Deutsches Diabetes-Zentrum
Leibniz-Zentrum für Diabetes-Forschung
an der Heinrich-Heine-Universität Düsseldorf
Institut für Biometrie und Epidemiologie

Direktor: Univ.-Prof. Dr. Oliver Kuß

Application of the illness-death model to estimate epidemiological measures for diabetes based on aggregated data

Dissertation

zur Erlangung des Grades eines Doktors in Public Health

der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Thaddäus Tönnies

2020

Als Inauguraldissertation gedruckt mit Genehmigung der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: PD Dr. Ralph Brinks

Zweitgutachter: Prof. Dr. Adrian Loerbroks

Teile dieser Arbeit wurden veröffentlicht:

Tönnies T., Imperatore G., Hoyer A., Saydah S.H., D'Agostino R.B., Divers J., Isom S., Dabelea D., Lawrence J.M., Mayer–Davis E.J., Pihoker C., Dolan L., Brinks R., (2019), Estimating prevalence of type I and type II diabetes using incidence rates: the SEARCH for diabetes in youth study. *Ann Epidemiol*, (37) 37-42

Tönnies T., Röckl S., Hoyer A., Heidemann C., Baumert J., Du Y., Scheidt-Nave C., Brinks R., (2019), Projected number of people with diagnosed Type 2 diabetes in Germany in 2040. *Diabet Med,* (36) 1217-1225

Tönnies T., Hoyer A., Brinks R., (2018), Excess mortality for people diagnosed with type 2 diabetes in 2012 – Estimates based on claims data from 70 million Germans. *Nutr Metab Cardiovasc Dis,* (28) 887-891

Zusammenfassung

Anwendung des Illness-Death Modells zur Schätzung epidemiologischer Kennzahlen des Diabetes auf Basis aggregierter Daten

Die Diabetesepidemie ist eine Bedrohung für die Gesundheit von Populationen weltweit. Diabetes Surveillance hat unter anderem zum Ziel, Informationen über die Diabeteshäufigkeit bezüglich Prävalenz und Inzidenz sowie über diabetesbedingte Komplikationen wie erhöhte Mortalität zu liefern. Um für die Zielgruppe von Nutzen zu sein, sollte Diabetes Surveillance Informationen auf Grundlage repräsentativer Daten zeitnah bereitstellen. Das übergeordnete Ziel dieser Arbeit ist es, eine Methode zur Schätzung epidemiologischer Maßzahlen zu veranschaulichen, die auf aggregierten Daten basiert.

Die Methode verwendet ein Illness-Death Modell und eine damit assoziierte partielle Differentialgleichung (PDE). Zur Veranschaulichung der Methode werden drei Anwendungen vorgestellt. Die jeweiligen Ziele der Anwendungen waren, (i) die Validität der mit Hilfe der PDE und beobachteten Inzidenzraten geschätzten aktuellen altersspezifischen Prävalenz von Diabetes bei Jugendlichen in den USA zu bewerten, (ii) die Anzahl der Menschen mit Typ-2-Diabetes in Deutschland zwischen 2015 und 2040 zu prognostizieren und (iii) die mit Typ-2-Diabetes assoziierte altersspezifische Exzess-Mortalität in Deutschland im Jahr 2012 zu schätzen.

Die mit der PDE und Inzidenzraten geschätzte Prävalenz von Typ-1- und Typ-2-Diabetes bei Jugendlichen in den USA stimmte gut mit der beobachteten Prävalenz überein. Die prognostizierte Anzahl der Menschen mit Typ-2-Diabetes in Deutschland im Jahr 2040 lag zwischen 10,7 und 12,3 Millionen. Gegenüber 2015 entspricht dies einem relativen Anstieg zwischen 54% und 77%. Bezüglich der Exzess-Mortalität wurde geschätzt, dass Frauen und Männer mit Typ-2-Diabetes älter als 65 Jahre im Vergleich zu Personen ohne Typ-2-Diabetes eine um das 3,0-fache bzw. 2,3-fache erhöhte Mortalitätsrate aufweisen.

Alle Anwendungen basieren auf publizierten und aggregierten Daten. Im Vergleich zu herkömmlichen Methoden, die auf Individualdaten basieren, birgt dies das Potenzial, die Diabetes Surveillance effizienter zu gestalten. Im Hinblick auf die Diabetesepidemie in Deutschland stellt diese Arbeit erstmals Schätzungen für die Anzahl der Menschen mit Typ-2-Diabetes im Jahr 2040 für die gesamte erwachsene Bevölkerung zur Verfügung. Im Vergleich zu einfacheren Methoden vorheriger Prognosen legen die Ergebnisse eine wesentlich höhere Anzahl Betroffener nahe. Die Ergebnisse zur altersspezifischen Exzess-Mortalität deuten auf eine deutlich erhöhte Mortalitätsrate bei Menschen mit Typ-2-Diabetes hin. Diese Ergebnisse stimmen mit der einzigen vorherigen deutschen, auf Individualdaten basierenden Studie zur altersspezifischen Exzess-Mortalität überein.

Summary

Application of the illness-death model to estimate epidemiological measures for diabetes based on aggregated data

The diabetes epidemic is a major threat to public health worldwide. Diabetes surveillance systems aim, among other things, to provide information on the frequency of diabetes in terms of prevalence and incidence as well as on diabetes-related complications, such as increased mortality. In order to be useful for the targeted audience, diabetes surveillance need to provide information based on representative data in a timely manner. The overarching aim of this work is to illustrate a method that estimates epidemiological measures based on aggregated data.

The method uses an illness-death model and an associated partial differential equation (PDE). For the illustration of the method, three applications are presented. The particular aims of the applications were (i) to assess the validity of using the PDE to estimate current age-specific prevalence of diabetes in U.S. youth using incidence rates, (ii) to project the number of people with type 2 diabetes in Germany between 2015 and 2040 and (iii) to estimate age-specific excess mortality associated with type 2 diabetes in Germany in 2012.

In the first application, the prevalence of type 1 and type 2 diabetes in the U.S. youth estimated with the PDE and the observed incidence rates was in good agreement with observed prevalence estimates. In the second application, it was projected that between 10.7 million and 12.3 million people will be affected with type 2 diabetes in Germany in 2040. Compared to 2015, this corresponds to a relative increase between 54% and 77%. With regard to excess mortality, it was estimated that women and men above the age of 65 years with type 2 diabetes compared to people without type 2 diabetes experience a 3.0-fold and 2.3-fold increased mortality rate, respectively.

All applications were based on published data on an aggregated level. Hence, the presented methodological approach bears the potential to conduct diabetes surveillance more efficiently, compared to traditional approaches based on primary data on the individual level. With regard to the diabetes epidemic in Germany, this work for the first time provided estimates for the number of people with type 2 diabetes in 2040 considering the whole adult population. The results suggest a substantially higher number of affected people, compared to results using simpler methods, often employed by previous projections. In addition, the results for age-specific excess mortality suggest a substantially increased mortality rate among people with type 2 diabetes. These findings are in line with the only previous German study reporting age-specific excess mortality based on individual data.

Abbreviations

CDC	Centers for Disease Control and Prevention
DALY	Disability adjusted life years
DEGS	German Health Interview and Examination Survey for Adults
GEDA	German Health Update
HLY	Healthy life years
HR	Hazard ratio
IDF	International Diabetes Federation
IDM	Illness-death model
NHANES	National Health and Nutrition Examination Survey
NHIS	National Health Interview Survey
PDE	Partial differential equation
SEARCH	SEARCH for Diabetes in Youth Study
YLD	Years lived with disability
YLL	Years of life lost

Contents

1 Introduction

One of the main goals of epidemiology is to describe the frequency and distribution of diseases in populations (1,2:p.62). This branch of epidemiology is often referred to as descriptive epidemiology (2:p.62,3:p.73). In order to reach this goal, various quantitative measures are used and estimated, usually based on individual data from primary studies or secondary data sources. However, the collection of primary data is costly and time consuming, particularly if representativeness for various subgroups is required. As an alternative, secondary data often provides a cheaper opportunity, because the data has already been collected, albeit for another purpose. However, access to secondary data from individuals for research purposes is sometimes limited due to strict data protection and security regulations.

This work illustrates practical applications of a method to estimate quantitative measures of descriptive epidemiology based on aggregated data. The method uses the well-known illness-death model (IDM) and an associated partial differential equation (PDE). The applications are exemplified in the context of diabetes in Germany and the U.S.

As an introduction, this chapter provides a background on the global diabetes epidemic, in order to illustrate the public health relevance of the disease and the resulting need for diabetes surveillance. Furthermore, it gives definitions of descriptive epidemiology and disease surveillance as well as a short overview of the current diabetes surveillance systems in Germany and the U.S. For a methodological background, some technical details on the IDM and the associated PDE are provided.

