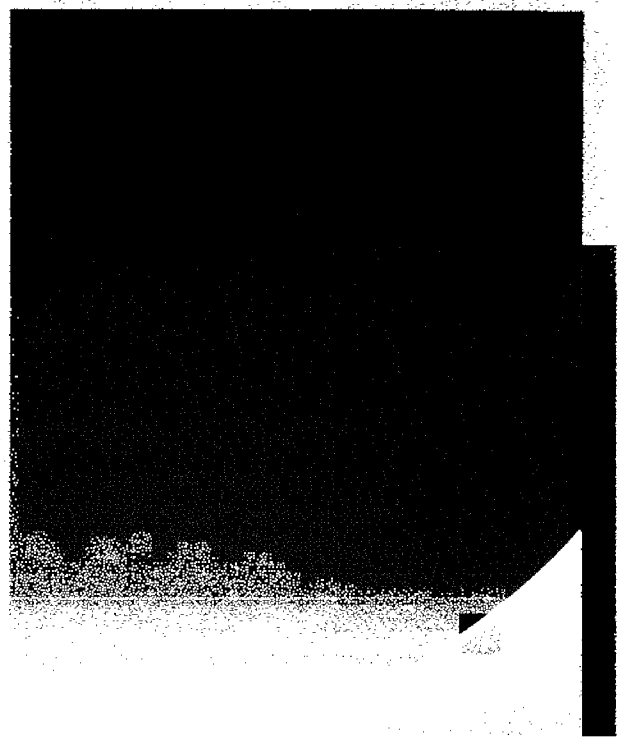
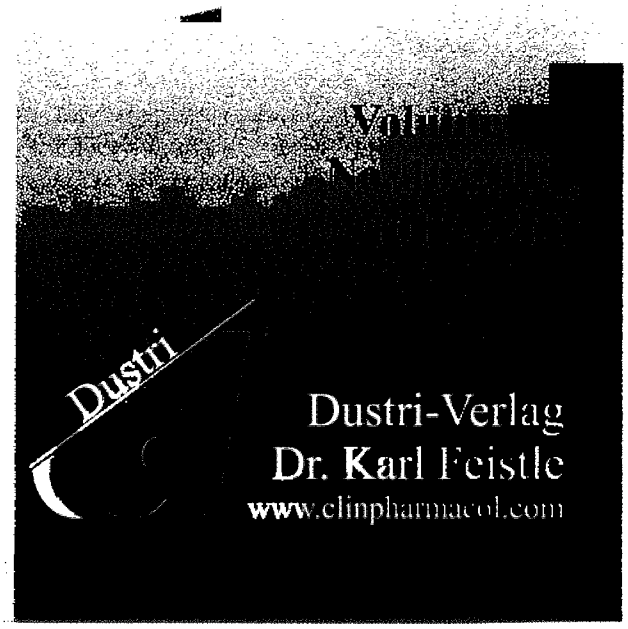
clinical pharmacology and therapeutics



Validity and representativeness of the "Disease Analyzer" patient database for use in pharmaco-epidemiological and pharmacoeconomic studies

H. Becher[1], K. Kostev[2] and D. Schröder-Bernhardi[2]

[1]*Ruprecht-Karls-University, Heidelberg and*
[2]*IMS Health, Frankfurt am Main, Germany*



International Journal of Clinical Pharmacology and Therapeutics, Vol. 47 – No. 10/2009 (617-626)

Original

ISSN 0946-1965

Validity and representativeness of the "Disease Analyzer" patient database for use in pharmacoepidemiological and pharmacoeconomic studies

H. Becher[1], K. Kostev[2] and D. Schröder-Bernhardi[2]

[1]Ruprecht-Karls-University, Heidelberg and [2]IMS Health, Frankfurt am Main, Germany

Key words
epidemiological studies – patient database – validation criteria

List of abbreviations:
ICD = International Classification of Diseases, ATC = (Anatomical Therapeutic Chemical) classification of medicines, RKI = Robert-Koch-Institute, COPD = Chronic Obstructive Pulmonary Disease, BMI = Body Mass Index, OAD = Oral Antidiabetic Drug, PPI = Proton Pump Inhibitors, N1, N2, N3 = Pack size coding used as ("N1" for the small pack size, "N2" for the medium pack size, "N3" for the large pack size), C02 = ATC-Code for Other Antihypertensives, C03 = ATC-Code for Diuretics, C07 = ATC-Code for Beta Blocking Agents, C08 = ATC-Code for Calcium Channel Blockers, C09 = ATC-Code for Agents acting on the Renin-Angiotesin System.

