Can Routine Data be used to Determine the Target Population of Patients with Type 2 Diabetes in Early Benefit Assessments in Germany?

Können Routinedaten zur Bestimmung der Zielpopulation in der Indikation Diabetes mellitus Typ 2 in frühen Nutzenbewertungen in Deutschland genutzt werden?

Authors
Corinna ten Thoren 1, Anja Schwalm 1, Sarah Mostardt 1, Dietmar Weber 2, Peter Ihle 3, Lutz Altenhofen 1

Affiliations
1 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen IQWiG, Gesundheitsökonomie, Köln, Germany
2 Diabetologische Schwerpunktpraxis, Köln, Germany
3 PMV Forschungsgruppe, Klinik und Poliklinik für Psychiatrie, Psychosomatik und Psychotherapie des Kindes- und Jugendalters des Universitätsklinikums Köln, Germany

Key words
Routine data, DIMDI, type 2 diabetes, early benefit assessment, target population

Schlüsselwörter
Routinedaten, DIMDI, Diabetes mellitus Typ 2, frühe Nutzenbewertung, Zielpopulation

Bibliography
DOI https://doi.org/10.1055/a-0948-5301
Online-Publikation: 2019
Gesundheitswesen
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 0941-3790

Correspondence
Dr. Corinna ten Thoren
Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen IQWiG, Versorgung und Gesundheitsökonomie, Im Mediapark 8
50670 Köln
Germany
corinna.ten-thoren@iqwig.de

ABSTRACT

Objectives Since 2011, early benefit assessment of all new drugs launched in Germany is mandatory. The exact determination of the appropriate target population (i.e. patients eligible for a drug) plays an important role for subsequent price negotiations. In type 2 diabetes, the size of the target population varies considerably between company dossiers submitted for assessment. Our aim was to explore whether routine data from all persons insured in German statutory health insurance (SHI) funds can be used to derive information on the size of the target population with type 2 diabetes.

Methods We explored how the data available at the German Institute of Medical Documentation and Information (DIMDI) can be used to obtain the information required. A data-based concept was chosen and the selection criteria were developed in a multidisciplinary project group. Before finalizing the database query, the criteria were evaluated in a test database and the database query was then repeatedly modified.

Results At the time of the design of our analysis in 2017, the most recent data available at DIMDI were for 2013. The algorithm we developed for identifying patients with type 2 diabetes and classifying them according to their medication, based primarily on the combination of ICD and ATC codes, enabled us to determine the size of target populations for different indications in diabetes mellitus type 2.

Conclusion Our methodological approach seems to be suitable to determine target populations in type 2 diabetes.
Introduction

Since 2011, an early benefit assessment has been required for all new drugs launched in Germany [1]. Evidence submitted in dossiers by pharmaceutical companies is assessed by the Institute for Quality and Efficiency in Health Care (IQWiG) Germany’s health technology agency, followed by an appraisal by the supreme decision-making body in the German health care system, the Federal Joint Committee (G-BA). The exact determination of the size of the appropriate target population (i.e., patients eligible for a specific drug within a one-year period) is important for subsequent price negotiations. The target population in type 2 diabetes is classified into several groups based on the type of antidiabetic medication prescribed. As a result, patients are assigned to the following, most common, groups: patients with antidiabetic monotherapy, with dual fixed- or loose-dose combination therapy, and with triple or more combination therapy. If insulin is part of the therapy regimen, classification of patients took place separately.

In type 2 diabetes, the size of the target population varies considerably between the company dossiers submitted for assessment [2]. In Germany, non-commercial public health surveillance databases are currently only in the process of being established [3]. Pharmaceutical companies therefore usually obtain information on the size of the target population reported in dossiers from routine data analyses of various commercial databases. These analyses differ with regard to the database, the inclusion criteria and the calculation methods applied. Furthermore, the reporting of these methods is often incomplete [2].