1.1 The diabetes epidemic

Diabetes mellitus is defined as a *"group of metabolic diseases characterized by hyperglycemia resulting from defects in insulin secretion, insulin action, or both."(4)*. The most frequent forms of diabetes mellitus are type 1 and type 2 diabetes. The former is characterized by an absolute deficiency of insulin due to autoimmune destruction of pancreatic β-cells. The latter is characterized by a relative insulin deficiency due to initial insulin resistance (4). Type 1 diabetes is strongly related to genetic predisposition with a peak incidence between age 10 and 14 years (5). Contrary, type 2 diabetes is mainly related to behavioral risk factors such as poor diet (6,7), with a peak incidence around age 85 years (8). The term “diabetes epidemic” usually refers to type 2 diabetes because 90% to 95% of people with diabetes are affected by this type (4) and the risk of type 2 diabetes is related to behavioral factors (6,7).

The first unambiguous description of diabetes is attributed to the physician Aretaeus of Cappadocia more than 2000 years ago (9). Back then, diabetes was considered an uncommon disease, leading to death quickly (9). Despite major advances in treatment of diabetes, people with diabetes still experience a higher mortality than people without diabetes (10-15). Besides death, most common consequences of diabetes are macro- and microvascular complications such as stroke, myocardial infarction, diabetic nephropathy and retinopathy (16).

Since its first description more than 2000 years ago, diabetes became one of the major public health threats worldwide (17,18). As a consequence, the International Diabetes Federation (IDF) was founded in 1950 with the aim to promote diabetes care, prevention and cure (19). In 2000, the IDF published its first estimate of the number of people with diabetes worldwide with 151 million cases in the age group 20 to 79 years (20). In the most recent IDF report, this number remarkably increased to 463 million in 2019 (21). Hence, within only 19 years, the number of people affected by diabetes more than tripled. Similarly, it is estimated that between 1980 and 2014 the worldwide age-standardized prevalence almost doubled among men from 4.8% to 9.0%. The increase in prevalence among women was somewhat lower from 5.0% to 7.9% (22).

Diabetes is associated with an increased risk of death (10-15), impaired quality of life (23) and severe disabilities. An example, diabetic kidney disease may require dialysis due to kidney failure (16). The rapid increase in the number of people with diabetes was accompanied by a substantial increase in disease burden due to diabetes. Table 1 shows results from the Global Burden of Disease Study, which quantifies disease burden in terms of years lived with disability (YLD) and years of life lost due to premature death (YLL). The sum of these measures are the disability adjusted life years (DALY) (24). Between 2000 and 2017 the number of YLD due to diabetes globally increased by 71%, from 23 million years to 39 million years. Similarly, YLL due to diabetes increased from 20 million years to 30 million years, resulting in a relative increase of 50%. Trends in YLD in Germany were similar, however the relative increase was somewhat lower. In contrast to global trends, YLL due to diabetes decreased by 6% in Germany between 2000 and 2017, perhaps reflecting advances in survival with diabetes that may not have taken place at the same level in other parts of the world, with presumably limited access to high quality health care. Overall, the disease burden due to diabetes in terms of DALYs increased globally and in Germany by 61% and 23%, respectively.

Table 1: Disease Burden due to diabetes globally and in Germany in 2000 and 2017.

	2000	**2017**	**% change**
		Years lived with disability	
Global	22,511	38,575	71.4
Germany	334	487	45.9
		Years of life lost due to premature death	
Global	19,565	29,300	49.8
Germany	271	256	-5.7
		Disability adjusted life years	
Global	42,076	67,875	61.3
Germany	606	744	22.8

Numbers in 1,000s, taken from (25)

Years lived with disability measures disease burden by years lived in less than ideal health based on the prevalence of the disease and the disability weight that quantifies the severity of the disease. Years of life lost due to premature death measures lost life years in a given year based on the number of deaths due to a disease as documented in death certificates and the remaining life expectancy at the age of death. Disability adjusted life years are the sum of years lived with disability and years of life lost due to premature death. See (24) for further details.

Historically, the enormous increase in the number of people with diabetes and in the associated disease burden was strongly underestimated (20). For instance, in 2003 the IDF projected that in 2025, there would be 333 million cases of diabetes worldwide (26). However, this number was already surpassed in 2011 (20,27). And still, it is expected that current estimates of global prevalence are probably an underestimation due to lack of data and poor quality of available data (17).

With regard to the future of the diabetes epidemic, there are vague hints that the incidence of diabetes plateaus or decreases in some countries (28). However, the major rise of diabetes already experienced by most high income countries is still expected in many low and middle income countries (29). For instance, it is projected that in Africa, the number of people with diabetes between 2019 and 2045 will increase by 143%, whereas in Europe, the projected increase amounts to 15% (21). Globally, the IDF projects that the number of people with diabetes will increase by 51%; from 463 million cases in 2019 to 700 million in 2045 (21). Given that the current global prevalence estimates are probably an underestimation (17) and that previous projections were proven far too low, one might expect that the future number of people with diabetes will be even higher.

1.2 Descriptive epidemiology

The numbers on the diabetes epidemic presented in the previous paragraph are examples of descriptive epidemiology. Although there are many definitions of epidemiology, many of them have in common that they distinguish two major branches of epidemiologic research (30). For instance, Greenland & Rothman’s (1) definition reads “*Epidemiology is the study of the distribution and determinants of disease frequency in human populations*”. Another definition from the International Epidemiological Association reads *“[Epidemiology is] the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control of health problems”* (2:p.62). Both definitions have in common that they refer to the study of distributions and determinants. The study of the distribution of disease and health is often referred to as descriptive epidemiology whereas the study of determinants of disease is often called analytic epidemiology (2:p.62,3:p.73). Hence, descriptive epidemiology could be understood as one of two broad fields within epidemiology and the description of disease frequency in a population as one of the overarching aims of epidemiology. Although dichotomizing epidemiology this way is sometimes criticized (3:p.73), the remainder of this work considers applications of descriptive epidemiology in the sense of the aforementioned definitions.

1.2.1 Disease surveillance

One application of the methods of descriptive epidemiology is disease surveillance. The aim of disease surveillance systems is to continuously monitor health and disease in defined populations and thereby inform about temporal changes of a health problem (31-35). The population under surveillance is mainly defined by person, place and time (36:pp.44,37). “Person” refers to the question which subgroups of a population are affected by a disease or health-related state, for example with regard to age, sex or socioeconomic characteristics. “Place” and “time” refer to the location and time period the disease or health-related state occurs, respectively. To obtain this information, surveillance activities include the collection, analysis and interpretation of data, as well as the dissemination of the results (31-35). The results are for instance used to contribute to policy and public health programs in order to aid planning, implementation and evaluation of public health practice (31-35,37).

In this regard, projections of disease frequency play a central role for planning and anticipating the future demand for health care resources and to quantify the future burden caused by the surveyed disease (32,34,37). This information might also be useful to inform about the need for preventive measures, in order to attenuate a projected increase in disease frequency (32,33,37). Besides describing current and projecting future disease frequency, surveillance may also identify potential risk factors for the disease and helps to identify high risk groups (32,33,37). Another aim is to describe the frequency of the outcomes of a disease, such as health-related quality of life, disease-specific complications and mortality (33,37).

In order to be of use for the targeted audience of the results, a surveillance system needs to be fast, timely and representative (31,35,37). In analytic epidemiology, the need for representativeness is subject to debate (38-40), since identification of causal effects is sometimes conceived more valid in homogeneous study samples due to higher internal validity (39,41). However, in descriptive epidemiology and disease surveillance, the need for representativeness is obvious, since the aim is the description of health and disease of the target population (31,37,38,41). These requirements of disease surveillance often make it costly, since it involves large sample sizes and is a continuous effort with regular updates on disease frequency (31-34). Hence, one major methodological challenge of disease surveillance is the balance between information needs, timely results and resources available for study design, data collection and analysis (31,37).

In surveillance systems, data often comes from population-based registries, surveys and secondary data (33,42). Among these methods, surveys are particularly time consuming and expensive, since they involve interviews or self-administered questionnaires as well as procedures to obtain clinical and laboratory measures of the participants. Secondary data, such as diagnoses data from health insurances, are more efficient, however often at the cost of information loss.

1.2.2 Diabetes surveillance in the U.S. and Germany

In the U.S. there is an established diabetes surveillance system, maintained by the Centers for Disease Control and Prevention (CDC). For the adult population, the data for diabetes surveillance mainly stems from the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS). The NHANES survey was first conducted in the 1960s and is performed annually since 1999. It includes a nationally representative sample of approximately 5,000 people and provides information on a variety of diseases and associated risk factors (43). Data is obtained via personal interviews at the participants’ home as well as via physical examinations. The physical examinations take place in mobile centers and provide physiological, clinical and laboratory data.

The NHIS survey is an annual household interview survey, first conducted in 1957. During one year, a representative sample of about 90,000 people from 35,000 households are interviewed (44). In contrast to NHANES, no physical examinations are accomplished in NHIS. However, because of the large sample size, NHIS enables to describe a wide range of health-related conditions by many demographic and socioeconomic strata with high precision. Both, NHANES and NHIS can be linked to death certificates in the National Death Index, which allows the estimation of mortality rates. Based on these data sources, diabetes surveillance at the CDC provides long-term temporal trends for instance in age-specific diabetes prevalence and incidence (45), excess mortality (11,12) and rates of diabetes complications (46).