Received
April 24, 2009;
accepted
July 1, 2009

Correspondence to
Dr. K. Kostev
Center of Excellence
Patient Data, Germany,
IMS Health GmbH & Co.
OHG, Darmstädter
Landstraße 108, 60598
Frankfurt am Main,
Germany
kkostev@
de.imshealth.com

Abstract. Objectives: Patient and health care databases are available in many countries. These are often based on routinely collected diagnosis and prescription data. Various research questions, such as those related to pharmacoepidemiological health services or drug supply, can be evaluated on the basis of these databases. In Germany, the Disease Analyzer patient database is the largest database of its kind. Using various validity criteria, the representativeness of this database is examined with respect to variables relevant to pharmacoepidemiological and pharmacoeconomic studies. Methods: The Disease Analyzer patient database contains data on diagnoses, prescriptions, risk factors (such as smoking and obesity), and laboratory values for approximately 10 million patients from Germany, the UK, France, and Austria. The database also contains data from various groups of specialist physicians as well as from general practitioners and specialists for internal medicine. Data from physicians' practices in Germany form the basis of this investigation. To check the validity and representativeness of the data, the distributions of several variables are analyzed. These variables refer partly to the physicians' practices participating in the study and partly to the patients in these practices. The factors observed include prescriptions for generic drugs, the distribution of diagnostic groups among participating physicians' practices, the distribution of patients according to health insurance fund, the most frequent products, the distribution of package sizes prescribed, and the age structure of patients with various incident cancer diagnoses. These factors were compared with available reference statistics. Results: The sampling methods for the selection of physicians' practices appear to be appropriate. Prescription statistics for several drugs were very similar to available data from the pharmaceutical prescriptions report (Arzneimittelverordnungsreport). The age structures for given diagnoses in Disease Analyzer also agreed well with those from corresponding disease registries. Additional comparisons were also in good agreement with data from available sources. Conclusion: The analyses carried out in comparison with reference statistics find no indication of lack of representativeness or validity of the Disease Analyzer database. In principle, the database appears suitable for pharmacoepidemiological and pharmacoeconomic studies. Development and maintenance of large pharmacoepidemiological databases is needed for modern health services. Such databases allow assessment of health care quality and rare adverse drug effects.

Introduction

Patient and health care databases are available in many countries and are often based on routinely collected diagnosis and prescription data. Over time, patient data from such databases have been linked with each other via pseudonyms and analyzed. Examples of these databases include the General Practice Research Database (GPRD), Prescription Pricing Authority (PPA), IMS Prescribing Insights (MIDAS), including the UK IMS Medical Data Index (MDI), and IMS Disease Analyzer. These databases can be used to evaluate important questions concerning health services, for example whether therapy regimens being applied reflect the current state of scientific knowledge, or whether supply shortages, surpluses, or mismatches occur. Using the above-mentioned databases, a great number of studies have been conducted to analyze the duration, adverse effects, success, costs, and courses of, as well as compliance with, therapies and therapy changes. The studies also play an important part in drug safety and risk prevention [Hasford et al.

2004]. To guarantee the scientific relevance of pharmacoepidemiological and pharmacoeconomic studies, a sufficiently valid database is required. The Disease Analyzer patient database is one of the most comprehensive pharmacoepidemiological databases in Europe. This database has been the basis of a number of studies and various peer-reviewed scientific publications in the fields of pharmacoepidemiology [Burkowitz and Brüggenjürgen 2004, Krobot et al. 1999a, 2004, Rathmann et al. 1999], health economics [Icks et al. 2006, Perez et al. 2002, Rathmann et al. 1998b], pharmacovigilance [Farmer et al. 1998a, b, Gaus et al. 2005, Mockenhaupt et al. 1998, 2005, Rathmann et al. 1998a], compliance/persistence [Hasford et al. 2002, Schröder-Bernhardi and Dietlein 2001], pharmaceutical guidelines [Hasford et al. 2007, Krobot 1999b, Schröder-Bernhardi et al. 2004], prescribing behavior [Schröder-Bernhardi and Dietlein 2003], Dietlein and Schröder-Bernhardi 2003], and drug application [Linsell et al. 2005, Sittl et al. 2005].

