depicted by the IDM (Figure 3). More specifically, it consists of the three states "Healthy" (with regards to the chronic disease of interest), “Ill” and “Dead” and it is assumed that each individual of a population is in one of the relevant disease states. Therein, $H(t, a)$ depicts the number of non-diseased aged a at time t, the disease state "Ill" comprises the number of ill people depicted by $I(t, a)$ and the death state indicating the number of deaths. It is assumed that at birth all individuals begin in the healthy state. From there on, they can either be diagnosed with a chronic disease and then die at some point in time, or they can transition directly to death state (without ever contracting the disease under consideration). The arrows in Figure 3 indicate the possible progressions between the states. These are modelled by transition rates which are all functions of calendar time t and age a. The transition rates are given by the incidence rate $IR(t, a)$ the mortality of the non-diseased $m_0(t, a)$ and the mortality of diseased people $m_1(t, a)$. Reflecting on chronic conditions, there is no remission from the diseased back to the healthy state.

2.1.2 Projecting Case Numbers via a Two-Dimensional Set of Partial Differential Equations

Building upon the theoretical background of the IDM and assuming its transition rates as constant in the short term (e.g., for the time period of a year) [45], Murray et al. [45-47] showed that the relation between prevalence, incidence and mortality can be expressed in terms of a set of ordinary differential equations (ODEs) [13]. An ODE is a differential equation dependent on only one single independent variable [48], which is, in the framework of Murray et al. [46,47], age a. Hence, the age variable a describes the (only) temporal progression. In the so-called incidence-prevalence model, Murray et al. [46,47] describe the evolution of the susceptible, i.e., healthy, population and infected or ill subpopulations. A system similar to the one presented by Murray et al. [46,47] is given by

$$\frac{\mathrm{d}H}{\mathrm{d}a} = -[IR(a) + m_0(a)] \times H(a) \tag{3}$$

$$\frac{\mathrm{d}I}{\mathrm{d}a} = IR(a) \times H(a) - m_1(a) \times I(a) \tag{4}$$

where $\frac{\mathrm{d}H}{\mathrm{d}a}$ and $\frac{\mathrm{d}I}{\mathrm{d}a}$ denote the change in the number of people without and with the disease under consideration, respectively. $IR(a)$ denotes the incidence rate, while $m_0(a)$ and $m_1(a)$ represent the mortality rate of the non-diseased and diseased, which are all dependent on age a. The set of ODEs presented in (3) and (4) is linear and of first order [49]. It assumes time-homogeneity, i.e., time dependence only with regards to temporal changes of age a, and that the transition rates (i.e., the incidence $IR(a)$ and mortality rates $m_0(a)$ and $m_1(a)$) are non-negative and sufficiently smooth [7,49]. The system of ODEs implies that the population is stationary and closed (i.e., no migration for instance), and that the birth rate is constant. The age-specific prevalence $p(a)$ is defined as

$$p(a) := \frac{I(a)}{I(a) + H(a)}. \tag{5}$$

The analytical solution of the corresponding initial value problem with initial conditions $H(0) = H_0 \geq 0$, $I(0) = I_0 \geq 0$ and $H_0 + I_0 > 0$ can be formulated as in [49]

$$H(a) = H_o \times \exp\left(-\int_0^a IR(\tau) + m_0(\tau)d\tau\right) \tag{6}$$

$$I(a) = \exp\left(-\int_0^a m_1(\tau)dt\right) \times \left(I_o + \int_0^a IR(\tau)H(\tau)\exp\left(\int_0^\tau m_1(t)dtd\tau\right)\right) \tag{7}$$

In reality however, the assumption of time-homogeneity is doubtworthy, and the dynamics of the IDM are more likely to depend on multiple time scales, such as age, calendar time and duration of a disease [13]. Apart from the age a, disease rates commonly also depend on the calendar time t. For instance, disease patterns of asthma and chronic obstructive pulmonary disease (e.g. in terms of prevalence and symptoms) were found to vary by season[50]. The dependence of the parameters on the two time scales of age a and calendar time t in the IDM can be illustrated in a Lexis diagram (Figure 4).

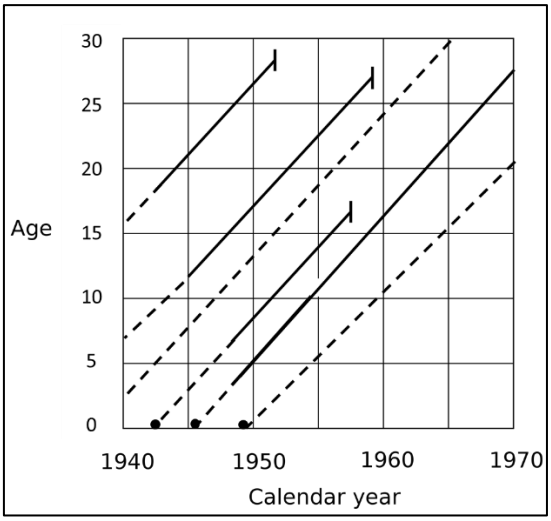

Figure 4: Lexis diagram of birth, onset of disease and death: birth •, healthy - - -, diseased -------, death |. Each line in the diagram represents the lifeline of one individual.

The lines depicted in the Lexis diagram represent the life course of individuals along the two time scales [35]. The end of such a life line indicates the point in time where the respective individual has left the

population under consideration, e.g., due to death or emigration [35]. As age a and calendar time t grow at the same pace, all life lines have a slope equal to 1. Assuming that the transition rates depend on more than one time scale, i.e., they are affected by changes in age a and calendar time t, Brunet et al. [51] have shown that and (3) and (4) can be rewritten as a two-dimensional system of PDEs [12,13,51]. In contrast to the ODEs defined above, a PDE is defined as equation that depends on partial derivatives of a multivariable function, i.e., one with two or more independent variables [48]. The rates $m_0(t,a), m_1(t,a)$ and $IR(t,a)$ henceforth depend on age a and calendar time t. The following system of PDEs describes the population flows of the IDM in the context of chronic diseases at any time t and age a:

$$\partial H(t,a) = -[IR(t,a) + m_0(t,a)] \times H(t,a) \tag{8}$$

$$\partial I(t,a) = IR(t,a) \times H(t,a) - m_1(t,a) \times I(t,a). \tag{9}$$

$\partial H(t,a)$ indicates age-, sex- and time-specific changes in the number of people without the disease of interest, while $\partial I(t,a)$ represents changes in the number of diseased. Equations (8) and (9) are now classified as PDEs, as they govern partial derivatives with respect to the two variables age a and calendar time t, instead of only one variable. As such, they describe the temporal changes of the number of diseased and non-diseased along the life lines in the Lexis diagram.

In epidemiology, it is typically the case that $m_0(t,a)$ and/or $m_1(t,a)$ are unknown [49]. However, the general mortality rate $m(t,a)$ of the overall population of interest (i.e., a convex combination of the mortality of diseased and non-diseased) is commonly observed and reported in life tables. As a remedy, it has been shown mathematically equivalent to incorporate the general mortality $m(t,a)$ and the mortality rate ratio (MRR), that is, the ratio of the mortality rates of people with versus without the disease under consideration $\left(MRR(t,a) = \frac{m_1(t,a)}{m_0(t,a)}\right)$ [34].

The general mortality can be computed from

$$m(t,a) = p(t,a) \times m_1(t,a) + \big(1 - p(t,a)\big) \times m_0(t,a) \tag{10}$$

$$= m_0(\mathrm{t},\mathrm{a}) \times [p \times MRR - 1) + 1]. \tag{11}$$

Implementing the prevalence $p(t,a)$, the general mortality (mt,a) and the $MRR(t,a)$ in

$$m_0(t,a) = \frac{m(t,a)}{(1 + p(t,a) \times (MRR(t,a) - 1))} \tag{12}$$

and

$$m_1(t,a) = m_0(t,a) \times MRR(t,a) \tag{13}$$

it is possible to derive $m_0(t,a)$ and $m_1(t,a)$. The estimates of the two mortality rates can then be used as input data to apply the two-dimensional set of PDEs shown in (8) and (9).

To derive and project the number of cases of a chronic disease via the two-dimensional set of PDE shown in Equation (8) and (9) in a practical example, the proposed approach requires disease-specific and demographic input data. The method relies on input values on the age- and sex-specific prevalence, the age- and sex-specific MRR and the age- and sex-specific general mortality for a base year. Aggregated data are sufficient. The latter are used to estimate the mortality of the non-diseased and the mortality of diseased people. Further, if in a projection it is of interest to model temporal trends in the incidence and/or mortality rate, this would require additional information on the rate

development over time. Projection results would be independent from future population estimates beyond the year used to determine initial values of $m_0(t,a)$ and $m_1(t,a)$.

2.1.3 Relating Prevalence, Incidence and Mortality in a Single Partial Differential Equation

In epidemiology, the prevalence is a fundamental measure to quantify disease frequency. Defined as the proportion of a specified population found to be affected by a (medical) condition at a certain point or period of time, it is often the preferred measure over absolute case numbers. Consequently, it is of interest and of great practical value to express (8) and (9) in terms of the age-specific prevalence instead of explicitly modelling the numbers of diseased and non-diseased. In 1999, and building upon the theoretical framework of the IDM in the context of chronic conditions, Brunet et al. [51] were among the first to present a differential equation that describes the relation of the prevalence odds and the incidence and mortality rates of the chronic disease under consideration. Their approach has been generalised, extended and reformulated, for instance to the context of infectious disease epidemiology, to allow for modelling remission rates or migration [7,9,13]. The following section shows how to derive the age-specific prevalence as solution of a PDE. For that means, it is possible to relate the theory of stochastic processes and (deterministic) differential equations [9].

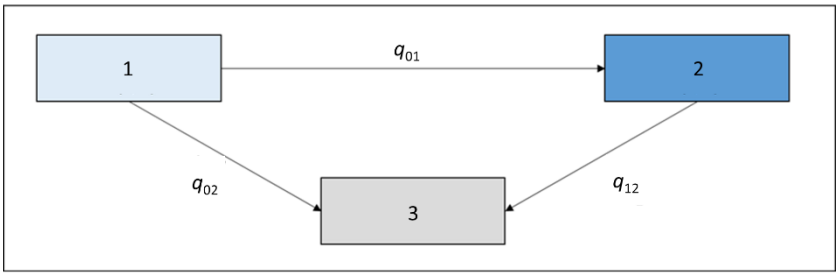

Figure 5: Theoretical Illness-Death Model expressed as Markov chain in continuous time. The boxes represent the states and the arrows depict possible transitions from one state into another.

Precisely, to derive the formulation of the prevalence in terms of a PDE, the IDM from Figure 3 is now considered in a more theoretical setting and expressed as Markov chain in continuous time $\{X_t; 0 \leq t < \infty\}$ (see Figure 5). This continuous-time stochastic process consists of a mathematical system that describes a sequence of possible transitions from one state into another in which the probability of each event depends only on the state attained in the previous period. In the context of chronic diseases and the IDM (Figure 3), the system is restricted to three states, i.e., a state space $S = \{1, 2, 3\}$. States are changed according to the probabilities of a stochastic matrix $Q(t) := (q_{jk}(t))_{j,k \in S} \geq 0$. The arrows in Figure 5 indicate the possible transitions from one state into another. The rate matrix, also-called intensity matrix, $Q(t)$ comprises the rates or intensity functions $q_{jk}(t)$ which denote the transitions from state j to a subsequent state k where $j, k \epsilon S, j \neq k$. Generally, the terms intensity and rate are used synonymously. The concrete associated intensity matrix to the current setting is

$$Q(t) = \begin{bmatrix} q_{00} & q_{01} & q_{02} \\ q_{10} & q_{11} & q_{12} \\ q_{20} & q_{21} & q_{22} \end{bmatrix} = \begin{bmatrix} -(q_{01} + q_{02}) & q_{01} & q_{02} \\ 0 & -q_{12} & q_{12} \\ 0 & 0 & 0 \end{bmatrix} \tag{14}$$

with $q_{10} = q_{20} = q_{21} = q_{22} = 0$. With regards to the rates,

$$\sum_k q_{jk}(t) = 0 \tag{15}$$

which implies that

$$q_{jj}(t) = -\sum_{j \neq k} q_{jk}(t) \quad \forall j \in S. \tag{16}$$

The elements q_{jk} are non-negative for $j \neq k$. Rates are chosen such that each row of the intensity matrix $Q(t)$ sum to zero. As usual in the theory of Markov models, the process is assumed time homogeneous, i.e., transition probabilities $P_{jk}(s,t)$ are independent of t. For an individual in state j at time s which transitions to state k at time t with $j,k \in S$ and $0 \leq s < t < \infty$, the transition probabilities $P_{jk}(s,t)$ equal

$$P_{jk}(s,t) = P(X_t = k \mid X_s = j). \tag{17}$$

Using Kolmogorow's forward differential equations to define the temporal change of the probability that a continuous-time Markov process is in a certain state, the transition probabilities are given by

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{jk}(s,t) = \sum_{k} P_{jk}(s,t) \times q_{jk}. \tag{18}$$

With (18) and for the present context, all relevant probabilities and their corresponding solutions can be defined as in the following ODEs:

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{00}(s,t) = -\big(q_{01}(t) + q_{02}(t)\big) \times P_{00}(s,t) \tag{19}$$

$$P_{00}(s,t) = \exp\left(-\int_s^t q_{01}(u) + q_{21}(u)du\right) \tag{20}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{11}(s,t) = -q_{12}(t) \times P_{11}(s,t) \tag{21}$$

$$P_{11}(s,t) = e\exp\left(-\int_s^t q_{12}(u)du\right) \tag{22}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{12}(s,t) = q_{12}(t) \times P_{11}(s,t) \tag{23}$$

$$P_{12}(s,t) = 1 - P_{11}(s,t) \tag{24}$$

$$= 1 - e\exp\left(-\int_s^t q_{12}(u)du\right) \tag{25}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{01}(s,t) = q_{01}(t) \times P_{00}(s,t) - q_{12}(t) \times P_{01}(s,t) \tag{26}$$

$$P_{01}(s,t) = \exp\left(-\int_s^t q_{12}(u)du\right) \int_s^t q_{01}(u) \times P_{00}(s,u)du \times \exp\left(\int_s^u q_{12}(v)dv\right) du \tag{27}$$

$$= \int_s^t P_{00}(s,u) \times q_{01}(u) \times \exp\left(\int_u^t q_{12}(v)dv\right) du \tag{28}$$

$$= \int_s^t P_{00}(s,u) \times q_{01}(u) \times P_{11}(u,t)du \tag{29}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} P_{02}(s,t) = q_{02}(t) \times P_{00}(s,t) + q_{12}(t) \times P_{01}(s,t) \tag{30}$$

$$P_{02}(s,t) \;\; = \; 1 - P_{00}(s,t) - P_{01}(s,t). \tag{31}$$

The probabilities P_{00} and P_{01} are unique solutions with initial conditions $P_{00}(s,s) = 1$ and $P_{01}(s,s) = 0$. Shown by Yang [52] and Brinks et al. [9], the prevalence $p(t)$ at time $t \geq 0$, the intensity functions q_{jk} and transition probabilities $P_{jk}(s,t)$ can be related as in

$$p(t) = \frac{P_{01}(0,t)}{P_{00}(0,t) + P_{01}(0,t)}. \tag{32}$$

$p(t)$ is mathematically well-defined as it holds that $P_{00}(0,t) + P_{01}(0,t) > 0$ for all $t \geq 0$. Since $P_{01}(0,t) = 0$ it follows that $p(t) = 0$ and $P_{01}(0,t) > 0$ implies that $p(t) = \frac{1}{1+\frac{P_{00}(0,t)}{P_{01}(0,t)}}$ where $\frac{P_{00}(0,t)}{P_{01}(0,t)} \geq 0$. Thus, $p(t) \in [0,1]$ for all $t \geq 0$. Further, (32) can be rearranged to

$$1 - p(t) = \frac{P_{00}(0,t)}{P_{00}(0,t) + P_{01}(0,t)}. \tag{33}$$

The temporal change of the prevalence with initial condition $p(0) = 0$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t) = \frac{\frac{\mathrm{d}}{\mathrm{d}t}P_{01}(0,t)[P_{00}(0,t) + P_{01}(0,t)] - P_{01}(0,t)\frac{\mathrm{d}}{\mathrm{d}t}[P_{00}(0,t) + P_{01}(0,t)]}{[P_{00}(0,t) + P_{01}(0,t)]^2} \tag{34}$$

This equation can be reformulated using Equation (32) and (33) as

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t) = \frac{(1-p(t))\frac{\mathrm{d}}{\mathrm{d}t}P_{01}(0,t) - p(t)\frac{\mathrm{d}}{\mathrm{d}t}P_{00}(0,t)}{P_{00}(0,t) + P_{01}(0,t)} \tag{35}$$

$$= \frac{\frac{\mathrm{d}}{\mathrm{d}t}P_{01}(0,t) - p(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}P_{00}(0,t) + \frac{\mathrm{d}}{\mathrm{d}t}P_{01}(0,t)\right]}{P_{00}(0,t) + P_{01}(0,t)}. \tag{36}$$

Using Equation (20) and (29), we get

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t)$$
$$= \frac{(q_{01}(0,t)P_{00}(0,t) + q_{11}(0,t)P_{01}(0,t)) \; - p(t)\big[P_{00}(0,t)\{q_{00}(0,t) + q_{01}(0,t)\} + P_{01}(0,t)\{q_{10}(0,t) + q_{11}(0,t)\big]}{P_{00}(0,t) + P_{01}(0,t)} \tag{37}$$

$$= \frac{\big(1-p(t)\big)P_{00}(0,t)q_{00}(0,t) + \big(1-p(t)\big)P_{01}(0,t)q_{11}(0,t) - p(t)(P_{00}(0,t)q_{00}(0,t) + P_{01}(0,t)q_{10}(0,t))}{P_{00}(0,t) + P_{01}(0,t)} \tag{38}$$

$$= (1-p(t))^2 q_{01}(0,t) + p(t)\big(1-p(t)\big)[q_{11}(0,t) - q_{00}(0,t)] - p(t)^2 q_{10}(0,t) \tag{39}$$

which can be rearranged using the properties $q_{11}(0,t) = -q_{10}(0,t) - q_{12}(0,t)$ and $q_{00}(0,t) = -q_{01}(0,t) - q_{02}(0,t)$ to

$$= (1-p(t))^2 q_{01}(0,t) + p(t)\big(1-p(t)\big)[-q_{10}(0,t) - q_{12}(0,t) + q_{01}(0,t) + q_{02}(0,t)] - p(t)^2 q_{10}(0,t) \tag{40}$$

such that finally the temporal change in the prevalence can be defined as

$$= (1 - p(t))q_{01}(0,t) - p(t)q_{10}(0,t) - p(t)\big(1 - p(t)\big)[q_{12}(0,t) - q_{02}(0,t)]. \quad (41)$$

For the context under consideration (see Figure 5), it is known that $q_{10}(0,t) = 0$, and thus (41) can be simplified to

$$\frac{\mathrm{d}}{\mathrm{d}t}p(t) = (1 - p(t))q_{01}(0,t) - p(t)\big(1 - p(t)\big)[q_{12}(0,t) - q_{02}(0,t)]. \quad (42)$$

Overall, this shows that the temporal change of the prevalence can be found as solution of the Kolmogorov equations (20) and (29). As discussed by Brinks et al. [9], the theory of ODEs proves that there exists a unique function to equation (42) that fulfils the initial condition $p(0) = 0$.

In the epidemiological context of chronic diseases and considering the IDM (Figure 3), Equation (42) can be rewritten to a more intuitive formulation. Henceforth, the parameter t is interpreted as time and the states $S = \{1, 2, 3\}$ denote the conditions "Healthy", "Ill" and "Dead". As discussed in section 2.1.2, it is more realistic to reflect on more than one time-scale. Consequently, in a practice, all states and transitions are considered as sex-specific functions of age a and calendar time t. The time scales of age a and calendar time t will however be omitted in the following equations for the purpose of better readability. It is assumed that all individuals begin in the healthy state at birth, hence, the initial conditions $P_{00}(s,s) = 1$ and $P_{01}(s,s) = 0$ are met. Thereafter, they can either be diagnosed with a chronic disease and die at a certain point of time, or they die without the disease of interest. Since death is considered the final state, it holds that $P_{22}(s,s) = 1$. Modelling a chronic disease with no remission, the process is irreversible which implies that $P_{jk}(s,t) = 0$ for $j > k$. Referring to the notation introduced in sections 2.1.1 and 2.1.2, $H(t,a)$ represents the number of healthy people, i.e., those without a specific disease in state "Healthy", and $I(t,a)$ represents the number of ill people, i.e., those in state "Ill", who have contracted the disease of interest. The transitions between the states are given by $m_0(t,a)$ and $m_1(t,a)$, that is, the mortality of the non-diseased and diseased, respectively, and the incidence rate $IR(t,a)$. In this setting, the age- and time-specific prevalence is defined as

$$p = \frac{I}{I + H}. \quad (43)$$

Since the states and rates depend on two time scales, the ODE defined in (42) becomes a PDE to express the change in the prevalence over time:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)p = (1 - p) \times [IR - p(m_1 - m_0)]. \quad (44)$$

As discussed, for the majority of diseases, the mortality of people without a disease of interest m_0 is commonly unknown. However, it is possible to incorporate the general mortality $m(t,a)$ as defined in (10) and the MRR, defined as mortality ratio of the diseased versus non-diseased. In that case, (44) can be reformulated to

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)p = (1 - p) \times \left[IR - \frac{p \times (MRR - 1) \times m}{p \times (MRR - 1) + 1)}\right]. \quad (45)$$

Integration of the PDE shown in (45) yields the estimated age-, sex- and calendar time-specific prevalence $\hat{p}(t,a)$.

Practical application requires disease-specific and demographic input data. Precisely, the PDE presented in this section relies on sex-, age- and calendar time-specific information on mortality,

prevalence, and incidence rates as well as input information on the population structure, i.e., sex-, age- and calendar time-specific population numbers. The PDE reflects on the complex interplay of the incidence and mortality rates, and thus, in a projection allows for incorporating temporal trends of these rates. Consequently, if the latter is of interest, it is necessary to provide information on the temporal development of the incidence and/or mortality.

The PDE is flexible enough to allow for possible extensions, other areas of application and reformulations that may be valuable for epidemiological research. With regards to potential extensions of the PDE, it is possible, for example, to include the duration of the disease [12], to transfer the PDE to a context of reversible diseases and model remission from a disease [9], or including a phase of undetected disease preceding the diagnosed stage [53]. Alternatively to estimating the prevalence, it may be of interest to estimate and/or project case numbers of a disease as discussed in Section 2.1.2 and shown in contribution 1. This can be achieved by firstly estimating the prevalence using the PDE. In a second step, the estimated prevalence is multiplied with the population counts which yields the number of people with a disease (e.g., diabetes or dementia) in the respective population [13,34,54]. Another area of application might be to reformulate the PDE and express it in terms of the incidence (see contribution 1 and 2). In epidemiology, it is typically more complicated to compute incidence of a disease compared to calculating the prevalence. The PDE offers a valuable and efficient approach for that purpose. Using information on the age-specific prevalence for two points in time, the general mortality of the population, as well as information on the MRR of the populations with and without the disease, it is possible to estimate the age- and sex-specific incidence of a chronic disease using the following equation:

$$IR(a,t) = \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p(a,t) + MRR(a,t) \times \frac{p(a,t) \times (m(a,t) - 1)}{p(a,t) \times m(a,t) + (1 - p(a,t))} \tag{46}$$

For practical application see e.g. contribution 2 or Hoyer et al. [55], who estimated the age-, sex-, and race/ethnicity-specific incidence of diabetes. Another application domain would be to estimate the relative mortality of a chronic disease. As discussed by Brinks et al. [9] and Egeberg et al. [56], this may be valuable in cases where information on the prevalence and incidence of the disease are registered but not followed-up for mortality. Furthermore, in view of the considerable increase in the number of people with chronic conditions and its associated costs, another context of application would be to use the PDE for estimating and projecting cost of illness information (see contribution 3). Cost of illness estimates are relevant from economic, medical and political point of view, and are necessary for effective healthcare management, resource planning, meeting future medical needs and evaluating measures for prevention and intervention [11,26]. In combination with healthcare cost data, the PDE allows to project age- and sex-specific healthcare expenses of chronic diseases considering future demographic, disease-specific and cost trends. To achieve this, per capita healthcare cost data are combined together with the demographic structure of the population under consideration, the disease prevalence, incidence and mortality. As a result, direct per capita costs, total annual costs, cost ratios for diseased versus non-diseased and attributable costs can be estimated and projected to the future. The proposed method is able to reflect on temporal trends in epidemiological, demographic and cost dynamics simultaneously. At the same time, the approach remains transparent, clear and understandable in its application. Further, the statistical method can be easily applied to other countries and chronic diseases and are flexible enough to anticipate impacts of alternative policy scenarios.

2.1.4 Future Epidemiological and Economic Burden of Diabetes Mellitus in Germany

For illustration purposes, the following section shows the practical application of the approaches presented in Section 2.1.2 and 2.1.3. First, this chapter provides background information on the global

diabetes epidemic, in order to illustrate the public health relevance of the disease and the resulting need for diabetes surveillance. Following, Section 2.1.4 briefly describes the employment of the statistical methods to quantify current and future epidemiological and economic measures in the context of the chronic disease diabetes in Germany.

Worldwide, diabetes mellitus is one of the most common chronic diseases, and hence, is a disease with high public health relevance [57-59]. In 2021, the global diabetes prevalence was estimated 10.5% among adults (20-79 years), i.e. 537 million people, and caused healthcare costs of at least $966 billion [60]. Most generally, diabetes refers to a disorder of carbohydrate metabolism [58]. As such, it comprises a group of diseases that are characterised by impaired ability of the body to produce or respond to insulin which in turn manifests in an incapability of the body to maintain proper levels of sugar, i.e., glucose, in the blood [58]. Therefore, the disease requires a structured self-management plan, monitoring blood glucose level, physical activity, diet and daily insulin treatment [57,58]. The increasing proportion of people suffering from diabetes is peculiarly worrying as the condition constrains activities of daily living, necessitates ongoing medical attention and hence, causes remarkably higher costs and healthcare expenses. Besides severe late complications, the disease is associated with considerably higher mortality and is estimated to lead to 1.5 to 4.4 times higher health-care costs compared to people without diabetes [26]. Diabetes has many subclassifications, including for instance T1D, T2D, maturity-onset diabetes of the young (MODY), gestational diabetes, neonatal diabetes or late onset (or latent) autoimmune diabetes in adults (LADA). Overall, diabetes is considered a chronic condition, with the exception of prediabetes and gestational diabetes which are potentially reversible diabetes conditions [58].

T1D and T2D are the major subtypes of the disease, however, the onset, pathogenesis and symptoms of T1D and T2D are considerably different [58,61]. Therefore, each type has its own risk factors, pathophysiology, presentations, aetiologies, and treatments. T1D usually arises in children and young adolescents, thus sometimes referred to as juvenile-onset diabetes [62]. Vice versa, T2D usually occurs after age 40 and becomes more common with increasing age and most often, is a result of poor lifestyle and dietary choices. T2D is far more prevalent than T1D, accounting for about 90% of all diabetes cases. Current T2D prevalence is estimated with 7.4% among men and 7.0% among women in Germany aged 40 years or older [34,63]. In 2009 in Germany, estimates of the prevalence of T1D was 0.17% for girls and 0.19% for boys younger than 19 years [64]. The prevalence among women and men (aged between 20 to 79 years) was estimated between 0.28% to 0.39% and among the elderly population (80+ years) between 0.47% and 0.50% [63,64]. The total number of people with T1D in Germany ranges from 256,000 to 373,000 in 2009 and 2016, while for T2D, more than 9 million people of the German population are estimated to be positively diagnosed [63,64]. As observed over the past decades, the incidence and prevalence of T1D and T2D are expected to rise further [34,58].

However, little is known about diabetes-related economic consequences and future healthcare expenses in Germany [57,64]. Further, the few studies investigating the current and future population with T1D or T2D in Germany as well as associated healthcare expenses are limited to certain ages or demographic cohorts, they ignore temporal trends in disease-related rates and/ or do not distinguish the diabetes types [11,34,65]. As a remedy, this thesis shows that using the statistical approaches presented in Section 2.1.2 and 2.1.3, it is possible to compute nationally representative estimates of the age- and sex-specific incidence, prevalence and the number of people diagnosed with T1D and T2D in Germany in 2010 until 2040. Further, this work shows that there exists a projection method that is capable of reflecting the complex interplay of future demographic, disease-specific and cost trends simultaneously to project diabetes-related healthcare expenses. Specifically, the approach enables to project sex-, age-, year- and diabetes-type specific per capita costs, total excess costs and cost ratios

of people with and without T1D and T2D, as well as attributable costs in Germany from 2010 until 2040.

The projection of the diabetes burden and its related healthcare costs in Germany requires several demographic, epidemiological and economic input data. All applications rely on anonymised, aggregated data, i.e., no individual data was required and data on population level is sufficient. In the best case, input information was obtained in differentiated manner with regards to age, sex and the specific disease, the data should comprise the time period considered in the estimation (or should at least be relatively recent or close to the time period of interest) and should be as reliable and representative as possible for the target population. Necessary input information may be obtained from secondary data which is commonly more resource saving compared to collecting primary data (e.g., with regards to time and costs). Further, the proposed methodologies allow for estimating and projecting the disease burden based on data derived from cross-sectional studies, i.e., collecting data from lengthy and costly longitudinal studies may be dispensable. Depending on the field of application and the concrete context, the required input data may be provided by governments, insurances or disease-specific registries, sometimes even publicly.

First, demographic input data in form of the observed and expected age- and sex-specific population distribution and the mortality of the general population in Germany for each year from 2010 to 2040 and all ages from 0 to 100 years are provided in the official population projection of the German Federal Statistical Office (FSO) [66]. The FSO focuses on demographic change in Germany and releases several variants wherein they assume different future birth rates, life expectancies at birth, and migration. For projection purposes, it may be recommendable to consider several variants to account for uncertainty in future demographic developments.

Second, epidemiological input information, i.e., starting values for the prevalence and incidence, were derived from claims data that have been used in previous projections of diabetes prevalence [63,67]. The data is taken from 65 million insures from all German statutory health insurances (SHI). Covering about 80% of the total population, the data set is considered representative for Germany [26,34,63]. As relevant information, the data set provides information on age, sex and diabetes diagnoses in 2010. To define T1D and T2D, the International Classification of Diseases (ICD) coding is used which is provided in the underlying claims data [26,34,58,61]. The ICD-codes E10.- to E14.- specify the different types of diabetes, with E10.- comprising all possible T1D diagnoses and E11.- denoting T2D. Information on the mortality of people with versus without T1D or T2D is lacking in Germany. As workaround, it is possible to use data on the age- and sex-specific MRR of people with versus without T1D or T2D. For T2D, Schmidt et al. [27] report nationally representative estimates of the MRR in Germany in 2014. Due to unavailability of the T1D-related MRR in Germany, the age- and sex-specific MRR from Denmark estimated by Carstensen et al. [32,33] were used to approximate the MRR of T1D in Germany. This is suggested by previous studies since the two countries' MRRs are claimed highly comparable [33,34,63,68].

Lastly, economic input values in terms of average per capita healthcare costs for people with and without T1D or T2D were obtained from aggregated claims data [63,67]. The claims data consists of a 6.8% random sample of all German people with statutory health insurance (SHI) (which covers almost 90% of the German population). Due to regulations on data protection, routine SHI data were provided in an anonymous and aggregated form. The cost data included direct per capita costs for physicians, dentists, pharmacies, hospitals, sick benefits and others in 2010 in Germany from payer perspective. Further, it contained ICD-10 codes which allowed to differentiate between costs associated with the respective type of diabetes. These data were used as starting values of per capita costs, cost ratios for people with versus without T1D and T2D as well as attributable costs in 2010.

Contribution 1 presents a detailed comparison of three different projection methods and their practical employment to project sex- and age-specific case numbers of people with diagnosed T2D for Germany between 2010 and 2040. The second contributing article first estimates the age- and sex-specific incidence and prevalence of T1D in Germany in 2010 using data from 65 million insurees of the German SHI and then projects the prevalence of type 1 diabetes until 2040 assuming several scenarios of the incidence and mortality. For the precise estimation methodology and projection results of future age- and sex-specific direct medical costs related to diagnosed T1D and T2D in Germany between 2010 and 2040, see contributing article 3. The freely available software R (R Core Team, 2021 [69]) was used for implementation of the analyses related to contribution 1, 2 and 3. The complete R-codes and relevant data sets for reproducing the analyses are publicly available.

Part II

2.2 Time-to-Event Analysis

"Don't know, really. In the light of some of the further results one knows since, I think I would normally want to tackle problems parametrically, so I would take the underlying hazard to be a Weibull or something. I'm not keen on nonparametric formulations usually." [20]
(David Cox)

Time-to-event analysis is a branch of statistics that allows appropriate handling of data where the variable under investigation involves times to some event of interest. It is commonly used to model the expected duration of time until an event occurs, to approximate transition rates or intensities from one state to another in a stochastic process with countably many states, or to estimate the effect of prognostic factors. The area of application for time-to-event models is diverse [70], ranging from e.g. finance (estimating companies survival in financial crisis or time to default in credit scoring), marketing (the probability of a customer leaving the company in the next months or the time until a customer makes a repeat purchase), engineering (component failures in machines or the reliability of a gas turbine), psychology (predictors of and time to potential criminal recidivism), logistics (lead times for metallic components in the aerospace industry) and, most commonly, healthcare. In medical research, the particular endpoint is commonly the time until occurrence of a particular disease, relapse from remission, recovery or the time until death. For this reason, the term "survival analysis" is often used synonymously for time-to-event analysis (or "failure time analysis" and "reliability analysis" in engineering).

When modelling time-to-event data, the response variable of interest usually measures the time until a specific endpoint. In its simplest form, time-to-event data consists of two tightly linked components, that is, an event indicator and a time variable that indicates the relevant points in time. Time-to-event data, i.e., data on the time between a starting point and an endpoint that refer to certain kind of events, are often affected by a peculiar kind of "partial missingness" [14,71]. This is due to truncation and censoring, which are two distinct phenomena that cause incompleteness in data samples. In the case of right-censoring, subjects do not experience the event of interest during the follow-up time. For example, if death is of interest, patients may not die during the overall study period. Vice versa, in left-censoring, the event has occurred before the data is collected or study has started, i.e., only the upper bound of time is known [21]. Truncation may arise due to a systematic selection process inherent to the study design. It occurs if a value below (left truncation) or above (right truncation) a certain truncation point is not recorded at all, one speaks of truncation. For instance, this may happen in cohort studies, in which study participants must not have the disease of interest at the start of the study. Knowing that traditional regression methods are not equipped to handle such data and that censoring and truncation mechanisms may affect the response, survival analysis requires special techniques to analyse the data properly and avoid biased estimates as well as incorrect conclusions.

The remaining part on modelling time-to-event data is structured as follows: Section 2.2.1 introduces a clinical data set from a real-world study which serves as practical example for the remaining chapter. Section 2.2.2 introduces basic mathematical terminology and notation used in time-to-event analysis. Section 2.2.3 is devoted to Cox's Proportional Hazards (PH) model and its generic effect estimate the HR. Section 2.2.4 describes alternative semi-parametric additive hazard models. The focus of Section 2.2.5 is on parametric additive modelling of time-to-event data and briefly derives the new approach. Building on Section 2.2, the fourth contributing article makes use of the relations and the background introduced and proposes a new parametric additive hazard model with time-independent covariates.

2.2.1 Survival of Non-Small-Cell Lung Cancer Patients

For illustration purposes, the subsequent sections consider an example based on a data set from the Halle Lung Cancer (HALLUCA) study [72]. The population-based multi-centre study investigated treatment options for non-small cell lung cancer patients in the region of Halle (Saale) in the eastern part of Germany between 1996 and 1999. In the HALLUCA study, the authors investigated the survival of 1,696 lung cancer patients depending on several prognostic factors. Thereof, 1,183 patients were diagnosed with non-small-cell lung cancer. 188 were in clinical stages I–IIIb (15.9%) and were treated with radiation therapy alone. 1,349 patients of the sample (79.5%) died until the end of follow-up. Minimal follow-up was 12 months, the median follow-up time 33 months. Survival time was defined as starting from the day of diagnosis of lung cancer. The end of survival time was either death or the end of the observation period planned for data collection in the HALLUCA study. The observed median survival time of all patients was 10.2 months with a 2-year overall survival of 15.8%. Besides tumour stage, radiation dose was found the most important prognostic factor for survival of patients after primary radiation therapy.

Generally, lung cancer belongs to the leading causes of cancer mortality, with about 2.1 million new cases and 1.8 million deaths in 2018 worldwide [73]. Non-small-cell carcinoma or non-small-cell lung cancer (NSCLC) comprises a heterogeneous class of tumours that represents approximately 80 - 85% of all lung cancers [74]. The disease is associated with a notably high proliferation, strong predilection for early metastasis and poor prognosis particularly in later stages [73]. It is confirmed that NSCLC has a strong epidemiological link to tobacco, as its prevalence seems to reflect the smoking prevalence with a lag time of about 30 years [75]. Internationally, NSCLC is grouped into five stages (stage 0 to IV) using the tumour-node-metastasis (TNM) classification [72]. The lower the staging, the less the cancer has spread with advanced disease represented by stage IV. Treatment and prognosis differ by stage. In the past decade, treatment options have improved markedly with wider lung cancer screening, improved radiation techniques, and treatment advances [73,74].

For the latter analysis, a dichotomized version of the TNM-scale was adapted for the classification of malignant tumours as a predictor of survival. 739 patients had a TNM-scale of III or smaller (TNM < IV), and 621 patients had a TNM-scale of IV (TNM IV). For the 336 remaining patients no TNM-scale was reported, and they were therefore excluded from the analysis. We are thus considering a problem that involves a binary covariate x (i.e., TNM IV versus TNM < IV, with TNM < IV as the reference) for each study participant i as predictor of observed survival time t.

2.2.2 Mathematical Relations in Survival Analysis

Let $T \geq 0$ be a random variable representing the time of the event under consideration, i.e., death in the HALLUCA study. Further referring to the data example of the HALLUCA study, the TNM classification of each respective patient (i.e., TNM < IV versus TNM IV) acts as predictor variable of a patient's survival time included in a covariate vector x. The probability of survival of a single individual with covariate vector x beyond time t can be modelled via the survival function $S_x(t)$ by

$$S_x(t) = \mathbb{P}(T > t), \qquad 0 < t < \infty. \tag{47}$$

This can also be expressed as

$$S_x(t) = 1 - F_x(t), \tag{48}$$

with $F_x(t) = \mathbb{P}(T \leq t)$ denoting the cumulative distribution function (CDF) which describes the probability of having experienced the outcome of interest before or exactly at time t. The probability density function (PDF) $f_x(t)$ describes the probability of the event occurring at exactly time t, given by

$$f_x(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t)}{\Delta t} \tag{49}$$

The PDF can be ascertained from the CDF [71,76] by

$$f_x(t) = \frac{\partial F(t_x)}{\partial t} \tag{50}$$

or vice versa it holds that

$$F(t_x) = \int_0^t f(u)du. \tag{51}$$

Integrating the PDF $f_x(t)$ from time t to $+\infty$ yields the survival function [76], i.e., $S_x(t)$ can be related to $f_x(t)$ as in

$$S(t_x) = \int_t^{+\infty} f(u)du \tag{52}$$

with

$$f(t_x) = -\frac{\partial S(t_x)}{\partial t} \tag{53}$$

as equivalent expression. Once the survival function and the PDF are specified, the corresponding hazard function $h_x(t)$ can be determined.

Most often in time-to-event analysis, it is the aim of modelling the instantaneous potential (or risk, with regards to death being the event) of having an event at a time t, given survival up to that time [71]. The hazard function $h_x(t)$ describes this instantaneous rate [76] and is defined as

$$h_x(t) = \lim_{\Delta t \to 0^+} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t\ )}{\Delta t} \tag{54}$$

or, expressed in terms of the survival function $S(t)$ and PDF $f_x(t)$, as

$$h_x(t) = \frac{f(t)}{S(t)} = \frac{-\partial S(t_x)/\partial t}{S(t)}. \tag{55}$$

Building on the example of the HALLUCA study, the relations between hazard, density, CDF and survival can be depicted as shown in Figure 6.

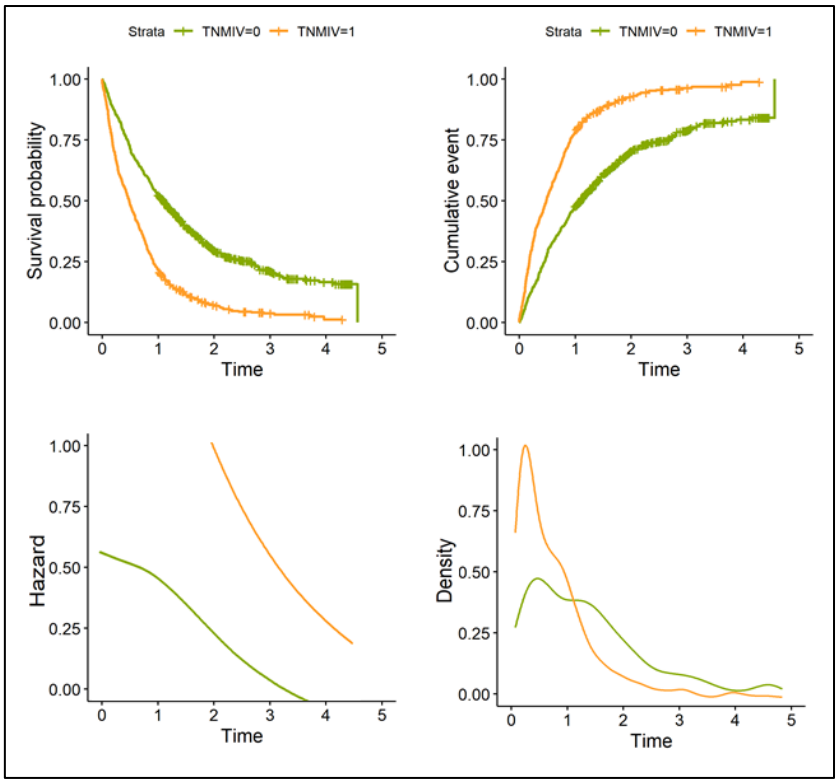

Figure 6: Relations between hazard, density, CDF and survival function shown for the example of the HALLUCA study. Using the Kaplan-Meier method, the plot top left depicts an estimate of the survival curve for the HALLUCA data, while the plot top right plots the cumulative event probability. The hazard function (bottom left) is modelled via the function bszahard() in R which estimates the hazard function non-parametrically from a survival object. The depicted smoothed estimate is based on B-splines. The plot bottom right visualises the distribution of the survival variable in the HALLUCA data, using geom_density() (a function from the ggplot2 package) that creates smooth density estimate plots.