The data described above focuses on the adult population. For youth aged less than 20 years, the SEARCH for Diabetes in Youth study is the primary source for diabetes surveillance. SEARCH is a population-based multi-centered study currently recruiting participants in five centers in five different states of the U.S. The study was initiated in the year 2000 with the aim to describe and understand diabetes in youth (47). Prevalent and incident cases of diagnosed diabetes in youth are identified through health care providers, electronic health records as well as administrative data systems (47). The geographic areas covered by the SEARCH centers results in approximately 3.5 million people under the age of 20 years under active surveillance with a completeness of case ascertainment larger than 90% (47). A subsample of incident cases is followed up prospectively, in order to provide information on the clinical course of diabetes in youth after diagnosis. Based on this data, SEARCH provides information on age-specific prevalence and incidence of diabetes in youth (47-50). Furthermore, the frequency of complications, such as death, retinopathy or neuropathy, can be assessed (51,52).

In contrast to the U.S., the German diabetes surveillance system is still under development and was initiated in 2015 at Robert Koch-Institute (RKI) (53). In 2018, a final set of 40 indicators was published as a basis for a standardized and continuous reporting system for diabetes surveillance (54). Among these indicators are prevalence of diagnosed and undiagnosed diabetes, incidence of (diagnosed) diabetes as well as diabetes related mortality (54). Currently, these indicators are only reported in the context of single research projects with limited comparability due to different methodological approaches, different data bases and different time periods considered. These fragmented approaches hamper a timely and evidence-based policy guidance as well as evaluation of long-term trends in indicators of diabetes surveillance (53). For instance, in a comprehensive literature review, Heidemann et al. (55) found that there is no data on long term trends of the age-specific incidence rate in Germany. Similarly, age and sex-specific excess mortality of people with vs. without diabetes was estimated for the first time on a national level only recently (56). As a comparison, in the U.S., data on age-specific incidence and excess mortality is available at least since 1997 (11).

The German diabetes surveillance system will partly be based on primary data from the representative health surveys 'German Health Interview and Examination Survey for Adults' (DEGS) and 'German Health Update' (GEDA). DEGS is comparable to NHANES as it combines personal interviews, questionnaires and clinical and laboratory examinations performed by mobile study teams (57). Data for the first DEGS wave was collected between 2008 and 2011 and included approximately 8,000 participants. Besides a representative cross-sectional sample, DEGS also has a longitudinal component (57). Data collection for the second DEGS wave is planned to begin in 2020 (58).

Besides primary data from DEGS and GEDA, secondary data from different institutions of the German health care system is an important source for diabetes surveillance (42). Among these sources, data from all statutory health insurance companies provided by the German Institute for Medical Documentation (8) and data from the Central Research Institute of Ambulatory Health Care in Germany (59) are particularly valuable, since they include data from all people covered by the statutory health insurance, which results in a sample size of about 70 million people. This corresponds to approximately 90% of the German population.

Overall, the U.S. and German diabetes surveillance systems have common aims with regard to quantification of epidemiological measures and approaches with regard to representative health surveys. A specialty of the German system is the explicit incorporation of routinely documented secondary data from the health care system. The largest difference however is the fact that, compared to the German system, the U.S. system is long established and provides long-term trends in basic epidemiological measures such as age-specific incidence rates (45) and excess mortality (11,12), whilst these are not available for Germany.

1.3 The illness-death model

As mentioned above, timeliness and efficiency are two problems related to disease surveillance. In this work, applications of the IDM that may improve both of these aspects are presented. As a background, this chapter briefly provides details on the IDM and an associated PDE, which are both central to the methodological approach.

The IDM is a special case of the broader class of multi-state models (60) and was described early by Fix & Neyman in 1951 (61). Multi-state models are a general framework to describe the movement of individuals between a finite number of states over time (62). As illustrated in fig. 1, the IDM divides the population into three mutually exclusive states, namely *healthy*, *diseased* and *dead*. The process of moving from one state to another is determined by the incidence rate i, the remission rate r and the mortality rates m_0 and m_1. The prevalence p of a disease in a population is in general defined as the proportion of people with the disease at a given point in time among all people alive at that point in time (1).

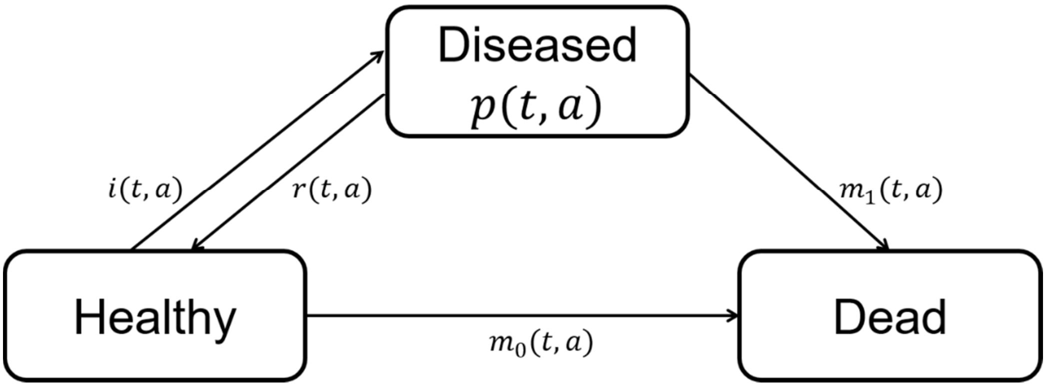

Fig. 1: Illness-death model with transitions rates i, r, m_0 and m_1 depending on age a and calendar time t

All parameters in the IDM depend on the two time scales calendar time t and age a. In the terminology of multi-state models, the state *dead* is an absorbing state, because a transition from this state to any other state is not possible (60). Graphically, an absorbing state has no arrows pointing outward from the state. All other states (from which transitions are possible) are called transient states (60).

In epidemiology, rates are defined as the number of events, i.e. transitions from one state to another, during a specified time period divided by the person time at risk for the transition during that period (1). The person time is often given in person-years and summarizes the time lived by all people at risk for the event in the population. As an example, a mortality rate is the number of deaths divided by the person time at risk for death.

One important feature of the IDM is that it accounts for competing risks. Competing risks remove people from the population at risk, such that they can no longer acquire the outcome under investigation (1). In the setting of fig. 1, dying without the disease is a competing risk for acquiring the disease. It follows that the prevalence of the disease partly depends on the mortality rate of the healthy population. Hence, ignoring competing risks may lead to biased results when estimating parameters of the IDM (63).

The dependence of the parameters on the two time scales age and time in the IDM can be illustrated in a Lexis diagram (fig. 2). The red lines in the Lexis diagram are called life lines and illustrate the life course of individuals along the two time scales t and a. All life lines have a slope of 1 because age and calendar time grow with the same pace. The end of a life line indicates the time point at which an individual left the population, for instance because the individual died or migrated. Brinks & Landwehr (64) have shown that the dynamics of the IDM in dependence of the two time scales calendar time and age in the Lexis diagram are governed by a PDE:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p=(1-p)\cdot[i-p\cdot(m_1-m_0)]-r\cdot p \tag{1}$$

Equation 1 describes the temporal change in prevalence p along a life line in the Lexis diagram (65). All parameters of the IDM can be found on the right hand side of equation 1. Equation 1 is classified as a PDE because it includes partial derivatives with respect to two variables, namely t and a (65). In case of chronic diseases without remission, i.e. $r=0$, equation 1 simplifies to:

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)p=(1-p)\cdot[i-p\cdot(m_1-m_0)] \tag{2}$$