Studies based on registry data – for our purposes, we include pharmacoepidemiological databases here – have fundamental advantages and disadvantages. The major advantages usually include the large sample size with which one can obtain precise estimates. Furthermore, it is good scientific practice to use existing databases for research questions wherever possible, particularly if the appropriate information cannot be obtained elsewhere or can only be obtained through disproportionate efforts. Health insurance data are another example of databases that can be suitable for addressing medical questions [Ahrens et al. 2007, Grimmsman et al. 2007, Hamann et al. 2003].

Disadvantages include the fact that the database and the data collection procedures are usually not designed with a specific research question in mind. One consequence of this is that variables desirable for a specific research question are often not available. In such cases, a careful analysis is required to determine whether a valid answer to the question can be achieved on the basis of the existing data. An example of this is a recently conducted study on breast cancer survival [von Zeppelin et al. 2007]. In the database, no information was available on the tumor stage (TNM status), a relevant predictor of survival. Plausibility checks were needed to assure that lack of this variable did not introduce any bias into the analysis. In this particular case, the plausibility checks did not indicate any potential bias. However, it may be that a question of interest cannot be answered using only the data available in the database.

Therefore, the number of variables or the length of the observation period is usually not a primary criterion with which to assess a database for use in pharmacoepidemiological studies. Instead, we present in this paper analyses of the Disease Analyzer database with regard to certain criteria that must be considered as prerequisites for valid studies. The authors are aware of the problem that such validity and plausibility analyses may give a good indication of the quality of the database while not allowing conclusions to be generalized for every possible research question. A thorough analysis is always required on a case-by-case basis to determine whether Disease Analyzer contains the necessary information to investigate a particular research question.

A primary prerequisite is that physicians' practices be selected such that the patients within a given medical field are representative of the general population regarding demographic characteristics, diagnoses, and therapies. This can be accomplished by an appropriate sampling scheme as described below.

A number of variables are investigated in this paper, permitting a statement regarding the representativeness and validity of the Disease Analyzer database, even under the low response rate. The details of these analyses are described more extensively in the chapters below. The basis for the analyses is the German part of the Disease Analyzer database.

Few other studies have investigated the quality of comparable databases. Jick et al. [2003] summarized investigations into the validity of the General Practice Research database in the UK and report, for example, comparisons with disease incidence rates from national data. Andersohn and Garbe [2008] underline the importance of valid data for pharmacoepidemiological research to detect rare adverse drug reactions. Motivated by these reports, this paper aims to evaluate the validity of the Disease Analyzer database in Germany.

Table 1. Baseline information from Disease Analyzer in Europe.

	UK	Germany	France	Austria
Patients (in millions)	4.2	17.2	5.2	1.0
Physicians	1,151	3,002	2091	237
Practices	218	2,357	2091	237
Prescriptions (in millions)	206	166	104	18
Prescriptions per patient	49.0	9.7	20.0	18.0
Data period	01/1991 – 07/2008	01/1992 – 08/2008	01/1997 – 08/2008	01/1995 – 08/2008

Table 2. Data on specialist practices in Germany in the period from 09/2005 to 08/2008.

Physician specialty	Physician numbers	No of patients (in thousands)
GP/Internists	1,061	5,300
Cardiologists	30	204
Diabetologists	84	428
Dermatologists	55	698
Gynaecologists	234	1,200
ENT-medicine	81	846
Neurologists/ Psychiatrists	84	404
Orthopaedists	106	939
Pediatricians	149	755
Urologists	72	460
Surgeons	54	529
total	2,010	11,763

Data and Methods

Characteristics of the Disease Analyzer database

In addition to the English GPRD database [Rodriguez and Gutthann 1998], Disease Analyzer is one of the major European patient databases. The database contains data from Germany, the UK, France, and Austria. It allows anonymous access to a selected panel of physicians' practices and patients.