Our aim was to explore whether routine data from all persons insured in German statutory health insurance (SHI) funds can be used to derive information on the size of the target population in type 2 diabetes. So far, no standard procedures are available to define different types of diabetes patients with the explicit use of routine data [4]. Reported possible misclassifications of type 1 and type 2 diabetes are the main challenge [5]. As an indicator of the validity of our methodology, we also decided to calculate the prevalence of type 1 and 2 diabetes in the German population to enable a comparison with published data. In this publication, we focus on the methodology of the approach. More information about the results, as compared to other approaches, will be available in subsequent publications as well as in the final report on the IQWiG website.

Methods

The prevalence of diabetes varies between different SHI funds in Germany due to differences in the characteristics of the SHI members (e.g., distribution of age, sex, social status, regions). We therefore decided to use data from the German Institute of Medical Documentation and Information (DIMDI), which has a database covering all of the approximately 70 million SHI fund members [6]. The DaTrAv (“Datentransparenzverordnung”, data transparency regulation) data are generated from the yearly national morbidity risk compensation scheme (Morbi-RSA) between the different SHI funds and are collected by the Federal Insurance Office. Since the implementation of the data transparency regulation in 2012, authorized organizations may use aggregated data to answer predefined research questions. In this process, to guarantee the anonymity of the SHI members, a group reported in the result tables has to consist of at least 30 patients. The DIMDI DaTrAv database contains, among other things, the following information [7]: administrative data of the SHI members (e.g., year of birth, sex, days insured), inpatient and outpatient diagnoses (International Classification of Diseases ICD-10 code), data on outpatient medication (pharmaceutical registration number). No information on specific medical parameters (e.g., HbA1c) is available. Data on SHI members whose membership terminated (for whatever reason, e.g., death) are eliminated from the database. At the time of the design of our analysis (2017), data of the year 2013 were the most recent available at DIMDI.

A comprehensive database query is necessary to access the DaTrAv data. The query is developed by the applicant or by the DaTrAv data analyst. For the generation of the database query and the analysis of the results, an in-depth understanding of routine data analyses and of diabetes and its treatment is required. The project team, which was formed at the beginning of 2017, comprised a diabetologist, an expert for routine data analyses of patient data (member of the research group “PMV forschungsgruppe”), and IQWiG researchers with expertise in early benefit assessments of antidiabetic drugs. Based on the structure of the DIMDI DaTrAv database, the team had to define selection criteria in order to

1. Reliably identify patients with diabetes,
2. Distinguish patients with type 2 diabetes from patients with other types of diabetes and
3. Classify patients into different medication groups.


Diskussion Der gewählte methodische Ansatz ermöglichte es, den Umfang von Zielpopulationen für verschiedene Anwendungsgebiete in der Indikation Diabetes mellitus Typ 2 zu bestimmen.
Moreover, the target population referred to patients with type 2 diabetes continuously treated within a period of exactly one year.

In several project meetings, we discussed different approaches regarding the choice and combination of selection criteria approaches like the one applied in the CoDiM-study analyzing the cost burden of diabetes mellitus [8, 9]. A data-based concept was chosen: Before finalizing the database query, we repeatedly tested our set of selection criteria in the database of a large SHI fund and examined the impact of modifications.

Results

1. Identification of patients with diabetes

As the target population had to have been continuously treated over a one-year period, only those patients insured continuously in 2013 and at least one day in 2014 were included. Moreover, detailed and consistent data on the age and sex of these patients had to be available in the database. To identify all patients with diabetes, we chose a combination of three criteria: a corresponding ICD diagnosis, repeated prescriptions of antidiabetic drugs, and exclusion criteria.

Table 1 ICD-10 diagnoses.

<table>
<thead>
<tr>
<th>ICD-10 code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E10.-</td>
<td>Type 1 diabetes mellitus</td>
</tr>
<tr>
<td>E11.-</td>
<td>Type 2 diabetes mellitus</td>
</tr>
<tr>
<td>E12.-</td>
<td>Malnutrition-related diabetes mellitus</td>
</tr>
<tr>
<td>E13.-</td>
<td>Other specified diabetes mellitus</td>
</tr>
<tr>
<td>E14.-</td>
<td>Unspecified diabetes mellitus</td>
</tr>
</tbody>
</table>

Table 2 ICD-10 diagnoses excluded.