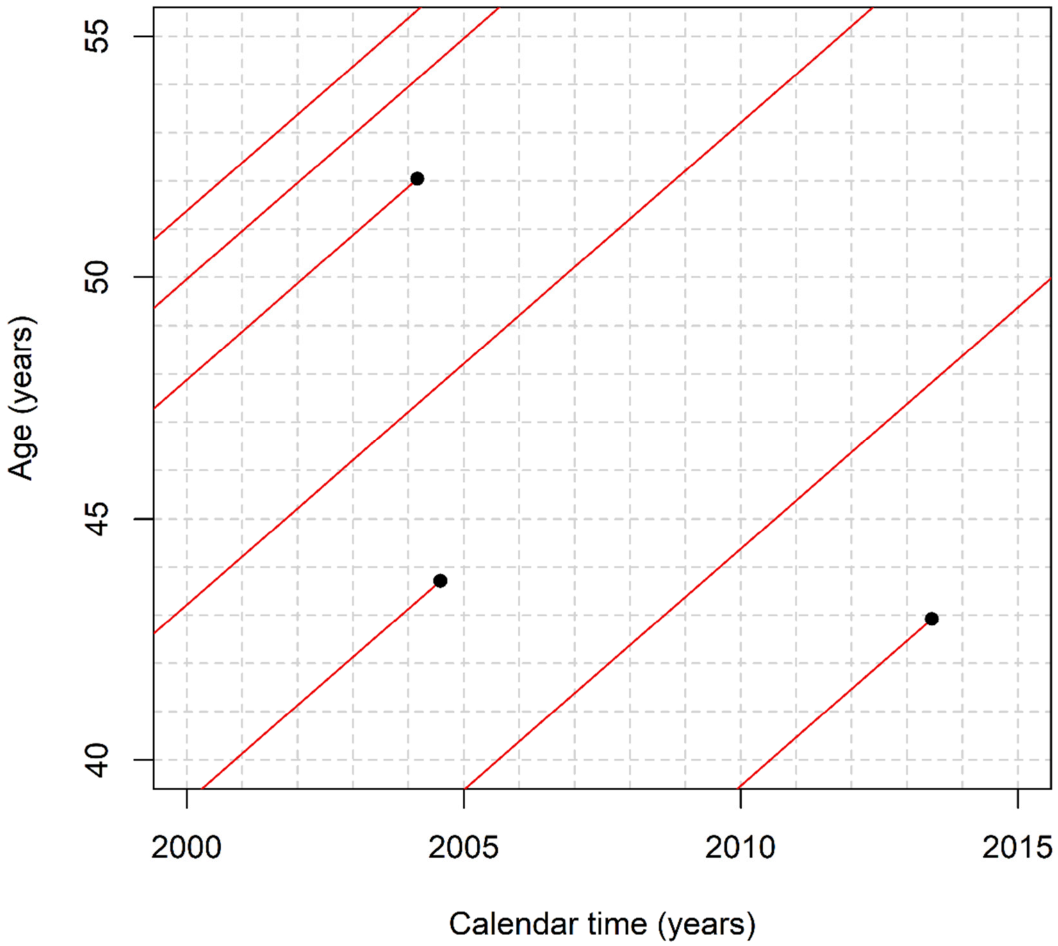

Fig. 2: Lexis diagram with life lines in red representing the life course of individuals along the time scales age and calendar time.

The remainder of this work is mainly concerned with the following three tasks of disease surveillance using the IDM for chronic diseases (equation 2):

(i) Estimation of current prevalence of a chronic disease
(ii) Projection of future number of cases of a chronic disease
(iii) Estimation of excess mortality associated with a chronic disease

For task (i), the age-specific prevalence is often of interest, since many chronic diseases are related to age. Formally, the prevalence p at a given point in time t at age a can be written as $p(t,a) = \frac{C(t,a)}{N(t,a)}$, where C and N denote the number of cases and people alive in the population, respectively. For task (ii), the target quantity is the number of cases $C(t,a) = p(t,a) \cdot N(t,a)$, for values of t that lie in future. The third point (iii), refers to an outcome of a chronic disease, namely death, and is based on the mortality rate of people with and without the disease. If the specified time period to which the mortality rate refers approaches the theoretical limit of zero, the mortality rate is also called hazard rate and can be interpreted as the instantaneous risk of death as opposed to the average mortality rate during the specified time period (60). One measure to quantify excess mortality due to a disease is the hazard ratio (HR), which is formally defined by $HR(t,a) = \frac{m_1(t,a)}{m_0(t,a)}$. Hence, the HR is a relative measure of association, quantifying the factor by which the hazard rate among people with the disease differs from those without the disease.

Brinks & Landwehr (66) showed that equation 2 can be expressed in a mathematically equivalent way using the mortality rate of the general population $m = p \cdot m_1 + (1-p) \cdot m_0$ and the HR:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p = (1-p) \cdot \left[i - m \cdot \frac{p \cdot (HR - 1)}{p \cdot (HR - 1) + 1} \right] \qquad (3)$$

Hence, if the HR is the quantity if interest, equation 3 may be used instead of equation 2. For practical applications in general, equation 3 may be preferred over equation 2, because empirical estimates of m and HR are usually more often available than estimates for m_0. The general procedure of the applications presented in this work follows three steps:

1. Solving equation 3 for the quantity of interest
2. Plugging in empirically motivated values for the remaining quantities
3. Calculate the quantity of interest

Hence, for estimating current prevalence and future case numbers, equation 3 can be used. In this case the temporal change in prevalence $\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p$ needs to be integrated to yield the age-specific prevalence at a certain time t. For this, a starting value for p at some time t_0 smaller than t as well as input values for i, m and HR for time points between t_0 and t are needed.

For the estimation of excess mortality, equation 3 needs to be solved for HR, which yields:

$$HR = 1 + \frac{1}{p} \cdot \frac{(1-p) \cdot i - \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p}{(1-p) \cdot (m-i) + \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p} \tag{4}$$

In this case, input values for the temporal change in prevalence $\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) p$, p, m and i are needed in order to calculate HR.

One important feature of these procedures is that all quantities in equation 3 are attributes of a population and not of individuals. This means, if all but one of the variables are known on an aggregated level, the remaining variable can be calculated without the need for individual data.

1.4 Aims of this thesis

As we have seen in the previous paragraphs, diabetes surveillance is of major importance to observe and project trends of the diabetes epidemic. However, efforts for a comprehensive surveillance system may be substantial and data availability is sometimes limited. The IDM and PDE introduced above may provide valuable opportunities to support efficiency and optimal use of available data in a diabetes surveillance system. Hence, the aim of this thesis is to illustrate three practical applications of the PDE that could be integrated into diabetes surveillance systems. These applications are illustrated in three papers. The specific aims of the papers are:

1. To assess the validity of using the PDE to estimate current age-specific prevalence of diabetes in U.S. youth using incidence rates (67)
2. To project the number of people with type 2 diabetes in Germany between 2015 and 2040 (68)
3. To estimate age-specific excess mortality associated with type 2 diabetes in Germany in 2012 (10)

2 Estimating prevalence of type I and type II diabetes using incidence rates: the SEARCH for diabetes in youth study, Tönnies T., Imperatore G., Hoyer A., Saydah S.H., D'Agostino R.B., Divers J., Isom S., Dabelea D., Lawrence J.M., Mayer–Davis E.J., Pihoker C., Dolan L., Brinks R., Annals of Epidemiology, 37: 37-42, (2019)

3 Projected number of people with diagnosed Type 2 diabetes in Germany in 2040, Tönnies T., Röckl S., Hoyer A., Heidemann C., Baumert J., Du Y., Scheidt-Nave C., Brinks R., Diabetic Medicine, 36: 1217-1225, (2019)

4 Excess mortality for people diagnosed with type 2 diabetes in 2012 – Estimates based on claims data from 70 million Germans, Tönnies T., Hoyer A., Brinks R., Nutrition, Metabolism & Cardiovascular Diseases, 28: 887-891, (2018)

5 Discussion

5.1 Main findings

The prevalence of type 1 and type 2 diabetes in the U.S. youth estimated with the PDE and the observed incidence rates in SEARCH was in good agreement with observed prevalence estimates (67). By integrating this new approach into the SEARCH surveillance activities, costs could potentially be reduced, because updates on current prevalence of diabetes in youth would not require collection of primary prevalence data if incidence rates are available.

In the context of projecting future case numbers of type 2 diabetes in Germany (68), it was shown that an increase between 54% and 77% between 2015 and 2040 seems likely. Furthermore, it was shown that the main driver of this increase will be trends in the incidence rate. Prevalence based projection methods used by previous studies, seem to underestimate the future number of cases.

With regard to estimating excess mortality, it was shown that women and men above the age of 65 years with type 2 diabetes compared to people without type 2 diabetes experience a 3.0-fold and 2.3-fold increased mortality rate, respectively. The excess mortality was higher in younger age groups and consistently decreased with age (10). These results are in line with the only previous German study reporting age-specific excess mortality based on individual data (56).

5.2 The methodological approach in the context of multi-state models

In chronic disease epidemiology, multi-state models are usually considered in the context of analysing individual time to event or survival data (60,62,69,70). The simplest setting is survival analysis with the two states *alive* and *dead* (i.e. the IDM without the *ill* state). Further uses with individual data refer to competing risk models with different causes of death and recurrent event models (60,69,70). The general approach is to fit a regression model to individual data to estimate population parameters of interest. Parameters of interest often are cumulative transition probabilities, such as cumulative incidence. Also the transition rates, such as the incidence rate, are commonly subject to statistical inference. These methods for analysing individual data are based on the theory of stochastic processes (65,71).

Between these methods and the approach used in this work, there are important differences. For instance, in most cases multi-state models are considered with regard to one time scale, e.g. age or follow-up time. In contrast, the approach of this work includes the two time scales age and calendar time. Furthermore, we used the IDM in combination with a PDE and aggregated population parameters as input data. Using differential equations in the context of multi-state models is more in the tradition of infectious disease epidemiology (65,72). For instance, Kermack & McKendrick (72) used differential equations to characterize the dynamics of infectious disease epidemics in a model including the states *susceptible*, *infected* and *recovered*. With regard to the IDM, the differences between the approaches based on the theory of stochastic processes and on differential equations was formally investigated by Brinks & Hoyer (65). It was shown that both approaches are related by the theory of Markov processes.