The data are generated directly from the computer in the physician's practice via standardized interfaces and provide daily routine information on patients' diseases and therapies. The practice transmits patient data stored in the physician's computer to IMS on a monthly basis. Before transmission, the data are encrypted for data protection and contain in similar scope and detail the information in the files of patients in the doctor's practice. Each month, the physician receives a so-called doctor feedback report. These reports reflect the physician's own prescription pattern and provide him or her a means to compare their patterns with those of collaborating colleagues in the IMS panel within their specialist group.

Patients and practices can be analyzed both in a cross-sectional and in a longitudinal way. Longitudinal data in Germany partly date back to 1 July 1989. Monthly updates are available 6 weeks after the end of a month. Within the past three years, approximately 10 million patients with 238 million prescriptions have become available at a pan-European level for longitudinal analyses (Table 1). This table shows that the number of prescriptions per patient per year varies largely between countries. This is partly due to the different health systems that offer patients varying levels of autonomy in changing physicians. It is also due to the differences in prescription package sizes between countries.

Altogether, the database contains data from 2,351 practices and approximately 17 million patients from Germany. This database only includes practices and patients for which data are available from the period of 08/2005 to 07/2008.

In addition to data from general practitioners and specialists in internal medicine, data for various specialist groups are also recorded in Germany (Table 2). The Disease Analyzer database provides a complete listing of all relevant patient details for each practice. Table 3 displays the variables collected for each of four subject groups: practices, patients, diag-noses, and therapies. Data on risk factors are only available if they are directly rele-

Table 3. Variables included in Disease Analyzer.

Physician data	**ID number**
	Age, sex
	Type & size of practice
	Region
	Speciality
Patient data	
General data	ID number
	Age, sex
	Insurance
	Risk Factors, e.g., BMI
	Lab tests and results
Diagnoses	Date of diagnosis
	ICD 10 and Read codes
	Original text
	Co-morbidity
	Treated/Untreated
	Referrals and hospitalizations
Therapy	Date of visit
	Product & quantity prescribed
	Molecule, ATC, NDF link
	Treatment Costs
	Therapy switch
	Dosage scheme
	Co-prescription

(ICD = International Classification of Diseases, BMI = Body Mass Index, ATC = Anatomical Therapeutic Classification, NDF = National Disease File).

vant to the diagnosis. For instance, smoking status is recorded for COPD patients, and BMI is recorded for fat metabolism disorder patients. However, the portion of missing risk factor entries is relatively high, so that studies using data on life-style factors are only possible to a limited extent. The data obtained directly from the practice computers are checked for plausibility, linked with relevant additional information such as the price of a medicinal product, ATC and ICD coded, saved, and updated on a monthly basis. The data bank includes only anonymized data in compliance with the regulations of the applicable data protection laws. For a more detailed description, please refer to [Dietlein and Schröder-Bernhardi 2002].

Selection of physicians' practices

The sampling method for the Disease Analyzer database is based on summary statistics of all doctors in Germany published yearly by the German Medical Association (Bundesärztekammer) [http://www.baek.de]. The statistical unit of IMS uses these statistics to determine the panel design according to the following strata: specialist group, German federal state, community size category, and age of physician.

This panel design forms the basis for the acquisition of the practices processed in Disease Analyzer. The acquisition of and support for the practices is performed by cooperating software companies with a standardized interface for IMS that enables the practices to collect the required data and send them to IMS in an anonymized form.

To account for natural fluctuation in the practices and an annual check of the summary statistics by the German Medical Association, the panel design is adjusted each year. Whenever a practice ends its collaboration with IMS, it is replaced by a new one. In 2007, 3.6% of the practices ended their collaboration. The reasons for terminating collaboration are not recorded by IMS, but a large proportion ceases due to the age of the practice's managing physician and retirement.

For their participation, physicians receive both compensation and an evaluation of their prescribing behavior in comparison to those of other physicians. The response rate of physicians asked to participate was 10% in 2007. In that year, a total of 3,200 practices were contacted, of which 321 agreed to participate. The differences in response rates between specialist fields and regions are small. See the discussion for possible implications of the response rate.

Altogether, 11 specialist fields are taken into account in the random sampling plan. For this purpose, the field of internal medicine has been subdivided into 5 subgroups. Furthermore, the field of neurology also includes pediatric and juvenile psychiatrists.