2.2.3 Cox Proportional Hazards Model and the Hazard Ratio

Often, one of the objectives of time-to-event analysis is to specify the potential effect of covariates (such as sex, age, treatment, diagnosis) on event times [76]. This can be solved using special regression techniques such as the popular Cox PH regression model. In order to assess the effect of multiple covariates on failure time of a system, Cox introduced his PH model in 1972 [16]. Precisely, Cox proposed that the hazard takes the form of

$$h(t_x) = h_0(t)\exp(\beta' x) \tag{56}$$

where $h_0(t)$ is the baseline hazard, x is a set of observed covariates and β is a vector of regression coefficients to be estimated, measuring the influence of each respective covariate x on the outcome [16]. In other words, the model decomposes the hazard into two distinct parts that act multiplicatively: First, into an arbitrary and unspecified, non-parametric baseline hazard that is shared across all individuals and dependent on time only and second, a linear predictor that consist of a functional term that is basically independent of time and that describes the effect that covariates have on the hazard. By that, Cox regression relies on the background assumptions of linearity and additivity of predictor variables. The model is essentially non-parametric as the functional form of the baseline hazard $h_0(t)$ remains unspecified [77,78]. However, the implicit PH assumption is an important feature of the model and implies that the HR, which measures the effect of the predictor, is constant over time for any two individuals or strata compared. In order to be correctly specified, it must satisfy that the baseline hazard is a function of time but does not involve the covariates [70,71,79,80].

The model's generic effect estimate, the HR, is defined as ratio of two distinct hazard rates and is used extensively as conventional effect measure in time-to-event analyses [79]. In medical research for instance, the HR is typically reported to evaluate treatment effects or other safety and efficacy outcomes of clinical trials comparing a treatment group and a control group [76]. Assuming PH, the hazard ratio for given values of a single binary covariate x_{TNM} (with referral to the HALLUCA study this may be the TNM classification) is given by [80]

$$\frac{h(t|x_{TNMIV=1})}{h(t|x_{TNMIV=0})} = \exp(\beta). \tag{57}$$

In the context of time-to-event data, the Cox PH regression model has become one of the most popular and widely-used regression techniques [77,81]. From its introduction, it has generated a great amount of interest and has become one of the most used and cited statistical models in applied research (the original article ranks among the top 100 papers in terms of citation frequency). Its predominance and popularity stem from several decades of application, together with the facts that (i) it allows for censored observations, (ii) it is possible to estimate survival curves, (iii) it guarantees interpretable hazard ratios as effect measures, (iv) it is implemented in standard statistical software, (v) it is a robust model and (vi) it is essentially distribution-free as it does not require the investigator to assume any underlying distribution of the baseline hazard function [76,77]. Despite these key strengths of the model, it does introduce some limitations. One inherent and overly strict assumption imposed by the Cox model, the PH assumption, is that the hazard functions of all individuals are strictly proportional and that the HR is thus constant over time [70]. In practice however, this very restrictive condition inhibits the model's usefulness [77]. In case of violation or misspecification, the Cox model may result in biased, potentially misleading estimates and false inference [80]. Considering the complexity of medical or biological relations and associations, this assumption can rarely be justified. Instead, hazards may vary because the susceptibility of a disease differs between patients or because a new treatment might change the pattern of mortality over time, rather than the overall mortality rate. Another example would be the decision between surgery and radiation treatments: the surgery might have a higher initial risk, but a better long-term prognosis while this is inversed for radiation therapy.

In addition, the frequent use of the Cox model is surprising because the HR, its generic effect estimate, has been repeatedly criticized in recent years [79]. Although the HR, may seem convenient, e.g., it is easily estimated and summarizes the treatment effect into a single number [79], it does continuously fall in disgrace in terms of interpretation, technical implementation, and flexibility. For instance, the HR may be hard to interpret (causally) and is commonly mistaken as relative risk although essentially, the HR is a ratio of rates, and not one of risks [17,79,80]. Another common point of critique is that even in randomized trials, the HR has a built-in selection bias because it is estimated conditionally on the survival of the set of observations which is still under risk [18,79,80]. This bias is even more severe in case of misspecification, e.g. in terms of an omitted variable that affects the outcome of interest. Recalling that the HR for a given value of a single binary covariate is given by (57) and now assuming an omitted covariate x_{OV}, the true conditional model would be correctly specified by

$$h(t|x) = h_0(t)\exp(\beta_{TNMIV}x_{TNMIV,i} + \beta_{OV}x_{OV}). \tag{58}$$

As the covariate x_{OV} represents an omitted prognostic input factor, only the marginal population-averaged HR $h\left(t|x_{TNMIV,i}\right)$ would be estimated, which is unequal to the true conditional and subject-specific treatment effect [80]. Though, it must be noted here that the presence of bias would only occur in the presence of large treatment effects together with an omitted covariate that largely impacts the outcome [80]. In case of a weak effect of treatment on survival and/ or weak impact of the omitted

covariate, the imbalance in treatment estimates is negligible. More virulent is that the HR suffers from non-collapsibility, i.e., adjusting for a covariate that is associated with the event will generally change the HR, even if the covariate is not related to the exposure [19,82]. Supposing that t denotes the survival time of interest and assuming a binary exposure variable x and third variable z, the hazard function is given by

$$h(t) = (t|X = x, Z = z) = h_0(t)\exp(\beta x + \beta z). \quad (59)$$

It follows that the marginal hazard function is equal to

$$h(t|x) = h_0(t)e^{\beta x} j\{x, \theta(t)\} \text{ with } j\{x, \theta(t)\} = \frac{E[S_1\{x,\theta(t);z\}|X=x]}{E[S_0\{x,\theta(t);z\}|X=x]} \quad (60)$$

where $S_0\{x, \theta(t); z\} = \exp(-\Lambda_0(t)\exp(\beta x + \beta z))$, $S_1\{x, \theta(t); z\} = S_0\{x, \theta(t); z\}\exp(\beta z)$ and $\theta(t) = \{\beta_x, \beta_z, \Lambda_0(t)\}'$ with Λ as cumulative hazard function. That being the case, the causal hazard function is then denoted by

$$h(t|\hat{x} = x) = h_0(t)\exp(\beta z)h\{x, \theta(t)\} \text{ where } h\{x, \theta(t)\} = \frac{E[S_1\{x,\theta(t);z]}{E[S_0\{x,\theta(t);z]}. \quad (61)$$

From there, the marginal exposure effect $\tau(t)$ becomes

$$\tau(t) = \beta_x + log\left[\frac{h\{1,\theta(t)\}}{h\{0,\theta(t)\}}\right]. \quad (62)$$

Generally, if the marginal exposure effect differs from the marginal association, i.e., $\tau(t) \neq \beta_x$, there is confounding. Hence, in the given case, and even when z is independent, non-linearity (or as commonly referred to, non-collapsibility) is present [82].

2.2.4 Semi-Parametric Additive Hazard Models

As the PH assumption may not always be true (for example, the effect of a treatment may change over time) and in response to the disadvantages of the HR, alternative additive frameworks which allow non-proportional hazards (i.e., time-varying covariate effects) have been proposed [83]. As indicated by the name, in additive hazards regression models, covariates act additively on the baseline hazard. Unlike the PH model which estimates HRs, additive models estimate the difference in hazards [84]. Therefore, additive approaches are noteworthy because they do not suffer from the beforementioned issues and instead offer additional pleasurable properties [83]. First, when defined in continuous settings, results do not suffer from non-collapsibility. Second, compared to HRs, hazard differences are more comprehensible and can be more easily interpreted [79]. For instance, additive measures provide information about a population's base rates of the outcome of interest which in turn, can inform clinical and policy decision-making [85]. Thus, hazard differences are useful in assessing risks without requiring additional data to estimate the baseline rates [86]. Third, results from additive hazard models can be translated to a relative survival scale. Relative survival quantifies the cumulative effect of covariates on the relative survival probability and thus incorporates information about the strength of the investigated association [85]. This can be nicely interpreted as the observed survival probability of the population studied, divided by the expected survival probability if the population was free of the disease of interest. Knowing that absolute and relative results may generate diverging evidence, proper interpretation of the effect of an exposure or treatment on survival times may require reporting both effect measures. Supported by the reporting guidelines from the Consolidated Standards of Reporting Trials (CONSORT) and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), it is advocated to report both, absolute and relative measures, because the complementary findings provide a more comprehensive risk picture [87,88]. Though it must be noted that the choice of the measure and whether to provide a combination of measures is still an ongoing debate in literature. An overview of potentially useful measures is provided in Table 1. In view of that, additive hazard

models may appear as preferable alternative to multiplicative models which only capture relative hazards [85].

Table 1: Overview of relative and absolute effect measures, together with their definition, interpretation, disadvantages and advantages.

Effect measure	Definition	Interpretation	Disadvantages	Advantages
Relative Measures				
Hazard ratio (HR) [89]	$\frac{h(t\|X_1)}{h(t\|X_0)}$	Hazard that the event time of a treated group exceeds the event time of a control group conditional on the other covariates	- Potentially misleading - May exaggerate effects	- Directly estimable from Cox regression model - Easy to compute
Odds ratio (OR) [89,90]	$\frac{\frac{F_1(t)}{1-F_1(t)}}{\frac{F_0(t)}{1-F_0(t)}}$	Ratio between the odds of the treated vs. the odds of the control group	- Frequently mislabelled or misused as RR - Hard to comprehend - May exaggerate effects	- Directly estimable from a logistic model and regardless of the study design (follow-up, case-control, cross sectional)
Relative risk/ Risk ratio (RR) [89-91]	$\frac{F_0(t)}{F_1(t)}$	Ratio of risks of the treated group vs. a control group	- The RR can be the same for very different clinical situations - Potentially misleading - May exaggerate effects	- Easy to compute - Easy to interpret - Included in standard statistical software
Relative risk reduction (RRR) [91]	$1 - RR$	Relative decrease in risk of an adverse event in a treated vs. a control group	- Potentially misleading	- Easy to compute
Absolute Measures				
Risk difference (RD)/ Absolute risk reduction (ARR)/ Absolute risk increase (ARI) [90,91]	$F_0(t) - F_1(t)$	Difference between the risk of an event in a control group and the risk of an event in a treated group	- Difference in risk may have greater importance when risks are close to 0 or 1 than when near the middle of the range	- Easy to compute - Easy to interpret - Clear meaning - Reflects both, the underlying risk without and with treatment
Number needed to treat (NNT) [89-91]	$\frac{1}{RD} = \frac{1}{ARR}$	Number of patients to be treated to achieve the desired outcome in one patient who would not have benefited otherwise	- NNT should be positive, only defined if RD > 0 - Its confidence interval does not include the point estimator	- Easily understood - Addresses statistical and clinical significance in a way that is easily interpreted
Number needed to harm (NNH) [89]	$\frac{1}{ARI}$	Number of patients needed to treat to experience a particular adverse outcome (e.g. a significant side effect)		

The first introduced and most known additive hazards model was proposed by Aalen in 1980 [92]. Thereafter, the approach of modelling time-to-event outcomes in an additive manner has been considered by numerous other authors (e.g., Aalen (1980, 1989, 1993), Breslow and Day (1980, 1987); Buckley (1984); Cox and Oakes (1984); Thomas (1986); Huffer and McKeague (1991); Andersen et al. (1993)) [93]. In his model from 1989, Aalen assumed that covariates act in an additive manner on an unknown baseline hazard rate [94]. Precisely, he proposed that the hazard function associated with a set of time-independent covariates is the sum, rather than the product, of the baseline hazard function and the regression function of covariates as in

$$h(t; X_{ij}) = h_0(t) + \beta_1(t)x_{j1}(t) + \ldots + \beta_p(t)x_{jp}(t) \text{ [94]}. \quad (63)$$

Therein, $h_0(t)$ represents the baseline hazard, X_i denotes the set of covariates and $\beta_i(t)$ the regression coefficients which are dependent on time t [83]. The estimated hazard differences can be

interpreted by plotting cumulative hazards over time. The unknown risk coefficients are allowed to be functions of time, so that the effect of a covariate can easily be detected at each distinct survival time. While this is a main advantage of the model, it also introduces its major drawback in terms of application and interpretation. With regards to application, the Aalen approach gives large flexibility in modelling, but this complicates parameter estimation considerably. Concerning interpretation, the time-dependence of covariate effects renders their description more complex [95]. As the model coefficients may change repeatedly over time (see Figure 7), they may be more complicated to understand and no single quantifiable effect size can be offered [96].

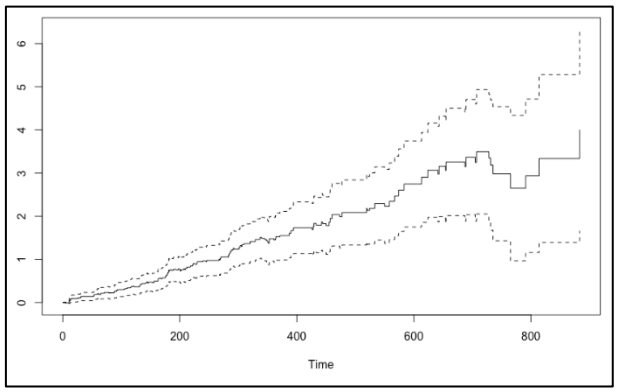

Figure 7: Estimated cumulative regression function based on the dataset "lung" contained in the R library "survival" and plotted using the function plot.aareg().

As a partial remedy, Lin and Ying [97] proposed an additive hazard model for time-independent covariates as a special case of Aalen's model [93]. Their semi-parametric model allows to reduce effects that are time-invariant to a parametric form which results in a simpler description to those effects [95]. Precisely, in addition to assuming the linearity of continuous covariates, Lin and Ying suggested that all regression functions, i.e., all covariate effects, except the baseline hazard are constant over time [97]. Thus, for a given subject *i*, the hazard is written as

$$h(t; X_{ij}) = h_0(t) + \beta_1 x_{j1}(t) + \ldots + \beta_p x_{jp}(t)^{98}. \quad (64)$$

In view of the discussed properties, additive hazards regression models could take priority over the popular Cox PH model if the absolute difference in hazard is of primary interest instead of modelling the HR, or if the proportionality assumption cannot be justified [84,93]. Additive approaches are said to provide a better fit for survival data and to be more useful to assess underlying biological interactions [86,99]. Though, it is important to recognize that these approaches are not new and despite their coverage in the literature, they are rarely applied in the routine analysis and reporting of medical data [84,100]. One reason for this is that using a semi-parametric hazard may require long computation times in certain contexts. Thus, they may be complicated in terms of implementation and interpretation. Another drawback is that some additive methods are not yet widely implemented and available in commonly used software [93].

2.2.5 A new Parametric Additive Hazard Model

An alternative branch of time-to-event analysis that bypasses many of the discussed shortcomings of non- and semi-parametric approaches, is fully parametric modelling. With parametric models, the outcome of interest (e.g. survival time) is postulated to follow a particular distributional form. Thus, the PDF $f_{x_i}(t_i)$ can be expressed in terms of unknown parameters. Once the PDF is specified, the corresponding survival and hazard functions can be determined. Many parametric models are accelerated failure time (AFT) models in which survival time is modelled as a function of predictor

variables. Whereas a PH model assumes a multiplicative effect of covariates on the hazard, an AFT model suggests that the effect of covariate is multiplicative, i.e., proportional, with respect to survival time.

Given that the parametric form is correctly specified, parametric models for time-to-event modelling were reported to be simpler, more informative, more robust, to have more validity and higher accuracy in parameter estimates [101]. Further, they allow to extrapolate beyond the available range of the data and estimated survival curves are smoother. Another strength of parametric models is that estimation is simplified using the maximum likelihood principle to estimate the parameters [102]. Further appeal of parametric approaches lies in the ease of interpreting the results: they have absolute effect measures and hazard function readily available, and fitted values from the model provide estimates of survival time, as well as the estimated coefficients or suitable transformations thereof function as clinically meaningful estimates of the investigated effects [103]. Albeit these numerous favourable properties, it remains a phenomenon that parametric survival models are rather underutilised in medical applications while semi- or non-parametric modelling persist being the more popular choice [104].

One aim of this dissertation is to show that it is possible to exploit both, the advantages of additive approaches as well as the advantages of parametric survival analysis. Evidently, the model assumes that the hazard $h_x(t)$ at time t is modelled additively as in

$$h_x(t) = h_{0,\theta}(t) + x\beta \, . \tag{65}$$

The parameters β are the regression coefficients that measure the impact of the covariates x and $h_{0,\theta}(t)$ denotes the parametric baseline hazard function which is independent of the covariates. The parameter θ denotes the distribution parameters. Depending on the choice of the baseline distribution, θ may differ in terms of number and commonly used notation. Using the relations between PDF, hazard and survival function (see Section 2.2.2) and reformulating (55) to $f_x(t) = h_x(x)S(t)$, the PDF associated to the new model can be expressed as

$$f_x(t) = \left(\frac{f_0(t)}{S_{0,\theta}(t)} + x\beta \right) S_x(t). \tag{66}$$

Substituting $S_x(t)$, this becomes

$$f_x(t) = \left(\frac{f_0(t)}{S_{0,\theta}(t)} + x\beta \right) \exp(-H_x(t)) \tag{67}$$

And thus

$$f_x(t) = \left(\frac{f_0(t)}{S_{0,\theta}(t)} + x\beta \right) \exp\left(-\int_0^t h_x(u)du\right) \tag{68}$$

Using Equation (65), the last part can be replaced by

$$f_t(t) = \left(\frac{f_0(t)}{S_{0,\theta}(t)} + x\beta \right) \exp\left(-\int_0^t h_{0,\theta}(u)du - \int_0^t x\beta du\right) \tag{69}$$

which is equal to

$$f_x(t) = \left(\frac{f_0(t)}{S_{0,\theta}(t)} + x\beta \right) \frac{\exp(-\int_0^t h_{0,\theta}(u)du)}{\exp(\int_0^t x\beta du)}. \tag{70}$$

Replacing the numerator by $S_{0,\theta}(t)$ and further solving the equation, (70) can be rewritten to

$$f_x(t) = \frac{f_0(t)}{\exp(tx\beta)} + \frac{S_{0,\theta}(t)}{\exp(tx\beta)} \tag{71}$$

Finally yielding

$$f_x(t) = \frac{f_0(t) + x\beta\, S_{0,\theta}(t_i)}{\exp(tx\beta)} \tag{72}$$

with the corresponding survival function $S_x(t)$ given by

$$S_x(t) = \frac{S_{0,\theta}(t)}{\exp(tx\beta)} = 1 - F_x(t) \tag{73}$$

where $F_x(t)$ depicts the model's CDF. Estimation is straightforward via maximising Equation (74) with respect to the unknown regression coefficients β and the parameters of the assumed baseline distribution using any software that allows maximising a hand-coded log-likelihood function. Precisely, the log-likelihood function of the model for a single observation i at time t_i and covariate vector x_i is given by

$$\ell_i = (1 - \delta_i) \log\left(f_{x_i}(t_i)\right) + \delta_i \log\left(S_{x_i}(t_i)\right). \tag{74}$$

Plugging in the derived expressions for $f_{x_i}(t_i)$ and $S_{x_i}(t_i)$ in Equation (66) and (73), respectively, finally this becomes

$$\ell_i = (1 - \delta_i)(\log\left(f_0(t_i) + x_i\beta S_{0,\theta}(t_i)\right) - t_i x_i \beta) + \delta_i\, (\log\left(S_{0,\theta}(t_i)\right) - t_i x_i \beta). \tag{75}$$

with δ_i as censoring indicator being $\delta_i = 0$ for observed events and $\delta_i = 1$ if an observation is censored. For practical application, it is possible to assume a wide range of baseline distributions whereby their fit is comparable via model selection criteria. In addition to the parameters of main interest, i.e., the regression coefficients β, the distribution parameters and more intuitive transformations thereof, the method is able to returns for instance relative survival or absolute measures such as the number needed to treat (NNT). Referring to the binary covariate X_{TNM} from the HALLUCA example and using the information from Table 1, the latter can be obtained via

$$NNT = \frac{1}{RD} = \frac{1}{F_0(t) - F_1(t)} = \frac{1}{S_1(t) - S_{0,\theta}(t)}. \tag{76}$$

Knowing that for $X_{TNM} = 1$ the survival function $S_1(t)$ is given by $\frac{S_{0,\theta}(t)}{\exp(t\beta)}$, this can be rewritten as

$$NNT = \frac{1}{\frac{S_0(t)}{\exp(t\beta)} - S_{0,\theta}(t)} = \frac{\exp(t\beta)}{S_{0,\theta}(t)(1 - \exp(t\beta)} \tag{77}$$

for the proposed additive hazard model.

Part III

2.3 Neutral Comparison Studies

"Statisticians, like artists, have the bad habit of falling in love with their models." [105]
(Brad Efron)

In its broadest sense, statistical modelling refers to the analysis, approximation, depiction and prediction of real-world events and dynamics in a formal and theoretical manner using mathematical relationships and statistical assumptions. In fields such as computational science, bioinformatics and medical research, the development of statistical models has always been a vivid research area [22,23,106]. New methods, i.e., approaches and procedures for analysing data, are expected to "improve the world", be it by simplifying processes or bringing results of statistical analyses closer to the truth. While the motivation for this phenomenon should be the aim for continuous scientific progress and improvement, it may also partly be attributable to pressuring publications policies and the prevailing publish-or-perish-culture [107]. In that regards it seems only logical that nowadays, most published articles are devoted to the development of new methods, as it is one of the most straightforward ways for a work to be considered as novel and innovative, a prerequisite for publication [22,23,106,108]. Consequently, new statistical methods are currently flooding fields, resulting in a quickly evolving situation with an ever-growing conglomeration of available models to be used.

When first presenting a new method, authors commonly perform comparative studies (i.e., the new versus established models or versus a gold standard method) as part of the introductory paper [23]. This is usually done with the (unconscious or deliberate) intention of demonstrating the dominance of the new method compared to existing ones. In the past, such original research articles presenting new methods have often been found to be biased in favour of the newly proposed method and over-optimistic, stressing the superiority of the new method [22,108]. Based on anecdotal evidence, statistician Efron claims that "new methods always look better than old ones. […] In fact it is very […] easy to inadvertently cheat by choosing favourable examples, or by not putting as much effort into optimizing the dull old standard as the exciting new challenger" [109]. Into the bargain, there is substantial empirical evidence that selective reporting, publication and optimistic biases are prevalent in diverse domains of science and that it is comparably easy to make a method appear better than it actually is [23,110]. Doubtlessly, this overoptimistic bias could become apparent without any intent to be fraudulent [111]. Nevertheless, it may arise through means of intentional bad scientific practice, such as "HARKing" (Hypothesizing after the Results are Known) [112], p-hacking [113], fishing for significance [114,115], data dredging [116], selective reporting [117], or "SOTA" (State of the Art) hacking [118]. Further, literature confirms that the author's level of expertise in the field of application and alternative statistical models, the optimisation and overfitting of new algorithms to the data sets considered during the development and introduction phase, the selection of favourable datasets and settings, a profitable choice of competing methods, optimal parameter tuning for the preferred method, post-hoc modification of specific design and/or analysis components, as well as selective reporting of method variants and analysis settings play a crucial role [108,110]. With trust in science, it may be important to note here that the superiority of a new method may obviously not necessarily be wrong and overt fraud is probably rare [23,111]. However, referring to Nuzzo who claims that "even an honest person is a master of self-deception" [119], one must acknowledge that everyone is at risk of (consciously or unconsciously) engaging in such questionable research practices [111,114]. The recurrent character of such claims of superiority may seem somewhat suspicious, also because it is not systematically assessed and validated from a neutral perspective.

As concern over the use of problematic research practices in academia has increased in the past decade, it is of interest to identify root causes of such behaviour. Cognitive biases (such as confirmation, experimenter or hindsight bias) are considered as risk factors, but it is foremost claimed that institutional and career-oriented incentives may encourage the use of bad scientific practices [120]. The pressure to publish novel and positive results combined with low requirements from journals (e.g., with regards to code sharing, documentation and guidance on designing a comparison study) may prompt researchers to being susceptible to engage in questionable research practices in comparison, benchmark and simulation studies [121]. Further, studies that focus on the review, evaluation, validation and comparison of only existing methods and that are written by neutral authors are generally appreciated by readers, but are given poor consideration by many journals, editors and researchers [23,108]. This is in strong contrast to clinical research where most published medical papers do not suggest new measures [22]. Instead, the vast majority focuses on many other types of clinical research projects such as large validation studies, phase IV clinical trials, or meta-analyses [22]. Vice versa in other fields, replication and comparison studies are often implicitly excluded from the journals' scopes, not deserving publication owing to lacking innovation and novelty [121].

This prevailing situation combined with publication, over-optimism and other reporting biases gives rise to several dilemmas: Firstly, over-optimism may largely detriment the credibility and value of research evidence in case subsequent comparison studies by independent authors fail to reproduce the superior performance of the method of interest [22,23,106,108,110,111]. Secondly, it seems that authors have been afraid in the past of not being published when reporting balanced (and thus potentially less exciting) results from comparison studies of existing studies without introducing any new features or models. The little acceptance, tolerance or interest from the sides of editors and reviewers in such studies results in a challenge for researchers as it remains open to debate on how to "properly" compare methods and make recommendations [23,122]. In recent years, pioneering works have put forward some general guidelines, frameworks and recommendation on how to conduct comparison studies [23,106,108,123-125]. Though, little thereof is put into practice, and despite this latest general advice, for many issues relevant for practical application in reality no concrete guidance or methodology can be found. Hence, researchers face a multiplicity of design and analysis options when conducting benchmarking and comparison studies [23]. Thirdly, the overload of available methods together with a lack of neutral guidance results in a challenge for practitioners with regards to an appropriate choice and correct application of a method. Often, it remains open to debate which method to use, or when and how to apply a method. Consequent inappropriate applications of statistical methods may return false and misleading, inaccurate or exaggerated findings [114]. Altogether, these issues combined with the publish-or-perish-attitude mentioned above contribute to an estimated waste of 85% of research resources and ultimately [114], perpetuate the ongoing methodological replication crisis which centres around problematic failures to replicate [22].

On this account and given the large set of publications in which researchers struggle with presenting fair comparisons, some scholars have made a plea for neutral comparison studies in computational sciences as possible solution [22]. Having shown the potential to make the establishment of standards more objective and to give fair chances to all methods, the need for and interest in studies that focus solely on neutral and fair comparisons seems to deserve more attention in the scientific community.

Section 2.3 is divided into the following subsections: The purpose of Section 2.3.1 and 2.3.2 is to define neutral comparison studies. Section 2.3.3 introduces DTA studies in general and gives information on the tasks and particular characteristics of meta-analysis of DTA studies. Section 2.3.4 provides a brief description of three different frequentist approaches previously proposed for the meta-analysis of ROC curves from DTA studies. As the field presented in Section 2.3.3 currently lacks guidance on the methodological choice, Section 2.3.5 describes the steps involved in generating a neutral comparison

study for the methods introduced in Section 2.3.4. Using the methodological knowledge and background information discussed in the overall Section 2.3, the fifth and final contribution follows the "plea for neutral comparison studies", and aims to neutrally and systematically compare three frequentist approaches for the meta-analysis of receiver-operating-characteristics (ROC) curves in a simulation study.

2.3.1 Neutral Comparison Studies

Neutral comparison studies are broadly defined and characterised by three main components which make them essentially unbiased [22]:

i. The primary research goal of a neutral comparison study is to conduct a comparison of methodological approaches proposed elsewhere instead of introducing a novel method [22].
ii. Neutral comparison studies should be performed by reasonably neutral authors, as non-neutrality of the authors may induce a bias in general. Further, the authors should be equally experienced with all methods under investigation and relevant in the field of application [22,23,126].
iii. Neutral comparison studies should be designed, analysed and reported in a systematic and rational way. That is, the selection of the dataset(s), methods, and evaluation criteria should be based on strictly pre-defined inclusion criteria [22,23,106,115].

Analogous to clinical research in drug development, the development stage of a method can be structured in four subsequent phases [120]. Whereas the first phases of methodological development focus on explorative, method-demonstrating and illustrative comparisons, later phases should consist of confirmative, neutral comparison studies. The framework consists four subsequent phases of methodological research defined as [120]:

Phase I: Initial introduction of a new methodological idea based on logical reasoning, proofs and the assessment of asymptotic properties.
Phase II: Provision of empirical evidence in a narrow target setting, i.e., an application to data from real-world practice.
Phase III: Extensive investigation of the method and its assumptions in a wide range of settings, across several data sets and for various outcomes demonstrating a method's validity and relative performance.
Phase IV: Examinations showing that a method itself, as well as its strengths and limitations are sufficiently understood.

This entails that comparison studies become important only after early phases of method introduction. Their aim is to firstly, validate and ensure the proper functionality of existing methods in concrete settings and secondly, provide recommendations and guidance to applied and methodological researchers [120,122]. The purpose here is to enable researchers, i.e., "method users", to find the "optimal" method for her or his application. At the same time, it may provide evidence-based guidance for "developers" by identifying potential limitations of existing methods and thus, shed light on the need for development of extensions or new approaches [122].

Regarding terminology, there is a plethora of related expressions in the context of methodological research and statistical reporting terms are used inconsistently in the literature. Generally, the main emphasis in method comparison studies rests on a direct comparison of alternative methods, their properties and performances. Often, the question of interest is whether the methods are comparable to the extent that one method is superior to other ones, i.e., results in sufficient accuracy for a certain research purpose. The overall heading of comparison studies includes different approaches, such as benchmarking experiments or simulation studies. Essentially, benchmarking and simulation studies

provide different approaches to a similar problem: using real or simulated data for evaluating the performance of several alternative methods [122]. Benchmark experiments or benchmark studies are defined as a "systematic comparison between computational methods, in which all of them are applied to a gold standard data set and the success of their [...] predictions are summarized in terms of quantitative metrics [...]" [127]. They can be described as data example demonstrating the application of proposed methods to real-world data. Thereby, benchmark studies allow for assessing of whether the choice of methods matters in practice. As benchmarking studies are based on practical data examples, they are likely to adequately reflect properties of real-world data and therefore allow for generalisation of results [122]. The term "benchmarking" originates from computer sciences [128] and in the context of e.g. artificial intelligence, machine learning and bioinformatics, scholars commonly refer to "benchmarking" [108,122,127]. Less often used terms are "empirical study" [129], "empirical evaluation" [130], or "empirical comparison" [131]. The main difference is that while benchmarking studies present a data-driven approach, simulation studies follow a theoretic approach. Simulation studies are a common statistical tool to complement the theoretical derivation of a statistical model. It is assumed that the underlying statistical model and some theoretical concepts of the data-generating process are known. The main advantage in simulation studies is that the "ground truth" is known which enables an accurate investigation of the methods as the underlying true values are known by design [122]. In contrast to analyses using real-world data, simulation studies impose almost no restrictions on sample size. Simulation studies are particularly useful when the aim is to (i) compare existing models' performances in a given setting, (ii) assess small sample properties and asymptotic results of a method, (iii) investigate a model's robustness in case of violated assumptions and (iv) study the model in contexts of complex study designs (such as adaptive designs).

2.3.2 Guidance and Practical Recommendations for Conducting Neutral Comparison Studies

Ideally, neutral method comparisons should essentially guarantee a fair and systematic comparison of existing methods across different scenarios. Although the three characteristics by Boulesteix et al. [22] provide a clear formal definition of neutral comparison studies, these requirements may sometimes be challenging to fulfil in practice. Moreover, as these requirements are relatively broad and leave room for interpretation, several scholars have proposed concrete guidelines and recommendations for the implementation of neutral comparison studies.

Condition (ii) of the definition of Boulesteix et al. [22] necessitates that the authors of a neutral study (at least as a collective) do not have any method preferences as well as they should be experienced with the methods of interest and their fields of application. In case of potential non-neutrality and consequential potential to exploit the multiplicity of possible scientific options, strategies inspired from blinding in clinical trials could present a remedy. For instance, in case simulated data is used for the comparison, researchers could be blinded to the data generating process, that is, an independent researcher would generate the data. This would render it substantially difficult if not impossible to tune the parameters of preferred methods according to the known ground truth [23,132]. Alternatively, blinding could be achieved by labelling the methods under investigation with non-informative names during the analysis and evaluation processes [23,133]. With regards to experience, it has been noted decades ago that the performance of some methods is inherently tied to the skill of the analyst who applies them [126]. There is substantial evidence that a low level of expertise is one of the mechanisms leading to deteriorating performances in subsequent papers [110]. An option to avoid such distorted effects is to involve the initial authors of the method as co-authors, letting them each implement their method themselves [133-135]. Alternatively, and potentially more easily feasible, another possibility would be to contact the authors or well-known experts with substantial knowledge in the field of application and who are familiar with the methods considered and ask for assurance that all methods are correctly implemented [110].

With regards to systematic and rational structure of a neutral comparison study stated in requirement (iii) by Boulesteix et al. [22], i.e., ensuring an adequate design, data choice, analysis and proper reporting of results, researchers encounter a multiplicity of choices. In recent years, the available literature has increased and now provides a modest foundation of guidelines on how to conduct comparison studies [106,133], statistical frameworks [106,123-125] (all focused on the field of supervised learning) as well as strategies to prevent over-optimistic result and avoid common pitfalls [23,110,135,136], only to name a few.

In pioneering works for instance, Boulesteix [22] published ten rules for the development and testing of new computational methods in methodological research to alleviate the considerable influence of over-optimism in practice. These rules promote to assess the methods in a suitable context (Rule 1) and comparing a new method to the best existing methods (Rule 2) in an appropriate manner (Rule 8). With regards to data, Boulesteix recommends to consider several independent, reasonably chosen datasets for evaluation and validation (Rule 3, 4, 7). In terms of reporting, Boulesteix urges to reporting all information related to the methods and study itself (Rule 9), discuss several performance measures instead of only single objective performance criterion (Rule 6) and particularly, clearly document any limitation that may become apparent (Rule 5).

For concrete information on the choices of design and analysis options faced in a neutral comparison study, see for example Nießl et al. [23]. They developed a framework that (i) gives a detailed overview of design and analysis options, (ii) shows how using alternative options for a specific choice affects the results and (iii) give effective strategies to prevent over-optimistic interpretations and biased conclusions. They found that particularly the choice of performance measure and data sets causes major variability in the methods' performances and that therefore, any related choice should be made with caution. They refer to Hoffmann et al. [137] for further strategies to prevent over-optimistic results. Hoffmann et al. [137] outlined six steps to increase the replicability and credibility of one's own methodological research. Essentially, the steps proposed by Hoffmann et al. [137] include that researchers should (i) be conscious of the flexibility and multiplicity of analysis options, (ii) reduce uncertainty, (iii) include uncertainty, (iv) report uncertainty, (v) acknowledge uncertainty and (vi) make all source code, data and research material publicly available. In a subsequent work, Nießl et al. [110] presented further strategies to prevent over-optimistic result and avoid common pitfalls in the development and initial introduction of a new method. For instance, they suggest to involve the authors of the investigated methods to overcome different and potentially insufficient levels of expertise. Most importantly, they advocate maximum transparency and comprehensibility with regards to documentation of (i) a method's properties, field of application, limitations and the like, as well as transparent reporting of (ii) the planning and execution of a method comparison.

With regards to the use of simulated or real-world data, advice can be found in a variety of works. See Friedrich et al. [122] for an extensive discussion of the advantages and disadvantages of real versus simulated data sets. They conclude that ideally, a combination of both approaches should be used for method evaluation wherever possible. In addition, they suggest to establish infrastructure, databases and gold standards to enable large-scale benchmarking and comparison studies, and encourage the conduct and publication of comparison studies. Pawel et al. [135] provide a wide body of recommendations involving various stakeholders in the research community to increase the number of well-designed simulation studies. Researchers should adopt pre-registered simulation protocols and good computational practices in terms of code review, packaging and unit-tests. Further, researchers are asked to maximise transparent reporting, and archive, time-stamp and share all code, data, documentation and materials. To disclose multiplicity and bias, Pawel et al. [135] recommend that researchers perform simulations in a blinded manner, assess uncertainty of results via sensitivity analyses and collaborate with other research groups. Further, they advise researchers to engage in teaching simulation study methodology in statistics (post)graduate courses. With regards to editors

and reviewers, Pawel et al. [135] put forward that they should encourage (neutral) method comparisons and (pre-registered) simulation protocols, as well as they should provide enough space for description of simulation methodology. Journals and funding bodies should provide incentives and funding for assessing and comparing methods in simulation studies. In addition, Pawel et al. [135] suggest that journals and funders should promote standardized reporting, require code and data and adopt reproducibility checks. Into the bargain, Morris et al. [134] published a tutorial that contains a wealth of guidance and advice on how to run simulation studies and how to use simulation studies to evaluate statistical methods. They introduced a structured approach (abbreviated as "ADEMP") which involves defining aims ("A"), data-generating mechanisms ("D"), estimands ("E"), methods ("M"), and performance measures (“P”). Beyond that, in a 16-point list with illustrating examples, Morris et al. [136] offer advice on how to check a simulation study for potential errors, and how to optimally design and conduct a simulation study to give results that are easier to check. For instance, they suggest to carefully examine for outliers and failed estimation, and inspect surprising results against Monte Carlo errors.

With respect to the evaluation of methods, Strobl et al. [121] argue that comparison studies should be enriched by shifting the focus away from finding an overall winner towards more differentiated results. They refer to the principle of “meta-learning” (or more specifically “algorithm recommendation”) taken from the field of machine learning. In the spirit of meta-learning, comparison studies should aim to predict which algorithm performs best on a specific data set based on its characteristics. They stress the importance of scholars to accept that there is no universally best method and strongly argue against the “one method fits all data sets” philosophy. Rather, Strobl et al. [121] advise to pursue the objective of offering comprehensive information about concrete method properties and about which methods performs well in which particular kind of data situation and choice of performance measures. Ultimately, this could lower the pressure inherent in the publication process by reducing the “necessity” to tweak parameters or cherry-pick data sets.

Lastly, there is general agreement that transparent and clear documentation is key [111,134,136]. While the mechanisms discussed above would reduce any deterioration of performance, there is no tradition of explicit neutral comparison or replication studies. In other words, despite some general advice on the design and analysis of neutral comparison and benchmark studies provided in recent literature, there remains a lack of concrete guidance for several issues faced in reality. Hence, the conduct of such studies still equals what statistician Gelman calls a "garden of forking paths" [138] and the researcher’s degree of freedom remains inherent in the scientific execution of comparison studies [113]. This stresses the importance of clear, precise and comprehensive documentation of any preliminary actions, and all choices made throughout and post-hoc to the execution of a study. Maximum transparency and availability, also with regards to sharing of data, protocols, materials, and software, are crucial to impede biased interpretations of results, over-optimistic conclusions and failures of validation or replication [110,134,135].

2.3.3 Meta-Analysis of Diagnostic Test Accuracy Trials

With the attempt to gather all available empirical evidence to obtain answers to a specific question, systematic reviews are a centrepiece of evidence-based medicine [139]. In the form of a meta-analysis, the results of all studies comprised in a certain systematic review can be combined into an overall result. Such meta-analyses offer the opportunity to combine and critically and systematically evaluate results of comparable studies [140]. Among other benefits, the statistical combination of results from two or more separate studies increases the numbers of observations, improves the statistical power and achieves higher certainty in estimates of the effect sizes of an association or intervention [140]. Meta-analyses continue to become increasingly popular and with regards to interventional studies, there is a range of well-established statistical methods combining the results from different studies [141].

Contrarily, the development of methods for meta-analysis of DTA studies is still an ongoing, vivid area of research [139,141]. This is particularly attributable to the increased complexity of the bivariate outcome of DTA studies [139,141]. Generally, the aim of DTA studies is to assess how accurate and well a diagnostic test (e.g., a score, a biomarker, or an imaging parameter) is able to detect or exclude a target condition of interest (i.e., in medical research this is often a disease of interest) at varying diagnostic thresholds [140,142]. This results in a bivariate outcome, i.e., two measures of probability, usually reported in terms of sensitivity and specificity [142]. Sensitivity refers to a test's ability to correctly appoint a diseased individual as positive. Vice versa, specificity quantifies the test's ability to correctly identify a non-diseased individual as negative. Meta-analyses in this context may even further be complicated, as each DTA study reports several pairs of sensitivity and specificity which all belong to a different diagnostic threshold [141]. ROC curves are frequently used to graphically show the connection or trade-off between sensitivity and specificity for different possible thresholds. Albeit there is a range of well-known and routinely used methods for single pairs of sensitivity and specificity, there is a lack of guidance on the methodological choice when meta-analysing ROC curves [139].

Among the most well-known methods for bivariate meta-analyses of DTA studies are the hierarchical summary receiver operating characteristic curve model proposed by Rutter and Gatsonis [143] and the bivariate model proposed by Reitsma et al. [144] and Chu and Cole [145] [139,141]. These approaches take heterogeneity and correlation between individual studies arising from different threshold values across studies into account, but do not explicitly consider study-specific threshold values. Further, they do not allow considering several and different thresholds per study. Though, accommodating only sensitivity and specificity from a single threshold from each study may result in heavy information loss [139]. In addition, as they do not use the actual numerical values of the thresholds, it remains unknown what threshold value any pooled estimate (or point on the summary ROC curve) corresponds to. To counteract these issues it has been suggested in the past to either use standard meta-analysis methodology and select only one pair of sensitivity and specificity or to perform several independent meta-analyses based on the same studies to estimate summary sensitivity and specificity at each threshold [146]. However, this is far from optimal, potentially resulting in neglected correlations between several thresholds, loss of knowledge due to ignorance of the full information from single studies and unreliable estimates [139]. Literature provides more advanced and specialized statistical approaches that are able to handle the full information in a unified analysis. Yet, they come with relevant limitations, as for instance, the requirement of identical thresholds across studies, overoptimistic findings or the ignorance of precise threshold values which turns inference on sensitivity and specificity at given thresholds impossible [139]. Other drawbacks are the assumption of fixed-effects, i.e., identical true underlying values for sensitivity and specificity across studies or that methods are inapplicable in extreme situations, e.g., with values of 100% for sensitivity or specificity (see [139] for a detailed discussion).

Recently, some advanced approaches with desirable properties have been proposed [141]. These allow handling data from multiple thresholds in a unified analysis and do not suffer from the beforementioned issues [141]. Three of the methods have been shown to work well in practice, and have been discussed as promising additions to the toolbox of meta-analysis of DTA studies. Precisely, the three frequentist approaches have been proposed by Steinhauser et al. [147], Hoyer et al. [139], and Frömke et al. [148] and are presented in the following Section 2.3.4.

2.3.4 Methods for Meta-Analysis of Diagnostic Test Accuracy Studies with Multiple Thresholds

The method proposed by Frömke et al. [148] is essentially non-parametric, i.e., does not assume a certain distribution. In contrast, the two other approaches that are introduced in this section assume that each study to be included in a meta-analysis returns test results that differ in their distribution for diseased and non-diseased. As outcome, the approaches model suitable transformed diagnostic test values of

the diseased and the non-diseased, providing summary sensitivities and specificities for the respective threshold values. In terms of notation and model definition, the following Sections 2.3.3 and 2.3.4 are aligned with Zapf et al. [141].