5.3 Implications for diabetes surveillance systems

Overall, the papers showed that the IDM in combination with the PDE may provide useful tools to potentially supplement diabetes surveillance activities. All approaches solely relied on published data on an aggregated level, which has several advantages. Surveillance systems could be organized less costly if collection of primary data could be partly replaced by the applications illustrated, using data readily available. If routinely documented secondary data is used as input for the model, as for instance in (10), this could also improve the timeliness of the surveillance system, since collection of primary data not only is expensive, but also time consuming.

Another advantage is that the data often includes large sample sizes and in the case of Germany almost the entire population with all age groups (8,59). This is a particularly important advantage because participation in health surveys declined in the past decades (73). In general, secondary data is considered a useful approach to counteract non-response in health surveys (74). Furthermore, the DEGS health survey in Germany is restricted to the age range 18 to 79 years (57). This is problematic because the projection of future prevalence showed that the peak in prevalence of type 2 diabetes in Germany will probably shift above the age of 79 years (68). Hence, excluding these age groups would ignore a highly relevant subgroup with regard to type 2 diabetes

Of course, in order to profit from the advantages of secondary data, it needs to be available for research purposes in a timely manner. Hence, it is required that the data is archived in a standardized way such that it can be used for diabetes surveillance purposes (74). The Danish (14) and Swedish (13,75) diabetes registries are good examples of efficient use of secondary data for surveillance purposes. These registries provide temporal trends on prevalence, incidence and excess mortality and include data from more than 90% of the population with diabetes. However, since the data contains sensitive individually identifiable information, a high level of security and protection is required (74), which may hamper data access in general and timely access in particular. The approaches described in this thesis have the additional advantage that access to individual data is not needed. As an example, if the institution collecting the secondary data regularly publishes age-specific diabetes incidence and prevalence on an aggregated level, one could incorporate the approach presented in (10) into the surveillance system to timely update trends in excess mortality. Through this, time that is usually spent for anonymization and secure delivery of the data could be saved. In case the data is not available outside the collecting institution at all, using prevalence and incidence might be the only option to estimate excess mortality based on this data.

The use of comparably cheap methods for disease surveillance could be particularly useful for surveillance systems in low and middle income countries, which often lack resources for primary data collection (33). Usually, estimating excess mortality in studies based on individual data requires long follow-up periods, which may lead to high costs. In the case of estimating excess mortality from incidence and prevalence (10,15), these costs could be avoided. Similarly, Brinks et al. (76) illustrated how the PDE could be used to estimate the incidence rate based on repeated prevalence studies in case of a poorly funded disease surveillance system. Since it is projected that the increase in the number of people with diabetes is highest in low and middle income countries (29), the cost saving potential of the method presented here might be particularly important for future surveillance systems.

Besides the applications shown in the papers above, there are several other potential uses for diabetes surveillance systems. In (10), the mortality burden associated with diabetes was quantified using the HR. Another way to quantify the mortality burden of diabetes is to estimate YLL associated with the disease. There are several definitions of YLL (77) and the approach used by the global burden of disease study probably is most widely known. This method is based on the number of deaths with diabetes as the documented cause of death in the death certificate and the remaining life expectancy at the age of death (24). Another approach is to define YLL as the difference in life expectancy of a person with diabetes compared to a same aged person without diabetes (78-80). Formally, this quantity is defined by:

$$YLL(t,a) = \int_0^\infty S_{D-}(t+u, a+u)) - S_{D+}(t+u, a+u)\mathrm{d}u \qquad (5)$$

where S_{D-} and S_{D+} are the survival functions of people without and with diabetes, respectively. The survival functions in equations 5 are based on the mortality rates m_0 and m_1. In order to estimate a population-wide measure of YLL, one can sum up all individual YLL based on the prevalence of diabetes and the population size (79). It is also possible to project future trends in YLL using the projected prevalence, for instance from (68), and projected mortality rates (79).

Another measure of interest for diabetes surveillance is the average life expectancy free of diabetes (healthy life years, HLY) (81,82). Sullivan provided a general formula for the lifetime free of a disease in 1971 (83):

$$HLY(t,a) = \frac{1}{S(t,a)} \int_0^\infty (1 - p(t+u, a+u)) \cdot S(t+u, a+u)\mathrm{d}u \qquad (6)$$

where S is the survival function of the general population. Using equation 6, the projected prevalence from Tönnies et al. (68) and the projected mortality rate of the general population, one can estimate HLY for a person alive in a certain year, while accounting for future trends in diabetes prevalence and general mortality. This might yield more valid results than using currently observed mortality and prevalence as has sometimes been done (81,82).

5.4 Implications for the diabetes epidemic in Germany

Besides the aforementioned advantages from a methodologic perspective, the presented applications also provide important information on the diabetes epidemic in Germany. In (68), it was estimated that in 2040 there will be between 10.7 million and 12.3 million people with type 2 diabetes. Based on approximately 6.9 million cases in 2015, this corresponds to an increase between 54% and 77%. The peak number of cases is expected approximately at age 75 (fig. 3). However, as can be seen from fig. 3, there is expected to be an increasingly large number of people with type 2 diabetes above the age of 79 years.

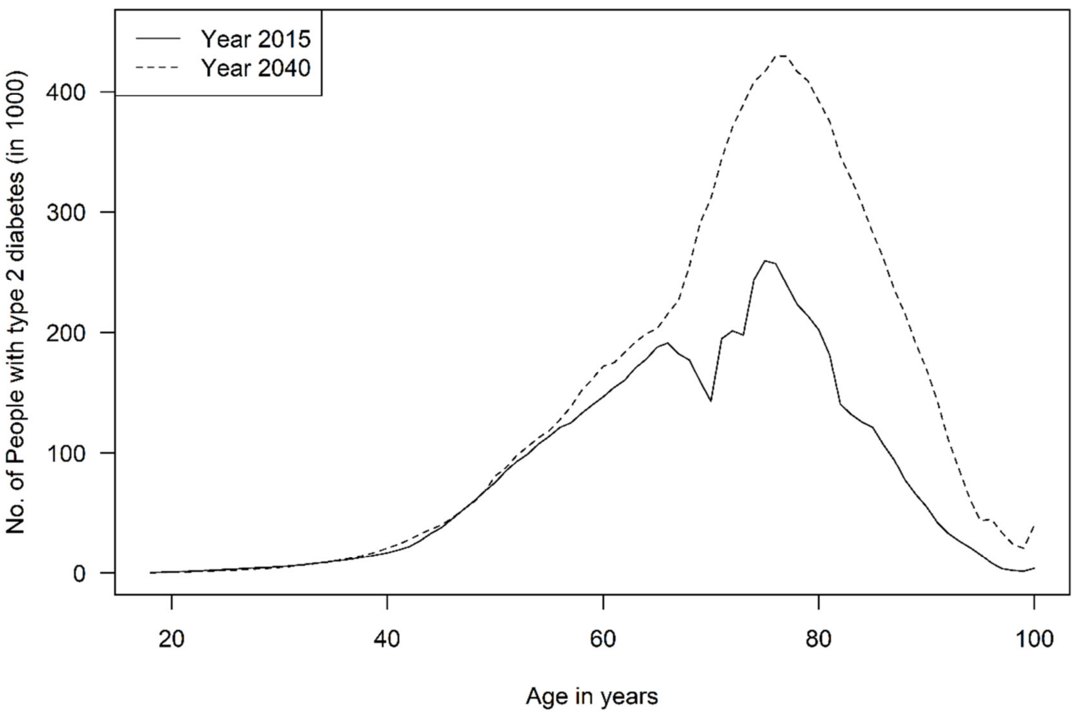

Fig. 3: Age-specific number of people with type 2 diabetes in Germany in 2015 and 2040. The results are based on Tönnies et al. (68) and refer to a scenario assuming excess mortality to decrease by 2% per year and no changes in the age-specific incidence rate. The projected population size was taken from variant 2 of the population projections of the Federal Statistical Office (84). See Tönnies et al. (68) for further details.

Accordingly, the number of people with type 2 diabetes in this age group is expected to increase 2.5-fold between 2015 and 2040, compared to a 1.7-fold increase in the whole adult German population (table 2). As a consequence, the proportion of people aged above 79 years among all people with type 2 diabetes is expected to increase from 21% to 32%. The increasing number and proportion of people with type 2 diabetes above the age of 79 years has implications for current practice of diabetes surveillance, since 79 years is the upper age limit in the DEGS health survey (57) as well as in the reports by the IDF (29). In the case of Germany, this means that in future, one third of the population affected by type 2 diabetes will not be covered by the DEGS survey, if this age restriction remains.

Table 2: Projected number of people with type 2 diabetes in 2015 and 2040 in different age groups.

Year	2015 (million)	2040 (million)	% change
Age 18->100	6.9	11.5	65
Age 80->100	1.5	3.7	154

The results are based on Tönnies et al. (68) and refer to a scenario assuming excess mortality to decrease by 2% per year and no changes in the age-specific incidence rate. The projected population size was taken from variant 2 of the population projections of the Federal Statistical Office (84). See Tönnies et al. (68) for further details.