The sampling plan is subdivided into 8 regions, which are summaries of the 16 German federal states. This stratification results in 176 cells derived from the summary statistics with regard to specialist fields and proportional to the summary statistics with regard to

Table 4. Age and regional distributions of general practitioners in comparison to national data.

Physician age	Germany (physicians in %)	Disease Analyzer (n = 497)	
		physicians in %	95% CI
≤ 34	0.5	1.6	(0.5 – 2.7)
35 – 39	6.1	4.2	(2.5 – 6.0)
40 – 49	33.5	31.4	(27.3 – 35.5)
50 – 59	39.3	41.3	(36.9 – 45.6)
60 – 65	15.3	17.9	(14.5 – 21.3)
> 65	5.3	3.6	(2.0 – 5.3)
Region			
I (S – H, HAM, N-S, BRE)	15.8	14.9	(11.8 – 18.0)
II (N-W)	19.7	21.1	(17.5 – 24.7)
III (HES, SAA, R-P)	14.1	14.3	(11.2 – 17.4)
IV (B-W)	13.4	14.3	(11.2 – 17.4)
V (BAY)	17.6	18.3	(14.9 – 21.7)
VI (B)	4.6	4.2	(2.5 – 6.0)
VII (MEC, BRA, S-A)	7.4	7.0	(4.8 – 9.3)
VIII (THÜ, SAC)	7.2	5.8	(3.8 – 7.9)

CI = confidence interval. Source: [Ärztestatistik der Bundesärztekammer 2007].

the German federal states. Within each specialist field, at least 30 doctors must be sampled. Within each region, a minimum of 7 physicians must be sampled within each specialist field to allow for estimates at the specialist field level for each region.

Methods to assess representativeness

Reference data that can be considered as a gold standard are needed to check representativeness. The variables from the Disease Analyzer database that are available for at least partial comparison with reference data are shown in Table 3.

If the response rate of physicians' practices had been high, the sampling scheme described above would already ensure the representativeness of the practices, thus resulting in a representative sample of patients according to specialist field.

In the present study, we apply a number of analyses to check the representativeness of the Disease Analyzer database given the low response rate of the practices. These analyses are:

- Analysis of the characteristics of the participating practices regarding the distributions of physicians' ages and regions. Reference data come from the physicians' statistical analysis (Ärztestatistik) of the German Medical Association from [2007]);
- Analysis of patients' demographic variables by diagnosis group; comparison of incidence and/or prevalence of selected disease groups;
- Comparison of the distributions of variables that are related to the prescription frequency of a drug in Germany. Here reference data come from the actual prescription frequency distribution in Germany from the pharmaceutical prescriptions report (Arzneimittelverordnungsreport) [Statistisches Bundesamt] and the GEK pharmaceutical products report (GEK-Arzneimittel-Report) by the health insurance fund Gmünder ErsatzKasse [Schwabe and Pfaffrath 2005].

Some of the analyses in this paper arose from a study of disease-free survival in patients with breast cancer [Krobot et al. 1999b]. That study investigated the influence of an isopropanolic Cimicifuga racemosa extract (iCR) on recurrence-free survival after breast cancer, including oestrogen-dependent tumors. At the time of the analysis, data from 2,000 practices were available in the German database. Of these, 1,511 practices provided information relevant to the study.

Results

Characteristics of the participating practices

The analysis of physicians' age and regional distributions among participating practices shows that the selection appears to be representative of the general physician population. The basis for comparison for this result is the physicians' statistical analysis of the German Medical Association from 2007. Table 4 illustrates demographic variables by diagnosis.

Table 5. Comparison of the age distribution (in percent) of cancer patients in Disease Analyzer with German registry data.