$k = 1, \dots, N$ denotes the studies included in a meta-analysis. The true disease state is captured by $d = 0$ (individuals without the target condition, i.e., non-diseased individuals) and $d = 1$ (individuals with the target condition, i.e., here diseased individuals). The total number of diseased and non-diseased individuals in study k is denoted by n_{kd}, while each individual study participant is indexed by $s = 1, \dots, n_{kd}$. For each individual s, the actual result of the continuous diagnostic test in disease state d and study k is labelled X_{kds}. The different thresholds at which data are available in each single study k are indexed by $i = 1, \dots, t_k$. The corresponding numerical threshold values to these indices are labelled c_{ki}. The two measures sensitivity se and specificity sp are used for final evaluation of the diagnostic test. The sensitivity se is defined as equivalent to the true positive fraction tpf, while the specificity sp equals the true negative fraction tn. Accordingly, $1 - sp$ yields the false positive fraction fpf. Lastly,

$$\text{logit}(x) = \log(x) - \log(1 - x) \tag{78}$$

is used for logit-transformation of the threshold values.

The Random Effects Model by Steinhauser et al. (2016)

The approach proposed by Steinhauser et al. [147] relies on a two-stage random effects model. The model assumes that at study level and for each value of the threshold, the observed true and false negative fraction (i.e., sp and $1 - se$, respectively) are transformed via a suitable quantile function f. Log-transformation was used for the threshold value and the logit-function defined in (78) was used for f. At meta-analysis level, Steinhauser et al. [147] apply a linear mixed model to fit the resulting values across studies. Aligning with the assumption of the underlying Log-Logistic distributions for X_{k0s} and X_{k1s}, this model is given by

$$\text{logit}(\, sp_{ki}) = \alpha_0 + a_{0k} + (\beta_0 + b_{0k}) \log(c_{ki}) + \varepsilon_{ki} \tag{79}$$

$$\text{logit}(1 - se_{ki}) = \alpha_1 + a_{1k} + (\beta_1 + b_{1k}) \log(c_{ki}) + \delta_{ki}. \tag{80}$$

$\widehat{sp_{ki}}$ and $\widehat{se_{ki}}$ denote the crude estimates of specificity sp and sensitivity $1 - se$ at threshold c_{ki} in study k. α_0 and α_1 are fixed intercepts, while β_0 and β_1 represent fixed slopes for the non-diseased and diseased individuals, respectively. Random intercepts and slopes are denoted by a_{0k}, a_{1k}, b_{0k} and b_{1k}, which are assumed to follow a common four-dimensional normal distribution allowing for correlation across studies. ε_{ki} and δ_{ki} are the within-study error terms. Each data point is weighted using the inverse variance of the corresponding logit-transformed proportion. Model-based distribution functions for the non-diseased and diseased individuals are obtained by back-transforming the fixed effects. Finally, these distributions are used to obtain estimates of the summary ROC curve with pointwise confidence regions. The area under the curve (AUC) can be estimated by numerical integration based on the trapezoidal rule.

The Steinhauser model was implemented in the R package diagmeta (version 0.5-0) [149] in the freely available software environment R (R Core Team, 2021 [69]).

The Time-to-Event Model by Hoyer et al. (2018)

The model proposed by Hoyer et al. [139] assumes that diagnostic test values can be seen as interval-censored since it is only known if test values exceed or fall below a predefined threshold. To model

these interval-censored diagnostic test values, it is possible to choose from various distributions (e.g., Weibull, Log-Normal, Log-Logistic). In the context of DTA studies, and in line with time-to-event models, the "events" of interest are here to be tested positive in the population of diseased or negative in the population of non-diseased. Further, the "time" is here indicated by the diagnostic test value. The event probability in the diseased population is thus reflected by the sensitivity, and the event probability in the non-diseased population is represented by 1- specificity. The outcome, i.e., the diagnostic test values, are log-transformed which results in an accelerated failure time model with a unified linear predictor. Formally, the model equations are

$$\log(x_{k0}) = b_0 + \varepsilon_0 + u_{k0} \tag{81}$$

$$\log(x_{k1}) = b_1 + \varepsilon_1 + u_{k1} \tag{82}$$

where

$$\begin{pmatrix} u_{k0} \\ u_{k1} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix}\right] \tag{83}$$

and with b_0 and $b_1$denoting the location parameters after log-transformation of the outcome of interest. ε_0 and ε_1 depict error terms with distributions corresponding to the log-transformed diagnostic test values x_{k0} and x_{k1} in the population of non-diseased and diseased, respectively. u_{k0} and u_{k1} are study-specific random effects which are assumed to be bivariate normally distributed with correlation parameter ρ and variances σ_0^2 and σ_1^2. These random effects account for potential between-study heterogeneity and correlations and are added to the location parameters after log-transformation of the outcome. Sensitivities and specificities are predicted at several thresholds based on the resulting survival function. The related AUC is obtained by use of the trapezoidal rule.

The approach was implemented in SAS 9.3 (SAS Institute Inc., Cary, NC, USA) with the source code available in the original article of Hoyer et al. [139].

The Non-Parametric Model by Frömke et al. (2020)

Frömke et al. [148] extended the non-parametric method for diagnostic studies by Konietschke and Brunner [150] and transferred it to the context of meta-analysis of DTA studies with multiple thresholds. Their proposed model explicitly avoids the concrete parametric distribution assumption for the test values among the diseased and non-diseased. The approach proposes that the AUC equals the relative effect

$$p = P(X_{k0s} < X_{k1s}) + \frac{1}{2}P(X_{k0s} = X_{k1s}). \tag{84}$$

To estimate the AUC, all measurements X_{kds} are replaced by their global mid-ranks R_{kds} and for all individuals with disease status d over all studies, the mean rank $\bar{R}_{.d.}$ is calculated. Then, it holds that the AUC is estimated by

$$\widehat{AUC} = \frac{1}{2} + \frac{1}{n}(\bar{R}_{.1.} - \bar{R}_{.0.}) \tag{85}$$

with $n = \sum_{k=1}^{N}\sum_{d\in[0,1]} n_{kd}$ as the total number of participants from all studies included. To calculate the ranks, this would require data on individual participant level X_{kds}. Though, for diagnostic studies only aggregated data are available. Therefore, the number of diseased and non-diseased individuals with diagnostic test values below or greater than the study-specific thresholds c_{ki} is used to generate fictious data between the thresholds c_{ki} and c_{ki+1} from a one-point or uniform distribution. Precisely,

data are drawn uniformly distributed from the interval from the smallest and largest threshold that is included in the meta-analysis for values below and above the first and last threshold. Specificity and sensitivity are estimated by replacing the observations of the diseased and non-diseased, respectively, by a one-point distribution. By means of the asymptotic equivalence theorem, it holds that $\sqrt{N}(\hat{p} - p) \sim N(0, \sigma^2)$. Wald confidence intervals for all estimates, i.e., AUC, specificity and sensitivity, can be derived. Applying logit transformation to the confidence interval ensures that the boundaries remain within the interval [0,1].

The model was implemented in R (R Core Team, 2021 [69]).

2.3.5 Neutral Comparison of Methods for Meta-Analysis of Diagnostic Test Accuracy Studies

Yet, there is a lack of (neutral) guidance on the usage of the different methods presented in Section 2.3.3 and it remains unclear which approach may be preferable in a given situation. This can be changed through means of a neutral comparison study and using simulated data to systematically compare the approaches and their respective performance in various scenarios when the ground truth is known. The following briefly outlines the carrying out of such a comparative simulation study (for details see Contribution 5). Incorporating the authors of the three approaches in the comparison study ensures neutrality (at least to a certain level) and guarantees the required expertise in each methodology and the field of application. Aligning the study with the guidelines provided by Morris et al.[134], Boulesteix et al.[22,106], Nießl et al.[23,110], Friedrich et al.[122] amongst others, warrants that the analysis can be considered a valid basis for a fair and neutral comparison of the approaches. For maximum transparency and to allow for replication, all data and code are publicly available on Zenodo[151].

Design and Setting

As means to combine simulations and benchmarking and thereby maximize the advantages and minimise the limitations of each type of data, Friedrich et al.[122] recommend to simulate data based on a real data example. Therefore, a case study meta-analysis from the field of nephrology conducted by Haase-Fielitz et al.[152] is used as an example to motivate the simulation setting. Haase-Fielitz et al.[152] aimed to evaluate the diagnostic accuracy of neutrophil gelatinase associated lipocalin (NGAL) as a biomarker for early prediction of acute kidney injury (AKI). The data has already been used as example in the practical benchmarking study by Zapf et al.[141]. It provides a suitable context for the implementation of the three methods and can be assumed representative for domain of interest as recommended by Boulesteix[22] ("it is important to make a selection of data sets that is "as representative as possible" to cover the domain of interest").

With regards to software, SAS 9.4 (SAS Institute Inc., Cary, NC, USA) is used to generate data. Estimation is conducted in the software in which the models were initially implemented, i.e., SAS 9.4 for the approach of Hoyer et al. (2018) and R[153] for the models of Steinhauser et al. (2016) and Frömke et al. (2022). For a unified evaluation, the results of all methods are summarised and visualised in R (R Core Team, 2021 [69]).

Data-Generating Mechanisms

The approach of Hoyer et al. (2018) is used as parametric data generating model to simulate diagnostic values of the diseased and non-diseased study participants. The Weibull-, Log-Logistic, Log-Normal and Normal distribution are assumed as true underlying distributions. The latter is notable as it results in misspecification of all three methods. To assess the approaches in a wide range of settings (108 different scenarios in total), several parameters are varied: The true underlying AUCs (0.7, 0.8 and 0.9), the true disease prevalence (0.02, 0.2 and 0.5) and the correlation between random effects to model

heterogeneity across studies (0.3, 0.6 and 0.9). Graphical and tabular representations are used to give precise information on how the model parameters were varied to mimic the true AUCs, to report the resulting true sensitivities and specificities for given values of thresholds at which the models are finally evaluated and to visualise the theoretical distribution curves of both populations per scenario. For each of the 108 scenarios, 1000 meta-analyses are generated. The uniform distribution is used and values are rounded to the nearest integer to sample the number of studies per meta-analysis (5 to 30), the number of participants per single study (30 to 300) and the number of thresholds per study (1 to 10). The size of the diseased population is derived by multiplying the true underlying prevalence by the total number of study participants. The number of non-diseased can then be determined as difference between the total number and number of diseased. To identify true positives and negatives, individual diagnostic test values are compared to the generated thresholds. This information is used to create a final diagnostic contingency table that serves as input data for the simulation.

Estimation Methods and Performance Measures

Data from each meta-analysis was incorporated in each of the three models as follows. Parameter estimation for the approach of Hoyer et al. (2018) is implemented in SAS PROC NLMIXED with the default settings and assuming that the interval-censored diagnostic test values are Weibull distributed. Input values for the fixed and random effect parameters were derived from univariate models. The logit-function defined in (78) is used to log-transform the threshold values for application of the method proposed by Steinhauser et al. (2016). For the approach of Frömke et al. (2022), no distribution specification is needed.

The performance of the three approaches is assessed by comparing the estimated sensitivities and specificities together with their 95% confidence intervals at the specified threshold values of 50, 100, 130, 150 and 200. As recommended, several measures of interest are used for a comprehensive evaluation of the approaches [134]. Precisely, the absolute bias, the mean squared error, the empirical coverage of the two-sided 95% confidence intervals and the model's convergence are computed. The latter is used to evaluate the robustness in terms of number of converged simulation runs [134].

3 Availability and Aspects related to Data, Code and Software

Without any doubt, the best statistical method is of little use if it cannot be used in practice and does not allow for application to an actual or simulated data set. Since one of the aims of this thesis is to make statistical modelling of chronic disease-related data more accessible and comprehensible, as well as to provide guidance for selecting and implementing suitable statistical approaches, all of the five contributions incorporate a practical application of the relevant statistical approaches.

Some of the analyses (related to contributing articles 4 and 5) were implemented in SAS (SAS Institute Inc., Cary, NC, USA). However, the open source software R (R Core Team, 2021 [69]) was used for the majority of the computational aspects related to perform the statistical analyses as well as for summarising and visualising the results. To make the methodology accessible, to maximise transparency and to allow for reproducing the analyses in the main articles and their supporting information, the source codes and underlying data sets of all contributing articles are publicly available on Zenodo or directly on the publisher's website:

Contribution 1 - https://doi.org/10.1371/journal.pone.0264739

Contribution 2 - 10.5281/zenodo.6799292

Contribution 3 - 10.5281/zenodo.8009685

Contribution 4 - 10.5281/zenodo.7124988

Contribution 5 - 10.5281/zenodo.7802089

Data are available either in a public, open access repository or have been published in previous publications of other scholars as referenced in the code and in the according contributing article. Due to data regulations, some data was not allowed for publication. However, in these cases only relevant parts of the data are uploaded or a simulated data set is provided to being able to run the analyses.

The software R consists of some core packages (known as "Base R") which are already included when installing the software. The core language of R is extended by numerous packages that contain reusable code and documentation and which are published online. These additional packages comprise further functions for analysing data with regard to research questions from various different contexts. The following essential packages (besides "Base R) have been used for this dissertation:

addhazard	functions for fitting additive hazard models for survival analysis
deSolve	functions for numerical treatment and solving of differential equations
diagmeta	functions for the meta-analysis of DTA studies with multiple thresholds
dplyr	functions for data transformations, such as tools for mutating, rearranging or joining data.
forcats	functions for manipulating data in the form of factors
ggplot2	functions for visualisations and graphics
gridExtra	functions for working with "grid" graphics, i.e., to arrange multiple grid-based plots on a page, and draw tables
haven	enables R to read and write various data formats used by other statistical packages

matrixcalc	functions to support matrix calculations for probability, econometric and numerical analysis
meta	functions for standard methods for meta-analysis
numDeriv	functions for calculating (usually) accurate numerical first and second order derivatives
optimx	general-purpose optimisation wrapper function that calls other R packages and functions for optimisation, such as the existing optim() function
purrr	tools for working with functions and vectors
readr	functions for importing data from different sources
stringr	functions for manipulating data in the form of strings
survival	functions for analysing survival data
tibble	functionalities for data wrangling and storing data in a tidy form
tidyr	functionalities for data wrangling and storing data in a tidy form
tidyverse	set of packages which help to transform and better present data. It assists with data import, tidying, manipulation, and data visualisation.

4 Discussion

"Statistics is the grammar of science." [154]
(Karl Pearson)

Statistics can be defined as mathematical discipline that concerns the collection, organisation, analysis, interpretation, and presentation of data[155]. Nowadays, due to the growing amount of electronic medical records, intensified data sharing and access, the available data from health research grow faster than ever[156]. Further, particularly in medical research, evidence-based decision making increasingly shifts to the centre of attention as it supplies the body of content that can lead to viable decisions[157]. This, along with the fact that that scientific knowledge continuously evolves as a valued resource in our society, the role of statistics becomes more and more crucial.

With regards to the increasing burden of chronic conditions, research on health and (chronic) diseases of a population, i.e., epidemiological research, continuously emerges as highly relevant scientific field[3-5]. Epidemiology has always had centrality in producing evidence to improve population health, to advance the quality of life and to extend life expectancy, amongst others[6]. Epidemiological findings and insights intrinsically matter because they shed light on what ameliorates or worsens population health and suggest ways to improve well-being[6].

Therefore, the central aim of this thesis lies in highlighting the relevance and usefulness of statistical methods for epidemiological research in the context of chronic disease modelling. This is achieved by (i) developing and introducing new methods, (ii) by raising awareness of useful statistical approaches and showing how to implement them in practice, (iii) by (neutrally and systematically) comparing existing and new models to advance methodological research and (iv) by applying and assessing the proposed statistical approaches to various diseases and contexts of chronic disease research to answer potential research questions that may arise in these fields.

Precisely, this dissertation investigated a variety of statistical modelling approaches from different areas of application in epidemiological research. The thesis is made up of three distinct parts that altogether highlight the importance and benefits of statistical methods for chronic disease modelling. Each part has a different focus and is accompanied by at least one contributed manuscript. Part I is devoted to epidemiological projections, e.g., projections of prevalence, case numbers or disease specific costs. Part II covers the analysis of time-to-event outcomes. It aims at justifying and motivating the need of a new statistical method. Precisely, its key contribution is to propose, implement and apply a new parametric additive hazard model. Part III focuses on the field of (neutral) comparison studies and provides guidance on suitable methodological choices by formally and systematically comparing three frequentist approaches for the meta-analysis of DTA studies in various scenarios.

Part I - Epidemiological Projection Models

Part I is concerned with statistical estimation and projection methods of epidemiological measures in the context of chronic diseases. Strengths and limitations of the different techniques are assessed and, for illustration purposes, the methods are applied in a practical example using aggregated and routinely collected data to estimate the current and future epidemiological and economic burden of diabetes in Germany. Overall, the findings of all contributing articles 1, 2 and 3 imply that Germany will possibly face greater demand for diabetes-related education, healthcare, and medical resources.

Altogether, the contributing articles highlight the need for urgent action to prepare for the potential development of diabetes and mitigate its consequences.

The first contribution focuses on a comparative analysis of existing chronic disease projection methods and introduces a new statistical modelling approach. Three methods are critically reviewed and compared in a practical application with the aim of estimating the number of men with diagnosed T2D in Germany between 2010 and 2040. The "status quo" approach simply combines the sex- and age-specific prevalence of T2D in 2010 with future population distributions[11,54]. Method 2) models the prevalence of T2D employing a scalar PDE (see Section 2.1.3) which incorporates incidence and mortality rates. Subsequently, the estimated prevalence is applied to the population projection of the FSO[54]. The newly proposed method 3) uses a two-dimensional system of PDEs which directly return future case numbers (see Section 2.1.2). The results of the three methods differ substantially: method 1 projects an increase by 29% in the number of men with diagnosed T2D in Germany in 2040 compared to 2010 (3.6 million versus 2.8 million in 2010). Methods 2) and 3) project increases by +104% (5.9 million men) and +116% (6.0 million men), respectively. It became evident that ignoring temporal trends in disease-specific rates, i.e., incidence and mortality, may result in misleading projections of future chronic disease numbers[34,54].

Contribution 2 provides estimates for the incidence, prevalence, and number of people with diagnosed type 1 diabetes for the whole German population between 2010 and 2040. Compared with 2010, the relative increase of the people with T1D ranges from 1% to 32% in 2040. A main driver of this considerable increase are temporal trends in the incidence. The work confirms that ignoring these trends, i.e., applying a constant prevalence to population projections, probably underestimates future chronic disease numbers. Further, the analysis showed that the peak of the age-specific prevalence is projected to shift toward older ages.

Based on nationwide representative routine data from 2010 from the SHI in Germany (almost 90% of the population's insurance)[26,63], contribution 3 projected age- and sex-specific healthcare expenses separately for T1D and T2D considering future demographic, disease-specific and cost trends. Currently, diabetes imposes a large economic burden on Germany which is projected to increase substantially until 2040. Total annual expenses were projected to rise remarkably until 2040 (versus 2010) by 1% to 281% for T1D (€1 to €4 billion) and by 8% to 364% for T2D (€30 to €131 billion). In 2040, and depending on annual cost growth (1% versus 5% p.a.), annual per capita costs were projected to rise to €6,581 to €12,057 for T1D and €5,245 to €8,999 for T2D. Temporal trends in the incidence and cost growth are main drivers of this increase.

With regards to methodology, the first part of this work contributes to the field of chronic disease modeling by presenting several statistical approaches for estimating and projecting epidemiological measures and healthcare costs. This is important because appropriate planning of economic and health care resources and the development of effective disease management programs require appropriate quantifications of current and future burden[11]. In order to be of use, epidemiological measures need to be representative, available in a fast and timely manner, estimated in a transparent way and must be communicated understandably[6]. However, many of the statistical principles and techniques that may be used to solve epidemiological problems require individual data from primary studies whose collection is time consuming and costly[9,49,158]. For instance, conducting longitudinal observational studies to survey long periods of time is a difficult task, if not even infeasible in some contexts. Secondary data provides a pleasing alternative, as the data has already been collected albeit for another purpose. Yet, due to strict data protection and security regulations, access to secondary data from individuals for research purposes is often limited or fully prohibited. In contrast to traditional approaches which are commonly based on individual, primary data, the presented methodological

approaches in Section 2.1.2 and 2.1.3 and in the associated contributing articles 1, 2 and 3 bear the potential to survey chronic diseases more efficiently. No individual data are needed, i.e., data in aggregated form as for instance routinely documented secondary data, are sufficient. Using secondary instead of primary data, the approaches are inherently resource-saving. For instance, the proposed methodologies show large potential to reduce costs and to improve timeliness of a surveillance system. In addition, the quality of epidemiological estimates and projections could improve and may even be considered representative, as such aggregated and routinely documented secondary data often comprise large sample sizes and evenly reflect on all age groups and other demographic factors. In view of the little time and cost expenses required for the application of the proposed methods, it would also be possible for poorer countries to develop and maintain disease surveillance activities.

Of course, in order to profit from the advantages of secondary data, it needs to be available for research purposes in a timely manner. Future valuable efforts should focus on collecting, integrating and archiving data in a standardised way and making it publicly available (or at least openly available for scientific research purposes). This is imperative for the data to be of actual use for surveillance purposes of diabetes or any other chronic disease. In Germany, there is room for improvement and it may be valuable to develop such collective systems and data bases. In countries such as Sweden, Denmark or Belgium, this is achieved through population-based registries that contain records for all individuals of a particular population diagnosed with a certain disease[32,33]. Such registries show high potential to enrich chronic disease surveillance activities. However, they are relatively costly to establish and maintain and as shown, the proposed PDE may actually serve as a valuable alternative to estimate, for instance, the incidence based on secondary data of a disease's prevalence and mortality only.

Unfortunately, the PDEs discussed are rarely used in epidemiology or public health contexts[54]. This infrequent use of such statistical models is in contrast to the high potential of compartment models, the use of mathematical relations and to the tremendous worldwide burden of chronic diseases[9]. One of the reasons may be that only a few researchers are aware of the equation and its potential. Further important avenues for future research would be to apply the proposed methods to other countries and/or other chronic diseases and to update previous epidemiological findings of chronic diseases as soon as more up-to-date data would be available, in order to raise awareness of these (more advanced) modelling approaches. Another advantage of the proposed mathematical relations is their flexibility, e.g. in terms of the possibility of simulating scenarios and assessing the effect of covariates or interventions on future case numbers. Future valuable efforts may concentrate on considering further reformulations or extensions that may be relevant in chronic disease modelling and epidemiology. Such investigations could yield valuable insights into the respective disease situations and could contribute for instance to future development of efficient disease management, appropriate resource planning, effective prevention activities or to the refinement of health policies.

Part II - Time-to-Event Models

The model presented in contribution 4 aims at providing an alternative statistical approach to already existing models used for time-to-event analyses. Implementing the new model in a practical application of a real-world example and a simulation study showed that the model works well in practice and yields valid results. The findings from the practical application to real-world data from the HALLUCA study investigating the survival of NSCLC patients indicate that per year, there are on an additional 84 deaths among cancer patients in TNM stage 4 per 100 person-years compared to those with a lower TNM-class, i.e., the increase in hazards is about 0.8. Particularly from a clinical point of view, it is key to being able to comprehensibly and clearly communicate results from any analysis. This

facilitated interpretation, for instance in terms of communicating outcomes on the original time scale and including time-independent covariates, is one of the model's most convincing advantages. The implications for medical care of lung cancer, however, are not the primary focus here. Instead, the goal was to derive the new approach itself and thereby advance statistical modelling techniques in the context of time-to-event data. In principle, the model could also be used in any other subject than the medical field.

The available source code implements the proposed parametric additive hazard model in a relatively simple setting. Though, the proposed approach is highly flexible and exciting direction for further research would be to focus on extending it and exploit all of its properties. Therefore, current work focuses on another application of the parametric additive hazard model with the ultimate aim of describing how to derive, compute and interpret the NNT. Furthermore, the method allows for instance for modelling correlated data by including random effects in the linear predictor or allows for specifying baseline distributions that have more than two parameters. Extending the algorithm in that terms would highly increase the flexibility in processing more complex time-to-event data and would contribute largely to further development and refinement of the model. Lastly, future valuable efforts could focus on creating and publishing an R-package for maximum user-friendliness and ease of practical application.

Part III - Comparison and Benchmark Studies

The goals of part III are twofold. The first aim was to fill the research gap indicated by Zapf et al.[141] and systematically assess the advantages and limitations of three different, existing statistical approaches for meta-analyses of DTA trials with multiple-threshold information. Therefore, the fifth contribution conducted a simulation study for a joint evaluation to allow for a fair comparison of the methods. Depending on the simulation scenario (108 were assessed) all methods could be presented as best or worst. Overall, the approach by Hoyer et al. (2018) can be recommended for most cases, returning smallest bias and empirical coverages closest to the specified ones. The model of Steinhauser et al. (2016) can be considered a suitable alternative as it leads to satisfactory results in many settings. In case of non-convergence, the approach by Frömke et al. (2022) could be used as back-up strategy.

The second aim was to further raise awareness and promote the conduct of neutral comparison studies to improve and support continuous innovation, development and improvement in methodological research[22]. In an ideal future world and in analogy to the concept of "evidence-based medicine", statistics should establish standards and guidelines based on the results of well-done, neutral comparative studies and consensus from independent teams[22,106]. Though, currently, multiple biases such as over-optimism prevail in academia and produce inefficiency in knowledge building, and are likely to continue to do so in the future[23]. Of crucial note is that the problem of over-optimism is partly caused by the contemporary publication system[22,23]. Future valuable efforts may thus focus on how to change publication policies and the attitude of journals, editors, funding agencies, institutions, regulators, the public, referees or a combination thereof, and in the future, promote an efficient, self-correcting research process[135]. For researchers, there are several additional possibilities to minimize the impact of biases in academia without disrupting innovation and instead maximise efficient development of a credible knowledge base and a reliable corpus of published research[23,80,110,121,134-136]. Most important is to be aware of the presence of such biases[23,110]. Further, an important avenue for future research is to focus on of neutral comparison studies of existing studies, rather than only on the proposition of new statistical approaches. Future works should integrate the strategies and potential correctives proposed by other scholars and ensure that more stringent requirements with regards to transparent documentation are met. Also, it should be discussed how to optimize the formation and efficient co-working of collaborative consortia of scholars.

5 Contributing Publications

5.1 Statistical Modelling and Projections of Chronic Disease-Related Burden

5.1.1 Contribution 1 – Future Prevalence of Type 2 Diabetes—A Comparative Analysis of Chronic Disease Projection Methods

Contributing Article Voeltz, D., Tönnies, T., Brinks, R., & Hoyer, A. (2022). Future prevalence of type 2 diabetes—A comparative analysis of chronic disease projection methods. *Plos one*, *17*(3), e0264739.

Code and Data https://doi.org/10.1371/journal.pone.0264739

Supplementary Material https://doi.org/10.1371/journal.pone.0264739

Author Contributions D.V. is the corresponding author of this work. D.V. performed the analytic calculations, i.e., implemented the methods, scenarios and analysed the data. Further, D.V. took the lead in writing the manuscript and, together with A.H., was in charge of overall direction and planning. T.T., R.B. and A.H. supported the derivation of the methodology. All co-authors verified the analytical methods, the results and the main conceptual ideas All authors provided critical feedback and helped shape the research, analysis and manuscript.

PLOS ONE

RESEARCH ARTICLE

Future prevalence of type 2 diabetes—A comparative analysis of chronic disease projection methods

Dina Voeltz[1,2]*, **Thaddäus Tönnies**[3], **Ralph Brinks**[2,3,4], **Annika Hoyer**[1]

1 Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany, **2** Department of Statistics, Ludwig-Maximilians-University Munich, München, Germany, **3** Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany, **4** Chair for Medical Biometry and Epidemiology, Faculty of Health/School of Medicine, Witten/Herdecke University, Witten, Germany

* dina.voeltz@uni-bielefeld.de

OPEN ACCESS

Citation: Voeltz D, Tönnies T, Brinks R, Hoyer A (2022) Future prevalence of type 2 diabetes—A comparative analysis of chronic disease projection methods. PLoS ONE 17(3): e0264739. https://doi.org/10.1371/journal.pone.0264739

Editor: Bedreddine Ainseba, Universite de Bordeaux, FRANCE

Received: December 1, 2021

Accepted: February 15, 2022

Published: March 7, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0264739

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the articles of Tamayo et al. [23], Schmidt et al. [2] and their respective supplementary materials. Other data about the

Abstract

Background

Accurate projections of the future number of people with chronic diseases are necessary for effective resource allocation and health care planning in response to changes in disease burden.

Aim

To introduce and compare different projection methods to estimate the number of people with diagnosed type 2 diabetes (T2D) in Germany in 2040.

Methods

We compare three methods to project the number of males with T2D in Germany in 2040. Method 1) simply combines the sex- and age-specific prevalence of T2D in 2010 with future population distributions projected by the German Federal Statistical Office (FSO). Methods 2) and 3) additionally account for the incidence of T2D and mortality rates using partial differential equations (PDEs). Method 2) models the prevalence of T2D employing a scalar PDE which incorporates incidence and mortality rates. Subsequently, the estimated prevalence is applied to the population projection of the FSO. Method 3) uses a two-dimensional system of PDEs and estimates future case numbers directly while future mortality of people with and without T2D is modelled independently from the projection of the FSO.

Results

Method 1) projects 3.6 million male people with diagnosed T2D in Germany in 2040. Compared to 2.8 million males in 2010, this equals an increase by 29%. Methods 2) and 3) project 5.9 million (+104% compared to 2010) and 6.0 million (+116%) male T2D patients, respectively.

population projections used in this study are openly available from the FSO [20]. We provide our complete R-code and underlying data sets for reproducing the analysis in the supporting information.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Conclusions

The results of the three methods differ substantially. It appears that ignoring temporal trends in incidence and mortality may result in misleading projections of the future number of people with chronic diseases. Hence, it is essential to include these rates as is done by method 2) and 3).

Introduction

The increasing proportion of people suffering from chronic diseases is peculiarly worrying as these conditions constrain activities of daily living, necessitate ongoing medical attention and hence, are associated with considerably higher costs and healthcare expenses [1]. Disease management activities have been shown to reduce the risk of acute complications and premature mortality caused by chronic diseases [2]. Though, effective responses require accurate estimates of current and future chronic disease burden to tailor health care planning and resource allocation [3]. Worldwide, diabetes mellitus is one of the most frequent chronic diseases, and thus, is a disease with high public health relevance [1, 2, 4]. Current type 2 diabetes (T2D) prevalence is estimated with 7.4% among men and 7.0% among women in Germany aged 40 years or older [4, 5]. Besides severe late complications, diabetes mellitus leads to significantly higher mortality and is associated with 1.5 to 4.4 times higher health-care costs compared to people without diabetes [1].

Projection models are powerful tools to estimate future case numbers of a disease in order to inform decision-makers and cost-bearers in the health care system. Consequently, further developing and spreading the knowledge about accurate projection methods is essential to counteract the ever-worsening disease situation. However, different models may vary in their outcomes and closeness to reality [6]. Nonetheless, to our best knowledge, there is no scientific work that systematically compares these different projections methods for the context of chronic diseases. Therefore, the aim of this work is to fill this gap and to introduce, describe, and critically discuss each of the methods individually and in comparison.

In general, projection methods are limited by the availability of epidemiological and demographic data. Consequently, it is common to project and compare several scenarios that reflect on possible future trends in these areas [3, 4, 6–8]. Besides, the choice of the method can considerably affect projection results. In the context of chronic diseases, there are several methodological approaches that have been advocated for case number projections. For example, most reports are based on a 'status quo approach' [6] which relies on a simple application of the current prevalence to population projections. This procedure has been used for instance in the contexts of pulmonology [9], Parkinson's disease [10] and diabetes [11] among others. Probably due to its simplicity, this method is most popular in projection contexts. However, approaches that aim to incorporate for example underlying disease-specific transition rates, i.e., incidence and mortality, seem more appropriate as they better capture the complex nature of chronic diseases [12]. Therefore, some studies rely on multistate models that incorporate and relate disease-specific transition rates. In this regard, multistate models are widely used in infectious and chronic disease epidemiology [13–15]. For example, Milan and Fetzer [6], Brinks et al. [8, 16] and Waldeyer et al. [3] used time-discrete Markov models in the context of dementia, lupus and diabetes. Another approach is used by Carstensen et al. [7] who used a Poisson regression to model disease-specific transition rates of diabetes. Thereof, they extrapolate the future trends of the rates by extending the trends observed in the past. A relatively

novel approach is reported by Tönnies et al. [4], who use a partial differential equation to project future diabetes prevalence in Germany. There remains, however, considerable debate about the methodological approach.

In the present article, we give an overview of possible methods and underlying data used for future case number projections in the context of chronic conditions. For this purpose, we will examine three projection methods in more detail. We discuss how to employ each of the approaches in a practical application to project sex- and age-specific case numbers of people with diagnosed T2D for Germany between 2010 and 2040.

Methods

In the following section, we describe different models for projecting chronic disease case numbers. All methods were implemented using the free software R, v.4.1.0 (The R Foundation for Statistical Computing).

We focus on three different approaches to project chronic disease case numbers. Previous studies have mostly used a very simple approach, commonly referred to as status quo method [6]. Due to its popularity, we include this procedure in our work, speaking of it as method 1). This approach solely relies on the age- and sex-specific prevalence from a base year, which is then applied to population projections. Other epidemiological factors, such as the incidence, are only incorporated implicitly in the prevalence. Thus, this method ignores the fact that prevalence is a consequence of incidence and mortality. Hence, it might be too simplistic to accurately mirror reality. The alternative methods 2) and 3) rely on demographic components as well as on various disease-specific information on prevalence, incidence, and mortality rates as input factors. Method 2) is aligned with the work of Tönnies et al. [4], who takes advantages of the theory of multistate models in chronic disease epidemiology and an associated partial differential equation (PDE). With method 3) we present a novel projection method, which consists of a two-dimensional system of PDEs. The theoretical background for the PDEs used with method 2) and 3) originates in the illness-death model (IDM) as depicted in Fig 1. The IDM is a multistate model that represents continuous-time stochastic processes. Thereby, it allows individuals to move between a finite number of states [17, 18]. The classical IDM consists of three states, i.e., "healthy" (number of healthy people aged a at time t $H(t, a)$) with regards to the disease of interest, the disease state "ill" (number of ill people $I(t, a)$) and the death state, i.e., "dead" (D). It is assumed that at birth all individuals start in the healthy state. From there on, they can either be diagnosed with a chronic disease like T2D and then die at some point in time, or they can transition directly to death state (without contracting diabetes). The arrows indicate the transition rates between the states which depend on age and time. Since diabetes is a chronic condition, we assume that there is no remission from the chronic

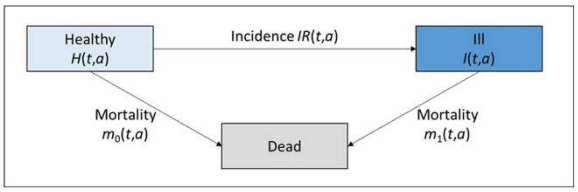

Fig 1. Illness-death model. All people in a population are in one of the three states: Healthy, Diseased, or Dead. It is assumed that at birth, all people start in the healthy state. Depending on time t and age a of each respective person, they will then transition to another state which is described by the IR, m_0, and m_1.

https://doi.org/10.1371/journal.pone.0264739.g001

condition back to the healthy state. The transition rates are given by the incidence rate $IR(t, a)$ the mortality of the non-diseased $m_0(t, a)$ and the mortality of diseased people $m_1(t, a)$, which are all sex-specific functions of calendar time t and age a. In epidemiological contexts, calendar time t is also denoted by period.

The following sections describe each of the three methods as sketched in Fig 2 and the required input data in more detail. Aggregated data about prevalence and incidence of the chronic condition of interest, as well as on the mortality of the general population and the excess mortality are sufficient, i.e., none of the methods require individual subject data.

Data

The starting point for the projections are disease-specific as well as demographic input factors. The required demographic information essentially comprises the expected age and sex-specific population distribution in Germany for each year that is to be included in the projection. The disease-specific input factors include the age- and sex-specific prevalence, incidence rate, and mortality rates of the diseased and non-diseased along with information on the age- and sex-specific excess mortality (mortality rate ratio, MRR).

We used published claims data on prevalence and incidence of diagnosed diabetes in 2009 and 2010 from 65 million people insured by the German public health insurance funds [19]. Diabetes was determined by the International Classification of Diseases-10 codes E10–E14. Overall, approximately 10.1% had any type of diabetes mellitus (excluding gestational diabetes), while 7.3% were diagnosed with T2D in 2010.

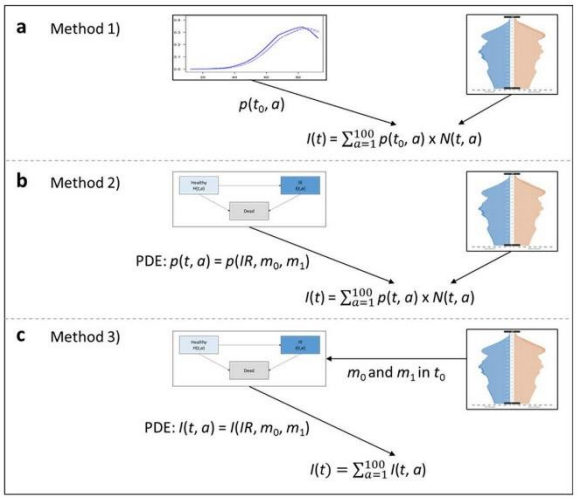

Fig 2. Overview of methods for future disease case projection. Illustrated in Fig 2A, method 1) uses the age-specific prevalence in base year t_0 which is applied to the population projections. Fig 2B depicts method 2). Using the theoretical background of the IDM, the age-specific prevalence for each year is derived by solving the PDE which requires input on the transition rates, namely the IR, m_0 and m_1. Method 3), as sketched in Fig 2C, calibrates m_0 and m_1 in the base year t_0 from the population projections. IR, m_0 and m_1 are inputs for the PDE which directly returns age-specific T2D case numbers for each year. Summing over all ages yields the total number of T2D cases for each year from 2010 to 2040 for each method.

https://doi.org/10.1371/journal.pone.0264739.g002

Concerning the population distribution, we used data from the German Federal Statistical Office (FSO) [20] that regularly issues updated projections of future population numbers in Germany. These projections include different scenarios regarding expected birth rate, life expectancy and migration. All projection scenarios contain sex-specific results for all ages from 0 to 100 years for the time horizon between 2010 and 2040. In our main analysis, we focus on one selected variant, namely B1L2M1. This variant assumes a birth rate (B) of 1.4 children per woman, a life expectancy (L) at birth in 2040 of 84.4 years for men and a long-term net migration (M) of 147,000 people. The motivation for this was twofold: First, we decided to consider more realistic variants instead of including more extreme options. Second, using variant B1L2M1 aligns with Tönnies et al. [4] who used the same variant for their projection. Nonetheless, in the supporting information we provide results for other variants as well, thereby showing that using different population projections of the FSO on the prevalence projection seems negligible.

Further, we obtain input values for the mortality rate of the general German population between 2010 and 2040 from the population projections of the German FSO [20]. Distinct information on the mortality rate of the healthy, i.e., non-diseased with regard to T2D, and for the mortality rate of the diseased would be of interest, but unfortunately, the mortality of the healthy population in Germany with respect to T2D is unknown. To cope with this lack of data, we substitute the missing mortality rates for method 2) with a mathematically equivalent expression based on the general mortality and the MRR as an alternative epidemiological measure. For method 3), we show how to calculate the mortality rates of the diseased and the healthy population.

The MRR is based on a similar, nationally representative dataset from 2014 reported by Schmidt et al. [2]. Unfortunately, the MRR is not differentiated by diabetes type. However, since T1D is frequent at ages younger than 20 and T2D is more common among older ages, the MRR is mostly driven by deaths among the latter. Moreover, most diabetes cases are attributable to T2D. Therefore, we use the MRR estimates provided by Schmidt et al. [2] as an approximation of the T2D-related MRR in Germany.

Further, reliable information on the temporal trend of the diabetes-specific MRR is relatively restricted in Germany. Following the work of Brinks et al. [8] and Tönnies et al. [4], we therefore refer to trends in the sex- and age-specific MRR observed in Denmark [21]. The motivation to do so is twofold. Firstly, it has been shown that for countries that are comparable in terms of their disease burden and health care systems, such as Denmark and Germany, the MRR settles in a similar range [22]. More precisely, a 2% decrease in the MRR per year is reported for Denmark [21]. Secondly, the same approach is used by Tönnies et al. [4], who also assume a decrease of 2% per year in the MRR as observed in other countries.

Method 1—The simple approach

Method 1) combines diabetes prevalence with population projections as depicted in Fig 2A. Specifically, to project the number of T2D cases until 2040, we multiplied German age- and sex-specific population projections provided by the FSO with age- and sex-specific T2D prevalence in 2010 from Tamayo et al. [23]. The latter is assumed to remain constant. The age- and sex-specific prevalence in 2010, $p(t, a)$ is determined by:

$$p(2010, a) = \frac{I(2010, a)}{H(2010, a) + I(2010, a)} \tag{1}$$

with $I(2010, a)$ being the number of people diagnosed with T2D in 2010 and $H(2010, a)$ the number of people without T2D in 2010. The underlying mathematical relation to calculate the

age- and sex-specific future number of cases $I(t, a)$ is then given by:

$$I(t,a) = N(t,a) \times p(2010,a) \tag{2}$$

where $N(t, a)$ denotes the age-, sex- and time-dependent total population number and $p(2010, a)$ the age-specific prevalence in the year 2010.

Method 2—The two-step multistate model

Method 2) refines the first method. To this end, we use mathematical relations to incorporate the relation between prevalence, incidence, and mortality, as well as temporal changes in the incidence and mortality [24]. More precisely, with method 2) we firstly model the temporal change in prevalence of T2D employing a PDE [25]. Secondly, we compute future T2D case numbers by multiplying the age- and sex-specific projected prevalence with the respective projected population size (Fig 2B).