These findings also have important implications for the future of diabetes care in Germany. It should be expected that health care costs will strongly increase since type 2 diabetes is associated with 1.7 fold higher direct health care costs compared to people without type 2 diabetes (85). In addition, the health care system should prepare to take care of more than twice as many people with type 2 diabetes patients above the age of 79 years. This might require an increased focus on diabetes specialists training, particularly for the elderly, since patients with type 2 diabetes often present with an increasing number of comorbidities with increasing age (46).

Furthermore, we found in (10) that men and women with type 2 diabetes above the age of 65 years in Germany have a 2.3-fold and 3.0-fold increased mortality rate, respectively. Hence, one might expect that the number of deaths due to type 2 diabetes and the mortality burden in terms of YLL will also strongly increase. However, the increasing number of people affected by excess mortality associated with type 2 diabetes could be partly offset by decreases in excess mortality, as shown in (79). Decreases in excess mortality were observed in several countries in past decades (11-14) and are probably due to improvements in diabetes care. As a consequence, changes in the occurrence and type of diabetes complications should be expected, since people with diabetes in the future will probably live longer than nowadays (46). Therefore, the disease burden due to type 2 diabetes will probably shift from mortality to morbidity.

As mentioned above, the global rapid rise of diabetes was strongly underestimated by the projections of the IDF. It is argued that this underestimation might be caused by the lack of standardization of the diagnostic criteria for diabetes, by using rather insensitive definitions of diabetes (20) and by poor quality of data (17).

However, the projections for Germany suggest that the projection method used by the IDF might be the main reason for the obvious gap between projected and observed case numbers. Basically, the IDF estimates age-specific prevalence in one year and applies this prevalence to the projected future age structure of a country or region. Using this method, the IDF projects that the number of people with diabetes in Europe in the age group 20-79 years will increase by approximately 19% between 2015 and 2040 (86). This increase is similar to the estimated increase of 21% in the German adult population between 2015 and 2040 using the same method in this work (68). However, this increase was far lower than in all other scenarios based on the PDE approach, even than in scenarios assuming a decreasing incidence rate.

The main reason probably is that the prevalence is influenced not only by the incidence rate, but also by the mortality rate of people with and without diabetes, as can be seen in the IDM and equation 2. In other words, the IDF method only yields valid projections, when temporal changes of the three rates in the IDM cancel each other out in a way that the age-specific prevalence remains constant. Given strong evidence of globally decreasing mortality rates (87), heterogeneous trends in the incidence rate (28) and decreasing excess mortality associated with type 2 diabetes, this seems an unlikely scenario with no support from empirical studies. In contrast, the use of the PDE for projections allows to incorporate different trends for all rates in the IDM and also evaluating different scenarios to provide a plausible range of projected case numbers. In general, it has been noted that mathematical models are increasingly used to project the distribution of disease in the context of disease surveillance (35). Hence, while the problems regarding diagnostics and data quality most certainly also hamper valid projections, the main reason for the underestimation in the projections by the IDF is probably caused by using the oversimplified method of applying current age-specific prevalence to the future age structure of the population. Although one could argue that this simple method might serve as a good starting point (37), surveillance systems should aim to provide the highest quality evidence possible (32). This might hold particularly true for the projection of future case numbers, which are potentially used for future planning of health care resources and prioritizing disease prevention activities.

5.5 Limitations

The applications of the PDE illustrated above provide potentially useful tools for disease surveillance and empirical research. However, the method is also subject to limitations. The model assumed that a transition from the diabetic state to the healthy state is not possible. However, on the population level, no studies provide evidence of a relevant remission rate. Hence, for the estimation of current prevalence (67) and excess mortality (10), these simplification probably did not bias the results. With regard to projection of future case numbers, one might need the extended model including a remission rate (equation 1), as soon as new interventions result in remission rates relevantly different from zero. However, as long as there is no evidence for a relevant remission rate of diabetes on the population level, it seems reasonable to project the number of future cases assuming a remission rate equal to zero.

Another concern is that all analyses only referred to diagnosed diabetes, although the prevalence of undiagnosed diabetes is substantial (86,88). The prevalence of undiagnosed diabetes is usually estimated in health surveys, which collect laboratory data (88). A case of undiagnosed diabetes is defined as a participant, who states that he or she never received a diagnosis, but at the same time has laboratory results fulfilling the criteria for diabetes. Hence, when estimating prevalence in the SEARCH study, the results only referred to the prevalence of clinically diagnosed diabetes, which might underestimate the true prevalence. Similarly, the projected number of future cases in Germany could be underestimated since the input data only referred to cases with diabetes documented as diagnostic codes in data from the statutory health insurance. With regard to estimating excess mortality based on incidence and prevalence of diagnosed diabetes, the consequence of ignoring undiagnosed cases is unclear. Here, the problem arises that trends in diagnosed prevalence, which are needed for this approach, might be influenced by changes in the detection of diabetes (89,90). Since changes in detection of diabetes do not reflect changes in true prevalence, the estimates for excess mortality could be biased by using only diagnosed diabetes as input data. However, it is unclear whether this bias would lead to over- or underestimation of excess mortality. Of note, the problem of undiagnosed diabetes is not a limitation of the modelling approach per se, but rather a problem due to the lack of input data. Studying the population of undiagnosed diabetes, e.g. with regard to mortality, is notoriously difficult because as soon as an undiagnosed case of diabetes is identified, he or she cannot be regarded as an undiagnosed case anymore. In order to deal with this problem, Brinks et al. (91) developed an extended model including a transient state of undiagnosed diabetes.

One limitation that applies to all three applications of the PDE refers to migration of people into the population during the study periods. This limitation might be particularly relevant for the projection of future case numbers, since the model implicitly assumes that the prevalence of people migrating into the population have the same prevalence as the resident population. Since the observation period in the projection comprises 25 years, the number of people that will have migrated to Germany is probably substantial. This might bias the projected prevalence, if the assumption of equal prevalence is not met. However, as Brinks & Landwehr (64) showed in case of dementia, even in extreme scenarios, this bias would be minor compared to the temporal trends projected by the PDE.

6 Conclusions

This work illustrated three applications of a PDE in the context of diabetes and provided evidence that the new methodological approach could be used to aid diabetes surveillance systems. All applications were based on publicly available data on an aggregated level. The use of aggregated data bears the potential to conduct diabetes surveillance more efficiently, compared to traditional approaches based on primary data on the individual level. Furthermore, results could potentially be delivered more timely.

Besides these methodological aspects, this work also provides important information on the diabetes epidemic in Germany. It was projected that in 2040, there will be between 10.7 million and 12.3 million people with type 2 diabetes. These numbers are substantially higher compared to the use of simpler methods, often employed by previous projections. Furthermore, it was estimated that men and women with type 2 diabetes above 65 years of age experience a 2.3-fold and 3.0-fold increased mortality rate, respectively, compared to people without type 2 diabetes.

References

1. Greenland S, Rothman KJ: Measures of Occurence. In *Modern Epidemiology* Rothman KJ, Greenland S, Lash TL, Eds. Philadelphia, USA, Lippincott Williams & Wilkins, 2008
2. Last J, Spasoff R, Harris S, Thuriaux M: *A Dictionary of Epidemiology*. New York, USA, Oxford University Press, 2001
3. Miettinen OS: *Epidemiological research: terms and concepts*. Springer Science & Business Media, 2011
4. American Diabetes Association: Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2011;34:S62-S69
5. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ: Epidemiology of type 1 diabetes. Endocrinol Metab Clin North Am 2010;39:481-497
6. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Möhlig M, Pfeiffer AF, Spranger J, Thamer C, Häring H-U: An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007;30:510-515
7. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, Keinänen-Kiukaanniemi S, Laakso M, Louheranta A, Rastas M, Salminen V, Aunola S, Cepaitis Z, Moltchanov V, Hakumäki M, Mannelin M, Martikkala V, Sundvall J, Uusitupa M: Prevention of Type 2 Diabetes Mellitus by Changes in Lifestyle among Subjects with Impaired Glucose Tolerance. N Engl J Med 2001;344:1343-1350
8. 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. Dtsch Arztebl Int 2016;113:177-182
9. Laios K, Karamanou M, Saridaki Z, Androutsos G: Aretaeus of Cappadocia and the first description of diabetes. Hormones 2012;11:109-113
10. Tönnies T, Hoyer A, Brinks R: Excess mortality for people diagnosed with type 2 diabetes in 2012 – Estimates based on claims data from 70 million Germans. Nutr Metab Cardiovasc Dis 2018;28:887-891
11. Gregg EW, Cheng YJ, Saydah S, Cowie C, Garfield S, Geiss L, Barker L: Trends in Death Rates Among U.S. Adults With and Without Diabetes