Age group	Skin cancer (ICD10: C43,44)		Cancer of prostate (ICD10: C61)		Cancer of dig. organs (ICD10: C15-26)		Cancer of breast (ICD10: C50)	
	RKI	Disease Analyzer (n = 2,370)	RKI	Disease Analyzer (n = 2,068)	RKI	Disease Analyzer (n = 3,867)	RKI	Disease Analyzer (n = 2,821)
< 40	4.6	8.9	0.0	0.7	1.1	2.7	5.4	3.4
40 - 44	2.9	4.3	0.1	0.9	1.1	2.2	5.4	4.6
45 - 49	4.2	4.1	0.5	1.5	2.5	3.2	10.7	6.3
50 - 54	4.6	5.0	2.2	2.8	4.1	4.8	10.5	8.9
55 - 59	6.4	5.8	5.6	6.4	6.1	7.1	9.3	9.8
60 - 64	11.1	10.0	15.5	13.0	12.0	12.9	12.1	14.7
65 - 69	14.5	14.1	24.5	20.8	14.4	16.2	12.1	14.8
70 - 74	15.0	12.9	21.4	18.6	15.8	15.3	11.6	12.4
75 - 79	15.2	13.0	16.4	18.3	17.9	15.3	12.6	10.3
> 80	21.5	22.0	13.8	17.0	25.0	20.3	10.2	14.8
total	100	100	100	100	100	100	100	100

Source: [Robert Koch Institut 2008].

Table 6. Comparison of the ratios of newly diagnosed cancer types.

Cancer type	Patients in %	
	RKI	DA (n = 10,877)
Oral cavity and pharynx	2	2
Digestive organs	26	19
Larynx, Trachea, bronchus, and lung	11	9
Melanoma of the skin	3	3
Breast (female)	13	13
Female genital organs	7	4
Male genital organs	12	11
Urinary tract	10	6
Thyroid	1	2
Non-Hodgkin lymphoma or leukemia	6	9

Source: [Robert Koch Institut 2008].

The next group of comparisons refers to the age distribution of patients with particular diagnoses. Selected diagnoses were those with comparatively high incidence as well as available reference statistics.

Table 5 compares the age distribution of incident cases of skin cancer, prostate cancer, cancer of the digestive organs, and breast cancer (female) from the Disease Analyzer database with data from the Robert-Koch-Institute (RKI). The table indicates good agreement in the age distributions.

The following results were obtained from a previous pharmacoepidemiological study [Krobot et al. 1999b]. The statistics on the incidence of cancer in Germany in 2004 published by the RKI were used as references [Robert Koch Institut]. For patients in the Disease Analyzer database, the first mention of a diagnosis of cancer in the period from 06/2005 to 04/2006 was extracted and used to calculate the incidence. The result was then used to calculate the relative incidence by cancer site. These statistics were compared with the corresponding data from [Robert Koch Institut]. There was good agreement in the proportion of incident cases of cancer by tumor site (Table 6).

Breast cancer patients comprised 13% of all incident cancers in 2004 according to the Disease Analyzer data, corresponding exactly to the proportion reported by the RKI for this period [Robert Koch Institut].

We also examined the proportion of newly diagnosed breast cancers classified as "primarily metastasizing mammary carcinoma". According to the Disease Analyzer database, 8.8% of patients were classified in this way, in line with estimates of 5 – 10% from the literature [Feige et al. 2001]. A certain inaccuracy must be assumed for the date

Table 7. Comparison of diabetes patients in 2005 by age group.

Patient age	GEK patients (%)	Disease Analyzer patients (n = 12,533)	
		[redacted]	95% CI
≤ 39	11.4	12.1	(10.7 – 13.4)
40 – 49	14.3	15.9	(14.4 – 17.4)
i050 – 59	23.3	24.1	(22.3 – 25.9)
60 – 69	27.0	27.0	(25.2 – 28.9)
70 – 79	17.8	16.5	(15.0 – 18.1)
> 80	6.3	4.4	(3.5 – 5.2)

CI = confidence interval. Source: [Glaeske and Janhsen 2007].

Table 8. Comparison of antihypertensive patients treated in 2005 by gender.

ATC class	GEK male patients (%)	Disease Analyzer patients (n = 2,156)	
		%	95% CI
C03A diuretics	54.6%	53.0%	(50.2 – 55.8)
C07 β-blockers	57.5%	56.4%	(54.7 – 58.2)
C08 calcium-antagonists	61.3%	58.7%	(56.0 – 61.4)
C09AB ACE-inhibitors	63.8%	63.8%	(62.0 – 65.7)
C09CD sartans	58.6%	57.4%	(54.6 – 60.2)

CI = confidence interval. Source: [Glaeske and Janhsen 2007].

Table 9. Comparison of shares of patients (in percent) by health insurance fund in 2005.