Brinks et al. [26] showed that the change in prevalence can be modelled by a PDE in case information about mortality, incidence and prevalence at a specific time point is given. The PDE is given by

$$\partial p = (1-p) \times [IR - p \times (m_1 - m_0)] \tag{3}$$

In other words, method 2) relies on the relation between prevalence $p(t, a)$, the incidence rate $IR(t, a)$ and the mortality rates $m_0(t, a)$ and $m_1(t, a)$. Unfortunately, $m_0(t, a)$ is unknown for most diseases. Therefore, we use $m(t, a)$ the mortality of the general population, and the age- and sex-specific $MRR(t, a)$ which denotes the ratio of the two mortality rates, i.e. $\frac{m_1(t,a)}{m_0(t,a)}$. The general mortality is defined as

$$\mathrm{m} = \mathrm{p} \times m_1 + (1-p)m_0 \tag{4}$$

The PDE can then be rewritten as

$$\partial p = (1-p) \times \left[IR - \frac{p \times (MRR-1) \times m}{p \times (MRR-1) + 1}\right] \tag{5}$$

For the present application to the context of T2D, this allows us to project the change in prevalence at a certain age and time even without data about $m_0(t, a)$. Accordingly, we integrate the PDE shown in Eq (5). Doing so, we use nationally representative input values for the projected general mortality of the German population as provided by the FSO and the observed age-specific prevalence of T2D in year 2010 among all individuals within the German statutory health insurance as initial prevalence. Recall that the equation is particularly favourable since the PDE allows for incorporating the complex interplay of the IR and MRR, as well as it is able to reflect on temporal trends of these rates. By application and integration of the PDE we derive the estimated age- and sex-specific prevalence $\hat{p}(t,\ a)$ for all following years considered in the time horizon of the projection, i.e., here from 2010 to 2040. In a final step, this estimated prevalence $\hat{p}(t,\ a)$ is multiplied with the population projections of the FSO [20] as follows

$$I = N \times \hat{p} \tag{6}$$

This yields case numbers of T2D in Germany $I(t, a)$ for each year between 2010 and 2040 for all ages ranging from 18 to 100 years. All ages below 18 years were considered negligible due to minimal prevalence at these ages [4, 23].

Method 3—The two-dimensional PDE model

Similar to method 2), the third method builds upon the IDM. Though, method 3) features two PDEs as sketched in Fig 2C to directly express changes in the age- and sex-specific numbers of diabetes cases, i.e., $I(t, a)$, and the changes in the healthy population $H(t, a)$. We can show that these two figures $I(t, a)$ and $H(t, a)$ are interrelated with the transition rates $IR(t, a)$, $m_0(t, a)$ and $m_1(t, a)$ via the following equations

$$\partial H = -(IR + m_0) \times H \tag{7}$$

$$\partial I = IR \times H - m_1 \times I \tag{8}$$

$\partial H(t, a)$ indicates age-, sex- and time-specific changes in the number of people without diabetes, while $\partial I(t, a)$ represents changes in the number of cases. As $m_0(t, a)$ is unknown for Germany, we use the T2D prevalence from 2010, the general mortality $m(t, a)$ and the $MRR(2014, a)$ to derive values for $m_0(t, a)$ and $m_1(t, a)$ [21, 22, 27]. Precisely, $m_0(t, a)$ and $m_1(t, a)$ are calculated by

$$m_0 = \frac{\mathrm{m}}{(1 + \mathrm{p} \times (MRR - 1))} \tag{9}$$

$$m_1 = m_0 \times MRR \tag{10}$$

In contrast to method 2), we only need data on the general mortality and MRR in 2010 for method 3) which is then used to determine $m_0(t, a)$ and $m_1(t, a)$. Besides, we derive information on the development of the future birth rate of the population projections. However, using only information on the annual number of new-borns is little critical for our projection period, as the projection ends in 2040. That means, someone born in 2020 will then be 20 years old and will thus not belong to the diabetes risk group. That means we do not need further information from population projections that are based on historical data and thus subject to retrograde developments as provided by the FSO [20]. This is favourable as these population projections do not explicitly take specific diseases into account. In contrast, using method 3) and a two-dimensional system of PDEs allows us to reflect on short-term changes especially with regards to relevant disease-specific alterations. As outcome, the PDEs directly describe the age- and sex-specific number of people with and without T2D for each year between 2010 and 2040.

Results

Based on the age-specific prevalence in 2010, we projected the age-specific case numbers for T2D in the German male population between 2010 and 2040 using the above-mentioned approaches. Remember that we assume a constant age-specific prevalence with method 1), i.e., we apply the prevalence from 2010 to our population projections. We further assumed that also the IR remains constant over time for method 2) and 3) in the main analysis. However, we consider temporal trends in the MRR for method 2) and 3) which affects, amongst others, the prevalence. Since for Germany, information on long-term trends in the latter are limited, we use speculative time-related developments. Current evidence undermines that the mortality rate among people with T2D is likely to decrease faster than among healthy people due to progresses in medical care [21]. Therefore, it seems plausible to assume a reduction in the MRR [28]. For simplicity, we restrict the main analysis to a baseline scenario for method 2) and 3) which thus comprises a constant IR and a decrease in the MRR of 2% per year. We reflect on

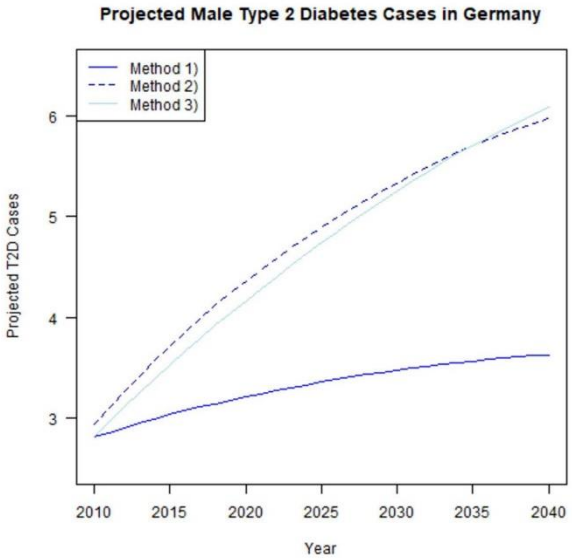

Fig 3. T2D case projection results. Projected male T2D cases in Germany for variant B1L2M1 of the population projections of the German FSO between 2010 and 2040. The population projection variant is based on constant birth rates, increases in life expectancy and a net migration of 147 000 people. Method 1) assumes constant age specific prevalence, method 2) and method 3) build on constant incidence rate and 2% decrease in the MRR.

https://doi.org/10.1371/journal.pone.0264739.g003

rather likely, as well as on relatively extreme scenarios for the MRR and IR in the sensitivity analyses discussed in the supporting information.

Fig 3 depicts the number of projected male T2D cases in Germany between 2010 and 2040 using the three different approaches. All the proposed methods suggest that the number of people with diabetes continues to grow in Germany. However, the projected number of male T2D cases differs substantially across the three methods. It is clearly visible that over time the difference between the projected values from each of the three methods enlarges. Since only the population composition changes over time, method 1) shows only slightly increasing future case numbers. In contrast, methods 2) and 3) also take into account the incidence and mortality, which influence the prevalence and number of cases. This, in addition to the demographic aging, is above all the reason for the higher number of projected T2D cases for these methods compared to method 1).

Table 1 displays the projected T2D case numbers among men in Germany in millions and the relative changes from 2010 compared to 2040 in percent. With method 1), we find an

Table 1. T2D case projection results (in million).

Method	2010	2020	2030	2040	Absolute difference	Relative difference
1)	2.81	3.21	3.47	3.63	0.82	29.09%
2)	2.93	4.36	5.33	5.97	3.04	103.60%
3)	2.81	4.16	5.25	6.08	3.27	116.34%

Results are based on variant B1L2M1 of the German FSO (absolute (in Million) and relative (in percent) difference with regards to 2010 vs. 2040).

https://doi.org/10.1371/journal.pone.0264739.t001

increase of male T2D case numbers by 0.8 million. This equals a total of almost 3.6 million T2D cases among males in Germany in 2040 (+29%). When presuming that the MRR decreases by 2% per year while the IR is assumed to remain constant, the number of cases is projected to increase by about 3.0 million (+104%) and 3.3 million (+116%) for method 2) and method 3), respectively. Apparently, the results show an obvious gap between method 1) and the other two methods. Interestingly, the difference between the projected case numbers of the latter methods is minor: During the entire time horizon of the projection until 2035, the future number of T2D cases for method 2) is slightly higher than the number projected by method 3). Thereafter, method 3) projects a higher number of male T2D cases in Germany.

S1 and S2 Figs in S2 File refer to the age-specific prevalence among males in Germany from 2010 to 2040 projected using the different methods. We find apparent discrepancies in the age-specific prevalence for the methods. Compared to the constant prevalence assumed by method 1), the prevalence computed with methods 2) and 3) is expected to increase considerably, particularly for people older than 60 years. Additionally, the peak prevalence is presumed to shift towards older age groups. We provide further details in the supporting information.

Sensitivity analyses

We performed several sensitivity analyses to assess inaccuracies and uncertainty in our results which we briefly discuss in this section. First, we assessed uncertainty due to unknown future trends in the model parameters. In order to inspect uncertainty that could arise due to sampling error of the input values, we perform a Monte Carlo simulation for all three methods. To evaluate the latter, i.e., uncertainty due to future trends, we analyse 14 different scenarios of declining or rising MRR and IR with methods 2) and 3) following the example of Tönnies et al. [4]. More detailed information on results and computation is given in the supplementary information.

General future epidemiological trends for the age-specific IR or the MRR are heterogeneous and precise information are not available for Germany [4, 5]. Therefore, we considered 14 different scenarios of declining or rising MRR and IR and examined the impact on future case numbers using method 2) and 3). We find considerably large deviations in the number of projected T2D cases: With the most extreme scenarios we project a decrease of approx. 0.3 million (-11%; stable MRR & IR -5%) or an increase of 6.3 million (216%; MRR -2% & IR +3%) male T2D cases between 2010 and 2040 for Germany (S3 Fig in S2 File). This is an essential finding, as it underlines how impactful disease-specific factors are in projection contexts. This result is strongly in favour of method 2) and 3), as these approaches incorporate the rates underlying the prevalence in the mathematical model.

In order to assess uncertainty in our results, we calculated 95% confidence intervals using a Monte Carlo simulation as shown in S4 Fig in S2 File. The sampling error of the input values led to deviations in the future number of cases by approximately 7.2% for method 2) and by about 3.5% for method 3). Hence, the Monte Carlo simulation shows that the uncertainty due to sampling error in the input values seems to be of limited relevance. We report further information about the calculation and the estimation results in the supporting information.

Discussion

The aim of this article is to assess and advance the usage and performance of different projection methods in chronic disease modelling. While there is a lack thereof in previous publications, we compared the underlying assumptions, mathematical details, strengths, and limitations of three distinct projection methods and demonstrated their usage. Precisely, in this article, we discussed three methods to project age- and sex-specific case numbers of a

chronic condition from age- and sex-specific prevalence, incidence, and mortality data. Further, we illustrated each method in a practical application in the context of T2D among males in Germany for the time period from 2010 to 2040. We found considerable differences between the results of the three methods.

Generally, all three methods can be easily adapted to other chronic diseases or countries. Additionally, for Germany, data about the future population number is available until 2060, i.e., the projection of T2D case numbers can be extended easily to the far-off future. Nonetheless, a projection is subject to some unforeseen changes in the economic, political, or disease-specific environment amongst others. Generally, the reliability of projections decreases over time. Therefore, instead of delving further into the future and thereby decreasing certainty, we restricted the projection until 2040.

In many previous projections on chronic diseases, age- and sex-specific prevalence from a base year is transferred to population projections as in method 1) [6, 9–11]. Population projections required for this purpose, generated using the so-called cohort component method [6, 29], are widely used and publicly available in Germany. For the calculation, the birth cohorts are hypothetically extrapolated from a base year under assumptions on the future development of fertility, mortality, and net migration. But just as the future population depends on several components, the development of a disease is also influenced by various disease-specific aspects [6]. However, the latter are ignored in method 1). Without a doubt, method 1) is the least reliable of the three discussed approaches. Assuming a constant prevalence and not reflecting on changes in disease-specific factors substantially limit the reliability of this method's results. Unfortunately, in some contexts, data on the latter may be rarely available. If that is the case, method 1) could at least provide a short-term indication of future case numbers. Furthermore, short-term projections could reflect a second context in which method 1) may be applicable. A projection is always subject to some unforeseen changes in the economic, political, or disease-specific environment amongst others. Consequently, when keeping the overall time-horizon to a minimum, changes in relevant factors such as the mortality rate ratio or the incidence rate may be minimal and therefore less impactful, too. Nonetheless, method 1) should be used with caution and whenever possible, method 2) or 3) should be the preferred options. Methods 2) and 3) include disease-specific information on mortality, prevalence, and incidence rates as input factors in addition to the demographic components mentioned above. Thus, as opposed to method 1), the PDEs used in method 2) and 3) account for temporal trends in the prevailing epidemiological situation. To do so, aggregated data are sufficient. Method 2) incorporates the MRR and IR, as opposed to solely using the prevalence as is done by method 1). Though, method 2) is, as method 1), still based on projections of the population and the general mortality in Germany. The future estimates of the German FSO do not consider potential epidemiological influences and simply extrapolate future values based on long-term trends of the past [20]. Contrarily, using method 3), we can take short-term fluctuations in the mortality into account. The results of method 3) are independent from projections of the German FSO beyond the year used to estimate m_0 and m_1. Instead, this method purely relies on input values on the age- and sex-specific prevalence, the MRR and the general mortality for a base year. The latter are used to estimate the mortality of the non-diseased and the mortality of diseased people. Furthermore, instead of calculating the prevalence in each year for each age group, method 3) directly returns future changes in absolute case and population numbers.

Besides, some health actions or interventions may not directly alter the prevalence of a disease but still have far-reaching consequences with regards to a particular disease situation [28, 30, 31]. Consequently, other health indicators such as life expectancy, years of life lost, or years of productivity lost may provide a more appropriate insight to anticipate the potential impact of such interventions. Nonetheless, in most cases, a disease's prevalence is a necessary input

factor for the computation of other health indicators. Accordingly, using the illness–death model and the corresponding PDE in a first step to estimate a future prevalence allows to derive other health characteristics of a population associated with T2D in Germany in 2020 and 2040 from an individual and population perspective in a second step [28, 31]. In fact, this reflects another drawback of method 1). In contrast to method 2) and 3), the approach assumes a stable prevalence, and consequently, this does not allow to easily infer the future development of other epidemiological indicators.

Lastly, different future scenarios that anticipate potential impacts of alternative scenarios of prevention and intervention or disease-dynamics can be computed and compared using method 2) or 3). The comparison of methods 2) and 3) with method 1) reveals the great impact of trends in incidence and mortality on future disease burden. Our findings confirm the suggestion that ignoring temporal trends in incidence and mortality provokes an underestimation of the actual number of cases [4]. Consequently, a future course can be better, if not optimal, approximated using method 2) and 3).

As a further remark, we are convinced that our work provides a good basis for future research. Our study discusses the large importance of choosing a suitable projection method which is supported by a practical application to the context of T2D in Germany. Since our approach to use a PDE that describes the IDM is not the only one that has been proposed and published [31, 32], it would be interesting to extend the comparison and focus on different approaches that are based on the IDM. Since this would have been beyond the scope of this article, this investigation could prove desirable for future work.

Limitations

A primary limitation for method 2) and 3) is that they need assumptions about the temporal changes in the IR and MRR. Unfortunately, information about future trends in Germany in the context of T2D is scarce. Therefore, we made assumptions about future trends of the IR and MRR that may be oversimplified. Further, it is noteworthy that our population projection methods are trend-based, hence, they may be less accurate for instance in periods of sudden and fast changes in the incidence. However, also the intensity, frequency and pace of changes have implications on epidemiologically relevant measures such as the prevalence and the number of cases. As discussed, variant projections, which are feasible with method 2) and 3), can provide additional information by illustrating potential alternative scenarios of future trends. Aside from that, rapid changes in chronic disease incidences are rare and the general impact of high-frequency distortions in future trends is often limited by the inertia in population change.

Another limitation of all three methods arises with a potential violation of the assumption that the prevalence in migrants is equal to the prevalence observed among German residents. Though, the T2D prevalence in people migrating from and to Germany is unknown. However, other examples show that this issue is minor and overall epidemiological measures are only negligibly affected [4, 25]. Though, one should keep in mind that it may remain an issue e.g., for small populations.

Another weakness of the methods arises with the ignored effect of potential covariates and their development. Examples for relevant covariates in the diabetes context might be a change of diagnostic criteria for T2D, the distribution of body weight, the impact of nutritional behaviour, or the presence of co-morbidities. In an epidemiological context, it may be essential to consider such covariates since they likely modify the transition rates between the states in the IDM. Although this is not done in this work, it is possible to account for the impact of possible covariates such as interventions and risk-factors in the PDEs [33].

Conclusion

Forecasts of the growing non-communicable disease burden will be key to guide future healthcare policies. With this work, we compare three projection methods and demonstrated how to apply each of them to quantify future T2D case numbers in Germany until 2040. The three methods uniformly confirm that there is a substantial increase in the number of males diagnosed with T2D ranging from 0.8 million (+29%) to 3.3 million (+116%) additional cases in Germany in 2040. Assessing the strengths and limitations of three different methods may help researchers to better apply statistical methods for projecting future case numbers of chronic diseases. We suggest that future projections should move away from blunt prevalence extrapolation and instead, employ methods that are based on theory from the IDM.

Supporting information

S1 File. R code.
(PDF)

S2 File.
(DOCX)

Author Contributions

Conceptualization: Annika Hoyer.

Methodology: Dina Voeltz, Ralph Brinks, Annika Hoyer.

Writing – original draft: Dina Voeltz.

Writing – review & editing: Dina Voeltz, Thaddäus Tönnies, Ralph Brinks, Annika Hoyer.

References

1. Jacobs E, Hoyer A, Brinks R, Icks A, Kuß O, Rathmann W. Healthcare costs of type 2 diabetes in Germany. Diabet Med 2017; 34(6):855–61. https://doi.org/10.1111/dme.13336 PMID: 28199029
2. Schmidt C, Reitzle L, Heidemann C, Paprott R, Ziese T, Scheidt-Nave C, et al. Excess mortality in adults with documented diabetes in Germany: routine data analysis of all insurance claims in Germany 2013–2014. BMJ Open 2021; 11(1):1–8. https://doi.org/10.1136/bmjopen-2020-038889corr1 PMID: 33472795
3. Waldeyer R, Brinks R, Rathmann W, Giani G, Icks A. Projection of the burden of type 2 diabetes mellitus in Germany: a demographic modelling approach to estimate the direct medical excess costs from 2010 to 2040. Diabet Med 2013; 30(8):999–1008. https://doi.org/10.1111/dme.12177 PMID: 23506452
4. Tönnies T, Röckl S, Hoyer A, Heidemann C, Baumert J, Du Y, et al. Projected number of people with diagnosed Type 2 diabetes in Germany in 2040. Diabet Med 2019; 36(10):1217–25. https://doi.org/10.1111/dme.13902 PMID: 30659656
5. Goffrier B, Schulz M, Bätzing-Feigenbaum J. Administrative Prävalenzen und Inzidenzen des Diabetes mellitus von 2009 bis 2015: Versorgungsatlas-Bericht. Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi) 2017(17):1–34.
6. Milan V, Fetzer S. Die zukünftige Entwicklung von Demenzerkrankungen in Deutschland–ein Vergleich unterschiedlicher Prognosemodelle. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2019; 62(8):993–1003. https://doi.org/10.1007/s00103-019-02981-3 PMID: 31243489
7. Carstensen B, Rønn PF, Jørgensen ME. Components of diabetes prevalence in Denmark 1996–2016 and future trends until 2030. BMJ Open Diabetes Res Care 2020; 8(1):1–7. https://doi.org/10.1136/bmjdrc-2019-001064 PMID: 32784246
8. Brinks R, Tamayo T, Kowall B, Rathmann W. Prevalence of type 2 diabetes in Germany in 2040: estimates from an epidemiological model. Eur J Epidemiol 2012; 27(10):791–7. https://doi.org/10.1007/s10654-012-9726-2 PMID: 22878939
9. Pritzkuleit R, Beske F, Katalinic A. Erkrankungszahlen in der Pneumologie—eine Projektion bis 2060. Pneumologie 2010; 64(9):535–40. https://doi.org/10.1055/s-0030-1255637 PMID: 20827635

10. Rossi A, Berger K, Chen H, Leslie D, Mailman RB, Huang X. Projection of the prevalence of Parkinson's disease in the coming decades: Revisited. Mov Disord 2018; 33(1):156–9. https://doi.org/10.1002/mds.27063 PMID: 28590580

11. Boyle JP, Honeycutt AA, Narayan KV, Hoerger TJ, Geiss LS, Chen H, et al. Projection of diabetes burden through 2050: impact of changing demography and disease prevalence in the U.S. Diabetes Care 2001(24):1936–40. https://doi.org/10.2337/diacare.24.11.1936 PMID: 11679460

12. Carstensen B, Kristensen JK, Ottosen P, Borch-Johnsen K. The Danish National Diabetes Register: trends in incidence, prevalence and mortality. Diabetologia 2008; 51(12):2187–96. https://doi.org/10.1007/s00125-008-1156-z PMID: 18815769

13. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london 1927(115):700–21.

14. Fix E, Neyman JA. A simple stochastic model of recovery, relapse, death and loss of patients. Human Biology 1951(23):205–241.

15. Keiding N. Age-specific incidence and prevalence: a statistical perspective. Royal Statistical Society 1991(154):371–412.

16. Brinks R, Fischer-Betz R, Sander O, Richter JG, Chehab G, Schneider M. Age-specific prevalence of diagnosed systemic lupus erythematosus in Germany 2002 and projection to 2030. Lupus 2014; 23(13):1407–11. https://doi.org/10.1177/0961203314540352 PMID: 24928831

17. Brinks R. Illness-Death Model in chronic disease epidemiology: characteristics of a related, differential equation and an inverse problem. Comput Math Methods Med 2018; 2018:1–6.

18. Murray CJL, Lopez AD. The global burden of disease. Cambridge: Harvard University Press. 27 p.

19. Deutsches Institut für Medizinische Dokumentation und Information (DIMDI). Aktuell Informationssystem Versorgungsdaten [Internet] [cited 2015 Aug 11]. Available from: www.dimdi.de/static/de/versorgungsdaten/aktuelles/news_0364.html_ 319159480.html.

20. Federal Statistical Office of Germany. 14. Coordinated population projection [Internet]; 2019 [cited 2022 Jan 21]. Available from: https://www.destatis.de/EN/Themes/Society-Environment/Population/Population-Projection/_node.html#sprg479856.

21. Carstensen B, Rønn PF, Jørgensen ME. Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996–2016. BMJ Open Diabetes Res Care 2020; 8(1):1–9. https://doi.org/10.1136/bmjdrc-2019-001071 PMID: 32475839

22. Brinks R, Tönnies T, Hoyer A. Impact of diagnostic accuracy on the estimation of excess mortality from incidence and prevalence: simulation study and application to diabetes in German men. F1000Res 2021; 10:1–16. https://doi.org/10.12688/f1000research.27868.3 PMID: 34316353

23. Tamayo T, Brinks R, Hoyer A, Kuß OS, Rathmann W. The prevalence and incidence of diabetes in Germany. Dtsch Arztebl Int 2016; 113(11):177–82. https://doi.org/10.3238/arztebl.2016.0177 PMID: 27118665

24. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data, Vol 360: John Wiley & Sons; 2011.

25. Brinks R, Landwehr S. Age-and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. Theoretical population biology 2014; 92:62–8. https://doi.org/10.1016/j.tpb.2013.11.006 PMID: 24333220

26. Brinks R, Tönnies T, Hoyer A. New ways of estimating excess mortality of chronic diseases from aggregated data: insights from the illness-death model. BMC Public Health 2019; 19(1):1–7. https://doi.org/10.1186/s12889-018-6343-3 PMID: 30606151

27. Brinks R, Hoyer A, Weber S, Fischer-Betz R, Sander O, Richter JG, et al. Age-specific and sex-specific incidence of systemic lupus erythematosus: an estimate from cross-sectional claims data of 2.3 million people in the German statutory health insurance 2002. Lupus Sci Med 2016; 3(1):1–3. https://doi.org/10.1136/lupus-2016-000181 PMID: 27933200

28. Tönnies T, Hoyer A, Brinks R. Productivity-adjusted life years lost due to type 2 diabetes in Germany in 2020 and 2040. Diabetologia 2021; 64(6):1288–97. https://doi.org/10.1007/s00125-021-05409-3 PMID: 33665686

29. Whelpton PK. An empirical method of calculating future population. Journal of the American Statistical Association 1936; 31(195):457–73.

30. Tönnies T, Baumert J, Heidemann C, Lippe E, Brinks R, Hoyer A. Diabetes free life expectancy and years of life lost associated with type 2 diabetes: projected trends in Germany between 2015 and 2040. Population health metrics 2021; 19(1):1–10. https://doi.org/10.1186/s12963-020-00244-x PMID: 33413469

31. Wanneveich M, Jacqmin-Gadda H, Dartigues JF, Joly P. Impact of intervention targeting risk factors on chronic disease burden. Statistical Methods in Medical Research 2018; 27(2):414–27. https://doi.org/10.1177/0962280216631360 PMID: 26988925

32. Arthur KC, Calvo A, Price TR, Geiger JT, Chiò A, Traynor BJ. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat Commun 2016; 7:1–6. https://doi.org/10.1038/ncomms12408 PMID: 27510634

33. Hoyer A, Kaufmann S, Brinks R. Risk factors in the illness-death model: Simulation study and the partial differential equation about incidence and prevalence. PLoS One 2019; 14(12):1–10. https://doi.org/10.1371/journal.pone.0226554 PMID: 31846478

5.1.2 Contribution 2 – Future Number of People with diagnosed Type 1 Diabetes in Germany until 2040: An Analysis based on Claims Data

Contributing Article 2. Voeltz, D., Brinks, R., Tönnies, T., Hoyer, A. Future number of people with diagnosed type 1 diabetes in Germany until 2040: an analysis based on claims data. *BMJ Open Diabetes Research and Care*. 2023; 11(2), e003156.

Code and Data 10.5281/zenodo.6799292

Supplementary Material https://drc.bmj.com/content/bmjdrc/11/2/e003156.full.pdf?with-ds=yes

Author Contributions D.V. is the corresponding author of this work. As such, D.V. performed the analytic calculations, i.e., implemented the methods and scenarios and analysed the data. Further, D.V. took the lead in writing the manuscript and, together with A.H., was in charge of overall direction and planning. All co-authors verified the analytical methods, the results and the main conceptual ideas. All authors provided critical feedback and helped shape the research, analysis and manuscript.

BMJ Open Diabetes Research & Care

Future number of people with diagnosed type 1 diabetes in Germany until 2040: an analysis based on claims data

Dina Voeltz [1,2] Ralph Brinks,[3,4] Thaddäus Tönnies [3] Annika Hoyer [1]

To cite: Voeltz D, Brinks R, Tönnies T, *et al*. Future number of people with diagnosed type 1 diabetes in Germany until 2040: an analysis based on claims data. *BMJ Open Diab Res Care* 2023;**11**:e003156. doi:10.1136/bmjdrc-2022-003156

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjdrc-2022-003156).

Received 29 September 2022
Accepted 24 March 2023

[1]Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany
[2]Department of Statistics, Ludwig Maximilians University Munich, Munchen, Germany
[3]Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany
[4]Chair for Medical Biometry and Epidemiology, Witten/Herdecke University, Faculty of Health/School of Medicine, Witten, Germany

Correspondence to
Dr Dina Voeltz;
dina.voeltz@uni-bielefeld.de

ABSTRACT

Introduction We aim to project the number of people with diagnosed type 1 diabetes in Germany between 2010 and 2040.

Research design and methods We first estimate the age-specific and sex-specific incidence and prevalence of type 1 diabetes in Germany in 2010 using data from 65 million insurees of the German statutory health insurance. Then, we use the illness-death model to project the prevalence of type 1 diabetes until 2040. We alter the incidence and mortality underlying the illness-death model in several scenarios to explore the impact of possible temporal trends on the number of people with type 1 diabetes.

Results Applying the prevalence from 2010 to the official population projections of Germany's Federal Statistical Office yields a total number of 252 000 people with type 1 diabetes in Germany in 2040 (+1% compared with 2010). Incorporating different annual trends of the incidence and mortality in the projection model results in a future number of people with type 1 diabetes between 292 000 (+18%) and 327 000 (+32%).

Conclusions For the first time in Germany, we provide estimates for the incidence, prevalence, and number of people with diagnosed type 1 diabetes for the whole German population between 2010 and 2040. The relative increase of the people with type 1 diabetes ranges from 1% to 32% in 2040 compared with 2010. The projected results are mainly influenced by temporal trends in the incidence. Ignoring these trends, that is, applying a constant prevalence to population projections, probably underestimates future chronic disease numbers.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The few studies investigating the current and future population with type 1 diabetes in Germany are limited to certain ages, cohorts, and ignore temporal trends in disease-related rates.

WHAT THIS STUDY ADDS

⇒ Based on data from 65 million insurees of all German statutory health insurances, we estimate an increase in the German population with type 1 diabetes ranging from 252 000 (+1% compared with 2010) to 327 000 (+32%) people in 2040 which is mainly driven by the incidence.

⇒ The peak of the age-specific prevalence is projected to shift toward older ages.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Germany will possibly face greater demand for diabetes-related education, healthcare, and medical resources.

INTRODUCTION

Worldwide, chronic disease burden rises dramatically.[1] As one of the most common chronic diseases in Germany, diabetes mellitus causes suffering among diagnosed people as well as tremendous costs for the overall population.[1–4] Without a structured self-management plan, monitoring blood glucose level, physical activity, diet and, most important, without daily insulin treatment, people with type 1 diabetes (T1D) are at high risk of premature death.[5] Although defined as autoimmune disease with an onset at ages younger than 35 years,[6 7] the disease is nowadays prevalent among the whole age range.[1 7 8] In 2009 in Germany, estimates of the prevalence of T1D for girls and boys younger than 19 years were 0.17% to 0.19%, respectively. The prevalence among women and men aged between 20 and 79 years was estimated between 0.28% and 0.39% and among elderly, that is, older than 80 years, between 0.47% and 0.50%.[2 9] Reports of the total number of people with T1D in Germany range from 256 000 to 373 000 in 2009 and 2016, respectively.[2 9] Approximately 7200 people per year are newly diagnosed.[2] Similar to trends observed over the past decades, the T1D incidence and prevalence are expected to rise further.[1]

Despite the high relevance of T1D and its considerable health and economic consequences,[5] representative studies that investigate the population-wide current and future disease burden in Germany are sparse.[1 2] The generalizability of previous studies to Germany is impaired due to their focus on particular age ranges and/or their restriction to data from selected health

insurances, and cohorts.[10 11] Additionally, although there is evidence that diabetes type-specific interventions help in reducing complications and mortality,[6] many existing analyses do not distinguish the diabetes types.[11 12] This makes the need for type-specific analyses of the current and future burden of T1D even more compelling.[3]

In this article, we provide nationally representative estimates of the age-specific and sex-specific T1D incidence in Germany in 2010. Furthermore, we project the age-specific and sex-specific prevalence and the number of people diagnosed with T1D from 2010 to 2040.

METHODS

Based on published data from 65 million insurees of the German statutory health insurance,[8 9 13 14] we first estimated the age-specific and sex-specific incidence of diagnosed T1D in Germany in 2010 for all ages between 0 and 100 years. Using mathematical relations between prevalence, incidence, and mortality based on the illness-death model (IDM), we projected the future age-specific prevalence of T1D between 2010 and 2040. In this second step, we compared several scenarios regarding possible temporal trends in the incidence and mortality. To calculate the number of people with diagnosed T1D in each year, we finally multiplied the projected prevalence to the projected distribution of the German population issued by the German Federal Statistical Office (FSO).[14] The following paragraphs provide more details about these three steps.

All methods were implemented using the free statistical software R, V.4.1.0 (The R Foundation for Statistical Computing). The source code and data for running the analysis are published in the open-access repository Zenodo.[15] We only used published data on aggregated level.

Definition of type 1 diabetes

To define T1D, we refer to the International Classification of Diseases (ICD) coding provided in the underlying claims data which is commonly used to specify a particular type of diabetes.[9 10 13] The aggregated prevalence data, provided by the German Institute for Medicine Documentation and Information in 2012, contains information from about 65 million policyholders from the inpatient and outpatient sector insured for at least 360 days per year by one of the German statutory health insurances. This represents >80% of Germany's overall resident population. Accordingly, the data can be considered a unique and suitable information source that enables nationwide representative estimates of diagnosed diabetes in Germany.[9 10] For instance, the Robert Koch Institute features the same data for estimating the overall incidence of diabetes in Germany.[16 17] Generally, the ICD-code E10.- specifies a T1D diagnosis. Since in some cases, the data report multiple diagnoses for the same individual, the combination of the ICD-codes E14.- (unspecified diabetes mellitus) and E10.- also resulted in the assignment to T1D. Any other ICD coding or unclear code combination were excluded.

Estimation of the age-specific T1D incidence in 2010

To estimate the age-specific and sex-specific T1D incidence in Germany in 2010, we applied a partial differential equation (PDE), which describes the relation between prevalence, incidence and mortality as a function of age and calendar time (see online supplemental material).[9 18] This requires prevalence input data for 2 years, information on the general mortality, and on the mortality rate ratio (MRR), that is, the ratio of the mortality rates of people with versus without T1D. The aforementioned claims data[9] provide aggregated, age-specific and sex-specific information on the number of people in Germany along with their respective diabetes status (ie, diagnosed with T1D vs no T1D) for the two consecutive years 2009 and 2010. Knowing that the prevalence is the proportion of a particular population found to be affected by a certain condition (here, T1D vs no T1D) at a specific time (here, 2009 and 2010), we analogously calculated the overall and the age-stratified T1D prevalence for each sex for 2009 and 2010 using our input data.[9] To do so, we divided the number of T1D diagnoses in Germany by the total German population number for both years on overall and age-specific level, respectively. We obtain the general mortality of the German population as of 2010 from the population projections of the German FSO.[14] Due to unavailability of the T1D-related MRR in Germany, we align with other diabetes-related work,[9 10] and refer to age-specific and sex-specific estimates from Denmark as the two countries' MRRs are claimed highly comparable.[8 9 19 20]

Due to inconsistencies in coding of diagnoses,[9] we approximated the course of the T1D incidence in Germany for ages 35+ years (see online supplemental material). For that purpose, and due to a lack of appropriate alternative data from Germany, we used information from the Danish National Diabetes Register. In Denmark, the T1D incidence rate in ages 35+ years decreases logarithmically with age. This resembles the German diabetes burden estimated in previous studies,[12] where the proportion of T1D among all diabetes types peaks among the youth and decreases with increasing age. Since the erroneous coding occurred among older ages and since often latent autoimmune diabetes in adults (LADA) commonly showing in patients over 35 years is diagnosed as T1D, the chosen age cut-off seems reasonable.[21]

Projection of the age-specific T1D prevalence and future number of people with T1D

We use the aforementioned PDE to project the age-specific and sex-specific prevalence of T1D in Germany between 2010 and 2040 (see online supplemental material). Solving the PDE returns the age-specific prevalence for each year depending on the incidence and MRR.[18] Since information on current and future trends

of the incidence and mortality are scarce in Germany,[10] we assumed five hypothetical scenarios that have been discussed in previous projections of T1D and T2D in Germany.[1 3 10 22] Scenario 1, a 'simple' sex-specific and age-specific prevalence extrapolation, assumes a constant age-specific prevalence between 2015 and 2040 and neither reflects on the incidence nor on the mortality. In scenarios 2–5, we alter the temporal trends of the incidence and MRR. We consider scenario 2 as baseline scenario, assuming constant incidence and mortality. In the future, it is likely that the mortality rate among people with T1D will decrease faster than among people without T1D due to probable advancements in medical care.[10 20] Consequently in scenarios 3–5, we anticipate that the MRR decreases by 2% per year as observed in Denmark.[10 20] Furthermore, scenario 3 assumes a constant incidence rate, while scenarios 4 and 5 assume an annual increase and decrease of 0.5%, respectively.

We obtain the number of people with T1D in Germany between 2010 and 2040 for all ages between 0 and 100 by multiplying the projected prevalence from the PDE with the number of people from the population projections of the FSO.[14] The FSO releases several variants wherein they assume different future birth rates, life expectancies at birth, and migration. We consider a rather realistic version (variant B1L2M1) which assumes declining fertility (B1: birth rate of 1.4 children per woman), moderate development in life expectancy (L2: life expectancy of 84.4 and 88.1 years for men and women, respectively), and low long-term net migration (M1: 147 000 people per year). This aligns with Tönnies *et al*[10] and Voeltz *et al*,[23] who also examine this variant in their projections of type 2 diabetes (T2D) in Germany.

Sensitivity analyses

Model assumptions and data sparsity might hamper the projection of future prevalence. Precise information about temporal trends of the incidence and mortality in Germany are lacking. Since these rates can emerge very heterogeneously, we assessed further, relatively extreme scenarios which may cover possible unexpected developments. First, we evaluated an annual decrease in the incidence of 5% as observed among older ages in Denmark between 1996 and 2016.[8] In contrast to a lower incidence, researchers postulated that the SARS-CoV-2 virus might damage cells in the pancreas that produce insulin, which may notably increase the risk of new-onset T1D.[24 25] Although data on a relation between SARS-CoV-2 and T1D remain heterogeneous,[25] we assessed a 5% annual increase in the incidence as observed in European children younger than 5 years between 2005 and 2020.[26] In addition, we performed a Monte Carlo simulation with 500 bootstraps to account for possible sampling error of the input values and to calculate 95% CIs (see online supplemental material). Furthermore, we assess the impact of alternative variants of the FSO population projections (see online supplemental material).

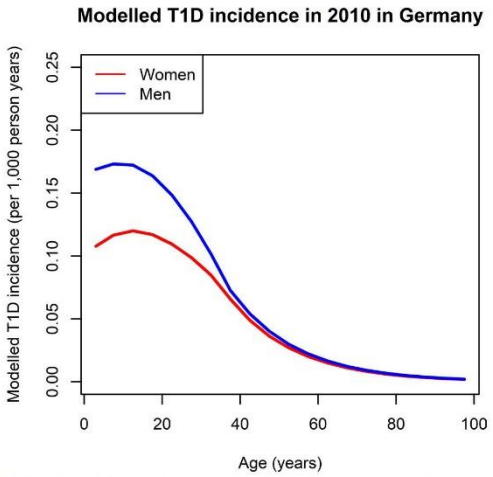

Figure 1 Estimated age-specific incidence of type 1 diabetes (T1D) in Germany among women (red) and men (blue) in 2010.

RESULTS

Estimated incidence

Figure 1 visualises the age-specific and sex-specific estimates of the German T1D incidence in 2010. Across all ages, the T1D incidence is estimated higher for men than women. For both sexes, the incidence peaks between 5 and 15 years of age with approximately 0.17 for men and 0.12 for women per 1000 person-years. For all following ages, the estimated course declines.

Projected prevalence

Figure 2 shows the estimated age-specific and sex-specific prevalence for Germany in 2009 and 2010. For any age below 80 years, the prevalence of T1D is higher among men compared with women. Among the elderly, the relationship between prevalence and sex does not remain that obvious.

Figure 3 shows the projected age-specific and sex-specific T1D prevalence for five scenarios of the incidence and MRR from 2010 to 2040 assuming moderately increasing life expectancy. For almost any age, the prevalence of T1D for men is slightly higher than for women. Particularly noteworthy is the shift in the course of the prevalence. In 2040, the prevalence is estimated to peak at around 40 years of age for both sexes, while the prevalence decreases among older age groups compared with 2010. Compared with scenario 1 assuming a constant prevalence, results for all other scenarios suggest a substantial increase in the overall prevalence for men and women (online supplemental figure S2). The different variants of the FSO had no remarkable influence on the prevalence projection (online supplemental figure S4).

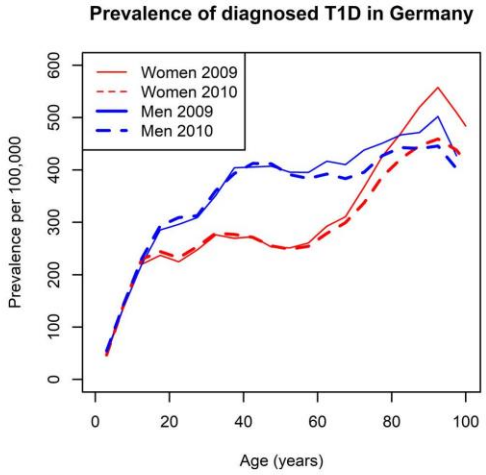

Figure 2 Prevalence of type 1 diabetes (T1D) among women (red lines) and men (blue lines) in Germany in 2009 and 2010.

Projected number of people with T1D

Figure 4 and table 1 show the projected number of people with T1D in Germany in 2010 and 2040 and approximate CIs for several scenarios of the incidence and MRR. Coherently, as for the prevalence, the number of men diagnosed with T1D is higher than the number of women throughout the overall projection horizon. Assuming that the prevalence remains as in 2010 (scenario 1), overall numbers are projected to increase by 1% to a total of 252 000 people with T1D in 2040. For scenarios 2–5, the number of people with T1D in Germany in 2040 is projected to increase by 44 000 people (+18%) to 78 800 people (+32%) compared with 2010. Compared with a relatively linear course for scenario 1, the course of the estimated case numbers for scenarios 2–5 resembles a J-shaped growth pattern in the first few years and then grows linearly for men and women alike. The FSO population variants had minor impact (see online supplemental material).

Sensitivity of the results

Additionally, we assessed an annual change of the incidence of −5% and +5%. Comparing 2010 and 2040, we obtain a decrease of the population with T1D of −27 500 (−11%) and an increase of almost 300 000 (+121%), respectively. Possible sampling error of the input values had minor impact and led to deviations of 5%–7% for women and men (see online supplemental material).

DISCUSSION

The aim of this study was to project the number of people with T1D in Germany between 2010 and 2040. For this purpose, we first estimated the age-specific and sex-specific incidence in 2010. Then, we projected the age-specific and sex-specific prevalence and the number of individuals with T1D for each year until 2040. Overall, our results indicate a general increase in the future burden of T1D.