<table>
<thead>
<tr>
<th>ICD-10 code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E28.2</td>
<td>Polycystic ovarian syndrome</td>
</tr>
<tr>
<td>E66.-</td>
<td>Obesity</td>
</tr>
<tr>
<td>O24.4 / O24.9</td>
<td>Diabetes mellitus arising in pregnancy</td>
</tr>
<tr>
<td>R73.-</td>
<td>Elevated blood glucose level</td>
</tr>
</tbody>
</table>

Table 3 Combination of the inclusion criteria.

<table>
<thead>
<tr>
<th>Included</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥ 2 quarters with a diagnosis of diabetes mellitus</td>
</tr>
<tr>
<td></td>
<td>1 quarter with a diagnosis of diabetes mellitus</td>
</tr>
<tr>
<td></td>
<td>AND ≥ 1 day with a prescription of an antidiabetic drug</td>
</tr>
<tr>
<td></td>
<td>≥ 2 days with a prescription of an antidiabetic drug</td>
</tr>
<tr>
<td></td>
<td>WITHOUT an exclusion diagnosis *</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Excluded</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 day with a prescription of an antidiabetic drug</td>
</tr>
<tr>
<td></td>
<td>WITHOUT a diagnosis of diabetes mellitus</td>
</tr>
<tr>
<td></td>
<td>≥ 2 days of prescriptions of an antidiabetic drug</td>
</tr>
<tr>
<td></td>
<td>WITHOUT a diagnosis of diabetes mellitus but</td>
</tr>
<tr>
<td></td>
<td>WITH an exclusion diagnosis *</td>
</tr>
<tr>
<td></td>
<td>1 quarter with a diagnosis of diabetes mellitus</td>
</tr>
<tr>
<td></td>
<td>WITHOUT a prescription of an antidiabetic drug</td>
</tr>
</tbody>
</table>

* see Table 2

All inpatients and outpatients for whom a repeated ICD diagnosis for diabetes (acc. to E10 to E14, > Table 1) could be found in at least two quarters of 2013 were included (so called M2Q criteria). For all outpatients, a specification of the diabetes diagnosis “G” (“gesichert”), i.e “confirmed” was essential to exclude those patients for whom the diagnosis had not been conclusively confirmed during the period considered.

In cases where a corresponding ICD diagnosis for diabetes existed in only one quarter, we also considered prescriptions of anti-diabetic drugs. To identify these prescriptions, at first, the pharmaceutical registration numbers in the database had to be linked to corresponding Anatomic-Therapeutic Chemical (ATC) classification system codes [10]. All ATC codes for antidiabetic drugs (A10) were included, with a separate consideration of insulin (A10A) and other antidiabetic drugs (A10B). The ATC code A10X (other antidiabetic drugs) was excluded, as the number of cases was too small to ensure the anonymity of the SHI members and no information could therefore be obtained from the DIMDI DaTraV database. In cases of no existing diabetes diagnosis, we included those cases with at least two days of prescriptions of antidiabetic drugs within the one-year period and no documentation of other specific diagnoses (ICD-10 diagnoses leading to exclusion are listed in > Table 2).

Table 3 presents an overview of the combination of the three inclusion criteria.

2. Classification of patients with type 2 diabetes

To identify all patients with type 2 diabetes as accurately as possible, we developed a methodological approach in cooperation with our diabetologist. Since patients could have visited different physicians in 2013 (e.g. general practitioner, diabetes specialist, hospital physicians) who made different diagnoses, it is possible that different ICD-10 diagnoses were available for the same patient within the course of the year analysed. To distinguish between patients with different types of diabetes, we developed a classification scheme based on a combination of ICD and ATC codes and particularly considered prescriptions of insulin and other antidiabetic drugs (> Fig. 1).