Health insurance fund	BKK Bundesverband (n = 70,500,457 pat.)	Disease Analyzer (n = 1,439,843 pat.)
AOK	35.9	36.3
EAN	31.0	31.3
BKK	20.7	19.9
IKK	6.7	6.4
EAR	2.2	2.3
BKN	2.0	2.3
LKK	1.3	1.3
SEE	0.1	0.1

Source: [BKK 2005].

of a diagnosis of a recurrent cancer. With this in mind, the yearly recurrence rate of 4.75% calculated from the Disease Analyzer database agrees well with the yearly rate of 5 – 8% indicated in the literature [Sauer et al. 2003].

Table 7 demonstrates diabetes patients from the Disease Analyzer database by age group in comparison with GIK patients [Glaeske and Janhsen 2007]. This table suggests that older patients are somewhat underrepresented in Disease Analyzer. Table 8 shows agreement in the proportion of male patients diagnosed with hypertension in 2005 and treated with antihypertensives in 2006, again in comparison with GEK data [Glaeske and Janhsen 2007] (ATC categories CO2, CO3, CO7, CO8, CO9).

Table 9 indicates that the distribution of patients by health insurance fund (non-private) in Disease Analyzer is also very similar to the overall distribution of patients by health insurance fund in Germany [BKK Gesundheitsreport].

Prescriptions

Statistics from the pharmaceutical prescriptions report (Arzneimittelverordnungsreport) from 2004 show that generic prescriptions filled with products from the three leading manufacturers (Ratiopharm Hexal and Stada) made up 22.2% of all such prescriptions. The corresponding proportion in Disease Analyzer is 21%. Good agreement was also found for the percentages of the prescription figures of the three leading manufacturers of prescription products (Sanofi-Aventis, Novartis Pharma and Pfizer Pharma). According to Disease Analyzer, the proportion was 11.8%; the pharmaceutical prescriptions report indicates 12.2% [Statistisches Bundesamt].

The German Diabetes Center (Deutsches Diabetes Zentrum) compared diabetes treatment in Germany according to different sources (Disease Analyzer, BGS, AOK and Praxisnetz) (Table 10). We added the corresponding figures from the Disease Analyzer database. According to data from the German health insurance fund AOK, 28% of patients treated with medicinal products received insulin therapy [Hauner et al. 2003], while the corresponding figure in Disease Analyzer was 26%. Oral antidiabetics (OAD) plus insulin were administered to 11% of patients according to AOK data and 10% according to Disease Analyzer.

Table 10. Treatment of diabetes patients (shares in percent).

Treatment	Disease Analyzer (2004)	BGS (1998)	AOK (2003)	Practice network (2000)
No diabetes drugs	32	30	28	32
Oral antidiabetics (OAD)	52	46		
Insulin	26	25	28	24
OAD only	43		44	
Insulin only	16		17	
OAD + insulin	10		11	

Source: [Altenhofen et al. 2006, Hauner et al. 2003].

Table 11. Comparison of patient shares (in percent) being treated with proton pump inhibitors.

Substance	GEK-patients (%)	Disease Analyzer (n = 2.234)	
		%	95% CI
Omeprazole	48.2	52.6	(50.8 – 54.5)
Pantoprazole	24.4	23.4	(21.9 – 25.0)
Lansoprazole	3.8	4.6	(3.8 – 5.3)
Rabeprazole	1.6	2.0	(1.5 – 2.5)
Esomeprazole	22.0	17.4	(16.0 – 18.8)

CI = confidence interval. Source: [Glaeske and Janhsen 2005].

Table 12. Comparison of prescription shares regarding package sizes.

Package size	Prescriptions in %	
	Arzneiverordnungsreport	Disease Analyzer
N1	29.7	29.4
N2	28.1	28.4
N3	41.8	42.2

Source: [Schwabe and Pfaffrath 2006].

An examination of the data from the Disease Management Program diabetes mellitus Type 2 (DMP) of the Central Research Institute of Ambulatory Health Care in Germany (Zentralinstitut für die kassenärztliche Versorgung) also reveals comparable results regarding data from the region of North Rhine (Nordrhein) [Altenhofen et al. 2006]. The rate of therapy without drugs was 26.1 – 30.6% in the DMP and 29.4% in Disease Analyzer. In the DMP, 46.6 – 46.9% of patients received treatment with OAD, whereas in Disease Analyzer the rate was 46.2%.