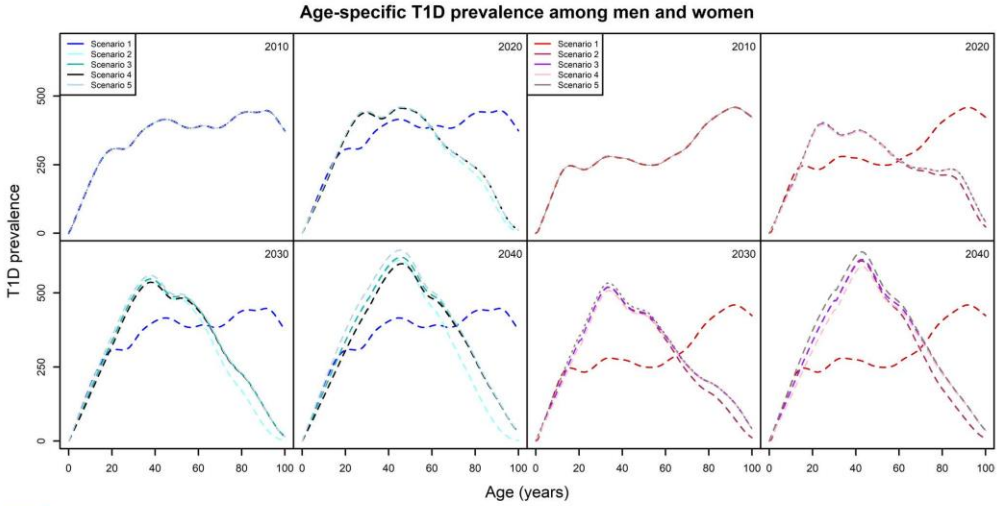

Figure 3 Projected age-specific and sex-specific prevalence of type 1 diabetes (T1D) in Germany for variant B1L2M1 of population projections of the German Federal Statistical Office. Blue lines depict the prevalence among men in 2010, 2020, 2030 and 2040, red lines visualise the prevalence among women in 2010, 2020, 2030 and 2040. The different line types represent the scenarios, that is, scenario 1, constant age-specific prevalence from 2010; scenario 2, constant T1D incidence rate (IR), constant mortality rate ratio (MRR); scenario 3, constant IR, MRR −2% per year; scenario 4, IR −0.5% per year, MRR −2% per year; scenario 5: IR +0.5% per year, MRR −2% per year.

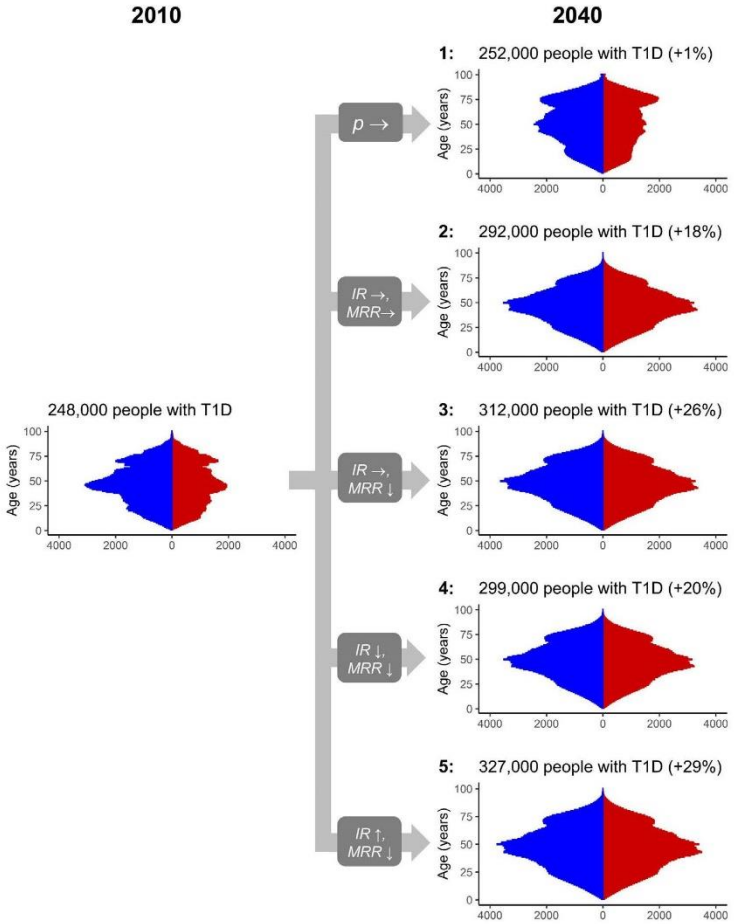

Figure 4 Projected number of type 1 diabetes (T1D) cases (in thousands) for men and women in Germany in 2010 and 2040 for variant B1L2M1 of the German Federal Statistical Office for scenarios 1–5. Scenario 1, constant age-specific prevalence from 2010; scenario 2, constant incidence rate (IR), constant mortality rate ratio (MRR); scenario 3, constant IR, MRR −2% per year; scenario 4, IR −0.5% per year, MRR −2% per year; scenario 5: IR +0.5% per year, MRR −2% per year.

Comparing different scenarios of disease-specific rates, we found considerable differences in the projected population with T1D. In scenario 1, we estimated an increase in the population with T1D of 1% in 2040 compared with 2010. In this scenario, we simply applied the constant sex-specific and age-specific prevalence of T2D from 2010 to future population distributions projected by the German FSO. Although this approach is commonly used for chronic disease projections,[23] it appears that ignoring temporal trends in the prevailing epidemiological situation likely leads to misleading results since the future development of a disease is influenced by various disease-specific aspects. More precisely, disregarding potential changes in mortality and incidence as is done in scenario 1 may provoke an underestimation of future T1D cases in Germany.[10 23] Contrastingly, reflecting on changes in epidemiological indicators such as mortality and incidence likely produces more reliable projection results.[23] This was done in scenarios 2–5, where the use of the PDE enables to additionally account for disease-specific information when modeling the future prevalence of T1D. Considering scenarios 2–5, which take temporal trends in the incidence and mortality into consideration, results ranged from 44 000 (+18%) to 78 800 (+32%) additional people with T1D in 2040. Unfortunately, there are no data available to validate our projected results and/or pinpoint the most plausible scenario. Nonetheless, the general increase observable across all scenarios highlights the importance of intensified diabetes-related public health management and the need for more resources

Table 1 Number of people with T1D (in thousands) in Germany in 2010 and 2040 for variant B1L2M1, 95% CIs for case numbers in 2040 as well as absolute and relative change comparing 2010 vs 2040

Scenario	Annual trend in MRR (%)	0	0	0	0	0	0	0	−2	−2	−2	−2	−2	−2	−2	(constant
	Annual trend in incidence (%)	0	−0.1	−0.5	+0.1	+0.5	+5	−5	0	−0.1	−0.5	+0.1	+0.5	+5	−5	prevalence)
Men	2010	137	137	137	137	137	137	137	137	137	137	137	137	137	137	137
	2040	149	148	143	151	156	266	104	162	160	155	163	169	281	115	139
	95% CI	133 to 154	131 to 152	127 to 147	134 to 155	139 to 160	237 to 260	92 to 112	145 to 160	143 to 159	139 to 154	146 to 162	151 to 167	251 to 268	103 to 118	139 to 139
	Absolute change	12	11	5	13	19	129	−33	24	23	18	26	32	144	−22	2
	Relative change	+9%	+8%	+4%	+10%	+14%	+94%	−24%	+18%	+17%	+13%	+19%	+23%	+105%	−16%	+1%
Women	2010	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
	2040	143	142	136	144	150	258	98	151	149	144	152	158	267	105	113
	95% CI	128 to 143	127 to 141	122 to 137	129 to 144	134 to 149	230 to 247	88 to 102	136 to 145	135 to 144	130 to 139	137 to 147	143 to 152	240 to 250	96 to 104	113 to 113
	Absolute change	32	31	26	33	39	147	−13	40	38	33	41	47	156	−6	2
	Relative change	+29%	+28%	+23%	+30%	+35%	+132%	−11%	+36%	+35%	+30%	+37%	+42%	+141%	−5%	+2%
Overall	2010	248	248	248	248	248	248	248	248	248	248	248	248	248	248	248
	2040	292	289	279	295	307	524	202	312	310	299	315	327	548	221	252
	95% CI	260 to 296	258 to 293	248 to 264	262 to 298	273 to 309	466 to 507	180 to 213	280 to 305	278 to 303	268 to 293	283 to 308	293 to 318	490 to 518	199 to 222	251 to 251
	Absolute change	44	41	31	47	58	276	−46	64	61	51	67	79	300	−28	4
	Relative change	+18%	+17%	+12%	+19%	+24%	+111%	−18%	+26%	+25%	+20%	+27%	+32%	+121%	−11%	+1%

MRR, mortality rate ratio; T1D, type 1 diabetes.

allocated to diabetes healthcare. Moreover, it emphasizes the need for research on primary prevention of T1D and effective interventions to reduce its incidence.[27]

Environmental and lifestyle changes may explain the rapid increase of the number of people with T1D worldwide.[1 5 28] The relatively short projection period makes changes in the genetic background of populations an unlikely cause. For instance, lifestyle factors that lead to rapid growth and large gains in weight particularly in the first years of life are more likely causing the increasing incidence. Factors such as accelerated autoimmune processes contribute to a growing T1D incidence in childhood.[22] Besides, factors that initiate the autoimmune destruction of beta cells are also discussed, for example, fetal and neonatal factors. Another hypothesis for the increase in the incidence is the reduced exposure to microbial antigens.[22] Overall, to gain a better understanding of drivers of the T1D burden, future studies could include these environmental factors in the evaluation to shed light on possible causes.

With regard to the methodological approach, our results underline the advantage of the IDM as theoretical background as it allows to incorporate temporal trends in disease-specific rates which impact the epidemiological trends of a disease.[10 23] Comparing multiple future scenarios, we showed that the incidence rate is a major driver of future numbers. A blunt extrapolation, that is, applying the observed prevalence in 2010 to the population projections, is unlikely to project accurate future numbers.[10 23] Ignoring disease-specific rates probably leads to underestimation of an upcoming disease situation, and consequently, underestimates the importance of public health management, healthcare facilities, education of specialists, and the like.

In addition to the increasing T1D burden, our projection indicates a shift of the prevalence peak toward older ages. Concomitant, although historically believed, T1D may not remain a disease of childhood. In line with Ostrauskas *et al*,[29] our results contradict a shift of the onset of T1D toward younger ages as explanation for the increasing temporal trend among children.

Comparison with previous studies

Worldwide, T1D incidence varies largely. In the great majority of countries, it has steadily increased over the past and its growth is expected to continue. In 2021, the International Diabetes Federation (IDF) estimates that 1 211 900 children and adolescents younger than 20 years were diagnosed with T1D and reports an annual growth of 149 500 children worldwide.[5] Direct health expenditures related to diabetes are close to US$1 trillion and are projected to exceed this by 2030. Strikingly, current evidence suggests that the IDF figures are still substantially underestimated.[30] Using data from the Childhood Diabetes Registry of Saxony, Germany, Manuwald *et al*[1] report an increase in the T1D incidence among children and adolescents in Germany from 17.1 to 24.7 per 100 000 person-years between 1999 and 2019. Continuing this trend, they project a growth to 34.8 per 100 000 person-years in 2030. Based on medical data from Baden-Wuerttemberg and Saxony and the national diabetes registry of North Rhine-Westphalia, Rosenbauer *et al*[2] report a national T1D incidence among adults of 6.1 per 100 000 person-years between 2014 and 2016. Their prevalence estimate equals 445 and 544 per 100 000 individuals with slightly lower figures for women compared with men. This equals a number of 341 000 adults with T1D in 2016, which is estimated to grow by 4150 adults annually. Notably, these studies neglect relevant age groups, the role of disease-specific rates or are based on limited data. Contrarily, one of the main strengths of our prevalence input data is its completeness in terms of covering all age groups as well as all health insurances in the German statutory health insurance.

Limitations

Although the reasons remain unclear, ICD codes are often inaccurately captured in administrative data in Germany (such as our prevalence input data),[9 12 13] impeding a valid differentiation of the diabetes types.[31] Indeed, research confirms that limited diagnostic accuracy, that is, sensitivity and specificity, from aggregated routine data may lead to unstable or implausible estimation results.[32 33] Potential misclassifications and double diagnoses in our setting may have occurred when a diagnosis was changed after more precise specification of a person's health state by different care providers or clinicians. For instance, and controversially discussed, the ICD coding specifies LADA as E10.-,[34–36] although its pathogenesis only gradually resembles the one of T1D. Moreover, since T1D is more common to develop at younger ages, it can be assumed that many unspecific cases among the older population could be attributed to T2D.[13 31] Consequently, these false positive diagnoses might explain questionable estimates among older ages. On the contrary, unspecific diabetes diagnosis among children very likely mask a T1D case.[13 31] To filter out such inaccuracies observed among older ages and since our focus is on T1D developed during childhood, we adapted the log-linear trend from Denmark to the German T1D incidence in 2010. Including misspecifications, unclear or double diagnoses, or cases masked by the unspecified diabetes coding E14.-, may lead to an additional increase in the prevalence of T1D.[13] Alternatively, previous studies advise the additional integration of medication to validate that the coding accurately reflects T1D prevalence rates from 2010,[31] since for patients with T1D it is necessary to inject insulin to control blood glucose levels. This method is of help for children and adolescents, in adulthood however it is inapplicable.[31] This falls beyond the scope of our article and is left open for future research.

Since the future course of the incidence and mortality is unknown, we explored the impact of trends in incidence and mortality by considering several scenarios. This highlights the need for long-term, country-specific and diabetes type-specific studies to obtain confidence

about future T1D cases. Assuming that medical aid and the treatment of diabetes will improve further, a decrease in the mortality seems likely. Evidence for continuous declines in mortality is found in several countries such as the USA, Canada, Finland, Sweden and Denmark.[8 37 38] Contrastingly, the T1D incidence and its development are heterogeneously distributed across countries.[1 5] Despite a universally reported growth, evidence from some regions suggests a slowing in this increase.[5 28] Against this reporting, pooled estimates for Europe show a continuous annual 3.4% growth, indicating a doubling of the incidence within 20 years.[28] In Germany, the T1D incidence among children has increased threefold in the past decades.[1] In line with the trend over the past 20 years, other studies assume a 40% increase in the incidence of T1D in children and adolescents for the next 10 years.[22] Overall, declining mortality and increasing incidence as in scenario 3 seems most plausible. Considering that the claims data reflect the medical situation a decade ago, it does neither comprise latest happenings such as the SARS-CoV-2 pandemic nor other recent developments that may have affected the disease situation.[39] It appears that a SARS-CoV-2 diagnosis is associated with an increased risk of new-onset T1D.[24 25] With our scenario in the sensitivity analysis assuming a 5% annual increase in the incidence, we aim to account for extreme events such as the impact of SARS-CoV-2. In that scenario, we project an increase in the future T1D burden in 2040 in Germany of up to 121% additional T1D cases compared with 2010. Evidently, other unexpected events such as the global climate, energy or inflation crisis, future developments of migration, or changes in diagnostic criteria may alter future trends of the T1D incidence and mortality.[40–43] Furthermore, there are medical and non-medical issues that make particular ethnic groups more or less vulnerable to diabetes. Routine data and particularly aggregated prevalence data from administrative origin do not provide any identification possibility of the individual's ethnicity.[13] Assessing the impact therefore goes beyond scope of our study, but could be addressed in the future.

Due to data scarcity in Germany, we referred to information on the incidence and MRR from Denmark for our initial T1D incidence estimation and prevalence projection. The motivation to use Danish data is threefold. First, it is based on the national diabetes register which has a coverage of 100% due to compulsory reporting and which accurately differentiates the diabetes types.[20] Second, previous studies concerning diabetes surveillance in Germany refer to Danish data.[9 10] Third, for countries that are comparable in terms of their disease burden and healthcare systems, such as Denmark and Germany, deviations of disease-specific rates should remain minimal.[9 10] Obviously, modelling the T1D situation over a 30-year horizon from 2010 until 2040 using prevalence data from 2010 and borrowed information from Denmark requires some assumptions. Our analysis and the reliability of our results would be strengthened if more recent, validated data from Germany were available. Moreover, information on T1D directly measured in the German population would be more appropriate to generate reliable national projections. As final remark with regard to an accurate diabetes surveillance in Germany, we claim that more detailed, precise, and timely data regarding diabetes in general and T1D in particular is urgently needed.

CONCLUSION

This is the first study addressing the current and future age-specific T1D incidence and prevalence for the whole German population for all ages from 0 to 100 years between 2010 and 2040. Using the IDM and data from 65 million people in Germany, we show that temporal trends in the incidence and mortality mainly drive the future number of individuals with T1D. Our projection suggests a substantial increase in the incidence, prevalence, and in the number of people diagnosed with T1D ranging from 4000 (+1%) to 78 800 (+32%) additional cases in Germany in 2040 vs 2010.

Contributors DV performed the analytic calculations, that is, implemented the methods and scenarios and analyzed the data. All coauthors verified the analytical methods, the results, and the main conceptual ideas. Furthermore, DV took the lead in writing the manuscript and, together with AH, was in charge of overall direction and planning. All authors provided critical feedback and helped shape the research, analysis, and manuscript. DV accepts full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish.

Funding The support for the publication costs was provided by the Open Access Publication Fund of Bielefeld University.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval An ethics committee approval was not required since no individual data on humans or animals were involved.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data and source file for estimation of the incidence of type 1 diabetes (T1D) and projection of the prevalence and number of people living with T1D in future Germany are freely available on Zenodo (see reference and link below). Prevalence input is based on a huge German claims data set and is uploaded as aggregated data. The population projections are publicly avilable on the website of the German Federal Statistical Office. Voeltz D, Brinks R, Tönnies T, et al. Future number of people with diagnosed type 1 diabetes in Germany until 2040: Zenodo 2022. Available at: 10.5281/zenodo.6799293 or https://zenodo.org/record/6799293#.ZCvgG4TP02w

ORCID iDs
Dina Voeltz http://orcid.org/0000-0002-7579-7144
Thaddäus Tönnies http://orcid.org/0000-0003-1577-4212
Annika Hoyer http://orcid.org/0000-0002-0241-9951

REFERENCES

1 Manuwald U, Schoffer O, Kugler J, *et al*. Trends in incidence and prevalence of type 1 diabetes between 1999 and 2019 based on the childhood diabetes registry of saxony, Germany. *PLOS ONE* 2021;16:e0262171.
2 Rosenbauer J, Neu A, Rothe U, *et al*. Types of diabetes are not limited to age groups: type 1 diabetes in adults and type 2 diabetes in children and adolescents. *J Health Monit* 2019;4:29–49.
3 Livingstone SJ, Levin D, Looker HC, *et al*. Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008-2010. *JAMA* 2015;313:37–44.
4 Milan V, Fetzer S, Hagist C. Healing, surviving, or dying?-projecting the German future disease burden using a Markov illness-death model. *BMC Public Health* 2021;21:123.
5 International Diabetes Federation. *IDF diabetes atlas*. Brussels, Belgium, 2021.
6 Rawshani A, Rawshani A, Franzén S, *et al*. Mortality and cardiovascular disease in type 1 and type 2 diabetes. *N Engl J Med* 2017;376:1407–18.
7 Palmer JP, Hampe CS, Chiu H, *et al*. Is latent autoimmune diabetes in adults distinct from type 1 diabetes or just type 1 diabetes at an older age? *Diabetes* 2005;54 Suppl 2:S62–7.
8 Carstensen B, Rønn PF, Jørgensen ME. Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996-2016. *BMJ Open Diabetes Res Care* 2020;8:1–9.
9 Tamayo T, Brinks R, Hoyer A, *et al*. Prävalenz und inzidenz von diabetes mellitus in Deutschland. *Deutsches Aerzteblatt Int* 2016;113:117–82.
10 Tönnies T, Röckl S, Hoyer A, *et al*. Projected number of people with diagnosed type 2 diabetes in Germany in 2040. *Diabet Med* 2019;36:1217–25.
11 Waldeyer R, Brinks R, Rathmann W, *et al*. Projection of the burden of type 2 diabetes mellitus in Germany: a demographic modelling approach to estimate the direct medical excess costs from 2010 to 2040. *Diabet Med* 2013;30:999–1008.
12 Reitzle L, Ihle P, Heidemann C, *et al*. Algorithmus Zur unterscheidung von diabetes mellitus typ 1 und typ 2 bei Der analyse von routinedaten. *Gesundheitswesen* 2022.
13 Schmidt C, Reitzle L, Heidemann C, *et al*. Excess mortality in adults with documented diabetes in Germany: routine data analysis of all insurance claims in Germany 2013-2014. *BMJ Open* 2021;11:e041508.
14 Federal Statistical Office of Germany. 14. coordinated population projection. 2019. Available: https://www.destatis.de/EN/Themes/Society-Environment/Population/Population-Projection/_node.html#sprg479856 [Accessed 21 Jan 2022].
15 Voeltz D, Brinks R, Tönnies T, *et al*. Future number of people with diagnosed type 1 diabetes in Germany until 2040: Zenodo; 2022.
16 Heidemann C, Paprott R, Schmidt C, *et al*. *Aufbau einer Diabetes-Surveillance in Deutschland – Ergebnisse der ersten Projektphase 2015 – 2019*. 2019.
17 Robert Koch-Institut. Diabetes surveillance. Available: https://diabsurv.rki.de/Webs/Diabsurv/DE/startseite/startseite-node.html [Accessed 13 Sep 2022].
18 Brinks R, Landwehr S, Icks A, *et al*. Deriving age-specific incidence from prevalence with an ordinary differential equation. *Stat Med* 2013;32:2070–8.
19 World Health Organization (WHO). European detailed mortality database. Available: data.euro.who.int/dmdb/ [Accessed 11 Aug 2015].
20 Carstensen B, Kristensen JK, Ottosen P, *et al*. The Danish national diabetes register: trends in incidence, prevalence and mortality. *Diabetologia* 2008;51:2187–96.
21 Groop L, Tuomi T, Rowley M, *et al*. Latent autoimmune diabetes in adults (LADA) -- more than a name. *Diabetologia* 2006;49:1996–8.
22 Rosenbauer J, Stahl A. Häufigkeit des diabetes mellitus Im kindes- und jugendalter in deutschland. *Diabetologe* 2010;6:177–89.
23 Voeltz D, Tönnies T, Brinks R, *et al*. Future prevalence of type 2 diabetes-A comparative analysis of chronic disease projection methods. *PLOS ONE* 2022;17:e0264739.
24 Qeadan F, Tingey B, Egbert J, *et al*. The associations between COVID-19 diagnosis, type 1 diabetes, and the risk of diabetic ketoacidosis: a nationwide cohort from the US using the cerner real-world data. *PLOS ONE* 2022;17:e0266809.
25 Watson C. Diabetes risk rises after COVID, massive study finds. *Nature* 2022.
26 Patterson CC, Dahlquist GG, Gyürüs E, *et al*. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study. *Lancet* 2009;373:2027–33.
27 Skyler JS. Primary and secondary prevention of type 1 diabetes. *Diabet Med* 2013;30:161–9.
28 Patterson CC, Harjutsalo V, Rosenbauer J, *et al*. Trends and cyclical variation in the incidence of childhood type 1 diabetes in 26 European centres in the 25 year period 1989-2013: a multicentre prospective registration study. *Diabetologia* 2019;62:408–17.
29 Ostrauskas R, Žalinkevičius R, Jurgevičienė N, *et al*. The incidence of type 1 diabetes mellitus among 15-34 years aged Lithuanian population: 18-year incidence study based on prospective databases. *BMC Public Health* 2011;11:813.
30 Tönnies T, Rathmann W, Hoyer A, *et al*. Quantifying the underestimation of projected global diabetes prevalence by the International diabetes Federation (IDF) diabetes atlas. *BMJ Open Diabetes Res Care* 2021;9:e002122.
31 Schmidt C, Reitzle L, Dreß J, *et al*. Prävalenz und Inzidenz des dokumentierten Diabetes mellitus – Referenzauswertung für die Diabetes-Surveillance auf Basis von Daten aller gesetzlich Krankenversicherten. *Bundesgesundheitsbl* 2020;63:93–102.
32 Brinks R, Tönnies T, Hoyer A. Impact of diagnostic accuracy on the estimation of excess mortality from incidence and prevalence: simulation study and application to diabetes in German men. *F1000Res* 2021;10:49.
33 Brinks R, Tönnies T, Hoyer A. Assessing two methods for estimating excess mortality of chronic diseases from aggregated data. *BMC Res Notes* 2020;13:216.
34 Monge L, Bruno G, Pinach S, *et al*. A clinically orientated approach increases the efficiency of screening for latent autoimmune diabetes in adults (LADA) in a large clinic-based cohort of patients with diabetes onset over 50 years. *Diabet Med* 2004;21:456–9.
35 Zimmet P, Turner R, McCarty D, *et al*. Crucial points at diagnosis. *Diabetes Care* 1999:B59–63. Available: https://www.proquest.com/openview/861c1cab74e5b7e47de6a947fd92a8de/1?pq-origsite=gscholar&cbl=47715
36 Biesenbach G, Auinger M, Clodi M, *et al*. Prevalence of LADA and frequency of GAD antibodies in diabetic patients with end-stage renal disease and dialysis treatment in Austria. *Nephrol Dial Transplant* 2005;20:559–65.
37 Gregg EW, Zhuo X, Cheng YJ, *et al*. Trends in lifetime risk and years of life lost due to diabetes in the USA, 1985-2011: a modelling study. *Lancet Diabetes Endocrinol* 2014;2:867–74.
38 Ringborg A, Lindgren P, Martinell M, *et al*. Prevalence and incidence of type 2 diabetes and its complications 1996-2003 -- estimates from a Swedish population-based study. *Diabet Med* 2008;25:1178–86.
39 Kamrath C, Rosenbauer J, Eckert AJ, *et al*. Incidence of type 1 diabetes in children and adolescents during the COVID-19 pandemic in Germany: results from the DPV registry. *Diabetes Care* 2022;45:1762–71.
40 Zilbermint M. Diabetes and climate change. *J Community Hosp Intern Med Perspect* 2020;10:409–12.
41 Vallianou NG, Geladari EV, Kounatidis D, *et al*. Diabetes mellitus in the era of climate change. *Diabetes Metab* 2021;47:101205.
42 Seligman HK, Bindman AB, Vittinghoff E, *et al*. Food insecurity is associated with diabetes mellitus: results from the National health examination and nutrition examination survey (NHANES) 1999-2002. *J Gen Intern Med* 2007;22:1018–23.
43 Anekwe TD, Rahkovsky I. The association between food prices and the blood glucose level of US adults with type 2 diabetes. *Am J Public Health* 2018;108:S475–82.

5.1.3 Contribution 3 – Projecting the Economic Burden of Type 1 and Type 2 Diabetes Mellitus in Germany from 2010 until 2040

Manuscript in Preparation Voeltz, D., Vetterer, M., Seidel-Jacobs, E., Brinks, R., Tönnies, T., Hoyer, A. Projecting the economic burden of type 1 and type 2 diabetes mellitus in Germany from 2010 until 2040. *Currently submitted at Population Health Metrics.*

Status Currently under review at Population Health Metrics.

Code and Data 10.5281/zenodo.8009685

Author Contributions D.V. is the guarantor and corresponding author of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. D.V. researched data, derived the methodology, performed the formal analysis, and wrote the manuscript. A.H. supervised the study and, together with D.V., developed the theoretical concept of the present study. All authors discussed the results, reviewed and edited the manuscript and the statistical analyses.

Projecting the economic burden of type 1 and type 2 diabetes mellitus in Germany from 2010 until 2040

Dina Voeltz, M.Sc. (dina.voeltz@uni-bielefeld.de, 0000-0002-7579-7144)[1,2], Maximilian Vetterer, B.Sc. (mvetterer@googlemail.com)[2], Dr. Esther Seidel-Jacobs (esther.seidel-jacobs@ddz.de, 0000-0002-2200-4024) [3,4], Prof. Dr. Ralph Brinks (ralph.brinks@uni-wh.de, 0000-0003-0961-6592)[3,5], Dr. Thaddäus Tönnies (thaddaeus.toennies@ddz.de, 0000-0003-1577-4212)[3,4], Prof. Dr. Annika Hoyer (annika.hoyer@uni-bielefeld.de, 0000-0002-0241-9951)[1]

[1] Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld

[2] Department of Statistics, Ludwig-Maximilians-University Munich

[3] Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Institute for Diabetes Research at Heinrich-Heine-University Düsseldorf

[4] German Center for Diabetes Research, Partner Düsseldorf, München-Neuherberg, Germany

[5] Chair for Medical Biometry and Epidemiology, Witten/Herdecke University, Faculty of Health/School of Medicine, Witten

Running title: Future costs of diabetes mellitus in Germany

Manuscript word count: 3700

Number of tables: 0

Number of figures: 4

All correspondence should be sent to:

Dina Voeltz
Biostatistics and Medical Biometry
Medical School OWL
Bielefeld University
Universitätsstr. 25
33615 Bielefeld
Germany
+49 521 106-86710
Dina.voeltz@uni-bielefeld.de

Abstract

Background: Our aim is to estimate age- and sex-specific direct medical costs related to diagnosed type 1 and type 2 diabetes in Germany between 2010 and 2040.

Methods: Based on nationwide representative routine data from 2010 from the statutory health insurance in Germany (almost 90% of the population's insurance) we projected age- and sex-specific healthcare expenses for type 1 and 2 diabetes considering future demographic, disease-specific and cost trends. We combine per capita healthcare cost data together with the demographic structure of the German population, diabetes prevalence, incidence and mortality. Direct per capita costs, total annual costs, cost ratios for people with versus without diabetes and attributable costs were estimated. The source code for running the analysis is publicly available in the open-access repository Zenodo [159].

Results: In 2010, total healthcare costs amounted to about €1.14 billion for type 1 and €28 billion for type 2 diabetes. Depending on the scenario, total annual expenses were projected to rise remarkably until 2040 compared to 2010, by 1% to 281% for type 1 (€1 to €4 billion) and by 8% to 364% for type 2 diabetes (€30 to €131 billion). Depending on annual cost growth (1% vs. 5% p.a.), we estimated annual per capita costs of €6,581 to €12,057 for type 1 and €5,245 to €8,999 for type 2 diabetes in 2040.

Conclusions: Diabetes imposes a large economic burden on Germany which is projected to increase substantially until 2040. *T*emporal trends in the incidence and cost growth are main drivers of this increase. This highlight the need for urgent action to prepare for the potential development and mitigate its consequences.

Keywords: Cost analysis, Economic Burden of Disease, Epidemiology, Healthcare costs, Projection, Type 1 diabetes, Type 2 diabetes.

Introduction

Worldwide, the number of people with diabetes and its associated costs have increased considerably[59]. In 2021, the global diabetes prevalence was estimated 10.5% among adults (20-79 years), i.e. 537 million people, and caused healthcare costs of at least $966 billion [60]. The absolute global economic burden is projected to grow to more than $2.1 trillion by 2030 [160]. In Germany, at least 7.2% of the population have diabetes, while this number is predicted to increase by up to 77% over the next 20 years [161]. This makes it a major health concern and challenge with regards to medical and economic resource planning. However, little is known about diabetes-related healthcare expenditures in Germany [11]. Cost of illness estimates are relevant from economic, medical and political point of view, and are necessary for effective healthcare management, meeting future medical needs and evaluating measures for prevention and intervention.

The only two available diabetes-related cost projections for Germany estimated an increase by 79% between 2010 and 2040 (from about €12 billion to €21 billion) [11], or healthcare expenditures of $30 to $56 billion in 2030 [162]. Though, the latter forecast lacks detailed country-specific input data, uses oversimplified methods and is based on a debatable assumption of constant diabetes prevalence over time. This also applies to international cost studies which are inherently limited by the methods applied [11]. For instance, they do not include an analysis of cost data or future cost trends such as inflation, ignore demographic changes, are limited to one sex, certain ages or other demographic input factors, do not differentiate the types of diabetes, or disregard the interplay of prevalence, incidence and mortality [10,11,63,161]. However, with regards to diabetes, the incidence and mortality majorly impact disease dynamics and ignorance leads to severe underestimation of future prevalence, which is one of the input factors for cost analyses [10,161,163]. Besides, the diabetes types are considerably different in their clinical representation, onset and progression [163]. Demographic shifts such as the aging population is relevant for type 2 diabetes, as its prevalence peaks amongst the cost-intensive older age groups[11]. Consequently, accurate projection methods of future chronic disease-specific costs need to be capable of reflecting the complex interplay simultaneously and differentiate where appropriate.

In this study, we use nationally representative data and a comprehensive and transparent method to project the sex-, age-, year- and diabetes-type specific per capita costs, total excess costs and cost ratios of people with and without type 1 and type 2 diabetes, as well as attributable costs in Germany from 2010 until 2040.

Methods

Data

We used the official population projection of the German Federal Statistical Office (FSO) [66], aggregated and published data on the incidence and mortality of individuals with versus without type 1 and 2 diabetes [10,27,33,63,161,163], as well as information on healthcare expenditures of the general German population (see Additional Table 1 in Additional file 1) [26,67].

Demographic inputs

We obtained the expected age- and sex-specific population distribution and the mortality of the general population in Germany for each year from 2010 to 2040 and all ages from 0 to 100 years from the FSO [66]. Six variants that represent rather realistic future developments instead of extreme assumptions [66] were considered to account for changes in migration, birth rates and population ageing. (see Additional Table 2 in Additional file 1). We used variant 2 (G2L2W2) as baseline, which assumes a birth rate (G2) of 1.55 children per woman, a life expectancy (L2) at birth in 2040 of 84.6 and 88.2 years for men and women, respectively, and a long-term net migration (W2) of annually 290,000 people. These assumptions align relatively well with values observed in Germany in 2020 and 2021 which include dynamics such as the COVID-19 pandemic or the Russo-Ukrainian war [164].

Disease-specific epidemiological inputs

Initial prevalence and incidence information were obtained from claims data that are considered representative for Germany and that were featured in recent projections of diabetes prevalence [10,63,161,163]. The data comprise information on the age and sex of 65 million insures in 2010 (about 80% of the total population) from all German statutory health insurances (SHI) and their diabetes diagnoses. We defined type 1 and type 2 diabetes based on the International Classification of Diseases-10 (ICD-

10) codes E10–E14. In 2010, about 7.3% of the German population were diagnosed with type 2 and 0.3% with type 1 diabetes. Regarding the age- and sex-specific mortality rate ratio (MRR) of people with versus without diabetes, we aligned with previous studies investigating diabetes in Germany [10,163] and included nationally representative MRR estimates of type 2 diabetes from 2014 provided by Schmidt et al. [27] and estimates from Carstensen et al. [32,33] to approximate the MRR of type 1 diabetes in Germany [163].

Cost inputs

The average per capita healthcare costs for people with and without diabetes were obtained from aggregated claims data from a 6.8% random sample of all German people with SHI [26,67], i.e., of almost 90% of the German population. The expenses include direct per capita costs for physicians, dentists, pharmacies, hospitals, sick benefits and others in 2010 in Germany from payer perspective. The included ICD-10 codes allowed to differentiate between costs related to type 1 or 2 diabetes. These data serve as starting values of per capita costs, cost ratios for people with versus without diabetes and attributable costs in 2010. Due to regulations on data protection, routine SHI data were provided in an anonymous and aggregated form (§5 Data Transparency Regulation, paragraph 4).

Projection model

As measures of interest, we modelled direct total and excess costs, cost ratios for people with versus without diabetes, and population attributable costs (see Additional Table 1 in Additional file 1) using the statistical software R, version 4.1.2 (R Foundation for Statistical Computing). The source code is published in the open-access repository Zenodo [159].

Epidemiological development

We first estimated the observed sex-specific prevalence of diagnosed type 1 and 2 diabetes in Germany in 2010 for all ages between 0 and 100 years. Second, we used a partial differential equation (PDE) that originates in the illness−death model (IDM) to project the age- and sex -specific prevalences until 2040 [10,158,161,163]. Our PDE describes the relation between prevalence, incidence and mortality as a function of age and calendar time, and thereby allows to incorporate future trends in the incidence

and mortality. Applying the projected prevalence to population counts yielded the number of individuals with type 1 or type 2 diabetes in Germany from 2010 to 2040 by

Cost projection

Using the aggregated cost data, we computed the total costs for each age and sex stratum by multiplying the number of people with or without diabetes and the respective per capita costs in 2010. We added these costs across all strata to obtain the total costs depending on diabetes status, i.e., diagnosed type 1, type 2 or no diabetes. We obtained the average per capita costs in 2010 by dividing the total costs and the age- and sex-specific number of people with or without diabetes. We stratified the average per capita costs by sex and diabetes status and interpolated between the age groups to calculate average per capita costs for all ages from 0 to 100 years. Diabetes-related excess costs were defined as costs of a person with type 1 or 2 diabetes that go beyond the costs of people without diabetes. To project the excess costs until 2040, we multiplied the excess costs with a mean annual growth rate of 0%, 1% or 5% depending on the respective scenario.

Total diabetes-related future costs from 2010 to 2040 were calculated as sum of the product of the projected population sizes, projected prevalence and projected average per capita costs for each year. Total annual excess costs are defined analogously, but include average per capita excess costs instead of average per capita costs. Using the projected cost data, we computed annual age- and sex-specific cost ratios (*R*) for people with type 1 or 2 diabetes relative to people without. The age- (*a*), sex- (*s*), diabetes type- (*d*) and time-specific (*t*) attributable costs (*PAC*) were defined as

$$PAC_{asdt} = \frac{p_{asdt} \times (R_{asdt} - 1)}{1 + p_{asdt} \times (R_{asdt} - 1)}$$

with *p* denoting the prevalence of the respective diabetes type [26]. To extrapolate from the data of the random sample (6.8 %) to the whole German population, we divided the sum of the total costs by the sample size. The resulting year-specific quotient was multiplied with the respective population attributable costs for each year.

Scenario analyses

To account for uncertainty, we modelled several demographic-, diabetes- and cost-related future dynamics. We constructed 16 scenarios motivated by previous papers from Waldeyer et al. [11], Tönnies et al. [161] and Voeltz et al. [10,163], who projected diabetes-related cost and prevalence in Germany for similar time horizons (details given in Table S2). Scenario 1 represents a base-case scenario. It assumes moderate demographic developments (variant G2L2W2) and is limited to an annual 2% decrease in the MRR with no changes made to any other cost or epidemiological input. Scenarios 2 to 8 account for potential epidemiological developments, scenarios 9 and 10 model the robustness of our results concerning changes in diabetes-related costs and scenarios 11 to 15 assess hypothetical demographic trends. Scenario 16 represents a most probable scenario assuming moderate demographic development, a mean annual cost inflation of 1%, an annual 1% increase in the incidence and a decrease of 2% in the MRR.

Sensitivity analyses

We reflect on potential error in the input values and future trends of the model parameters using the relatively extreme scenarios 6, 7, 8, and 10 which return upper and lower projection bounds.

Results

Per capita costs

In the reference year 2010, average per capita healthcare expenses of people insured in the SHI in Germany amounted to €4,285 for men and €4,889 for women with type 1 diabetes, and €3,868 for men and €3,889 for women with diagnosed type 2 diabetes. Assuming a moderate cost growth of 1% per year as in scenario 9 and 16, annual per capita costs reached on average €6,581 for type 1 and €5,245 for type 2 diabetes in 2040 (Figure 1 and Additional Figure 1 in Additional file 1). Accordingly, the increase was markedly higher when assuming an annual cost growth rate of 5% (scenario 10) to approximately €12,057 for type 1 and €8,999 for type 2 diabetes.

Total annual healthcare costs

All scenarios with only one exception projected rising total annual healthcare expenses over time (Additional Table 3 and 4 in Additional file 1). In 2010, the observed total annual costs in Germany

amounted to €1.1 and €28.8 billion for type 1 and type 2 diabetes, respectively. In 2040 and depending on the scenario, costs were projected to exceed this by 1% to 281% for type 1 and by 8% to 364% for type 2 diabetes (Figure 2). Our baseline scenario 1 resulted in total annual costs of about €1.5 billion for type 1 and €60.4 billion for type 2 diabetes in 2040. The most probable scenario 16 returned total costs of €2.2 and €79.2 billion related to type 1 or type 2 diabetes in 2040, respectively.

Cost ratio

Most likely (scenario 16), annual healthcare expenses in 2040 will be 2.8- and 3.2-fold higher for men and women with type 1, and 3.8- and 3.3-fold higher for men and women with type 2 diabetes compared to an insured person without diabetes, respectively (Figure 3 and Additional Figure 2 in Additional file 1). Generally, cost ratios were largest among younger ages and decreased with increasing age. For people with type 1 diabetes aged between 0 and 10 years, the discrepancy is highest, results showed 7-fold higher costs in 2010 and up to 15-fold higher costs in 2040. Regarding type 2 diabetes, the cost imbalance between people with versus without diabetes is less high and peaks around the age of 20 to 30 years (about 3- to 4-fold higher in 2010 and 6-fold higher in 2040).

Excess costs

In 2010, diabetes caused total medical excess costs of €0.3 billion for men and €0.3 billion for women with type 1 diabetes, and €5.8 billion for men and €5.0 billion for women with type 2 diabetes (Additional Table 3 and 4 in Additional file 1). Our base-case scenario predicted an increase to a total of €0.8 billion for type 1 and €21.2 billion for type 2 diabetes in 2040, while scenario 16 projected excess costs of €1.5 and €40.1 billion, respectively.

Population attributable costs and cost extrapolation

In Germany in 2010, SHI expenses amounted to ~€160 billion. On average, 3.8% and 10.2% thereof are attributable to the medical care of type 1 and type 2 diabetes, respectively. For scenario 16 in 2040, attributable costs of type 1 diabetes will be 7.7%, and 26.3% of type 2 diabetes. Extrapolating from our sample to the whole population showed that this corresponds to total direct costs of €0.55 and €14.4 billion for type 1 and type 2 diabetes in 2010 and €1.1 and €37.0 billion in 2040, respectively.

Attributable costs of type 2 versus type 1 diabetes are markedly higher (Figure 4). For type 1 diabetes, attributable costs are highest among younger ages. Vice versa, for type 2, these increase with increasing age. We projected slightly higher attributable costs for men compared to women with type 2 diabetes.

Discussion

Principal findings

In Germany in 2010, direct total annual healthcare costs amounted to about €1.14 billion for type 1 and €28 billion for type 2 diabetes. Expenditures of SHI are projected to rise remarkably until 2040, to about €1 to €4 billion and €30 to €131 billion, respectively. Our results show that the future prevalence will increase strongest and peak around the ages of 20 to 40 years for type 1 and 60+ years for type 2 diabetes, which each represents the costliest age groups for the corresponding diabetes type. The combination of these dynamics might explain a large part of the cost growth.

Scenario comparison

Forecasting is fraught with uncertainty as unexpected cultural, political, economic or medical shifts may prompt change and current trends develop differently than assumed. Therefore, we assessed several scenarios and can only speculate which scenario is most likely. Comparing all scenarios (Figure 2) showed that the cost growth is mainly attributable to rising incidence and cost rates, with little impact of population ageing.

Generally, demographic changes had little influence on the projection. With regards to migration, Germany reports a positive and rising migration balance since several decades making scenario 12 (assuming high migration) likelier than scenario 11 (low migration) [164]. Population ageing, analysed in scenario 14 versus 15, is related to an increased number of high-risk individuals and an increased life expectancy of people with diabetes and therefore, contributes to a higher number of diseased people. The combination of population ageing and per capita costs that increase with age might explain the observed minor impact on future cost growth.

The projected economic burden was highly sensitive to future cost growth rates. If direct medical per capita excess costs would inflate by 1% annually (scenario 9), the total costs of type and type 2 diabetes in Germany would increase by 100% for type 1 and 180% for type 2 diabetes. By contrast, assuming a cost growth of 5% (scenario 10), total annual costs in 2040 were projected to increase by 281% and 364%, respectively.