To classify patients as having type 2 diabetes, the following constellations needed to apply:

- an ICD-10 diagnosis of E11.- or E12.- and no ICD-10 diagnosis of E10.-
- a combination of the ICD-10 diagnosis E10.- and E11.-/E12.- and prescriptions of other antidiabetic drugs than insulin (insulin prescriptions were also possible)
- a combination of the ICD-10 diagnosis E10.- and E11.-/E12.- and no antidiabetic prescriptions
- an ICD-10 diagnosis of E10.- and no insulin prescriptions
- an ICD-10 diagnosis of E14.- alone and prescriptions of other antidiabetic drugs than insulin (insulin prescriptions were also possible)
- an ICD-10 diagnosis of E14.- alone and no prescription of an antidiabetic drug

patients without the ICD-10 diagnoses E10.- to E14.- but with prescriptions of other antidiabetic drugs than insulin (insulin prescriptions were also possible)
To classify patients as having type 1 diabetes, the following constellations needed to apply:

- An ICD-10 diagnosis of E10.- and no ICD-10 diagnosis of E11.- or E12.- and at least one prescription of insulin
- A combination of the ICD-10 diagnoses E10.- and E11.-/E12.- and prescriptions of insulin without prescriptions of other antidiabetic drugs.

Patients who had an ICD-10 diagnosis of E13.- without an ICD-10 diagnosis of E10.- or E11.- were classified as “type 3 diabetes” (other specified type) and patients who had an ICD-10 diagnosis of E14.- without an ICD-10 diagnosis of E10.- to E13.- and with prescriptions of insulin without prescriptions of other antidiabetic drugs were classified as having an “unspecified type of diabetes”.

### 3. Classification of patients into medication groups

In early benefit assessments, the target population in type 2 diabetes consists of the following groups:

- Patients with antidiabetic monotherapy (e.g. metformin)
- Dual fixed- or loose-dose combination therapy (e.g. metformin + another oral antidiabetic)
- Triple or more combination therapy.

If insulin is part of the therapy regimen, a separate classification of patients takes place:

- Monotherapy with insulin
- Dual combination therapy with insulin as a combination drug
- Triple or more combination therapy with insulin as a combination drug

Patients treated continuously with one antidiabetic drug in the course of the reference year 2013 can simply be classified as patients receiving monotherapy. However, difficulties arise with the documentation of prescriptions of 2 (or more) different antidiabetic drugs within the reference year. As packages of antidiabetic drugs can contain different numbers of tablets with different dosages (and different dosages can be prescribed for different patients), packages can last for different treatment periods. In combination therapy, recurrent prescriptions are therefore not necessarily issued on the same day. In this case, it is difficult to distinguish between patients with a dual combination therapy and patients with a trial therapy (e.g. the drug was changed immediately after initiation because of intolerance) or patients with a switch from one monotherapy to another after a longer period of time. We therefore decided to take the individual last date of the prescription of an antidiabetic drug and then look for a previous prescription within the same group of drugs (▶ Table 4) in the course of the refer-
Thoren C et al. Can Routine Data be … Gesundheitswesen

Table 4  Groups of anti-diabetic drugs that are considered in the approach of the individual last prescription.

<table>
<thead>
<tr>
<th>Group of drug</th>
<th>ATC codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulins</td>
<td>A10A-</td>
</tr>
<tr>
<td>Bolus insulin</td>
<td>A10AB</td>
</tr>
<tr>
<td>Basal insulin</td>
<td>A10AC, A10AE</td>
</tr>
<tr>
<td>Premixed insulin</td>
<td>A10AD</td>
</tr>
<tr>
<td>Inhaled insulin</td>
<td>A10AF</td>
</tr>
<tr>
<td>Metformin</td>
<td>A10BA02-A10BD03, -05, -07, -08, -10</td>
</tr>
<tr>
<td>Sulfonylureas</td>
<td>A10BB-A10BD04, -06, -15</td>
</tr>
<tr>
<td>DPP-4 inhibitors</td>
<td>A10BH-A10BD07, -08, -10</td>
</tr>
<tr>
<td>Glinides</td>
<td>A10BX02, -03</td>
</tr>
<tr>
<td>SGLT2 inhibitors</td>
<td>A10BX09</td>
</tr>
<tr>
<td>Thiazolidinediones</td>
<td>A10BG02, -03 A10BD03 to 06</td>
</tr>
<tr>
<td>GLP-1 receptor agonists</td>
<td>A10BX04, -07, -10</td>
</tr>
<tr>
<td>Other antidiabetic drugs</td>
<td>A10BF01, -02</td>
</tr>
</tbody>
</table>