Between 1997 and 2006 - Findings from the National Health Interview Survey. Diabetes Care 2012;35:1252-1257
12. Gregg EW, Cheng YJ, Srinivasan M, Lin J, Geiss LS, Albright AL, Imperatore G: Trends in cause-specific mortality among adults with and without diagnosed diabetes in the USA: an epidemiological analysis of linked national survey and vital statistics data. Lancet 2018;391:2430-2440
13. Tancredi M, Rosengren A, Svensson A, Kosiborod M, Pivodic A, Gudbjörnsdottir S, Wedel H, Clements M, Dahlqvist S, Lind M: Excess Mortality among Persons with Type 2 Diabetes. N Engl J Med 2015;373:1720-1732
14. Carstensen B, Kristensen JK, Ottosen P, Borch-Johnsen K, Steering Group of the National Diabetes Register: The Danish National Diabetes Register: trends in incidence, prevalence and mortality. Diabetologia 2008;51:2187-2196
15. 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;19:844
16. Zheng Y, Ley SH, Hu FB: Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2017;14:88-98
17. Zimmet PZ, Alberti KG, Magliano DJ, Bennett PH: Diabetes mellitus statistics on prevalence and mortality: facts and fallacies. Nat Rev Endocrinol 2016;12:616-622
18. Zimmet PZ, Magliano DJ, Herman WH, Shaw JE: Diabetes: a 21st century challenge. Lancet Diabetes Endocrinol 2014;2:56-64
19. International Diabetes Federation: Mission & strategy, Online: https://www.idf.org/who-we-are/mission-strategy.html (Accessed 03 November 2019).
20. Zimmet PZ: Diabetes and its drivers: the largest epidemic in human history? Clin Diabetes Endocrinol 2017;3:1
21. International Diabetes Federation: *IDF Diabetes Atlas - Ninth edition*. Brussels, Belgium, International Diabetes Federation, 2019

22. Zhou B, Lu Y, Hajifathalian K, Bentham J, Di Cesare M, Danaei G, Bixby H, Cowan MJ, Ali MK, Taddei C, Lo WC, Reis-Santos B, Stevens GA, Riley LM, Miranda JJ, Bjerregaard P, Rivera JA, Fouad HM, Ma G, Mbanya JC, et al.: Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016;387:1513-1530
23. Rubin RR, Peyrot M: Quality of life and diabetes. Diabetes Metab Res Rev 1999;15:205-218
24. Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A, Abdollahpour I, Abdulkader RS, Abebe M, Abebe Z, Abil OZ, Aboyans V, Abrham AR, Abu-Raddad LJ, Abu-Rmeileh NME, Accrombessi MMK, et al.: Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1859-1922
25. Institute for Health Metrics and Evaluation (IHME): GBD Results Tool, Online: http://ghdx.healthdata.org/gbd-results-tool (Accessed 03 November 2019).
26. International Diabetes Federation: *IDF Diabetes Atlas - Second edition*. Brussels, Belgium, International Diabetes Federation, 2003
27. International Diabetes Federation: *IDF Diabetes Atlas - Fifth edition*. Brussels, Belgium, International Diabetes Federation, 2011
28. Magliano DJ, Islam RM, Barr ELM, Gregg EW, Pavkov ME, Harding JL, Tabesh M, Koye DN, Shaw JE: Trends in incidence of total or type 2 diabetes: systematic review. BMJ 2019;366:l5003
29. International Diabetes Federation: *IDF Diabetes Atlas - Eighth edition*. Brussels, Belgium, International Diabetes Federation, 2017
30. Frérot M, Lefebvre A, Aho S, Callier P, Astruc K, Aho Glélé LS: What is epidemiology? Changing definitions of epidemiology 1978-2017. PLoS One 2018;13:e0208442
31. Ebrahim S: Surveillance and monitoring: a vital investment for the changing burdens of disease. Int J Epidemiol 2011;40:1139-1143

32. Groseclose SL, Buckeridge DL: Public Health Surveillance Systems: Recent Advances in Their Use and Evaluation. Annu Rev Public Health 2017;38:57-79

33. Berkelman R, Stroup D, Buehler J: Publich health surveillance. In *Oxford Textbook of Public Health: The Methods of Public Health* Detels R, McEwen J, Beaglehole R, Tanaka H, Eds. Oxford, England, Oxford University Press, 2002

34. Thacker SB, Berkelman RL, Stroup DF: The Science of Public Health Surveillance. J Public Health Policy 1989;10:187-203

35. Declich S, Carter AO: Public health surveillance: historical origins, methods and evaluation. Bull World Health Organ 1994;72:285-304

36. Bhopal R: *Concepts of Epidemiology, An integrated introduction to the ideas, theories, principles and methods of epidemiology*. New York, USA, Oxford University Press, 2002

37. Buehler J: Surveillance. In *Modern Epidemiology* Rothman KJ, Greenland S, Lash TL, Eds. Philadelphia, USA, Wolters Kluwer Health, Lippincott Williams & Wilkins, 2008

38. Ebrahim S, Davey Smith G: Commentary: Should we always deliberately be non-representative? Int J Epidemiol 2013;42:1022-1026

39. Richiardi L, Pizzi C, Pearce N: Commentary: Representativeness is usually not necessary and often should be avoided. Int J Epidemiol 2013;42:1018-1022

40. Rothman KJ, Gallacher JE, Hatch EE: Rebuttal: When it comes to scientific inference, sometimes a cigar is just a cigar. Int J Epidemiol 2013;42:1026-1028

41. Rothman KJ, Gallacher JE, Hatch EE: Why representativeness should be avoided. Int J Epidemiol 2013;42:1012-1014

42. Schmidt C, Bätzing-Feigenbaum J, Bestmann A, Brinks R, Dreß J, Goffrier B, Hagen B, Laux G, Pollmanns J, Schröder H, Stahl T, Baumert J, Du Y, Gabrys L, Heidemann C, Paprott R, Scheidt-Nave C, Teti A, Ziese T: Integration von Sekundärdaten in die Nationale Diabetes-Surveillance. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2017;60:656-661

43. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics: National Health and Nutrition Examination Survey, Overview, Online: https://wwwn.cdc.gov/nchs/data/nhanes/2017-2018/documents/2017_nhanes_overview_brochure.pdf (Accessed 04 November 2019).
44. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics: About the National Health Interview Survey, Online: https://www.cdc.gov/nchs/nhis/about_nhis.htm (Accessed 09 September 2019).
45. Benoit SR, Hora I, Albright AL, Gregg EW: New directions in incidence and prevalence of diagnosed diabetes in the USA. BMJ Open Diabetes Res Care 2019;7:e000657
46. Gregg EW, Sattar N, Ali MK: The changing face of diabetes complications. Lancet Diabetes Endocrinol 2016;4:537-547
47. Hamman RF, Bell RA, Dabelea D, D'Agostino RB, Dolan L, Imperatore G, Lawrence JM, Linder B, Marcovina SM, Mayer-Davis EJ, Pihoker C, Rodriguez BL, Saydah S: The SEARCH for Diabetes in Youth Study: Rationale, Findings, and Future Directions. Diabetes Care 2014;37:3336-3344
48. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, Bell R, Badaru A, Talton JW, Crume T, Liese AD, Merchant AT, Lawrence JM, Reynolds K, Dolan L, Liu LL, Hamman RF, for the SEARCH for Diabetes in Youth Study: Prevalence of Type 1 and Type 2 Diabetes Among Children and Adolescents From 2001 to 2009. JAMA 2014;311:1778-1786
49. Pettitt DJ, Talton J, Dabelea D, Divers J, Imperatore G, Lawrence JM, Liese AD, Linder B, Mayer-Davis EJ, Pihoker C, Saydah SH, Standiford DA, Hamman RF: Prevalence of Diabetes in U.S. Youth in 2009: The SEARCH for Diabetes in Youth Study. Diabetes Care 2014;37:402-408
50. Mayer-Davis EJ, Lawrence JM, Dabelea D, Divers J, Isom S, Dolan L, Imperatore G, Linder B, Marcovina S, Pettitt DJ, Pihoker C, Saydah S, Wagenknecht L: Incidence Trends of Type 1 and Type 2 Diabetes among Youths, 2002–2012. N Engl J Med 2017;376:1419-1429