Another large part of the rising costs is possibly explained by the increase in prevalence attributable to future disease-specific dynamics [10,161,163]. Scenario 8 assumes that the age-and sex-specific prevalence remains constant as in 2010 and results in an increase of total costs in 2040 compared to 2010 of 1% and 29% for type 1 and 2 diabetes. Though, this probably severely underestimates the actual disease and cost burden [10,161,163]. More realistically are changing incidence rates (as in scenario 2 to 5), which resulted in increased annual costs by 28% to 41% for type 1 and by 109% to 128% for type 2 diabetes. Scenario 6 and 7, assuming a 5% annual increase or decrease in the incidence, account for extreme events such as the impact of the SARS-CoV-2 pandemic, which is associated with an increased risk of new-onset diabetes [10,163]. In 2040 compared to 2010, these scenarios lead to changes of the total cost by -1% or 149% for type 1 and by 6% or 322% for type 2 diabetes. This variability highlights the consequences of future incidence trends on upcoming healthcare expenditures.

Strengths and weaknesses

While our analysis provides important information for chronic-disease related research, future diabetes healthcare and policy interventions, it does have limitations.

Using mostly nationwide data, the risk of bias is considerably low. Though, one weakness of our data arises with the time that elapsed since the data collection in 2010. The current situation might differ and profound and unexpected shifts, such as the COVID-pandemic or the Russo-Ukrainian War, lead to discrepancy between current and past trends.

In addition to direct medical costs, diabetes is associated with indirect social and productivity costs such as premature mortality, disability, and higher rates of lost work time [11]. It is beyond the scope of

our study because precise information on the associated indirect costs is limited. The estimation of starting values for a projection requires additional methods and assumptions, for instance age restrictions because indirect costs only accrue from productivity losses arising from diabetes in working-age people [165]. Resulting, uncertain estimates would render any projection intrinsically highly speculative. We recommend future research to investigate this knowledge gap.

In spite of the high probability that costs for people without diabetes are likely to increase, they are kept constant in our analysis due to a lack of information on their precise development. The inclusion of doubtful cost trends might have impaired the trust in any cost estimate and its projection. Therefore, in line with Waldeyer et al. [11], we focused on an excess cost approach and limited the analysis to changes in costs of people with diabetes.

Lastly, it might seem critical that for simplicity and due to data unavailability, we assumed the same prevalence for German inhabitants and migrants. Though, Waldeyer et al. [11] showed that the impact of migration is negligible even when varying the prevalence.

Despite limitations, our study provides novel insights into the current and future economic burden of the two main types of diabetes in Germany. Although, type 1 and type 2 diabetes largely differ in their causes, symptoms, treatment and costs [166], previous studies rarely distinguished the types of diabetes [65]. For the first time, we projected type-specific direct medical cost, cost ratios and attributable costs related to the two main types of diabetes in Germany from 2010 until 2040 for both sex and all ages between 0 and 100 years based representative national routine data. Another advantage is the use of our forecasting model. Our method is able to reflect on temporal trends in epidemiological, demographic and cost dynamics simultaneously, but at the same time, remains transparent, clear and understandable in its application. Although our data and some assumptions may not be transferable, the statistical methods can be easily applied to other countries and chronic diseases and are flexible enough to anticipate impacts of alternative policy scenarios.

Comparison with previous studies

Worldwide, studies of diabetes-related health expenditure report large disparities [58]. The few existing studies of diabetes-related costs in Germany are consistent in estimating a large economic burden, but results are heterogeneous [11,26]. For our base year 2010 in Germany, studies reported diabetes-related per capita costs of €2,761 or €5,239 [167,168]. Type-specific per capita healthcare costs associated with type 2 diabetes were estimated at €3,352, €4,377 and €5,146 [26,169,170]. Current per capita cost estimates issued by the International Diabetes Federation (IDF) amount to $6,661 for Germany in 2021 [58]. The studies differ largely with regards to their statistical methods, the cost components included, the inclusion criteria and encoding for diabetes diagnosis, the type of diabetes considered and the representativeness of their study population which renders comparison complicated if not impossible.

For Germany, only two forecasts of diabetes-related costs are available [11]. One reports total annual costs of $30–56 billion in 2030 for type 1 and 2 of diabetes [58]. Due to lacking country-specific input data, an imbalanced ratio of individuals with versus without diabetes, the assumption of constant prevalence over time and the use of oversimplified methods, results may be overestimated. Using a time-discrete Markov model with locally limited data from the KORA (Cooperative Health Research in the Region of Augsburg) survey, Waldeyer et al. [11] projected total annual excess costs attributable to type 2 diabetes of €14,93 to €29.01 billion in 2040, with a baseline scenario yielding €21.1 billion. These findings are comparable to our baseline scenario, projecting excess costs of €21.9 billion for type 2 diabetes in 2040.

Implications

We found that the incidence and cost growth rate majorly drive future healthcare costs. This highlights the importance of population-based prevention and the need for *supporting investments* and cost policies, for instance in *health* infrastructure, new medications and technologies such as e-health for increased cost-efficiency.

In Germany, type 2 diabetes accounts for about 95% of all diabetes mellitus cases and in contrast to type 1 diabetes, can be prevented at little expense in the everyday life through cost effective and

structured interventions [171]. To reduce the projected diabetes epidemic and its associated costs, the government, healthcare providers and institutions should put efforts to raise awareness and to intervene to prevent the onset of type 2 diabetes. This can be achieved through changing behaviour, norms and structures in order to support healthy living. Although these recommendations include *upfront costs* to the German government and healthcare system, we feel these investments are necessary.

Conclusion

In summary, by 2040, type 1 and 2 diabetes will likely pose an even larger economic burden to the individuals suffering from the disease, as well as to the German health system, its economy and with that, the whole German population. Projected healthcare costs are highly sensitive to variation in cost growth and the incidence. Better and more timely epidemiological and economic diabetes-related data are needed in order to improve forecasting, to advance efforts at public awareness, and to support effective management and coordinated action for preventing and preparing for this development.

Supporting Information

The online version contains the following supplementary material provided in Additional File 1:

Additional Table 1: Overview of input data, variables and outcome measures

Additional Table 2: Overview of our projection scenarios

Additional Figure 1: Projected average annual per capita costs

Additional Figure 2: Projected cost ratios

Additional Table 3: Annual projected total costs of type 1 diabetes in Germany from 2010 until 2040 by sex (in millions)

Additional Table 4: Annual projected total costs of type 2 diabetes in Germany from 2010 until 2040 by sex (in millions)

Additional Table 5: Annual projected excess costs of type 1 diabetes in Germany from 2010 until 2040 by sex (in millions)

Additional Table 6: Annual projected excess costs of type 1 diabetes in Germany from 2010 until 2040 by sex (in millions)

Availability of data and materials

The statistical analysis was carried out using the free statistical software R, version 4.1.2 (R Foundation for Statistical Computing). The source code and most of the data for running the analysis of the current study are published in the open-access repository Zenodo [159]. Due to regulations on data protection, routine SHI data were only available for our study and were provided in an anonymous and aggregated form (§5 Data Transparency Regulation, paragraph 4). Therefore, these data are not publicly available. An ethics committee approval was not required since no individual data on humans or animals were involved.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Competing interests

The authors declare that there is no conflict of interest.

Author contributions

D.V. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. *D.V.* researched data, derived the methodology, performed the formal analysis, and wrote the manuscript. *A.H. supervised the study and, together with D.V., developed the theoretical concept of the present study.* All authors discussed the results, reviewed and edited the manuscript and approved the statistical analyses.

Figures

Figure 1: Projected average per capita costs

Projected average per capita costs (in €) of people with type 1 or 2 diabetes in the statutory health insurance in Germany between 2010 and 2040 (stratified by sex) assuming annual cost growth rates of 1% or 5%. Panel A shows average per-capita costs of men with type 1 or 2 diabetes, panel B displays the projected per-capita costs for women, respectively.

Figure 2: Projected annual total healthcare expenses

Projected annual total healthcare expenses for people with diagnosed diabetes type 1 or type 2 from 2010 until 2040 in Germany for different epidemiological, demographic or cost development scenarios. Results for type 1 diabetes are shown in panel A (epidemiological trends), B (demographic trends) and C (economic trends), while panel D (epidemiological trends), E (demographic trends) and F (economic trends) display results for type 2 diabetes.

Figure 3: Projected age-specific cost ratios of the total healthcare expenses

Age-specific cost ratios of the total healthcare expenses (in €) between men and women with diagnosed type 1 or 2 diabetes versus men and women without in the statutory health insurance in

Germany in 2010, 2020, 2030 and 2040. Panel A and B show results for men and women with type 1 diabetes, panel C and D display the projected cost ratios for men and women with type 2 diabetes, respectively.

Figure 4: Projected age- and sex-specific attributable costs

Age- and sex-specific attributable costs (%) of diagnosed type 1 or 2 diabetes in the statutory health insurance in Germany in 2010, 2020, 2030 and 2040. Panel A and B show results for men and women with type 1 diabetes, panel C and D display the projected attributable costs for men and women with type 2 diabetes, respectively.

References

1. Voeltz D. Projecting the economic burden of type 1 and type 2 diabetes mellitus in Germany from 2010 until 2040 2023: Zenodo. doi:10.5281/zenodo.8009686.
2. Zilbermint M. Diabetes and climate change. Journal of community hospital internal medicine perspectives. 2020;10:409–12.
3. Sun H, Saeedi P, Karuanga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes research and clinical practice. 2022;183:109119.
4. Bommer C, Sagalova V, Heesemann E, Manne-Goehler J, Atun R, Bärnighausen T, et al. Global economic burden of diabetes in adults: projections from 2015 to 2030. Diabetes care. 2018;41:963–70.
5. Tönnies T, Röckl S, Hoyer A, Heidemann C, Baumert J, Du Y, Scheidt-Nave C, Brinks R. Projected number of people with diagnosed Type 2 diabetes in Germany in 2040. Diabetic Medicine. 2019;36:1217–25.

6. Waldeyer R, Brinks R, Rathmann W, Giani G, Icks A. Projection of the burden of type 2 diabetes mellitus in Germany: a demographic modelling approach to estimate the direct medical excess costs from 2010 to 2040. Diabetic Medicine. 2013;30:999–1008.
7. Zhang P, Zhang X, Brown J, Vistisen D, Sicree R, Shaw J, Nichols G. Global healthcare expenditure on diabetes for 2010 and 2030. Diabetes research and clinical practice. 2010;87:293–301.
8. Voeltz D, Tönnies T, Brinks R, Hoyer A. Future prevalence of type 2 diabetes—A comparative analysis of chronic disease projection methods. Plos one. 2022;17:e0264739.
9. Tamayo T, Brinks R, Hoyer A, Kuß O, Rathmann W. The prevalence and incidence of diabetes in Germany: an analysis of statutory health insurance data on 65 million individuals from the years 2009 and 2010. Deutsches Ärzteblatt International. 2016;113:177.
10. Voeltz D, Brinks R, Tönnies T, Hoyer A. Future number of people with diagnosed type 1 diabetes in Germany until 2040: an analysis based on claims data. BMJ Open Diabetes Research and Care. 2023;11:e003156.
11. Federal Statistical Office of Germany. 14. Coordinated population projection. 2019. https://www.destatis.de/EN/Themes/Society-Environment/Population/. Accessed 12 Apr 2023.
12. Carstensen B, Rønn PF, Jørgensen ME. Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996-2016. BMJ Open Diabetes Research and Care. 2020;8:e001071.
13. Schmidt C, Reitzle L, Heidemann C, Paprott R, Ziese T, Scheidt-Nave C, Baumert J. Excess mortality in adults with documented diabetes in Germany: routine data analysis of all insurance claims in Germany 2013-2014. BMJ open. 2021;11:e041508.
14. Jacobs E, Hoyer A, Brinks R, Icks A, Kuß O, Rathmann W. Healthcare costs of Type 2 diabetes in Germany. Diabetic Medicine. 2017;34:855–61.
15. German Institute of Medical Documentation and Information. Information system for health care data. 2014. https://www.dimdi.de/static/en/versorgungsdaten/index.htm. Accessed 18 Jun 2016.
16. Federal Statistical Office of Germany. 15. Coordinated population projection. 2023. https://www.destatis.de/EN/Themes/Society-Environment/Population/. Accessed 12 Apr 2023.

17. Carstensen B, Kristensen JK, Ottosen P, Borch-Johnsen K. The Danish National Diabetes Register: trends in incidence, prevalence and mortality. Diabetologia. 2008;51:2187–96.
18. Brinks R, Landwehr S. Age-and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. Theoretical population biology. 2014;92:62–8.
19. Tönnies T, Hoyer A, Brinks R. Productivity-adjusted life years lost due to type 2 diabetes in Germany in 2020 and 2040. Diabetologia. 2021;64:1288–97.
20. Zucchini S. Editorial on Clinical Aspects of Different Forms of Diabetes in Children and Adolescents. Frontiers in Endocrinology. 2022;13:3396.
21. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. Diabetic Medicine. 2012;29:855–62.
22. International Diabetes Federation. IDF Diabetes Atlas. 2021. https://www.diabetesatlas.org/. Accessed 2 May 2023.
23. Köster I, Huppertz E, Hauner H, Schubert I. Costs of Diabetes Mellitus (CoDiM) in Germany, direct per-capita costs of managing hyperglycaemia and diabetes complications in 2010 compared to 2001. Experimental and Clinical Endocrinology & Diabetes. 2014;122:510–6.
24. Bächle C, Claessen H, Andrich S, Brüne M, Dintsios CM, Slomiany U, et al. Direct costs in impaired glucose regulation: results from the population-based Heinz Nixdorf Recall study. BMJ Open Diabetes Research and Care. 2016;4:e000172.
25. Müller N, Heller T, Freitag MH, Gerste B, Haupt CM, Wolf G, Müller UA. Healthcare utilization of people with type 2 diabetes in Germany: an analysis based on health insurance data. Diabetic Medicine. 2015;32:951–7.
26. Ulrich S, Holle R, Wacker M, Stark R, Icks A, Thorand B, et al. Cost burden of type 2 diabetes in Germany: results from the population-based KORA studies. BMJ open. 2016;6:e012527.

27. Ely EK, Gruss SM, Luman ET, Gregg EW, Ali MK, Nhim K, et al. A national effort to prevent type 2 diabetes: participant-level evaluation of CDC's National Diabetes Prevention Program. Diabetes care. 2017;40:1331–41.

5.2 Statistical Modelling of Time-To-Event Data

5.2.1 Contribution 4 – A Parametric Additive Hazard Model for Time-to-Event Analysis

Manuscript in Preparation Voeltz, D., Hoyer, A., Forkel, A., Schwandt, A., Kuss, O. A parametric additive hazard model for time-to-event analysis. *Currently submitted at BMC Medical Research Methodology.*

Status Currently under review at BMC Medical Research Methodology.

Code and Data 10.5281/zenodo.7124988

Author Contributions D.V. is the corresponding of this work and, as such, had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the analysis. A.H. and O.K. derived the methodology. A.H. supervised the study and, together with D.V., developed the theoretical concept and outline of the present study. With support of A.K. and A.F., A.H. and D.V. performed the formal analysis. D.V. and A.H. wrote the manuscript. All authors discussed the results, reviewed and edited the manuscript and the statistical analyses.

A parametric additive hazard model for time-to-event analysis

Dina Voeltz[1,2*][0000-0002-7579-7144], Annika Hoyer[1*][0000-0002-0241-9951], Amelie Forkel[2], Anke Schwandt[3][0000-0003-3041-0759], and Oliver Kuss[4][0000-0003-3301-5869]

[1] Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Bielefeld, Germany
[2] Department of Statistics, Ludwig-Maximilians-University München, München, Germany
[3] Nuvisan GmbH, Neu-Ulm, Germany
[4] Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany
* These authors contributed equally to this work.
Corresponding author: Dina Voeltz, dina.voeltz@uni-bielefeld.de, Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany.

Abstract. Background: In recent years, the use of non- and semi-parametric models which estimate hazard ratios for analysing time-to-event outcomes is continuously criticized in terms of interpretation, technical implementation, and flexibility. Hazard ratios in particular are critically discussed for their misleading interpretation as relative risks and their non-collapsibility. Additive hazard models do not have these drawbacks but are rarely used because they assume a non- or semi-parametric additive hazard which renders computation and interpretation complicated.
Methods: As a remedy, we propose a new parametric additive hazard model that allows results to be reported on the original time rather than on the hazard scale. Being an essentially parametric model, survival, hazard and probability density functions are directly available. Parameter estimation is straightforward by maximizing the log-likelihood function.
Results: Applying the model to different parametric distributions in a simulation study and in an exemplary application using data from a study investigating medical care to lung cancer patients, we show that the approach works well in practice.
Conclusions: Our proposed parametric additive hazard model can serve as a powerful tool to analyze time-to-event outcomes due to its simple interpretation, flexibility and facilitated parameter estimation.

Keywords: Additive hazard · Parametric modeling · Survival analysis · Time-to-event model.

1 Background

Regression models for time-to-event outcomes are preferentially fitted by the proportional hazard model [8]. This is surprising because the hazard ratio, which is the generic effect estimate of that model, has been repeatedly criticized in recent years. Most relevant points of concern were that the hazard ratio (*i*) is interpretable only when mistaken as a relative risk [24], (*ii*) has a built-in selection or left truncation bias even in randomized trials [13,2,23], and (*iii*) is non-collapsible [11], meaning that adjusting for a covariate that is associated with the event will in general change the hazard ratio, even if this covariate is not associated with the exposure. To conquer these shortcomings, parametric survival models have been recommended as they are simpler, more informative, more robust, and have hazard, survival and probability density function directly available [20,17].

A hazard-based model that does not suffer from the above mentioned problems is the additive hazard model introduced by Aalen [1]. Results from additive hazard models can be translated to a relative survival scale which is, for example, routinely used in cancer epidemiology when a cohort of cancer patients is compared to the general population in terms of expected survival [10]. Moreover, recent works showed that parameters of the additive hazard model do not suffer from non-collapsibility [18], at least when defined in continuous time [22]. However, the Aalen model is rarely used in applied research because it assumes a non-parametric hazard as well as time-dependent covariates. Of course, these properties give large flexibility in modeling, but also complicate parameter estimation considerably. In addition, Bradburn et al. [4] state that "[...] the model coefficients are not easy to understand, and as they change repeatedly over time, can offer no single quantifiable effect size." As a partial solution for that problem, Lin et al. [16] proposed an additive hazard model for (constant,) time-independent covariates, but still use a semi-parametric hazard which may require lengthy computations for some statistical standard software [21].

As a remedy, in this article we combine the additive hazard model with time-independent (, i.e., constant) covariates and a parametric assumption for the baseline hazard. This will result in a number of advantages in terms of interpretation, possible model extensions, and also enables parameter estimation using every software that allows maximizing a hand-coded likelihood function. We proceed by introducing the formal notation and deriving the model's equations in Section 2. Then, we report the settings of our simulation study in Section 2. We illustrate the model using a study which investigated provision of medical care to lung cancer patients in the eastern part of Germany [3] and present all results in Section 3. Finally, we conclude with a discussion (Section 4).

2 Methods

Parametric additive hazard model

We start with the assumption that the hazard $h_x(t)$ with covariate vector x at time t can be expressed as

$$h_x(t) = h_{0,\theta}(t) + x\beta, \tag{1}$$

with a parametric baseline hazard function $h_{0,\theta}(t)$ which is independent of the covariates. The parameter θ denotes the distribution parameters, which differ in terms of number and commonly used notation depending on the choice of the baseline distribution. The parameters β are regression coefficients that measure the additive impact of covariates.

Using the well-known relations between hazard, probability density function (pdf) ($f(t)$) and survival function $S(t)$, the corresponding pdf of model (1), $f_x(t)$, can be expressed in terms of the baseline pdf $f_0(t)$, the baseline survival function $S_{0,\theta}(t)$, and the covariates by

$$f_x(t) = \frac{f_0(t) + x\beta S_{0,\theta}(t)}{\exp(tx\beta)}. \tag{2}$$

The complete derivation of the additive hazard model equation in terms of the pdf can be found in Additional file 1.

The corresponding survival function $S_x(t)$ is given by

$$S_x(t) = \frac{S_{0,\theta}(t)}{\exp(tx_i\beta)} = 1 - F_x(t), \tag{3}$$

with $F_x(t)$ denoting the cumulative distribution function (cdf).

Using an additive hazard regression model allows to estimate relative survival instead of hazard ratios [26]. Contrarily to hazard ratios, which quantify the average or constant effect of covariates on the hazard function, relative survival measures the cumulative effect of covariates on the relative survival probability. Within this context and in case $x = 0$ represents the absence of disease, the relative survival probability can be interpreted as the observed survival probability of the population studied, divided by the expected survival probability if the population was free of the disease of interest. Using (3), we can confirm the relative survival interpretation of our additive hazard model:

$$\frac{S_x(t)}{S_{0,\theta}(t)} = \frac{\frac{S_{0,\theta}(t)}{\exp(tx\beta)}}{S_{0,\theta}(t)} = \frac{1}{\exp(tx\beta)}. \tag{4}$$

It should be noted that this relative survival interpretation is independent of the baseline distribution.

The likelihood function for the model can be derived by accounting for the fact that observations with an event contribute the logarithm of the pdf, and censored observations the logarithm of the survival function [6]. The contribution

of a single observation i ($i = 1, ..., N$) with covariate vector x_i and observation time t_i to the log-likelihood function ℓ_i is therefore

$$\begin{aligned}\ell_i &= (1-\delta_i)\log(f_{x_i}(t_i)) + \delta_i \log(S_{x_i}(t_i)) \\ &= (1-\delta_i)\log\left(\frac{f_0(t_i) + x_i\beta S_{0,\theta}(t)(t_i)}{\exp(t_i x_i \beta)}\right) + \delta_i \log\left(\frac{S_{0,\theta}(t)(t_i)}{\exp(t_i x_i\beta)}\right) \\ &= (1-\delta_i)(\log(f_0(t_i) + x_i\beta S_{0,\theta}(t)(t_i)) - t_i x_i\beta) + \delta_i(\log(S_{0,\theta}(t)(t_i)) - t_i x_i\beta).\end{aligned}$$

δ_i is the censoring indicator with $\delta_i = 1$ if an observation is censored, and $\delta_i = 0$ if an event has been observed. Parameter estimation is straightforward by maximizing the log-likelihood function with respect to the unknown regression coefficients β and the parameters of the assumed baseline distribution. Each software that allows coding and maximizing such function, as for example SAS via the NLMIXED procedure or R via the `optim`-function, can be used to this task.

For practical application, it is possible to assume a wide range of baseline distributions, including for example the Exponential, Weibull, Gamma, Gompertz, Log-Normal and Log-Logistic distribution. For instance, assuming the Weibull distribution as baseline distribution and including a covariate x_i, the cdf is then given by:

$$F_{x_i}(t_i) = 1 - S_{x_i}(t_i) = 1 - \exp\left(-\left(\frac{t_i}{b_{WB}}\right)^{a_{WB}} - t_i x_i \beta\right). \tag{5}$$

Alternatively, using a Log-Logistic distribution as baseline distribution, this leads to

$$F_{x_i}(t_i) = 1 - S_{x_i}(t_i) = 1 - \frac{\left(\left(\frac{t_i}{a_{LL}}\right)^{b_{LL}} + 1\right)^{-1}}{\exp(t_i x_i \beta)}. \tag{6}$$

a_{WB}, b_{WB}, a_{LL} and b_{LL} denote the distribution-specific parameters.

Results from applying the different baseline distributions can be compared via model selection criteria as the AIC or BIC. Parameters of main interest are finally the estimated regression coefficients and suitable transformations of the distribution parameters that have more intuitive interpretations, as the baseline mean or median of the assumed distribution.

Simulation Study

To evaluate the parametric additive hazard model, we conducted a simulation study. For comparison, we included the semi-parametric additive hazard model of Lin et al. [16]. The simulation study was implemented using R (version 4.1.2), the full code is publicly available on Zenodo [25].

Setting Our parameter settings were motivated by the Halle Lung Carcinoma (HALLUCA) study which we also use as exemplary application in Section 3 [3]. The data from the HALLUCA study has been used in previous work proposing an extension of relative survival models for clustered responses [15] and are also suitable for our purpose. Accordingly, we focused on simulating a single binary covariate which can be interpreted for example as exposure in an observational study. Survival time is measured beginning with the day of diagnosis of lung cancer. As true underlying models used for data generation we assumed (*i*) the Weibull additive hazard model and (*ii*) the Log-Logistic additive hazard model as shown in equation (5) and (6). The true parameters of the distributions equaled their estimates from the HALLUCA study

- with $a_{WB} = 0.86, b_{WB} = 1.77$ for the Weibull distribution, and
- with $a_{LL} = 1.06, b_{LL} = 1.14$ for the Log-Logistic distribution.

In addition, we varied the number of participants per study which were set to 50 or 200. Moreover, we distinguish between a smaller number of observed events per study of 60% (40% censoring) and a higher number of 80% observed events (20% censoring). Further, we evaluated the true effect of the binary covariate for $\beta = 0$, $\beta = 0.8$ and $\beta = 1.6$.

Data generation Combining all parameters led to 24 different settings, i.e., 12 settings for each of (*i*) and (*ii*), for which data were generated. For every setting, we simulated 1000 data sets. Participants of each study were randomly allocated into two groups, as indicated by the binary covariate to which the finally estimated regression coefficient β corresponds. We assumed that our two groups are of equal size. Survival times for each participant were generated from the true underlying distribution using inverse transform sampling. In case the covariate takes the value of zero, the cdf $F_{x_i}(t_i)$ from which we had to sample simply equals the Weibull or Log-Logistic distribution. Contrarily, if the covariate takes a value of one, the procedure is less trivial. In that case, it is not possible to invert the cdf analytically and we used numerical inversion to obtain random numbers from the respective function [5]. The likelihood of whether a study participant experienced an event or if the event is censored was generated from a Bernoulli distribution with respective success probability. To guarantee uninformative censoring, we multiplied the original survival time with a uniformly distributed random number in case the observation was censored.

Estimation methods and outcomes For parameter estimation we used the Weibull and Log-Logistic additive hazard model. The corresponding likelihood functions were manually coded and implemented in the statistical software R. Optimization was done using the `optim`-function. True parameter values served as starting values for this procedure. As described above, we also applied the semi-parametric additive hazard model of Lin et al. [16] as comparative model. For this aim, we used the R-package `addhazard` where the Lin and Ying model is already implemented. Outcome of primary interest was the estimated regression

Setting Our parameter settings were motivated by the Halle Lung Carcinoma (HALLUCA) study which we also use as exemplary application in Section 3 [3]. The data from the HALLUCA study has been used in previous work proposing an extension of relative survival models for clustered responses [15] and are also suitable for our purpose. Accordingly, we focused on simulating a single binary covariate which can be interpreted for example as exposure in an observational study. Survival time is measured beginning with the day of diagnosis of lung cancer. As true underlying models used for data generation we assumed (*i*) the Weibull additive hazard model and (*ii*) the Log-Logistic additive hazard model as shown in equation (5) and (6). The true parameters of the distributions equaled their estimates from the HALLUCA study

- with $a_{WB} = 0.86, b_{WB} = 1.77$ for the Weibull distribution, and
- with $a_{LL} = 1.06, b_{LL} = 1.14$ for the Log-Logistic distribution.

In addition, we varied the number of participants per study which were set to 50 or 200. Moreover, we distinguish between a smaller number of observed events per study of 60% (40% censoring) and a higher number of 80% observed events (20% censoring). Further, we evaluated the true effect of the binary covariate for $\beta = 0$, $\beta = 0.8$ and $\beta = 1.6$.

Data generation Combining all parameters led to 24 different settings, i.e., 12 settings for each of (*i*) and (*ii*), for which data were generated. For every setting, we simulated 1000 data sets. Participants of each study were randomly allocated into two groups, as indicated by the binary covariate to which the finally estimated regression coefficient β corresponds. We assumed that our two groups are of equal size. Survival times for each participant were generated from the true underlying distribution using inverse transform sampling. In case the covariate takes the value of zero, the cdf $F_{x_i}(t_i)$ from which we had to sample simply equals the Weibull or Log-Logistic distribution. Contrarily, if the covariate takes a value of one, the procedure is less trivial. In that case, it is not possible to invert the cdf analytically and we used numerical inversion to obtain random numbers from the respective function [5]. The likelihood of whether a study participant experienced an event or if the event is censored was generated from a Bernoulli distribution with respective success probability. To guarantee uninformative censoring, we multiplied the original survival time with a uniformly distributed random number in case the observation was censored.

Estimation methods and outcomes For parameter estimation we used the Weibull and Log-Logistic additive hazard model. The corresponding likelihood functions were manually coded and implemented in the statistical software R. Optimization was done using the `optim`-function. True parameter values served as starting values for this procedure. As described above, we also applied the semi-parametric additive hazard model of Lin et al. [16] as comparative model. For this aim, we used the R-package `addhazard` where the Lin and Ying model is already implemented. Outcome of primary interest was the estimated regression

coefficient β as this parameter is estimated by the parametric as well as by the semi-parametric model. The estimated parameters were compared in terms of bias, mean squared error (MSE) and empirical coverage. Moreover, we identified the number of converged simulation runs to assess numerical robustness.

3 Results

Simulation Study

In reporting the results of the simulation study, we confine ourselves to the outcome of main interest, the estimated regression coefficient β. In the following, we give a brief overview of the results concerning bias, MSE, empirical coverage and numerical robustness. The complete simulation results can be found in Appendix II to V of the Additional file 1.

Bias Appendix II of the Additional file 1 provides the numerical results in terms of bias. Figure 1 illustrates the corresponding estimates. If the estimated additive hazard model is consistent with the true underlying distribution, we observe the best performance, i.e., smallest median bias, with some exceptions. Further, also considering that the estimated additive hazard model is consistent with the true underlying distribution, the additive hazard model slightly outperforms the semi-parametric model by Lin and Ying. This holds true for all estimated models, leading to a bias between -0.35 and 0.17. For most settings, the treatment effect is slightly underestimated which leads to a negative bias. The observed underestimation also holds for the Lin-Ying model which results in underestimation in more than 65% of all cases. We observe a change in the sign of the bias for settings where the Weibull-model is estimated using the new additive hazard model and relying on data that was generated via a Log-Logistic distribution. This underlines that in these settings the Weibull distribution is not an appropriate choice for modeling data that follows a Log-Logistic distribution, and an alternative choice of the underlying distribution should be made. The most extreme positive bias of 0.17 is observed for the estimated Weibull model when the Log-Logistic distribution, a true β of 0.8, 20% censoring and 200 patients were assumed for data generation. The most extreme negative bias of -0.35 is observed for the estimated Weibull and Log-Logistic model in the setting with data generated from the Weibull additive hazard model, a true β of 1.6, 40% censoring and 200 observations per sample. Generally, when data was generated assuming a Weibull distribution, all three models perform similarly. Though, if a Log-Logistic distribution is assumed as true, the estimated Log-Logistic and Lin-Ying model perform similarly, while results from the Weibull model slightly differ. The number of participants modeled per study minorly influences the median bias. The variability of the bias is smaller for higher numbers of observations (200 versus 50). Furthermore, the estimated models are sensitive towards the true value of β, where generally the bias is closer to zero and its variability is reduced for smaller values of β. In most cases, the bias is smaller in settings

with more events (80% compared to 60%). With regards to the Weibull model, this only holds true if the estimated and true distribution are consistent.

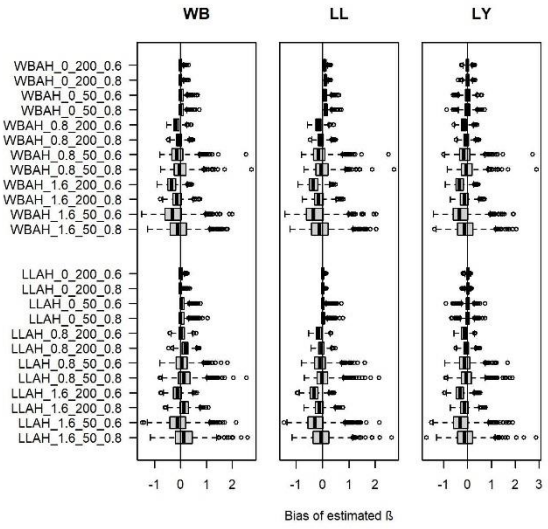

Fig. 1. Box plots of estimated median bias for the regression coefficient β over 1000 simulations for each setting. Y-axis denotes each setting with an ID consisting of the abbreviated true model (WBAH = Weibull additive hazard, LLAH = Log-Logistic additive hazard), the true β, the number of participants per study and number of events (e.g., Weibull additive hazard model, true $\beta = 0$, number of patients = 50, number of events = 60% results in "WBAH_0_50_0.6"). Left-most plot shows results for the Weibull (WB) additive hazard model, middle plot shows results for the Log-Logistic (LL) additive hazard model and right plot shows results for the Lin-Ying (LY) model.

MSE Appendix III of the Additional file 1 shows the detailed results with regards to the MSE. Figure 2 visualizes the findings. Generally, there are hardly any differences for the MSE across the three estimated models for the respective settings. In some settings we observe an MSE of 0.00. These results indicate that the model works very well at minimum in these contexts. Contrarily, we observe

a maximum error of 0.45 for the Lin-Ying model in one setting (true model based on Log-Logistic distribution, true $\beta = 1.6$, number of observations = 50, number of events 60%). As for the bias, in most cases the MSE is slightly smaller when the estimated model is consistent with the true underlying distribution. For almost any setting, it holds true that the median MSE of the Weibull and Log-Logistic model is better than the MSE of the respective Lin-Ying model. The MSE seems to be negatively correlated with the number of participants per study, as for settings with 200 observations per sample (versus 50 observations) the MSE is smaller. Similar results are found with regards to the assumed number of events: the MSE decreases when more events (80% compared to 60%) are observed. Furthermore, for larger values of β, the MSE and its variability increase for each of the estimated models.

Empirical coverage Appendix IV of the Additional file 1 provides detailed results about the models' empirical coverage. Figure 3 illustrates the corresponding estimates. In terms of empirical coverage (on the 95% level), the performances of the three estimated models are relatively similar with some exceptions for a handful of settings. For the majority of the settings, the models again perform best when the estimated model and the true underlying distribution are equal. If the distributions are consistent, the Weibull and Log-Logistic model mostly outperform the Lin-Ying model, respectively. Overall, the empirical coverage often falls below 95%. Precisely, this is the case for 75% of the settings for the Weibull, for 71% of settings for the Log-Logistic and for 71% of all settings for the Lin-Ying model. The highest coverage of 100% is observed for the Log-Logistic model when data was generated from a Log-Logistic distribution with β equal to 0, 60% events and 200 observations. On the other hand, the smallest coverage of 57.7% is observed for the Log-Logistic for data generated from a Weibull distribution with β equal to 0, 80% events and 200 observations. Similar to bias and MSE, results depend on the true value of β. Coverage results for the estimated models seem to improve when β is closer to zero. With regards to the number of events, results correlate positively with the event probability. Precisely, coverage results improve in settings where the event probability is higher (80% versus 60%). The number of participants per sample seems to have a slight influence on coverage results. However, the correlation of coverage and sample size is less clear for all three models compared to the correlation of the bias and study size.

Numerical robustness Appendix V of the Additional file 1 shows the results concerning the models' convergence. The numerical robustness is very satisfactory for all estimated time-to-event models. The Lin-Ying model performs most stable in terms of numerical robustness, always returning 1000 converged simulation runs. With respect to the Weibull model, the results are also very satisfactory. Only for two settings where data was generated from a Log-Logistic distribution, the model converges for 999 instead of 1000 simulation runs. The Log-Logistic model performs slightly less numerically robust with a minimum

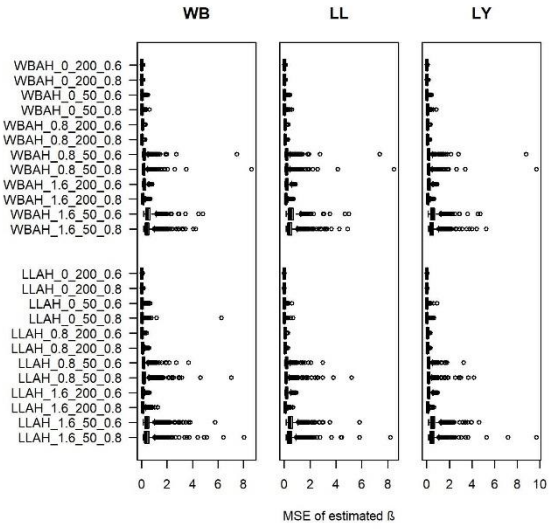

Fig. 2. Box plots of estimated median mean squared error (MSE) for the regression coefficient β over 1000 simulations for each setting. Y-axis denotes each setting with an ID consisting of the abbreviated true model (WBAH = Weibull additive hazard, LLAH = Log-Logistic additive hazard), the true β, the number of participants per study and number of events (e.g., Weibull additive hazard model, true $\beta = 0$, number of patients = 50, number of events = 60% results in "WBAH_0_50_0.6"). Left-most plot shows results for the Weibull (WB) additive hazard model, middle plot shows results for the Log-Logistic (LL) additive hazard model and right plot shows results for the Lin-Ying (LY) model.

of 983 converged runs. In 75% of the settings, the Log-Logistic model achieves convergence for all 1000 generated data sets.

Example: The HALLUCA study

For illustration purposes, we used data from the HALLUCA study, an epidemiological study that investigated medical care for lung cancer patients in the region of Halle (Saale) in the eastern part of Germany [3,25]. In the HALLUCA study, a total of 1696 lung cancer patients were observed between April 1996 and September 2000. 1349 patients (79.5%) died until the end of follow-up. The

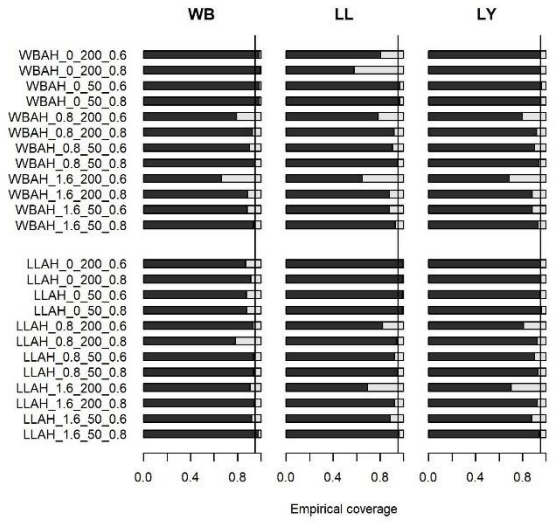

Fig. 3. Bar plot showing relative frequency of coverage for the regression coefficient β over 1000 simulations for each setting. Y-axis denotes each setting with an ID consisting of the abbreviated true model (WBAH = Weibull additive hazard, LLAH = Log-Logistic additive hazard), the true β, the number of participants per study and number of events (e.g., Weibull additive hazard model, true $\beta = 0$, number of patients = 50, number of events = 60% results in "WBAH_0_50_0.6"). Left-most plot shows results for the Weibull (WB) additive hazard model, middle plot shows results for the Log-Logistic (LL) additive hazard model and right plot shows results for the Lin-Ying (LY) model.

median survival time in the population was 284 days (0.78 years). For the analysis reported here, we focused on a dichotomized version of the TNM-scale for the classification of malignant tumors as a predictor, where 739 patients had a TNM-scale of IIIb or smaller (TNM < IV), and 621 patients had a TNM-scale of IV (TNM IV). The 336 remaining patients for whom the TNM-scale was not reported, were deleted. The starting point for survival definition relates to the day of diagnosis of lung cancer.

The Kaplan-Meier estimates for the two TNM classes and the estimated survival probabilities using the parametric additive hazard model with various distributions are given in figure 4. For the reference group (TNM < IV), for which

we can estimate the parametric baseline distribution, the observed mean survival time is 1.69 years (95% confidence interval [1.56; 1.81]). The observed median survival time is 1.10 years [0.96; 1.24]. The figure shows that the estimated survival curves from the additive hazard model fit well with the Kaplan-Meier curves. For an increasing time t there are greater differences between the curves, especially for the Exponential, Weibull, Gamma and Gompertz model. However, for the Log-Normal and Log-Logistic distribution the fit is also quite well for larger survival times.

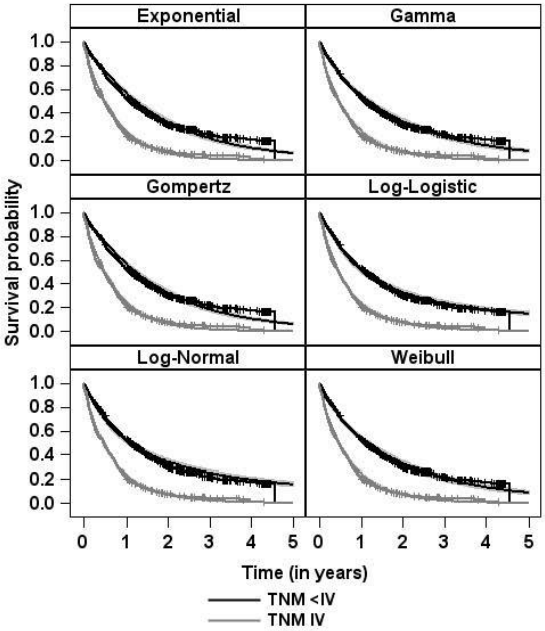

Fig. 4. Estimated survival probabilities from the parametric additive hazard model using different distributions compared to the Kaplan-Meier curves

In table 1, the results of the additive hazard model for six different baseline distributions are given. All models include a single binary covariate TNMIV, with (TNM < IV) being the reference group. Thus, we model the additive haz-

ard of being in TNM stage IV. To compare the model fit, we also report the Bayesian information criterion (BIC), where a smaller value indicates a better fit. The estimated regression coefficients for TNMIV show that being in stage IV increases the hazard by approximately 0.8. Being in an additive setting, the estimated β-value is added to the baseline distribution. That means being in the worse tumour group, the risk of dying rises because our model describes by how much the risk of dying changes in additive manner, not multiplicative. Alternatively, the coefficient could be interpreted in absolute terms, saying that when observing 100 people for one year, there are an additional 80 deaths among cancer patients in TNM class 4 compared to those with a lower TNM class. The results change slightly depending on the different distributions. Overall however, estimates from the parametric models are similar. In terms of the BIC, the Log-Logistic, Weibull and Gamma model fit best, returning almost identical parameter estimates and confidence intervals for β_{TNMIV}. With respect to the baseline location measures, the differences between the distributions are rather large, ranging from 1.77 to 7.90 for the baseline mean, and from 1.06 to 1.22 for the baseline median. For example, for the Log-Logistic approach, the model with the best BIC, we get a large baseline mean of 7.90 with a broad 95%-confidence interval of [3.39; 12.40]. With reference to figure 4, this can be explained by the flat survival curves at the end of the prediction range. Therefore, the baseline mean is necessarily larger with a broader confidence interval. However, regarding the baseline median, the Log-Logistic model yields an estimate of 1.06 which is in line with the empirical median of 1.10 from the Kaplan-Meier approach. The same holds true for the Log-Normal and Weibull model which achieve estimates that are also close to the empirical median.