In this context year (validation of prescriptions). With the documentation of another prescription of an antidiabetic drug that belonged to the same group of drugs in this period, the patient was classified in the target population (according to his/her medication). If no further prescription of an antidiabetic drug belonging to the same group of drugs existed in the given period, the patient was not included. By this method of validation of the prescriptions, we only included continuously treated patients. In respect of insulin, we separately considered treatment with basal, bolus and inhaled insulin as well as the fixed-dose combination of basal and bolus insulin (premixed insulin). By doing this, we were able to distinguish between different treatment regimens such as conventional therapy (CT – fixed-dose combination of basal and bolus insulin) and intensified conventional therapy (ICT – loose-dose combination of basal and bolus insulin).

To determine the appropriate period for validation, we chose a data-based approach: Within our test database, we analysed different periods (from 30 to 365 days before the last prescription). The median period was 90 to 100 days before the last individual prescription. Within a 90-day period, up to 75 % of the recurrent prescription cases would be covered; in a 180-day period, it would be over 90 %. The results of the 365-day period showed little difference from results of the 180-day period. To cover 100 % of all cases with a recurrent prescription, a 2-year period would be necessary. We decided to work with the 180-day period, as it covered the vast majority of recurrent prescriptions within one year. This allowed us to use data solely from the reference year 2013, thus avoiding inconsistencies of methodological approaches.

By proceeding this way, we classified patients with continuous prescriptions of one group of antidiabetic drugs as patients with monotherapy, while patients with continuous prescriptions of 2 groups of antidiabetic drugs received the classification as patients with dual combination therapy, and so forth.

Data processing at DIMDI

Based on the above considerations, the expert for routine data analyses (PMV) developed the database query, which was subsequently submitted to DIMDI. Both project partners, IQWiG and PMV, made use of the possibility to apply as partners. This allowed us to share technical information between all involved institutions (IQWiG, PMV, DIMDI) during the data processing procedure. As only information on the drug ID code (“Pharmazentralnummer”, PZN) is available in the DIMDI DaTraV database, we obtained permission from the Research Institute of the Local Health Care Funds (WidO) to use their main drug database of the year 2013, containing information on the relationship between the ATC and PZN codes. DIMDI screened the resulting tables created by the analysis scripts (written in Structured Query Language, SQL) to ensure the anonymity of the SHI members. As mentioned previously, no results were available for groups with fewer than 30 cases in at least one cell of the tables. The project partners (IQWiG and PMV) checked the preliminary results twice at the DIMDI guest workstation. Through this review, it was possible to assess whether an additional aggregation of the data due to data protection issues and revision of the script was required.

Discussion

The main aim of our study was to explore whether routine data from German statutory SHI funds is a sufficient basis to derive information on the size of the target population in type 2 diabetes, while excluding patients with other types of diabetes and categorizing those with type 2 diabetes according to their medication. Moreover, the study was conducted to obtain deeper insight into the specific methodological challenges and their impact on the size of the target population.

In contrast to other diabetes studies, which focus primarily on the determination of the prevalence and incidence of diabetes mellitus (e. g. [11, 12]) we considered not only the diagnosis, but also the antidiabetic medication of the patients as selection criteria. By combining ICD diagnoses and medication, we tried to detect as many patients with diabetes as possible. To reduce misclassifications, some indications were regarded as exclusion criteria, namely indications in which patients do not seem to suffer from diabetes, but still receive an antidiabetic medication. Based on their antidiabetic therapy we also classified some patients as having diabetes even if they did not have a corresponding ICD diagnosis. Even under considering the exclusion criteria, this could still be a source of overestimation of patient numbers.