51. Dabelea D, Stafford JM, Mayer-Davis EJ, D'Agostino R, Jr, Dolan L, Imperatore G, Linder B, Lawrence JM, Marcovina SM, Mottl AK, Black MH, Pop-Busui R, Saydah S, Hamman RF, Pihoker C, for the SEARCH for Diabetes in Youth Research Group: Association of Type 1 Diabetes vs Type 2 Diabetes Diagnosed During Childhood and Adolescence With Complications During Teenage Years and Young Adulthood. JAMA 2017;317:825-835
52. Reynolds K, Saydah SH, Isom S, Divers J, Lawrence JM, Dabelea D, Mayer-Davis EJ, Imperatore G, Bell RA, Hamman RF: Mortality in youth-onset type 1 and type 2 diabetes: The SEARCH for Diabetes in Youth study. J Diabetes Complications 2018;32:545-549
53. Gabrys L, Schmidt C, Heidemann C, Baumert J, Du Y, Paprott R, Teti A, Wolf I-K, Ziese T, Scheidt-Nave C: Diabetes Surveillance in Germany – Background, concept and prospects. Journal of Health Monitoring 2017;2:83-95
54. Gabrys L, Heidemann C, Baumert J, Teti A, Du Y, Paprott R, Ziese T, Banzer W, Böhme M, Borrmann B, Busse R, Freitag M, Hagen B, Holl R, Icks A, Kaltheuner M, Koch K, Kümmel S, Kuhn J, Kuß O, et al.: Selecting and defining indicators for diabetes surveillance in Germany. Journal of Health Monitoring 2018;3:3-21
55. Heidemann C, Scheidt-Nave C: Prevalence, incidence and mortality of diabetes mellitus in adults in Germany – A review in the framework of the Diabetes Surveillance. Journal of Health Monitoring 2017;2:98-121
56. Röckl S, Brinks R, Baumert J, Paprott R, Du Y, Heidemann C, Scheidt-Nave C: All-cause mortality in adults with and without type 2 diabetes: findings from the national health monitoring in Germany. BMJ Open Diabetes Res Care 2017;5:e000451
57. Scheidt-Nave C, Kamtsiuris P, Gößwald A, Hölling H, Lange M, Busch MA, Dahm S, Dölle R, Ellert U, Fuchs J, Hapke U, Heidemann C, Knopf H, Laussmann D, Mensink GBM, Neuhauser H, Richter A, Sass A-C, Rosario AS, Stolzenberg H, et al.: German health interview and examination survey for adults (DEGS) - design, objectives and implementation of the first data collection wave. BMC Public Health 2012;12:730

58. Robert Koch-Institut: DEGS: Studie zur Gesundheit Erwachsener in Deutschland, Online: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Studien/Degs/degs_node.html (Accessed 04 November 2019).
59. Goffrier B, Schulz M, Bätzing-Feigenbaum J: Administrative Prävalenzen und Inzidenzen des Diabetes mellitus von 2009 bis 2015. Versorgungsatlas-Bericht 2017:10.20364/VA-20317.20303
60. Andersen PK, Keiding N: Multi-state models for event history analysis. Stat Methods Med Res 2002;11:91-115
61. Fix E, Neyman J: A simple stochastic model of recovery, relapse, death and loss of patients. Hum Biol 1951;23:205-241
62. Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen PK: Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 2009;18:195-222
63. Andersen PK, Geskus RB, de Witte T, Putter H: Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol 2012;41:861-870
64. Brinks R, Landwehr S: Age- and time-dependent model of the prevalence of non-communicable diseases and application to dementia in Germany. Theor Popul Biol 2014;92:62-68
65. Brinks R, Hoyer A: Illness-death model: statistical perspective and differential equations. Lifetime Data Anal 2018;24:743-754
66. Brinks R, Landwehr S: A new relation between prevalence and incidence of a chronic disease. Math Med Biol 2015;32:425-435
67. Tönnies T, Imperatore G, Hoyer A, Saydah SH, D'Agostino RB, Divers J, Isom S, Dabelea D, Lawrence JM, Mayer–Davis EJ, Pihoker C, Dolan L, Brinks R: Estimating prevalence of type I and type II diabetes using incidence rates: the SEARCH for diabetes in youth study. Ann Epidemiol 2019;37:37-42
68. 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. Diabet Med 2019;36:1217-1225
69. Beyersmann J, Allignol A, Schumacher M: *Competing risks and multistate models with R*. Springer Science & Business Media, 2011

70. Amorim LD, Cai J: Modelling recurrent events: a tutorial for analysis in epidemiology. Int J Epidemiol 2014;44:324-333
71. Andersen PK, Borgan Ø, Gill RD, Keiding N: *Statistical models based on counting processes*. Berlin, Germany, Springer, 1992
72. Kermack WO, McKendrick AG: A contribution to the mathematical theory of epidemics. Proc R Soc Lond A Math Phys Eng Sci 1927;115:700-721
73. Mindell JS, Giampaoli S, Goesswald A, Kamtsiuris P, Mann C, Männistö S, Morgan K, Shelton NJ, Verschuren WMM, Tolonen H, on behalf of the HES Response Rate Group: Sample selection, recruitment and participation rates in health examination surveys in Europe – experience from seven national surveys. BMC Med Res Methodol 2015;15:78
74. Olsen J: Using Secondary Data. In *Modern Epidemiology* Rothman KJ, Greenland S, Lash TL, Eds. Philadelphia, USA, Wolters Kluiwer Health, Lippincott Williams & Wilkins, 2008
75. Rawshani A, Rawshani A, Franzén S, Eliasson B, Svensson A-M, Miftaraj M, McGuire DK, Sattar N, Rosengren A, Gudbjörnsdottir S: Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes. N Engl J Med 2017;376:1407-1418
76. Brinks R, Hoyer A, Landwehr S: Surveillance of the Incidence of Non-Communicable Diseases (NCDs) with Sparse Resources: A Simulation Study Using Data from a National Diabetes Registry, Denmark, 1995–2004. PLoS One 2016;11:e0152046
77. Murray CJL: Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Organ 1994;72:429-445
78. Hansen MB, Jensen ML, Carstensen B: Causes of death among diabetic patients in Denmark. Diabetologia 2012;55:294-302
79. Tönnies T, Brinks R, Hoyer A: Lebenszeit ohne Typ 2 Diabetes und verlorene Lebensjahre durch Typ 2 Diabetes in Deutschland im Jahr 2015 und 2040. Abstract from the 14th annual meeting of the German Society for Epidemiology 2019 Online: https://www.eventclass.org/contxt_dgepi2019/online-program/session?s=AG07%7C2#e106 (Accessed 14 November 2019)

80. Alva ML, Hoerger TJ, Zhang P, Cheng YJ: State-level diabetes-attributable mortality and years of life lost in the United States. Ann Epidemiol 2018;28:790-795
81. Magliano DJ, Shaw JE, Shortreed SM, Nusselder WJ, Liew D, Barr ELM, Zimmet PZ, Peeters A: Lifetime risk and projected population prevalence of diabetes. Diabetologia 2008;51:2179-2186
82. Muschik D, Tetzlaff J, Lange K, Epping J, Eberhard S, Geyer S: Change in life expectancy with type 2 diabetes: a study using claims data from lower Saxony, Germany. Popul Health Metr 2017;15:5
83. Sullivan DF: A single index of mortality and morbidity. HSMHA Health Reports 1971;86:347-354
84. Statistisches Bundesamt: Bevölkerung Deutschlands bis 2060 - Tabellenband - Ergebnisse der 13. koordinierten Bevölkerungsvorausberechnung, Online: https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060.html (Accessed 14 November 2019).
85. Jacobs E, Hoyer A, Brinks R, Icks A, Kuß O, Rathmann W: Healthcare costs of Type 2 diabetes in Germany. Diabet Med 2017;34:855-861
86. International Diabetes Federation: *IDF Diabetes Atlas - Seventh edition*. Brussels, Belgium, International Diabetes Federation, 2015
87. United Nations, Department of Economic and Social Affairs, Population Division: World Mortality 2017 - Data Booklet, Online: https://www.un.org/en/development/desa/population/publications/pdf/mortality/World-Mortality-2017-Data-Booklet.pdf (Accessed 14 November 2019).
88. Heidemann C, Du Y, Paprott R, Haftenberger M, Rathmann W, Scheidt-Nave C: Temporal changes in the prevalence of diagnosed diabetes, undiagnosed diabetes and prediabetes: findings from the German Health Interview and Examination Surveys in 1997–1999 and 2008–2011. Diabet Med 2016;33:1406-1414
89. Brinks R, Hoyer A, Rolka DB, Kuss O, Gregg EW: Comparison of surveillance-based metrics for the assessment and monitoring of disease

detection: simulation study about type 2 diabetes. BMC Med Res Methodol 2017;17:54-54
90. Geiss LS, Bullard KM, Brinks R, Hoyer A, Gregg EW: Trends in type 2 diabetes detection among adults in the USA, 1999-2014. BMJ Open Diabetes Res Care 2018;6:e000487

91. 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 Med Res Methodol 2015;15:98

Appendix

Appendix from *Estimating prevalence of type I and type II diabetes using incidence rates: the SEARCH for diabetes in youth study, Tönnies T., Imperatore G., Hoyer A., Saydah S.H., D'Agostino R.B., Divers J., Isom S., Dabelea D., Lawrence J.M., Mayer–Davis E.J., Pihoker C., Dolan L., Brinks R., Annals of Epidemiology, 37: 37-42, (2019)*

Appendix from *Projected number of people with diagnosed Type 2 diabetes in Germany in 2040, Tönnies T., Röckl S., Hoyer A., Heidemann C., Baumert J., Du Y., Scheidt-Nave C., Brinks R., Diabetic Medicine, 36: 1217-1225, (2019)*

Appendix from *Excess mortality for people diagnosed with type 2 diabetes in 2012 – Estimates based on claims data from 70 million Germans, Tönnies T., Hoyer A., Brinks R., Nutrition, Metabolism & Cardiovascular Diseases, 28: 887-891, (2018)e*