Table 1. Estimates for the HALLUCA data set with 95% confidence intervals

Distribution	β_{TNMIV}	Baseline Mean	Baseline Median	No. of distribution parameters	BIC
Exponential	0.84 [0.72; 0.97]	1.77 [1.62; 1.92]	1.22 [1.12; 1.33]	1	2441.9
Weibull	0.78 [0.65; 0.91]	1.91 [1.70; 2.11]	1.15 [1.03; 1.26]	2	2427.1
Gamma	0.79 [0.67; 0.92]	1.85 [1.66; 2.03]	–	2	2431.8
Gompertz	0.84 [0.72; 0.97]	–	1.22 [1.12; 1.33]	2	2449.3
Log-Normal	0.80 [0.68; 0.92]	3.44 [2.75; 4.13]	1.07 [0.95; 1.19]	2	2450.3
Log-Logistic	0.78 [0.65; 0.91]	7.90 [3.39; 12.40]	1.06 [0.94; 1.18]	2	2417.6

Figure 5 compares the relative survival probability from the parametric additive hazard model assuming different baseline distributions. The relative survival curves, taken as the ratio of survival in a group of individuals with TNM stage IV in comparison to the survival of a corresponding population with TNM $<$ IV (see (4)), are rather similar between the various baseline distributions. It can

easily be seen that survival strictly decreases over time, with a relative survival probability almost equal to zero after five years.

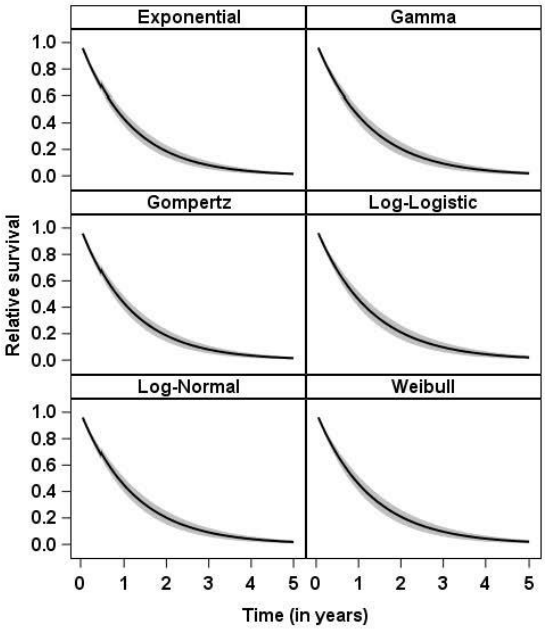

Fig. 5. Relative survival probabilities from the parametric additive hazard model using different distributions. The gray shaded area depicts estimated confidence intervals.

4 Discussion

In this article, we propose a new parametric additive hazard model for time-to-event outcomes. Generally, parametric survival models are advisable as they are simpler, more informative and robust than non- or semi-parametric models [19,9,7]. In addition, they have hazard, survival and probability density functions directly available. We showed that the model is valid from a theoretical point

of view. Further, in a simulation study as well as for an example from lung cancer research, we demonstrated that the model showed that the model works in practice. Convincingly, our proposed model can be implemented using any standard software that allows coding and maximizing a likelihood function. The corresponding SAS and R codes of our applications are publicly available on Zenodo [25].

The facilitated interpretation of our model is one of its most convincing advantages and is important for example from a clinical point of view. For instance, allowing to communicate outcomes of the additive hazard model in relative and absolute terms increases the comprehensibility of its results. Moreover, including time-independent (constant) covariates, our approach is advantageous over the additive hazard model of Aalen, since there, model coefficients often are difficult to understand as they change repeatedly over time and cannot be summarised easily. Furthermore, by modeling regression coefficients and distribution parameters that have more intuitive interpretations, our approach overcomes interpretational, mathematical and technical problems that arise when estimating hazard ratios as generic effect estimates. For practical application, several different distributions such as the Exponential, Weibull, Gamma, Gompertz, Log-Normal and Log-Logistic distribution can be implemented and compared via model selection criteria as the AIC or BIC. Moreover, parameter estimation of our model by the maximum likelihood principle is straightforward and can be accomplished with standard statistical software that allows for maximizing a likelihood function. Additionally, the proposed model is highly flexible in terms of possible extensions such as modeling correlated data by including random effects in the linear predictor, modeling non-linear covariate effects by splines, or by specifying baseline distributions that have more than two parameters. Maximum flexibility with respect to the baseline distribution can be obtained by using a piecewise-constant model, i.e., by dividing the observation period a-priori into intervals and assuming an exponential distribution in each of these intervals.

Evidently, there are some limitations with regards to our simulation study. The simulation was motivated by a real-world data example and thus offers a realistic setting, which we varied in a certain range of potential scenarios. However, these scenarios consider relatively small sample sizes and also rather moderate number of events. Consequently, it is not unlikely that the results of all models investigated in the simulation are biased in a certain extent and no optimal solution was found. For practical application, it should be considered positively that the treatment effect is more likely to be underestimated (versus overestimated). Thus, slightly biased inference in practical research may be less problematic than in a case of overestimation. Relying on the asymptotical properties and a correct implemented maximum likelihood estimation, we are confident that with an increased number of events and larger sample sizes, i.e., larger number of participants per study, the bias and its variability would asymptotically approach 0. Further, in practical research, it would evidently be an option to tune the parameters and settings of software functions used in the estimation to increase robustness and ensure convergence. In sum, further proofs and investigations

of other settings and data examples, asymptotic properties and comparisons to other existing methods are necessary to confirm our findings. Yet, we are confident that our proposed additive hazard model can be used (with caution) in applied settings.

With regards to the weaknesses of the proposed additive hazard model, our method suffers from the drawbacks that limits parametric versus semi- or non-parametric models. Non-parametric approaches generally require fewer assumptions about the data. Therefore, these may prove better when the true distribution is unknown and/or cannot be easily approximated. Vice versa, parametric methods are inherently dependent on the distribution chosen for estimation and on the assumption that this distribution is correctly specified. Further, it must be noted that parametric procedures require starting values for the optimization which may sometimes be complicated or problematic to define in practice. In case a distribution can be confidently specified to the data, parametric models will usually be more informative than semi- or non-parametric approaches. However, if this is not the case and the assumed distribution is false, results and conclusions are likely to be biased. For practical application, we recommend that the choice of the baseline distribution should be based on evidence from literature and previous research. However, if this is not applicable, and selection criteria such as the BIC or AIC is used for a data-based decision, this must be acknowledged in the reporting of the study and should be critically discussed. In that context, users should be cautious with any interpretation and should be are aware that confidence intervals for instance are potentially estimated too narrow. Besides, and by definition, our model assumes that the covariates act in an additive way to the baseline hazard function. Apparently, this assumption may be doubtful in some applications and in that case, other models may be more appropriate.

With regards to the framework of phases of methodological research Heinze et al. [12] , the proposed additive hazard model currently belongs to early stages of methodological development, i.e., phase I or II. The aim here was to introduce a new idea, as well as to demonstrate its validity and its potential to improve on existing methods. Into the bargain, we derived the new methodological idea while providing, logical reasoning and proofs of empirical evidence through a real-world data example and a simulation study in a (yet) relatively narrow but suitable target setting. Carefully planned method comparison studies that investigate the model in future works could advance the proposed additive hazard model to later stages of methodological development and would be of great value for future users and the scientific community as a whole. Therefore, there is a need for future studies that explore the empirical properties of our model in a wider range of problems, highlight its advantages and limitations, and possibly uncover previously unknown behavior (e.g., in simulations with wide range of scenarios and different outcome types and realistic or complex comparative example data analyses).

5 Conclusion

To summarize, the proposed parametric additive hazard model can serve as a powerful tool to analyze time-to-event outcomes. By definition, it simplifies interpretation, facilitates parameter estimation and permits greater flexibility than most existing and commonly used methods.

6 Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Zenodo repository, https://doi.org/10.5281/zenodo.7124989 [25].

Competing interests

The authors declare that there is no conflict of interest.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Authors' contributions

DV is the corresponding author of this work and, as such, had full access to all data in the study and takes responsibility for the integrity of the data and the accuracy of the analysis. AH and OK derived the methodology. AH supervised the study and, together with DV, developed the theoretical concept and outline of the present study. With support of AK and AF, AH and DV performed the formal analysis. DV and AH wrote the manuscript. All authors discussed the results, reviewed and edited the manuscript and the statistical analyses.

Acknowledgements

Not applicable

7 Additional files

The following additional information are available as part of the online article:

(i) Appendix I - Derivation of the additive hazard model
(ii) Appendix II - Median bias for the estimated regression coefficient β
(iii) Appendix III - Median mean squared error (MSE) for the estimated regression coefficient β
(iv) Appendix IV - Empirical coverage (in %) for the estimated regression coefficient β
(v) Appendix V - Number of converged simulation runs per setting (maximum number: 1000)

References

1. Aalen, O.O.: A linear regression model for the analysis of life times. Statistics in medicine **8**(8), 907–925 (1989)
2. Aalen, O.O., Cook, R.J., Røysland, K.: Does cox analysis of a randomized survival study yield a causal treatment effect? Lifetime data analysis **21**(4), 579–593 (2015)
3. Bollmann, A., Blankenburg, T., Haerting, J., Kuss, O., Schütte, W., Dunst, J., Neef, H.: Survival of patients in clinical stages i–iiib of non-small-cell lung cancer treated with radiation therapy alone. Strahlentherapie und Onkologie **180**(8), 488–496 (2004)
4. Bradburn, M.J., Clark, T.G., Love, S.B., Altman, D.G.: Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods. British journal of cancer **89**(3), 431–436 (2003)
5. Brinks, R., Landwehr, S., Fischer-Betz, R., Schneider, M., Giani, G.: Lexis diagram and illness-death model: simulating populations in chronic disease epidemiology. PLoS one **9**(9), e106043 (2014)
6. Collett, D.: Modelling survival data in medical research. CRC press (2015)
7. Columbia: (Oct 2022), https://www.publichealth.columbia.edu/research/population-health-methods/time-event-data-analysis#
8. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)
9. Crowther, M.J., Lambert, P.C.: A general framework for parametric survival analysis. Statistics in medicine **33**(30), 5280–5297 (2014)
10. Dickman, P.W., Sloggett, A., Hills, M., Hakulinen, T.: Regression models for relative survival. Statistics in medicine **23**(1), 51–64 (2004)
11. Greenland, S.: Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. Epidemiology pp. 498–501 (1996)
12. Heinze, G., Boulesteix, A.L., Kammer, M., Morris, T.P., White, I.R., of the STRATOS initiative, S.P.: Phases of methodological research in biostatistics—building the evidence base for new methods. Biometrical Journal p. 2200222 (2022)
13. Hernán, M.A.: The hazards of hazard ratios. Epidemiology (Cambridge, Mass.) **21**(1), 13 (2010)
14. Hu, J.: addhazard: Fit additive hazards models for survival analysis; 2020. R package version **1**(0)

15. Kuss, O., Blankenburg, T., Haerting, J.: A relative survival model for clustered responses. Biometrical Journal: Journal of Mathematical Methods in Biosciences **50**(3), 408–418 (2008)
16. Lin, D.Y., Ying, Z.: Semiparametric analysis of the additive risk model. Biometrika **81**(1), 61–71 (1994)
17. Lindsey, J.: A study of interval censoring in parametric regression models. Lifetime data analysis **4**(4), 329–354 (1998)
18. Martinussen, T., Vansteelandt, S.: On collapsibility and confounding bias in cox and aalen regression models. Lifetime data analysis **19**(3), 279–296 (2013)
19. Nardi, A., Schemper, M.: Comparing cox and parametric models in clinical studies. Statistics in medicine **22**(23), 3597–3610 (2003)
20. Reid, N.: A conversation with sir david cox. Statistical Science **9**(3), 439–455 (1994)
21. Schaubel, D.E., Wei, G.: Fitting semiparametric additive hazards models using standard statistical software. Biometrical Journal **49**(5), 719–730 (2007)
22. Sjölander, A., Dahlqwist, E., Zetterqvist, J.: A note on the noncollapsibility of rate differences and rate ratios. Epidemiology **27**(3), 356–359 (2016)
23. Stensrud, M.J.: Interpreting hazard ratios: Insights from frailty models. arXiv preprint arXiv:1701.06014 (2017)
24. Sutradhar, R., Austin, P.C.: Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. Annals of epidemiology **28**(1), 54–57 (2018)
25. Voeltz, D., Hoyer, A.: Parametric additive hazard model for time-to-event analysis (2022), https://doi.org/10.5281/zenodo.7124989
26. Zahl, P.H., Aalen, O.O.: Adjusting and comparing survival curves by means of an additive risk model. Lifetime Data Analysis **4**(2), 149–168 (1998)

5.3 Statistical Modelling and the Plea for Neutral Comparison Studies

5.3.1 Contribution 5 – A systematic Comparison of Different Approaches for Meta-Analysis of Diagnostic Accuracy Studies with multiple Thresholds - A Simulation Study

Manuscript in Preparation Zapf, A., Frömke, C., Hardt, J., Rücker, G., Voeltz, D., Hoyer, A. A systematic comparison of different approaches for meta-analysis of diagnostic accuracy studies with multiple thresholds - A simulation study. *Currently submitted at Biometrical Journal.*

Status Currently under review at Biometrical Journal.

Code and Data 10.5281/zenodo.7802089

Author Contributions A.Z., C.F., J.H., G.R. and A.H. developed the concept of the simulation study. AH simulated the data. A.Z., C.F., G.R. and A.H. wrote code to apply the models to the simulated data. A.H., D.V. and G.R. programmed and analysed the simulation results. D.V. created the figures and tables. A.Z., D.V. and A.H. wrote the first draft of the manuscript. All authors discussed the results, revised the manuscript and approved the final version.

Biometrical Journal

A systematic comparison of different approaches for meta-analysis of diagnostic accuracy studies with multiple thresholds - A simulation study

Journal:	*Biometrical Journal*
Manuscript ID	bimj.202300101
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	05-Apr-2023
Complete List of Authors:	Zapf, Antonia; Universitatsklinikum Hamburg-Eppendorf Frömke, Cornelia; Hochschule Hannover Fakultat III Medien Information und Design Hardt, Juliane; Tierarztliche Hochschule Hannover Rücker, Gerta; Universitatsklinikum Freiburg Voeltz, Dina; Universitat Bielefeld Medizinische Fakultat OWL; Ludwig Maximilians University Munich, Department of Statistics Hoyer, Annika; Universitat Bielefeld Medizinische Fakultat OWL
Keywords:	Diagnostic test accuracy studies, Meta-analysis, Multiple thresholds, ROC curve, Simulation study

A systematic comparison of different approaches for meta-analysis of diagnostic accuracy studies with multiple thresholds - A simulation study

Antonia Zapf[1], Cornelia Frömke[2], Juliane Hardt[3], Gerta Rücker[4], Dina Voeltz[5,6*], Annika Hoyer[5*]

[1] Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Christoph-Probst-Weg 1, Hamburg, Germany

[2] Department of Information and Communication, Faculty for Media, Information and Design, University of Applied Sciences and Arts Hannover, Expo Plaza 12, Hannover, Germany

[3] Department of Biometry, Epidemiology and Information Processing, WHO Collaborating Centre for Research and Training for Health at the Human-Animal-Environment Interface, University of Veterinary Medicine Hannover, Hannover, Germany

[4] Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Stefan-Meier-Straße 26, Freiburg, Germany

[5] Biostatistics and Medical Biometry, Medical School OWL, Bielefeld University, Universitätsstr. 25, Bielefeld, Germany

[6] Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstr. 33, Munich, Germany

* These authors have contributed equally to this work.

ORCIDs:
Antonia Zapf: https://orcid.org/0000-0002-8467-0508
Cornelia Frömke: https://orcid.org/0000-0002-4028-0604
Juliane Hardt: https://orcid.org/0000-0002-4514-5698
Gerta Rücker: https://orcid.org/0000-0002-2192-2560
Dina Voeltz: https://orcid.org/0000-0002-7579-7144
Annika Hoyer: https://orcid.org/0000-0002-0241-9951

Correspondence: Antonia Zapf, Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Christoph-Probst-Weg 1, Hamburg, Germany, a.zapf@uke.de

Abstract

The development of methods for the meta-analysis of diagnostic test accuracy (DTA) studies is still an active area of research. While methods for the standard case where each study reports a single pair of sensitivity and specificity are nearly routinely used today, there is a lack of recommendations which method should be used to meta-analyze ROC curves. This situation is more complex, as each DTA study may report on several pairs of sensitivity and specificity, each corresponding to a different threshold. Various methods have been published to accomplish this task and their applicability to analyze real-world data has already been shown. However, no simulation study exists that systematically compares different approaches with respect to their performance in various scenarios when the truth is known. In this article, we aim to fill this gap and show the results of a simulation study that compares three frequentist approaches for the meta-analysis of ROC curves. We assessed different scenarios that are motivated by an example from medical research. All three approaches worked partially well. The approach by Hoyer and colleagues was slightly superior in most scenarios and is recommended in practice.

Key words: Diagnostic test accuracy studies; Meta-analysis; Multiple thresholds; ROC curve; Simulation study.

Word count: 4552

1. Introduction

Systematic reviews in medical research aim to summarize the results of individual studies and thus to create a common body of evidence using meta-analysis. Diagnostic test accuracy (DTA) studies aim to evaluate how well a diagnostic test (for example, a score, a biomarker, or an imaging parameter) correctly distinguishes between two groups which are characterized by having a certain target condition of interest (for example, a disease) or not. This results in a bivariate outcome, which is composed of the probability of a positive test result for the individuals with the target condition (the sensitivity) and the probability of a negative test result for the individuals without the target condition (the specificity). Methods for bivariate meta-analyses of DTA studies were developed some time ago, the most well-known approaches being those of Rutter and Gatsonis (2001) and of Reitsma et al. (2005) and Chu and Cole (2006). These approaches take heterogeneity and correlation between individual studies that result from different threshold values across studies into account. However, they do not explicitly consider study-specific threshold values. Therefore, a summary receiver operating characteristic (SROC) curve cannot uniquely be estimated (Arends et al., 2008). Furthermore, these approaches do not allow considering several and different numbers of threshold values per study.

To overcome these limitations, special approaches have been developed: a two-stage random-effects model by Steinhauser et at. (2016), a time-to-event model by Hoyer et al. (2018), a Bayesian model with multinomial likelihoods by Jones et al. (2019), and a semiparametric model by Frömke et al. (2022). Zapf et al. (2021) demonstrated, using a meta-analysis on the evaluation of the diagnostic accuracy of a biomarker for the detection of acute kidney injury (Albert et al., 2018), that these four methods may lead to different results. The differences in threshold-specific estimates of sensitivity and specificity were partly considerable (up to 43% difference in specificity). By contrast, Benedetti et al. (2020), comparing the methods by Steinhauser and Jones to the standard bivariate approach for investigating the diagnostic accuracy of the Patient Health Questionnaire-9 (PHQ-9) depression screening tool, found no large differences between methods. Since in a real data example the truth is not known, no recommendation could be made which approach should be used in practice and under which circumstances.

This motivated us to conduct a simulation study that allows for a fair comparison of the frequentist approaches by Steinhauser et al. (2016), Hoyer et al. (2018), and Frömke et al. (2022) under various scenarios where the ground truth is known. We do not consider the approach by Jones et al. (2019), as it is a Bayesian approach and the author informed us that the run times would be too large to make the method feasible for our simulation study (Hayley Jones, personal communication). We finally aim to give recommendations which method should be used in which situation. In Section 2, the meta-analysis approaches are briefly presented. Section 3 describes the simulation study in detail. The results are presented in Section 4 and discussed in Section 5. In Section 6 we summarize the results and make recommendations.

2. Methods

This section briefly introduces the different approaches that are compared. For more details, we refer to the original publications.

2.1 Notation

We index the studies in a meta-analysis by $k = 1, ..., N$ and the true disease status of the participants by $d = 0$ (individuals without the target condition, "non-diseased") and $d = 1$ (individuals with the target condition, "diseased"). The number of diseased and non-diseased individuals in study k is denoted by n_{kd}. Individual study participants are indexed by $s = 1, ..., n_{kd}$. The exact result of the continuous diagnostic test per individual s in disease state d in study k is labeled X_{kds}. Different diagnostic thresholds given by the single studies are indexed by $i = 1, ..., t_k$. The numerical threshold values that correspond to these indices are denoted by c_{ki}.

The diagnostic test is finally evaluated in terms of the two measures sensitivity (se) and specificity (sp).

2.2 The random effects model by Steinhauser et al. (2016)

Steinhauser et al. (2016) proposed a two-stage random effects model for the meta-analysis of diagnostic studies accounting for varying thresholds. They assumed at the study level that, for each value of the threshold, the observed sp and $1 - se$ is transformed based on an appropriate quantile function f. In line with Zapf et al. (2021), we used a log-transformation of the threshold values and the logit-function for f, defined as

$$\text{logit}(x) = \log(x) - \log(1 - x) \quad . \qquad (1)$$

At the meta-analysis level, a linear mixed model is applied to fit the resulting values across studies. For our simulation study and conforming to Zapf et al. (2021), we used the model given by

$$\text{logit}(sp_{ki}) = \alpha_0 + a_{0k} + (\beta_0 + b_{0k}) \log(c_{ki}) + \epsilon_{ki} \qquad (2)$$

$$\text{logit}(1 - se_{ki}) = \alpha_1 + a_{1k} + (\beta_1 + b_{1k}) \log(c_{ki}) + \delta_{ki}. \qquad (3)$$

This model corresponds to the assumption of log-logistic distributions for X_{k0s} and X_{k1s}. The crude estimates of the observed values of sensitivity and specificity at threshold c_{ki} in study k are denoted by

$\widehat{se}_{ki}$ and $\widehat{sp}_{ki}$. The terms α_0 and α_1 represent fixed intercepts, whereas β_0 and β_1 are fixed slopes in the group of diseased and non-diseased participants, respectively. Random intercepts and slopes are denoted by a_{0k}, a_{1k}, b_{0k}, and b_{1k}. They are assumed to follow a common four-dimensional normal distribution, allowing for correlation across studies. ε_{ki} and δ_{ki} are the within-study random error terms. Data points are weighted using the inverse variance of the corresponding logit-transformed proportion. Model-based distribution functions for the diseased and non-diseased are then provided by back-transforming the fixed effects. Finally, the obtained distributions are used to estimate the summary ROC curve with pointwise confidence regions. The area under the curve (AUC) can be obtained by applying the trapezoidal rule.

The model is implemented in the `R` package `diagmeta` (version 0.5-0, Rücker et al., 2022) in the freely available software environment `R` (R Development Core Team, 2021).

2.3 The time-to-event model by Hoyer et al. (2018)

The model developed by Hoyer et al. (2018) can be embedded in the class of time-to-event models. The central idea is that the diagnostic test values can be assumed to be interval-censored as it is only known if these values (and how many of them) lie above or below a predefined threshold. To model the interval-censored diagnostic test values, it is possible to assume various distributions (Hoyer and Kuss, 2020), for example, the Weibull, log-normal or log-logistic distribution. According to time-to-event models, the 'events' of interest in the setting of diagnostic accuracy studies are to be tested positive or negative in the population of diseased or non-diseased, respectively. In addition, the diagnostic test values indicate the 'time' scale. Sensitivity and 1-specificity are then the event probabilities in the population of diseased and non-diseased, respectively. The outcome of interest, i.e. the diagnostic test values, is log-transformed which leads to an accelerated failure time model with a unified linear predictor. Formally, the model is given by

$$\log(x_{k0}) = b_0 + \varepsilon_0 + u_{k0} \quad (4)$$

$$\log(x_{k1}) = b_1 + \varepsilon_1 + u_{k1} \quad (5)$$

with

$$\begin{pmatrix} u_{k0} \\ u_{k1} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix}\right]. \quad (6)$$

Location parameters after the log-transformation of the outcome are denoted by b_0 and b_1, ε_0 and ε_1 indicate error terms with distributions corresponding to the log-transformed diagnostic test values x_{k0} and x_{k1} in the population of non-diseased and diseased, respectively. Study-specific random effects u_{k0} and

u_{k1} are assumed to be bivariate normally distributed with correlation parameter ρ and variances σ_0^2 and σ_1^2. The random effects are included to account for potential between-study heterogeneity and correlations. They are added to the location parameters after log-transforming the outcome. Finally, sensitivities and specificities are predicted at several thresholds based on the estimated survival function leading to the SROC curve. The related AUC is estimated using the trapezoidal rule.

The approach was originally implemented in SAS (SAS Institute Inc., Cary, NC, USA). Source code is available in Hoyer et al. (2018).

2.4 The semiparametric model by Frömke et al. (2022)

Frömke et al. (2022) proposed an alternative semiparametric approach for the meta-analysis of diagnostic test accuracy studies that explicitly avoids the assumption of parametric distributions of the test values in the population of diseased and non-diseased. Essentially, the model is an extension of a nonparametric method for diagnostic studies (Konietschke and Brunner, 2009), which was transferred to meta-analysis of diagnostic accuracy studies (Zapf et al., 2015).

In terms of the semiparametric approach by Frömke et al. (2022), the AUC equals the relative measure $p = P(X_{k0s} < X_{k1s}) + \frac{1}{2}P(X_{k0s} = X_{k1s})$. The AUC is estimated by replacing all measurements X_{kds} by their global mid-ranks R_{kds} and calculating the mean rank $\overline{R}_{.d.}$ of all individuals with disease status d over all studies. Then, it holds $\widehat{AUC} = \frac{1}{2} + \frac{1}{n}(\overline{R}_{.1.} - \overline{R}_{.0.})$ with $n = \sum_{k=1}^{N}\sum_{d \in [0,1]} n_{kd}$ as the total number of diseased and non-diseased participants across all studies.

In case of diagnostic studies only aggregated data are available whereas actually information on individual participant level X_{kds} is needed to calculate the ranks. Instead, the number of diseased and non-diseased individuals with diagnostic test values below or greater than the study-specific thresholds c_{ki} is known. Using this information, fictitious data from the uniform distribution can be generated between the thresholds c_{ki} and c_{ki+1}. Below and above the first/last threshold, the data are drawn uniformly distributed from the interval from the smallest/largest threshold used in the whole meta-analysis. It should be noted that in Zapf et al. (2021), the one-point distribution was used to generate fictitious data. During the review process of the article about the semiparametric approach (Frömke et al., 2022), however, it turned out that the uniform distribution leads to better results than the one-point distribution. Accordingly, we use the uniform distribution.

To estimate sensitivity and specificity at a specific threshold, the procedure is the same, but transformed data is needed. In terms of sensitivity, observations of the non-diseased individuals have to be replaced by a one-point distribution. Accordingly, for estimating specificity, observations of diseased individuals are replaced by a one-point distribution (Lange and Brunner, 2012). Based on the asymptotic equivalence theorem it holds that $\sqrt{N}(\hat{p} - p) \sim N(0,\sigma^2)$. Consequently, Wald confidence intervals for AUC, sensitivity and specificity can be derived, applying logit transformations to ensure an interval with appropriate boundaries of $[0,1]$.

The model is implemented in R (R Development Core Team, 2021) and the semiparametric approach was computed with the package diagnostic 0.4.4 (Rooney, 2017).

3. Simulation study

To enable a fair comparison between the three approaches presented in Section 2, we conducted a simulation study. We used SAS 9.4 (SAS Institute Inc., Cary, NC, USA) for the data generation whereas for estimation, we used the software in which the models were originally implemented. The results of the approaches of Steinhauser et al. (2016) and Frömke et al. (2022) were obtained using R (R Development Core Team, 2021), and results of the approach of Hoyer et al. (2018) using SAS 9.4. Results were then summarized and visualized in R (R Development Core Team, 2021). Data and code are openly available in Zenodo at https://zenodo.org/record/7802090#.ZC1i8oTP02w (Voeltz et al., 2023).

3.1 Design and setting

The example on kidney disease presented in Zapf et al. (2021) motivates the setting of our simulation study. We generated data from a Weibull-, log-logistic and log-normal distribution for the diagnostic test values in the population of diseased and non-diseased using the time-to-event model of Hoyer et al. (2018) as true underlying model. Additionally, we assumed a normal distribution, which results in all estimated models being misspecified. We varied the true underlying AUCs (0.7, 0.8 and 0.9), the true disease prevalence (0.02, 0.2 and 0.5) and the correlation between the random effects which models heterogeneity across studies (0.3, 0.6 and 0.9). Table A1 in the Appendix shows how the different model parameters were varied to mimic the true AUCs and plausible distributions of the diagnostic test values. Thereby, b_0, b_1, ϕ_0 and ϕ_1 indicate the location and scale parameter of the assumed distributions. σ_0^2, σ_1^2 and ρ correspond to the values of the covariance matrix of the random effects as given in equation (6). The four distributions are illustrated for the first and last scenario (from Table A1) in Figure A1. Table 1 indicates the resulting true sensitivities and specificities for the given values of the AUC at which we evaluated the various models.

Table 1. True simulated sensitivities and specificities at the thresholds used for model evaluation.

Model	Threshold	Sensitivity, % (AUC=0.7)	Sensitivity, % (AUC=0.8)	Sensitivity, % (AUC=0.9)	Specificity, %
Weibull	50	99.4	99.7	99.9	1.7

	100	69.9	82.0	92.1	64.3
	130	18.8	39.5	68.1	99.2
	150	2.1	11.6	41.0	100
	200	0	0	0.8	100
Log-normal	50	100	100	100	0.7
	100	53.0	71.7	89.0	84.8
	130	10.8	23.0	46.5	99.0
	150	2.5	7.3	21.1	99.9
	200	0	0.2	1.2	100
Log-logistic	50	99.8	99.9	100	1.3
	100	53.3	73.4	89.7	86.0
	130	10.1	21.3	46.1	98.4
	150	3..1	7.1	19.4	99.6
	200	0.2	0.6	1.9	99.6
Normal	50	93.0	97.1	99.3	15.9
	100	58.9	74.2	89.0	74.7
	130	30.0	46.0	68.3	95.2
	150	15.3	27.4	49.0	99.0
	200	1.1	3.2	10.1	100

3.2 Data generation

All parameter variations led to 108 different simulation scenarios. For each scenario, we generated 1000 meta-analyses. The number of studies per meta-analysis ranged from 5 to 30 and was sampled from a uniform distribution. The number of participants per single study was drawn from a uniform distribution with a minimum of 30 and a maximum of 300 participants. The number of diseased individuals was equal to the true underlying prevalence multiplied by the total number of study participants. The number of non-diseased individuals was obtained as the difference between the total number and the number of diseased study participants. To mimic real world scenarios, we additionally varied the number of thresholds per study from 1 to 10 which was generated from a uniform distribution. Final threshold values were drawn from uniform distributions depending on the number of simulated thresholds as shown in Table A2. We rounded the numbers of studies per meta-analysis, participants per study and thresholds to the nearest integer.

The exact diagnostic test values of each study participant were generated from the assumed true distributions (Weibull, log-normal, log-logistic, normal). For this, we first simulated the random effect (equation (6)) with respect to the true values of σ_0^2, σ_1^2 and ρ. This random effect was added to the true b_0 and b_1 (equations (4) and (5)). Considering the scale parameters of the different distributions ϕ_0 and ϕ_1, the diagnostic test values of each study participant were generated on the original scale using inverse transform sampling. Finally, the individual diagnostic test values were compared to the generated thresholds to classify potential true positives or true negatives. From this information, we constructed the diagnostic contingency table as input data for our simulation.

3.3 Estimation methods and outcomes

For each meta-analysis, we estimated the models proposed in Section 2. For the approach of Hoyer et al. (2018), we assumed that the interval-censored diagnostic test values are Weibull distributed. For parameter estimation, we used `SAS PROC NLMIXED` with the default settings. Initial values for the fixed effect parameter were achieved from two univariate models without random effects using `SAS PROC LIFEREG`. The raw estimates of σ_0^2, σ_1^2 and ρ from the univariate fits serve as starting values for the random effects parameters. For the approach of Steinhauser et al. (2016), we used a log-transformation of the threshold values and the logit-function as given in equation (1). The model proposed by Frömke et al. (2022) does not need any further specification of a distribution.

For model evaluation and comparison, we estimated sensitivities and specificities at the predefined threshold values of 50, 100, 130, 150 and 200 together with their 95% confidence intervals. As measures of interest, we computed the absolute bias, the mean squared error (MSE) and the empirical coverage of the two-sided 95% confidence intervals at each threshold, as well as the number of converged simulation runs to quantify numerical robustness (Burton et al., 2006; Morris et al., 2019).

4. Results

In reporting our results, we confine ourselves to the bias, coverage and convergence. The Appendix provides additional results.

4.1. Bias

Figures 1, 2, A2 and A3 show the distribution of the bias of the estimated sensitivities and specificities at different thresholds for all models. In many cases, the performances of the estimated models were similar when comparing them for the same scenario. Predominantly, the method of Hoyer et al. (2018) slightly outperforms the other two approaches.

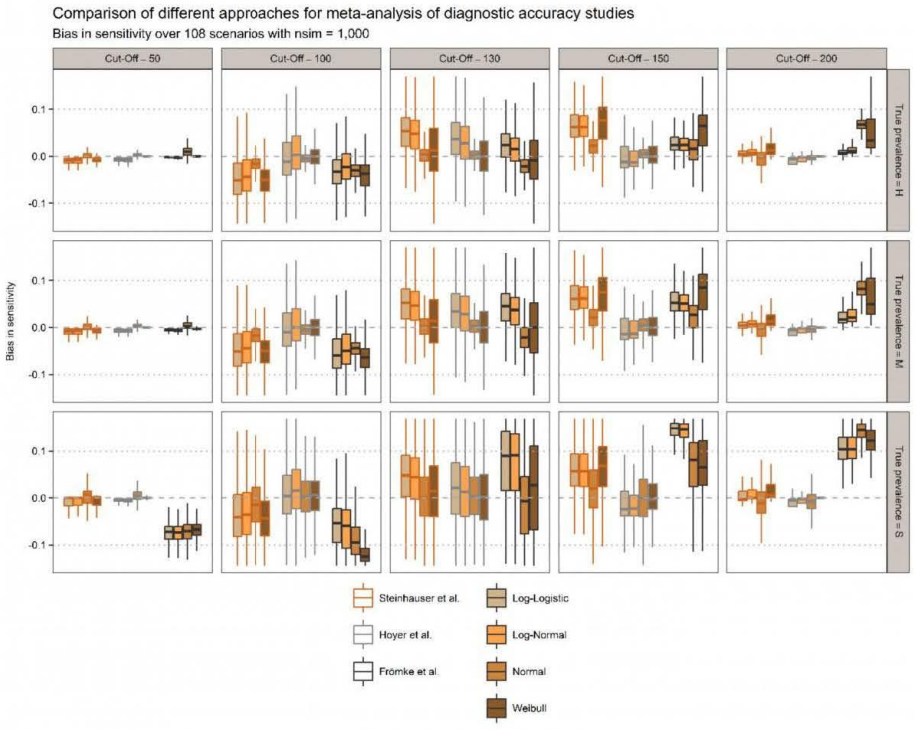

Figure 1. Distribution of the bias of sensitivity depending on the true underlying distribution and the true prevalence for the different models. The boxplots represent the median, the upper and lower quartile and

the two whiskers, which extend from the lower/upper quartile to the lowest/highest value within 1.5 times the inter-quartile range.

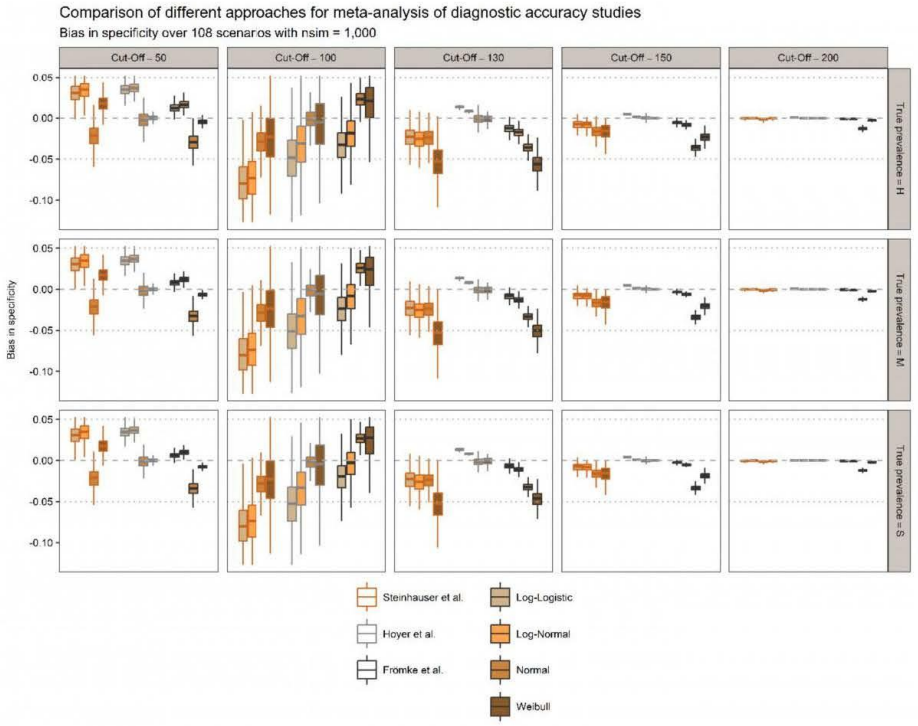

Figure 2. Distribution of the bias of specificity depending on the true underlying distribution and the true prevalence for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, which extend from the lower/upper quartile to the lowest/highest value within 1.5 times the inter-quartile range.

Regarding sensitivity, the method of Frömke et al. (2022) performed inferior compared to the other approaches. Often, its results were more biased and showed larger variability. Considering the underlying distribution, all methods were least biased in scenarios assuming a normal distribution. For these scenarios, particularly the methods of Hoyer et al. (2018) and Steinhauser et al. (2016) showed satisfactory results with almost no bias. Assuming very small or very large thresholds (i.e., 50 or 200), sensitivity results were again satisfactory for both methods with almost no bias and minimal variability. This was not the case with the semiparametric approach. Overall, results seemed to improve and become more stable when the prevalence increased. The impact of the correlation on the bias in sensitivity seemed negligible.

In terms of specificity and particularly for thresholds 50 and 100 the methods of Hoyer et al. (2018) and Steinhauser et al. (2016) returned rather insufficient values. Nonetheless, results of all methods improved and became more stable for higher threshold values. However, this pattern did not occur for threshold values of 50, where the bias and its variability were indeed smaller than for scenarios with a threshold of

100. Only for high thresholds, the method of Hoyer et al. (2018) and Steinhauser et al. (2016) provided promising results with almost no bias and outperform the slightly more biased approach of Frömke et al. (2022). Particularly noteworthy is the threshold of 100, where all methods were most variable and biased to their maximal degree. As for specificity, the methods of Hoyer et al. (2018) and Steinhauser et al. (2016) performed satisfactorily with almost no bias when assuming a normal distribution. Otherwise, the relation between the true distribution and the bias in specificity remains unclear. The effect of the underlying correlation seemed negligible.

Figures A4 and A5 display the bias in the estimated AUC. Due to some large outliers, the graphs are limited to the 1st to 99th percentile of all simulation results. Therefore, they do not always contain information on estimated minima and maxima. The highest positive bias found across all 108 scenarios was approximately 0.3 for the method of Hoyer et al. (2018), and 0.2 for Steinhauser et al. (2016) and for Frömke et al. (2022). The highest negative bias was estimated -0.7, -0.4 and -0.3, respectively. Overall, results were satisfactory for the approach of Hoyer et al. (2018) and only slightly inferior for the method of Steinhauser et al. (2016). The semiparametric approach returned comparable, but more variable, results. Generally, scenarios assuming higher prevalence were least variable. The bias of the AUC and its variability improved for the methods of Hoyer et al. (2018) and Steinhauser et al. (2016) for higher values of the prevalence and higher values of the true AUC. Vice versa, small true values of the AUC and small prevalence values particularly impaired results of the method of Frömke et al. (2022). Further, results improved across all methods when assuming a normal or Weibull distribution. As for sensitivity, the approach of Hoyer et al. (2018) returned promising, almost unbiased results for scenarios based on a normal distribution. Contrarily, for scenarios based on the log-logistic or log-normal distribution, the bias and variability were higher. Regarding correlation, no clear relation was apparent, neither on the bias itself nor on its variability.

Besides visualizing the distribution of the bias, Figures 1 and 2 provide information on the variance of the estimation. In addition, Figures A6 to A11 show the distribution of the squared bias and thus gives a flavor of the MSE.

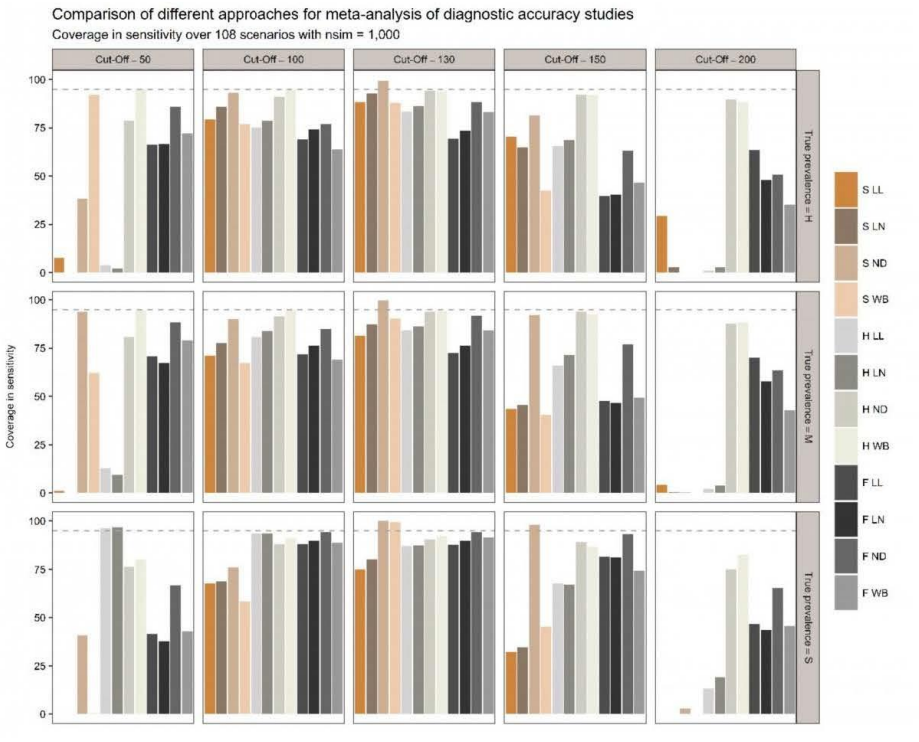

Figure 3. Coverage to the two-sided 95% level of sensitivity depending on the true underlying distribution and the true prevalence for the different models.