It was possible to develop a methodological approach that allowed categorizing patients with type 2 diabetes into respective medication groups according to their prescriptions. For the target population, only patients with type 2 diabetes and continuous therapy were eligible. Moreover, the target population referred to exactly one year. We therefore decided to use only information in the database from the year 2013. This approach implies that persons who terminated their SHI membership in 2013 are not included in our dataset, as automatic deletion of such data from the DIMDI DaTraV database takes place in the year of termination. To obtain information on the number and treatment of these persons in the year 2013, information from the previous year 2012 would have been required. We decided against this approach, as in 2012 changes to the ATC classification were made, which would have led to methodological problems. Moreover, the classification of patients into the medication groups (individual last prescription and another prescription within the same group of drugs within 180 days be-
Can Routine Data be …

The fact that SHI members with a terminated membership in the year analysed were not included in the medication groups implies that there will be some underestimation of the size of the target population. The reason for this is that the prevalence of diabetes might have been higher in these patients than in the remaining SHI members. An underestimation could also occur from our approach to take the individual last prescription and look for another prescription within the same group of drugs within the course of the year. This implies exclusion of incident patients diagnosed in the last quarter, as no further diagnosis or prescription of an anti-diabetic drug was available. In total, the results on the numbers of patients in the medication groups were largely consistent with the findings in company dossiers submitted to the G-BA [2]. According to our diabetologist, they are plausible from a medical point of view.

Compared to other routine data sources, the DIMDI DaTraV database has the advantage of including information on all SHI fund members in Germany [6]. No information is lost when persons switch from one SHI fund to another. Moreover, no adjustments of the results are necessary to ensure appropriate distributions of key parameters such as age and sex. This would have been necessary with the use of data from only one SHI fund.

Limitations arise from the fact that pharmaceutical companies have no authorization to access the DaTraV data; they cannot use the DaTraV data for early benefit assessments on their own. Further limitations relating to most types of routine data analyses apply: Without further detailed medical information, only diagnosed patients can be identified, while no information on the estimated number of unknown cases with diabetes is available. Moreover, the data on ICD diagnoses, medications, dates for doctor’s appointments and prescriptions can contain mistakes. These arise, among other things, from coding procedures and different standards for defining the type of diabetes. Results based on the DIMDI DaTraV database therefore reflect more real-world situations. We tried to address these problems by implementing validation steps; however, it is still possible that some patients were wrongly included in our dataset.

Generation of the DIMDI DaTraV database is based on the data of the Morbi-RSA and contains only a selection of SHI data [6]. It therefore contains information more relevant to costs than to medical aspects. Information on inpatient and outpatient care is restricted as, for example, no information is available about compensations concerning the so called German Uniform Assessment Standard (“Einheitlicher Bewertungsmaßstab”, EBM) or Diagnosis Related Groups (DRG) [6]. Nevertheless, the information on the indication of diabetes available in the database was sufficient to determine the target population. Further limitations can arise with the use of the DIMDI DaTraV database for other indications: Because of the restricted actuality, the database is more suited to address these problems by implementing validation steps; however, it is still possible that some patients were wrongly included in our dataset.

In summary, DIMDI data are a valuable source for obtaining information on the target population in type 2 diabetes in early benefit assessments. The methodological approach seems to be suited to determine the target population in type 2 diabetes. For the comparison of routine data analyses within the same indication, a detailed and standardized description of the methodological approach is important [13].

Acknowledgements

With special thanks to the members of the German Institute of Medical Documentation and Information (DIMDI) for their support. The authors thank Consuela Jacobi-Yniguez and Natalie McGauran for the editorial support.

Conflict of Interest

Authors declare that they have no conflict of interest.

References

[7] Deutsches Institut für Medizinische Dokumentation und Information. Informationssystem Versorgungsdaten (Datentransparenz); Datensatzbeschreibung; Version 02.09.000. Im Internet: https://www.dimdi.de/static/downloads/deutsch/versorgungsdaten-datensatzbeschreibung.pdf Stand: 10.09.2018