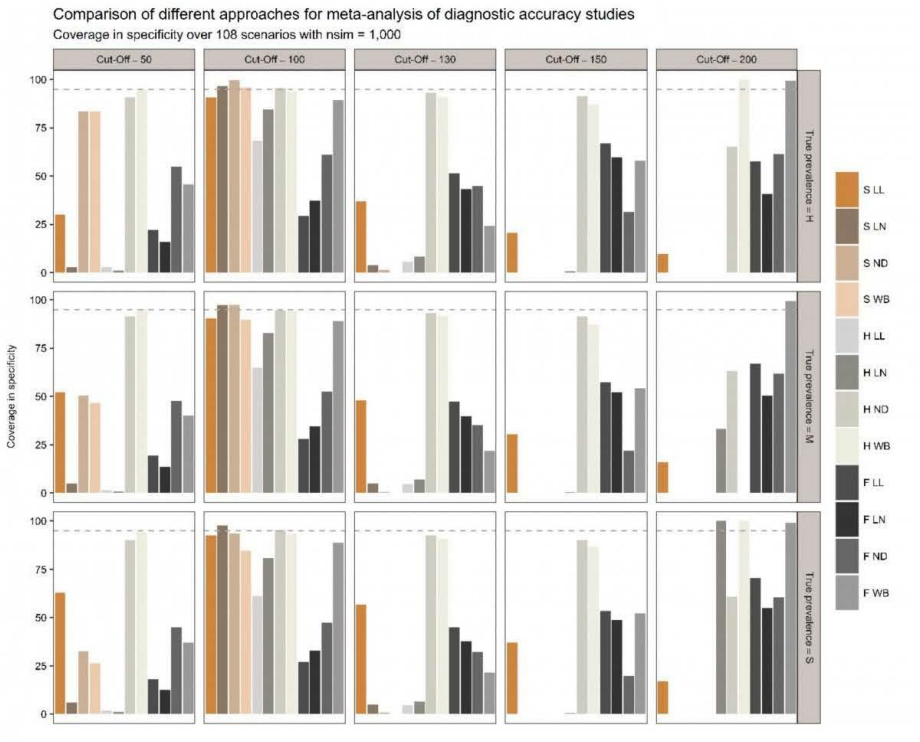

Figure 4. Coverage to the two-sided 95% level of specificity depending on the true underlying distribution and the true prevalence for the different models.

4.2. Coverage

Figure 3 and 4 display coverage results regarding sensitivity. For most scenarios, results were comparable for all methods and with few exceptions always below the nominal level. Results were quite similar across methods for medium threshold values (100, 130 and 150). By contrast, in scenarios with low or high threshold values (50 and 200), results were considerably lower. Comparing the approaches, the method of Frömke et al. (2016) was least affected by changes in the threshold. In most cases, coverage results were best in scenarios based on the normal or Weibull distribution. Regarding the prevalence, results of the method of Steinhauser et al. (2016) and Hoyer et al. (2016) improved for higher levels of the prevalence. The relation between the empirical coverage for sensitivity and the prevalence was less obvious for the remaining two approaches. Neither Figure 3 nor Figure 4 reveals any relation between the sensitivity coverage and the correlation.

For most scenarios, the empirical coverage for specificity was below the nominal level for all approaches (Figures A12 and A13). Solely for scenarios based on a Weibull or normal distribution, the approach by Hoyer et al. (2018) returned satisfactory coverage results close to 95%. Otherwise, no clear relation with the underlying distribution was visible. For the method of Steinhauser et al. (2016), coverage results were promising and close to the 95% level only for thresholds close to 100. Contrarily, results of the approach of Frömke et al. (2022) seemed to improve when the threshold increased. Coverage results for the method of Hoyer et al. (2018) remained mostly unaffected when changing the threshold. Nonetheless, the relation between the empirical coverage for specificity and the threshold seems less obvious and less impactful compared to results for sensitivity. Again, the effect of the underlying correlation seemed negligible. The results did not show any clear impact of the prevalence.

Figure A14 and A15 illustrate the estimates regarding the empirical coverage for the AUC. A comparison with the approach of Hoyer et al. (2018) is not possible because runtime issues rendered the computation of confidence interval estimators too time-consuming as they are based on bootstraps. In most cases, the AUC coverage for the illustrated two methods was comparable. Though, with only few exceptions, the parametric model by Steinhauser et al. (2016) was superior to the non-parametric model by Frömke et al. (2022). Particularly in scenarios with medium prevalence and medium or small values of the true AUC, the approach of Steinhauser et al. (2016) performed satisfactorily. For the remaining scenarios, the method of Steinhauser et al. (2016) still outperformed the method of Frömke et al. (2022) when assuming a normal or Weibull distribution. This does not hold true for scenarios based on the log-logistic or log-normal distribution. On the contrary, in case of a small prevalence and high true AUC, the method of Frömke et al. (2022) was close to 95% coverage for all distributions, while the method of Steinhauser et al. (2016) returned poor results ranging from about 48% (for the Weibull distribution) to 82% (for the normal distribution). The impact of the correlation seemed minor.

4.3. Convergence

Table A3 provides information on the models' convergence. Per definition, the semiparametric method of Frömke et al. (2022) is numerically robust, always returning 100% converged simulation runs. With respect to the approach of Steinhauser et al. (2016), results are also adequate. Across all scenarios, this method converges for at least 79% and a maximum of 93% of the simulation runs. The approach of Hoyer et al. (2018) performs less numerically robust, with a minimum of 22% converged for normally distributed data and a maximum of 92% converged runs.

5. Discussion

In the article by Zapf et al. (2021), four different approaches for the meta-analysis of diagnostic studies with multiple or different thresholds were compared using an example meta-analysis. Because some of the results differed substantially, the goal of our comprehensive and systematic simulation study was to

investigate which approach has the best statistical properties under which conditions. The approach by Jones et al. (2019) could not be included due to its long runtime in combination with the very large simulation data set, which is a limitation of this study. For the three included approaches by Steinhauser et al. (2016), Hoyer et al. (2018), and Frömke et al. (2022), it has been shown that, with exceptions, the approach by Hoyer et al. (2018) has the smallest bias and the empirical coverage probability is closest to 95%. A further advantage of this approach, as well as the approach from Jones et al. (2019), is the great flexibility, which also allows meta-analyses based on non-normally distributed individual data. In many scenarios, the results of the approaches by Steinhauser et al. (2016) and by Frömke et al. (2022) were also satisfactory. The advantage of the approach by Frömke et al. (2022) is that, unlike the other approaches, there are no convergence problems because it is not an iterative procedure. Thus, this approach could be planned as a fallback strategy in case of non-convergence. For a comparison of the characteristics of the approaches in terms of requirements, possible results, advantages, and limitations, we refer to Zapf et al. (2021).

Besides the limitation mentioned above that the approach by Jones et al. (2019) could not be considered, the validity of a simulation study is of course always limited by the representativeness of the scenarios presented. Accordingly, a simulation study can never lead to generally valid statements. This includes the limitation that the simulation data were generated on the basis of the model of Hoyer et al. (2018), which could mean an advantage for this approach. Furthermore, the general recommendation to use the approach of Hoyer et al. (2018) is a rough summary. Of course, there are also scenarios in which the approach of Steinhauser et al. (2016) or that of Frömke et al. (2022) showed the best results.

We simulated data under various distributional assumptions. In reality, however, a meta-analyst does not know the true distribution which might even be one that is not implemented in any package. This means that the model will often be misspecified. Reflecting this, we analysed all scenarios in a way blind to the true distribution. For example, for the Steinhauser model, we evaluated all data based on the log-logistic distribution, and consequently the analysis model was misspecified for all other scenarios. In practice, however, users would probably visually inspect their data and check them for a suitable distribution assumption, thus possibly ending up with a better model fit – which is not possible to mimic in a simulation study. We did not investigate more general model selection approaches, such as those based on the generalized F distribution (Hoyer and Kuss, 2020) or the Box-Cox transformation (Jones et al., 2019). Finally, we note that the semiparametric model by Frömke et al. (2022) only uses uniformly distributed data and is robust in this respect. However, when using more informative distributions for the fictitious data the results might be better.

The strength of this simulation study is that a wide range of scenarios was covered and the three approaches were directly compared. We also tried to create comparable conditions for all approaches by using the different underlying distributions for data generation. By making the programs and simulation data available, researchers can also compare the approaches for their own questions with the help of a simulation study.

6. Conclusion

Recently, three frequentist methods for bivariate meta-analyses of DTA studies have been developed that allow using full threshold information provided in primary studies. So far, a joint evaluation of these new methods in a simulation study was lacking. We conducted a simulation study to allow for a fair comparison of the approaches by Steinhauser et at. (2016), Hoyer et al. (2018), and Frömke et al. (2022). Results from 108 different simulation scenarios show that all methods can perform best or worst depending on the respective scenario. Generally, we can recommend the approach by Hoyer et al. (2018) for most scenarios, as it returns the smallest bias and empirical coverages closest to the specified ones. As the model of Steinhauser et al. (2016) also leads to satisfactory results in many cases, it is a suitable alternative, The approach by Frömke et al. (2022) could be used as fallback strategy in case of non-convergence because it always yields results.

Author Contributions

AZ, CF, JH, GR and AH developed the concept of the simulation study. AH simulated the data. AZ, CF, GR and AH wrote code to apply the models to the simulated data. AH, DV and GR analysed the simulation results. DV created the figures. AZ, DV and AH wrote the first draft of the manuscript. All authors discussed the results, revised the manuscript and approved the final version.

Conflict of Interest

AZ and CF developed the Frömke approach, GR was the senior author of the Steinhauser approach, AH developed the Hoyer approach. JH and DV have declared no conflict of interest.

Data availability: The data and code that support the findings of this study are openly available in Zenodo at https://zenodo.org/record/7802090#.ZC1i8oTP02w.

Appendix

Table A1. Parameters set for the different simulation scenarios.

Table A2. Simulated thresholds per study using uniform distributions.

Table A3. Number of converged simulation runs of the different models.

Figure A1. Theoretical distribution curves for the diseased and non-diseased individuals of the four simulated distributions for the first (left side) and last scenario (right side) per distribution from Table A1.

Figure A2. Distribution of the bias of sensitivity depending on the true underlying distribution and the true correlation for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A3. Distribution of the bias of specificity depending on the true underlying distribution and the true correlation for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A4. Distribution of the bias of AUC depending on the true correlation. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A5. Distribution of the bias of AUC depending on the true prevalence. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A6. Distribution of the squared error of sensitivity depending on the true underlying distribution and the true correlation for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A7. Distribution of the squared error of sensitivity depending on the true underlying distribution and the true prevalence for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A8. Distribution of the squared error of specificity depending on the true underlying distribution and the true correlation for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A9. Distribution of the squared error of specificity depending on the true underlying distribution and the true prevalence for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A10. Distribution of the squared error of AUC depending on the true underlying distribution and the true correlation for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A11. Distribution of the squared error of AUC depending on the true underlying distribution and the true prevalence for the different models. The boxplots represent the median, the upper and lower quartile and the two whiskers, i.e., the interquartile range multiplied by 1.5.

Figure A12. Coverage to the two-sided 95% level of sensitivity depending on the true underlying distribution and the true correlation for the different models.

Figure A13. Coverage to the two-sided 95% level of specificity depending on the true underlying distribution and the true correlation for the different models.

Figure A14. Coverage to the two-sided 95% level of AUC depending on the true correlation. Results for Hoyer et al. (2018) are not included because runtime issues rendered the computation of confidence interval estimators too time-consuming.

Figure A15. Coverage to the two-sided 95% level of AUC depending on the true prevalence. Results for Hoyer et al. (2018) are not included because runtime issues rendered the computation of confidence interval estimators too time-consuming.

References

- Albert, C., Albert, A., Bellomo, R., Kropf, S., Devarajan, P., Westphal, S., ... Haase-Fielitz, A. (2018). Urinary neutrophil gelatinase-associated lipocalin-guided risk assessment for major adverse kidney events after open-heart surgery. Biomarkers Medicine, 12, 975–985.
- Arends, L.R., Hamza, T.H., van Houwelingen, J.C., Heijenbrok-Kal, M.H., Hunink, M.G., & Stijnen, T. (2008). Bivariate random effects meta-analysis of ROC curves. Medical Decision Making, 28(5), 621-38.
- Benedetti, A., Levis, B., Rücker, G., Jones, H.E., Schumacher, M., Ioannidis, J.P.A., & Thombs, B.; DEPRESsion Screening Data (DEPRESSD) Collaboration (2020). An empirical comparison of three methods for multiple cutoff diagnostic test meta-analysis of the Patient Health Questionnaire-9 (PHQ-9) depression screening tool using published data vs individual level data. Res Synth Methods, 11(6), 833-848.
- Burton, A., Altman, D.G., Royston, P., & Holder, R.L. (2006). The design of simulation studies in medical statistics. Stat Med, 25(24), 4279-4292.
- Chu, H., & Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: A generalized linear mixed approach. Journal of Clinical Epidemiology, 59, 1331–1332.
- Frömke, C., Kirstein, M., & Zapf, A. (2022). A semiparametric approach for meta-analysis of diagnostic accuracy studies with multiple thresholds. Research synthesis methods, 13(5), 612–621.
- Hoyer, A., Hirt, S., & Kuss, O. (2018). Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data. Research Synthesis Methods, 9(1), 62–72.
- Hoyer, A., & Kuss, O. (2020). Meta-analysis of full ROC curves with flexible parametric distributions of diagnostic test values. Research Synthesis Methods, 11(2), 301–313.
- Jones, H. E., Gatsonsis, C. A., Trikalinos, T. A., Welton, N. J., & Ades, A. E. (2019). Quantifying how diagnostic test accuracy depends on threshold in a meta-analysis. Statistics Medicine, 38(24), 4789–4803.
- Konietschke, F., & Brunner, E. (2009). Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes. CSDA, 53, 730–741.
- Lange, K., & Brunner, E. (2012). Sensitivity, specificity and ROC-curves in multiple reader diagnostic trials— A unified, nonparametric approach. Statistical Methodology, 9, 490–490.
- Morris, T.P., White, I.R., & Crowther, M.J. (2019). Using simulation studies to evaluate statistical methods. Statistics in Medicine 38, 2074–2102.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

- Reitsma, J. B., Glas, A. S., Rutjes, A. W.S., Scholten, R. J.P.M., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. Journal of Clinical Epidemiology, 58(10), 982–990.
- Rooney, D. (2017). Covariate adjusted nonparametric estimation of sensitivity and specificity in clustered data [Master Thesis]. Heidelberg: Ruprecht-Karls-Universität Heidelberg.
- Rücker, G., Steinhauser, S., Kolampally, S., & Schwarzer, S. (2022). diagmeta: Meta-Analysis of Diagnostic Accuracy Studies with Several Cutpoints. R package version 0.5-0. https://CRAN.R-project.org/package=diagmeta.
- Rutter, C. M., & Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med, 20(19), 2865–2884.
- Steinhauser, S., Schumacher,M., & Rücker, G. (2016). Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. BMC Medical Research Methodology, 16, 97.
- Voeltz, D., Zapf, A., Frömke, C., Hardt, J., Rücker, G., & Hoyer, A.. (2023). Simulation study to: A systematic comparison of different approaches for meta-analysis of diagnostic accuracy studies with multiple thresholds. https://doi.org/10.5281/zenodo.7802090.
- Zapf, A., Albert, C., Frömke, C., Haase, M., Hoyer, A., Jones, H. E., & Rücker, G. (2021). Meta-analysis of diagnostic accuracy studies with multiple thresholds: Comparison of different approaches. Biometrical Journal, 63(4), 699–711.
- Zapf, A., Hoyer, A., Kramer, K., & Kuss, O. (2015). Nonparametric meta-analysis for diagnostic accuracy studies. Statistics in Medicine, 34(29), 3831–3841.

6 Appendix

References

1. University of Colorado Anschutz Medical Campus. From smoking decrees to COVID-19: Epidemiology has changed the world. https://news.cuanschutz.edu/news-stories/from-smoking-decrees-to-window-screens-epidemiology-has-changed-the-world
2. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019). Accessed June 24, 2023. http://ghdx.healthdata.org/gbd-results-tool
3. Australian Institute of Health and Welfare. Chronic condition multimorbidity. Published May 13, 2021. Accessed June 24, 2023. https://www.aihw.gov.au/reports/chronic-disease/chronic-condition-multimorbidity/contents/chronic-conditions-and-multimorbidity
4. World Health Organisation. Noncommunicable diseases. Accessed October 25, 2023. https://www.who.int/data/gho/data/themes/noncommunicable-diseases
5. World Health Organisation. The top 10 causes of death. Published December 9, 2020. Accessed June 24, 2023. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death
6. Woodward M. *Epidemiology: Study Design and Data Analysis*. CRC press; 2013.
7. Brinks R, Landwehr S, Icks A, Koch M, Giani G. Deriving age-specific incidence from prevalence with an ordinary differential equation. *Statistics in Medicine*. 2013;32(12):2070-2078.
8. World Health Organisation for Animal Health. Manual 5 Surveillance and Epidemiology. 2018. Accessed October 22, 2023. https://rr-asia.woah.org/wp-content/uploads/2019/09/seacfmd-manual-5.pdf
9. Brinks R, Hoyer A. Illness-death model: statistical perspective and differential equations. *Lifetime Data Analysis*. 2018;24:743-754.
10. Voeltz D, Tönnies T, Brinks R, Hoyer A. Future prevalence of type 2 diabetes—A comparative analysis of chronic disease projection methods. *PloS One*. 2022;17(3):e0264739.
11. Waldeyer R, Brinks R, Rathmann W, Giani G, Icks A. Projection of the burden of type 2 diabetes mellitus in Germany: a demographic modelling approach to estimate the direct medical excess costs from 2010 to 2040. *Diabetic Medicine*. 2013;30(8):999-1008.
12. Brinks R, Landwehr S. A new relation between prevalence and incidence of a chronic disease. *Mathematical Medicine and Biology: a Journal of the IMA*. 2015;32(4):425-435.
13. Brinks R, Landwehr S. Age-and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. *Theoretical population biology*. 2014;92:62-68.
14. Altman DG, Bland JM. Time to event (survival) data. *BMJ*. 1998;317(7156):468-469. doi:10.1136/bmj.317.7156.468
15. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*. 2007;8(1):16. doi:10.1186/1745-6215-8-16
16. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1972;34(2):187-202.
17. Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. *Annals of Epidemiology*. 2018;28(1):54-57.
18. Hernán MA. The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*. 2010;21(1):13.
19. Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*. 1996;7(5):498-501.
20. Reid N. A conversation with Sir David Cox. *Statistical Science*. 1994;9(3):439-455.
21. Lindsey JK. A study of interval censoring in parametric regression models. *Lifetime Data Analysis*. 1998;4:329-354.

22. Boulesteix A-L, Lauer S, Eugster MJA. A plea for neutral comparison studies in computational sciences. *PloS One*. 2013;8(4):e61562.
23. Nießl C, Herrmann M, Wiedemann C, Casalicchio G, Boulesteix A-L. Over-optimism in benchmark studies and the multiplicity of design and analysis options when interpreting their results. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2022;12(2):e1441.
24. Henri Poincare. Henri Poincare Quotes. Accessed October 21, 2023. https://www.brainyquote.com/quotes/henri_poincare_130303
25. Noordzij M, Dekker FW, Zoccali C, Jager KJ. Measures of disease frequency: prevalence and incidence. *Nephron Clinical Practice*. 2010;115(1):c17-c20.
26. Jacobs E, Hoyer A, Brinks R, Icks A, Kuß O, Rathmann W. Healthcare costs of Type 2 diabetes in Germany. *Diabetic Medicine*. 2017;34(6):855-861.
27. Schmidt C, Reitzle L, Heidemann C, et al. Excess mortality in adults with documented diabetes in Germany: routine data analysis of all insurance claims in Germany 2013-2014. *BMJ open*. 2021;11(1):e041508.
28. Milan V, Fetzer S. Die zukünftige Entwicklung von Demenzerkrankungen in Deutschland-ein Vergleich unterschiedlicher Prognosemodelle. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*. 2019;62:993-1003.
29. Rossi A, Berger K, Chen H, Leslie D, Mailman RB, Huang X. Projection of the prevalence of Parkinson's disease in the coming decades: Revisited. *Movement Disorders*. 2018;33(1):156-159.
30. Pritzkuleit R, Beske F, Katalinic A. Erkrankungszahlen in der Pneumologie-eine Projektion bis 2060. *Pneumologie*. 2010;64(09):535-540.
31. Boyle JP, Honeycutt AA, Narayan KV, et al. Projection of diabetes burden through 2050: impact of changing demography and disease prevalence in the US. *Diabetes care*. 2001;24(11):1936-1940.
32. Carstensen B, Kristensen JK, Ottosen P, Borch-Johnsen K. The Danish National Diabetes Register: trends in incidence, prevalence and mortality. *Diabetologia*. 2008;51:2187-2196.
33. Carstensen B, Rønn PF, Jørgensen ME. Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996-2016. *BMJ Open Diabetes Research and Care*. 2020;8(1):e001071.
34. Tönnies T, Röckl S, Hoyer A, et al. Projected number of people with diagnosed Type 2 diabetes in Germany in 2040. *Diabetic Medicine*. 2019;36(10):1217-1225.
35. Keiding N. Age-specific incidence and prevalence: a statistical perspective. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 1991;154(3):371-396.
36. Heggland T. *Estimating Transition Probabilities for the Illness-Death Model*; 2015.
37. Ross R. An application of the theory of probabilities to the study of a priori pathometry.—Part I. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1916;92(638):204-230.
38. Ross R, Hudson HP. An application of the theory of probabilities to the study of a priori pathometry.—part iii. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1917;93(650):225-240.
39. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1927;115(772):700-721.
40. Kendall DG. Deterministic and stochastic epidemics in closed populations. In: Proceedings of the third Berkeley symposium on mathematical statistics and probability. Vol. 4; 1956:149-165.
41. Fix E, Neyman J. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology*. 1951;23(3):205-241.
42. Andersen PK, Geskus RB, Witte T de, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *International journal of epidemiology*. 2012;41(3):861-870.

43. Scheike TH, Zhang M-J. Analyzing Competing Risk Data Using the R timereg Package. *J Stat Softw*. 2011;38(2).
44. Lau B, Cole SR, Gange SJ. Competing Risk Regression Models for Epidemiologic Data. *American Journal of Epidemiology*. 2009;170(2):244-256. doi:10.1093/aje/kwp107
45. Murray CJL, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *The Lancet*. 1997;349(9061):1269-1276.
46. Murray CJ, Lopez AD. Quantifying disability: data, methods and results. *Bulletin of the World Health Organization*. 1994;72(3):481.
47. Murray CJL, Lopez AD. Global and regional descriptive epidemiology of disability: incidence, prevalence, health expectancies and years lived with disability. *The global burden of disease*. 1996;1:201-246.
48. Lewis BJ, Onder EN, Prudil A. *Advanced Mathematics for Engineering Students: The Essential Toolbox*. Butterworth-Heinemann; 2021.
49. Brinks R. On characteristics of an ordinary differential equation and a related inverse problem in epidemiology. *arXiv preprint arXiv:1208.5476 [Titel anhand dieser ArXiv-ID in Citavi-Projekt übernehmen]*. 2012.
50. Khan MA. Monthly and seasonal prevalence of asthma and chronic obstructive pulmonary disease in the District Dera Ismail Khan, Khyber Pakhtunkhwa, Pakistan. *The Egyptian Journal of Bronchology*. 2022;16(1):63.
51. Brunet RC, Struchiner CJ. A non-parametric method for the reconstruction of age-and time-dependent incidence from the prevalence data of irreversible diseases with differential mortality. *Theoretical population biology*. 1999;56(1):76-90.
52. Yang G. Neyman, Markov processes and survival analysis. *Lifetime Data Analysis*. 2013;19(3):393-411. doi:10.1007/s10985-013-9250-z
53. Brinks R, Bardenheier BH, Hoyer A, Lin J, Landwehr S, Gregg EW. Development and demonstration of a state model for the estimation of incidence of partly undetected chronic diseases. *BMC Medical Research Methodology*. 2015;15(1):1-11.
54. Voeltz D, Tönnies T, Brinks R, Hoyer A. Future prevalence of type 2 diabetes—A comparative analysis of chronic disease projection methods. *PloS One*. 2022;17(3):e0264739.
55. Hoyer A, Brinks R, Tönnies T, et al. Estimating incidence of type 1 and type 2 diabetes using prevalence data: the SEARCH for Diabetes in Youth study. *BMC Medical Research Methodology*. 2023;23(1):1-8.
56. Egeberg A, Kristensen LE. Impact of age and sex on the incidence and prevalence of psoriatic arthritis. *Annals of the rheumatic diseases*. 2018;77(4):e19-e19.
57. Manuwald U, Schoffer O, Kugler J, et al. Trends in incidence and prevalence of type 1 diabetes between 1999 and 2019 based on the Childhood Diabetes Registry of Saxony, Germany. *PloS One*. 2021;16(12):e0262171.
58. International Diabetes Federation. IDF Diabetes Atlas. Accessed May 2, 2023. https://www.diabetesatlas.org/
59. Zilbermint M. Diabetes and climate change. *Journal of community hospital internal medicine perspectives*. 2020;10(5):409-412.
60. Sun H, Saeedi P, Karuanga S, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes research and clinical practice*. 2022;183:109119.
61. Reitzle L, Ihle P, Heidemann C, Paprott R, Köster I, Schmidt C. Algorithmus Zur unterscheidung von diabetes mellitus typ 1 und typ 2 bei Der analyse von routinedaten. *Das Gesundheitswesen*. 2022;84.

62. Groop L, Tuomi T, Rowley M, Zimmet P, Mackay IR. Latent autoimmune diabetes in adults (LADA)—more than a name. *Diabetologia*. 2006;49:1996-1998.
63. Tamayo T, Brinks R, Hoyer A, Kuß O, Rathmann W. The prevalence and incidence of diabetes in Germany: an analysis of statutory health insurance data on 65 million individuals from the years 2009 and 2010. *Deutsches Ärzteblatt International*. 2016;113(11):177.
64. Rosenbauer J, Neu A, Rothe U, Seufert J, Holl RW. Diabetestypen sind nicht auf Altersgruppen beschränkt: Typ-1-Diabetes bei Erwachsenen und Typ-2-Diabetes bei Kindern und Jugendlichen. 2019.
65. Hex N, Bartlett C, Wright D, Taylor M, Varley D. Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs. *Diabetic Medicine*. 2012;29(7):855-862.
66. Federal Statistical Office of Germany. 14. Coordinated population projection. Accessed April 12, 2023. https://www.destatis.de/EN/Themes/Society-Environment/Population/
67. German Institute of Medical Documentation and Information. Information system for health care data. Accessed June 18, 2016. https://www.dimdi.de/static/en/versorgungsdaten/index.htm
68. World Health Organization. European detailed mortality database. Accessed August 11, 2015. data.euro.who.int/dmdb/
69. R Core Team. *R: A Language and Environment for Statistical Computing*. 2021. https://www.R-project.org/
70. Kumar D, Klefsjö B. Proportional hazards model: a review. *Reliability Engineering & System Safety*. 1994;44(2):177-188.
71. Kartsonaki C. Survival analysis. *Diagnostic Histopathology*. 2016;22(7):263-270.
72. Bollmann A, Blankenburg T, Haerting J, et al. Survival of Patients in Clinical Stages I–IIIb of Non-Small-Cell Lung Cancer Treated with Radiation Therapy Alone. *Strahlentherapie und Onkologie*. 2004;180(8):488-496. doi:10.1007/s00066-004-1184-7
73. Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. *Nature Reviews Disease Primers*. 2021;7(1):3.
74. Gridelli C, Rossi A, Carbone DP, et al. Non-small-cell lung cancer. *Nature Reviews Disease Primers*. 2015;1(1):15009. doi:10.1038/nrdp.2015.9
75. Huang R, Wei Y, Hung RJ, et al. Associated links among smoking, chronic obstructive pulmonary disease, and small cell lung cancer: a pooled analysis in the International Lung Cancer Consortium. *EBioMedicine*. 2015;2(11):1677-1685.
76. Sestelo M. *A Short Course on Survival Analysis Applied to the Financial Industry.* Available at: https://bookdown.org/sestelo. 2017.
77. Kuss O, Hoyer A. A proportional risk model for time-to-event analysis in randomized controlled trials. *Statistical Methods in Medical Research*. 2021;30(2):411-424.
78. Ameis L, Kuß O, Hoyer A, Möllenhoff K. A non-parametric proportional risk model to assess a treatment effect in time-to-event data. *arXiv preprint arXiv:2303.07479*. 2023.
79. Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. *European Heart Journal*. 2019;40(17):1378-1383.
80. Strobel A, Wienke A, Kuss O. How hazardous are hazard ratios? An empirical investigation of individual patient data from 27 large randomized clinical trials. *European Journal of Epidemiology*. 2023:1-9.
81. Carroll KJ. On the use and utility of the Weibull model in the analysis of survival data. *Controlled Clinical Trials*. 2003;24(6):682-701.
82. Martinussen T, Vansteelandt S. On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Analysis*. 2013;19(3):279-296. doi:10.1007/s10985-013-9242-z

83. Başar E. Aalen's additive, cox proportional hazards and the Cox-Aalen Model: Application to kidney transplant data. *Sains Malaysiana*. 2017;46(3):469-476.
84. Madadizadeh F, Ghanbarnejad A, Ghavami V, Bandamiri MZ, Mohammadianpanah M. Applying additive hazards models for analyzing survival in patients with colorectal cancer in Fars Province, Southern Iran. *Asian Pacific Journal of Cancer Prevention: APJCP*. 2017;18(4):1077.
85. Mahar AL, Davis LE, Kurdyak P, Hanna TP, Coburn NG, Groome PA. Using additive and relative hazards to quantify colorectal survival inequalities for patients with a severe psychiatric illness. *Annals of Epidemiology*. 2021;56:70-74.
86. Wang Y, Lee M, Liu P, et al. Doubly robust additive hazards models to estimate effects of a continuous exposure on survival. *Epidemiology (Cambridge, Mass.)*. 2017;28(6):771.
87. Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*. 2007;370(9596):1453-1457.
88. Schulz KF, Altman DG, Moher D, CONSORT Group*. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*. 2010;152(11):726-732.
89. Citrome L. Relative vs. absolute measures of benefit and risk: what's the difference? *Acta Psychiatrica Scandinavica*. 2010;121(2):94-102.
90. Schechtman E. Odds ratio, relative risk, absolute risk reduction, and the number needed to treat—which of these should we use? *Value in health*. 2002;5(5):431-436.
91. Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology*. 1994;47(8):881-889.
92. Aalen O. A model for nonparametric regression analysis of counting processes. In: Mathematical Statistics and Probability Theory: Proceedings, Sixth International Conference, Wisła (Poland), 1978; 1980:1-25.
93. Schaubel DE, Wei G. Fitting semiparametric additive hazards models using standard statistical software. *Biometrical Journal*. 2007;49(5):719-730.
94. Aalen OO. A linear regression model for the analysis of life times. *Statistics in Medicine*. 1989;8(8):907-925.
95. Aalen OO, Scheike TH. Aalen's additive regression model. *Encyclopedia of Biostatistics*. 2005. doi:10.1002/0470011815.b2a11002
96. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: multivariate data analysis-an introduction to concepts and methods. *British Journal of Cancer*. 2003;89(3):431-436.
97. Lin DY, Ying Z. Semiparametric analysis of the additive risk model. *Biometrika*. 1994;81(1):61-71.
98. Hjort NL, Stoltenberg EA. The partly parametric and partly nonparametric additive risk model. *Lifetime Data Analysis*. 2021:1-31.
99. Hougaard P. *Analysis of Multivariate Survival Data*. Springer; 2000.
100. Hosmer DW, Royston P. Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model. *The Stata Journal*. 2002;2(4):331-350.
101. Teshnizi SH, Ayatollahi SMT. Comparison of cox regression and parametric models: Application for assessment of survival of pediatric cases of acute leukemia in Southern Iran. *Asian Pacific Journal of Cancer Prevention: APJCP*. 2017;18(4):981.
102. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. John Wiley & Sons; 2013.
103. Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *Journal of Clinical Epidemiology*. 2010;63(1):46-55.
104. Zaman Q, Pfeiffer KP. Survival analysis in medical research. *Interstat*. 2011;17(4):1-36.

105. Gorunescu F. *Data Mining: Concepts, Models and Techniques*. Springer Science & Business Media; 2011.
106. Boulesteix A-L. *Ten Simple Rules for Reducing Overoptimistic Reporting in Methodological Computational Research.* Public Library of Science San Francisco, CA USA. *PLoS Computational Biology*. 2015;11.
107. Coolidge HJ. *Archibald Cary Coolidge*. Houghton Mifflin Boston, Massachusetts, United States; 1932.
108. Buchka S, Hapfelmeier A, Gardner PP, Wilson R, Boulesteix A-L. On the optimistic performance evaluation of newly introduced bioinformatic methods. *Genome Biology*. 2021;22(1):152.
109. Efron B. [Statistical modeling: The two cultures]: Comment. *Statistical Science*. 2001;16(3):218-219.
110. Nießl C, Hoffmann S, Ullmann T, Boulesteix A-L. Explaining the optimistic performance evaluation of newly proposed methods: A cross-design validation experiment. *Biometrical Journal*. 2023:2200238.
111. Ioannidis JPA, Munafo MR, Fusar-Poli P, Nosek BA, David SP. Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. *Trends in Cognitive Sciences*. 2014;18(5):235-241.
112. Kerr NL. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*. 1998;2(3):196-217.
113. Simonsohn U, Nelson LD, Simmons JP. P-curve: a key to the file-drawer. *Journal of Experimental Psychology*. 2014;143(2):534.
114. Ioannidis JPA. Why most published research findings are false. *PLoS Medicine*. 2005;2(8):e124.
115. Boulesteix A-L, Hornung R, Sauerbrei W. On fishing for significance and statistician's degree of freedom in the era of big molecular data. *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. 2017:155-170.
116. Wasserstein RL, Lazar NA. *The ASA Statement on P-Values: Context, Process, and Purpose.* Taylor & Francis. *The American Statistician*. 2016;70.
117. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PloS One*. 2008;3(8):e3081.
118. Rogers A. Peer review in NLP: reject-if-not-SOTA. *Hacking Semantics*. 2020.
119. Nuzzo R. How scientists fool themselves-and how they can stop. *Nature*. 2015;526(7572):182-185.
120. Heinze G, Boulesteix A-L, Kammer M, Morris TP, White IR, Simulation Panel of the STRATOS initiative. Phases of methodological research in biostatistics—Building the evidence base for new methods. *Biometrical Journal*. 2022:2200222.
121. Strobl C, Leisch F. Against the "one method fits all data sets" philosophy for comparison studies in methodological research. *Biometrical Journal*. 2022.
122. Friedrich S, Friede T. On the role of benchmarking data sets and simulations in method comparison studies. *Biometrical Journal*. 2023:2200212.
123. Demšar J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research*. 2006;7:1-30.
124. Eugster MJA, Hothorn T, Leisch F. Domain-based benchmark experiments: Exploratory and inferential analysis. *Austrian Journal of Statistics*. 2012;41(1):5-26.
125. Hothorn T, Leisch F, Zeileis A, Hornik K. The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics*. 2005;14(3):675-699.
126. Duin RPW. A note on comparing classifiers. *Pattern Recognition Letters*. 1996;17(5):529-536.

127. Xie C, Jauhari S, Mora A. Popularity and performance of bioinformatics software: the case of gene set analysis. *BMC Bioinformatics*. 2021;22(1):1-16.
128. Raji ID, Bender EM, Paullada A, Denton E, Hanna A. AI and the everything in the whole wide world benchmark. *arXiv preprint arXiv:2111.15366*. 2021.
129. Seide SE, Röver C, Friede T. Likelihood-based random-effects meta-analysis with few studies: empirical and simulation studies. *BMC Medical Research Methodology*. 2019;19:1-14.
130. Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, McKenzie JE. Comparison of six statistical methods for interrupted time series studies: empirical evaluation of 190 published series. *BMC Medical Research Methodology*. 2021;21(1):1-19.
131. Wiksten A, Rücker G, Schwarzer G. Hartung-Knapp method is not always conservative compared with fixed-effect meta-analysis. *Statistics in Medicine*. 2016;35(15):2503-2515.
132. Kreutz C. Guidelines for benchmarking of optimization approaches for fitting mathematical models. *arXiv preprint arXiv:1907.03427*. 2019.
133. Boulesteix A-L, Wilson R, Hapfelmeier A. Towards evidence-based computational statistics: lessons from clinical research on the role and design of real-data benchmark studies. *BMC Medical Research Methodology*. 2017;17(1):1-12.
134. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 2019;38(11):2074-2102.
135. Pawel S, Kook L, Reeve K. Pitfalls and potentials in simulation studies: Questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. *Biometrical Journal*. 2023.
136. Morris T, White I, Pham TM, Quartagno M. How to check a simulation study. 2023.
137. Hoffmann S, Schönbrodt F, Elsas R, Wilson R, Strasser U, Boulesteix A-L. The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *Royal Society Open Science*. 2021;8(4):201925.
138. Gelman A, Loken E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*. 2013;348:1-17.
139. Hoyer A, Hirt S, Kuss O. Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data. *Research Synthesis Methods*. 2018;9(1):62-72.
140. Fagard RH, Staessen JA, Thijs L. Advantages and disadvantages of the meta-analysis approach. *Journal of Hypertension*. 1996;14:S9-S13.
141. Zapf A, Albert C, Frömke C, et al. Meta-analysis of diagnostic accuracy studies with multiple thresholds: Comparison of different approaches. *Biometrical Journal*. 2021;63(4):699-711.
142. Takwoingi Y, Quinn TJ. Review of Diagnostic Test Accuracy (DTA) studies in older people. *Age and Ageing*. 2018;47(3):349-355.
143. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*. 2001;20(19):2865-2884.
144. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*. 2005;58(10):982-990.
145. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *Journal of Clinical Epidemiology*. 2006;59(12):1331-1332.
146. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.* Version. 2010.
147. Steinhauser S, Schumacher M, Rücker G. Modelling multiple thresholds in meta-analysis of diagnostic test accuracy studies. *BMC Medical Research Methodology*. 2016;16(1):1-15.

148. Frömke C, Kirstein M, Zapf A. A semiparametric approach for meta-analysis of diagnostic accuracy studies with multiple cut-offs. *Research Synthesis Methods*. 2022;13(5):612-621.
149. Rücker G, Steinhauser S, Kolampally S, Schwarzer G. *R Package Diagmeta Version 0.5-1*. 2022.
150. Konietschke F, Brunner E. Nonparametric analysis of clustered data in diagnostic trials: Estimation problems in small sample sizes. *Computational Statistics & Data Analysis*. 2009;53(3):730-741.
151. Voeltz D, Zapf A, Frömke C, Hardt J, Rücker G, Hoyer A. *Simulation Study to: A Systematic Comparison of Different Approaches for Meta-Analysis of Diagnostic Accuracy Studies with Multiple Thresholds.* Zenodo. 2023.
152. Haase-Fielitz A, Haase M, Devarajan P. Neutrophil gelatinase-associated lipocalin as a biomarker of acute kidney injury: a critical evaluation of current status. *Annals of clinical biochemistry*. 2014;51(3):335-351.
153. R Developement Core. R: A language and environment for statistical computing. 2010.
154. Türkmen Göksel. Introductory Chapter: Statistics: 1. In: Türkmen Göksel, ed. *Statistics*. IntechOpen; 2018.
155. Agrawal A, Gopal K. Principles of Statistics and Reporting of Data. *Biomonitoring of Water and Waste Water*. 2013:93-99.
156. Groves P, Kayyali B, Knott D, van Kuiken S. The'big data'revolution in healthcare: Accelerating value and innovation. 2016.
157. Rubin O, Errett NA, Upshur R, Baekkeskov E. The challenges facing evidence-based decision making in the initial response to COVID-19. *Scandinavian Journal of Public Health*. 2021;49(7):790-796.
158. Brinks R, Landwehr S. Age-and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. *Theoretical population biology*. 2014;92:62-68.
159. Voeltz D. *Projecting the Economic Burden of Type 1 and Type 2 Diabetes Mellitus in Germany from 2010 Until 2040.* Zenodo. 2023. 10.5281/zenodo.8009686
160. Bommer C, Sagalova V, Heesemann E, et al. Global economic burden of diabetes in adults: projections from 2015 to 2030. *Diabetes care*. 2018;41(5):963-970.
161. Tönnies T, Röckl S, Hoyer A, Heidemann C, Baumert J, Du Y, Scheidt-Nave C, Brinks R. Projected number of people with diagnosed Type 2 diabetes in Germany in 2040. *Diabetic Medicine*. 2019;36(10):1217-1225.
162. Zhang P, Zhang X, Brown J, et al. Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes research and clinical practice*. 2010;87(3):293-301.
163. Voeltz D, Brinks R, Tönnies T, Hoyer A. Future number of people with diagnosed type 1 diabetes in Germany until 2040: an analysis based on claims data. *BMJ Open Diabetes Research and Care*. 2023;11(2):e003156.
164. Federal Statistical Office of Germany. 15. Coordinated population projection. Accessed April 12, 2023. https://www.destatis.de/EN/Themes/Society-Environment/Population/
165. Tönnies T, Hoyer A, Brinks R. Productivity-adjusted life years lost due to type 2 diabetes in Germany in 2020 and 2040. *Diabetologia*. 2021;64:1288-1297.
166. Zucchini S. Editorial on Clinical Aspects of Different Forms of Diabetes in Children and Adolescents. *Frontiers in Endocrinology*. 2022;13:3396.
167. Köster I, Huppertz E, Hauner H, Schubert I. Costs of Diabetes Mellitus (CoDiM) in Germany, direct per-capita costs of managing hyperglycaemia and diabetes complications in 2010 compared to 2001. *Experimental and Clinical Endocrinology & Diabetes*. 2014;122(09):510-516.

168. Bächle C, Claessen H, Andrich S, et al. Direct costs in impaired glucose regulation: results from the population-based Heinz Nixdorf Recall study. *BMJ Open Diabetes Research and Care*. 2016;4(1):e000172.
169. Müller N, Heller T, Freitag MH, et al. Healthcare utilization of people with type 2 diabetes in Germany: an analysis based on health insurance data. *Diabetic Medicine*. 2015;32(7):951-957.
170. Ulrich S, Holle R, Wacker M, et al. Cost burden of type 2 diabetes in Germany: results from the population-based KORA studies. *BMJ open*. 2016;6(11):e012527.
171. Ely EK, Gruss SM, Luman ET, et al. A national effort to prevent type 2 diabetes: participant-level evaluation of CDC's National Diabetes Prevention Program. *Diabetes care*. 2017;40(10):1331-1341.

Affidavit

Eidesstaatliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, 15.11.2023	Dina Voeltz
_____	_____
Ort, Datum	Dina Voeltz