A comparison with the GEK pharmaceutical products report [Glaeske and Janhsen 2005] regarding patients undergoing PPI therapy demonstrates some deviations in prescription rates of the products Esomeprazol and Omeprazol (Table 11). Overall, however, the concordance was good.

Finally, the distributions of prescriptions by package sizes N1-N3 (Table 12) in the Disease Analyzer database and the pharmaceutical prescriptions report agree well [Schwabe and Pfaffrath 2005].

Discussion

Pharmacoepidemiological databases are a very useful tool, without which well-directed quality improvements and targeted risk management of adverse drug effects would not be possible [Hasford et al. 2004]. Data from such retrospective databases allow investigations into specific subpopulations – e.g., groups with specific diagnoses – thanks to their size and duration of observation. Regarding the quality of such data, studies have shown that carefully planned observational studies can produce results comparable to those of randomized controlled trials [Benson and Hartz 2000, Cancato et al. 2000].

The results of the analyses in the present study give no reason to conclude that the validity of the Disease Analyzer database is insufficient. We found an agreement - we used the word "agreement" not to indicate agreement in the individual observation, but in the distribution or in the percentages – in most of the variables under consideration. Therefore, the database is generally suitable for use in pharmacoepidemiological studies. Of course, suitability must be re-examined on a case-to-case basis, since each research question demands specific criteria that cannot be considered in an entirely generalized manner in a study like this.

Regarding the age distribution of cancer patients, we noted that the cases in the database are somewhat younger than in the general population. A possible explanation is that older cancer patients may be referred more often directly to hospitals.

The results from the aforementioned study with Cimicifuga racemosa [von Zeppelin 2007] made an important contribution to the safety of applied therapies. At the same time, the results may also serve as the basis for a discussion concerning safety matters with the responsible regulatory authority. The database may therefore be suitable to investigate rare events, risks, and safety concerns regarding prescription pharmaceuticals. It appears plausible that Disease Analyzer is also suitable for other research questions.

The low response rate among physicians is certainly a point of concern and warrants a thorough investigation of the potential for bias when using the Disease Analyzer database in a study. Of course, it is not possible to exclude the possibility of some bias. It is important, however, to determine whether participating doctors demonstrate different prescription patterns or whether patients in participating practices differ from non-participants. Since the main reason for non-participation is reluctance to perform additional administrative tasks that are financially unattractive, we do not believe that a major bias has been introduced. The results from our study support this conclusion.

The selection of analyses performed in this paper may appear arbitrary. However, since there is neither a formal checklist of appropriate analyses nor defined criteria, and since a large number of analyses may be possible, this appearance can hardly be avoided. Within the context of a specific research question, it would be comparatively straightforward to define the validity checks appropriate for exactly that question.

We conclude that the Disease Analyzer database provides a good basis for pharmacoepidemiological studies. Compared with other patient databases [Schröder-Bernhardi et al. 2004], Disease Analyzer is unique, being the only database available that includes and combines all information on physicians, patients, diagnoses, and courses of therapy relevant for the decision-making process.

Conclusion

Comparisons of Disease Analyzer data with standard reference data suggest that the Disease Analyzer database is sufficiently representative and valid. The results of this paper thus support the validity of pharmacoepidemiological and pharmacoeconomic studies based on this database. Establishing and maintaining pharmacoepidemiological databases is necessary for modern health care management. Without such extensive databases, well-directed quality improvements and targeted risk management of adverse drug effects would not be possible.

Acknowledgment

The research of Heiko Becher was supported by IMS Health.

References

Ahrens W, Hagemeier C, Mühlbauer B, Pigeot I, Püntmann I, Reineke A, Steinbach M, Timm J. Hospitalization rates of generic metoprolol compared with the original beta-blocker in an epidemiological database study. Pharmacoepidemiol Drug Saf. 2007; *16:* 1298-1307.

Altenhofen L, Hagen B, Haß W. DMP Diabetes mellitus Typ 2: Was zeigen die bisherigen Daten aus Nordrhein. Kirchheimforum Diabetes 2006 in Münster, 2006.

Andersohn F, Garbe E. Pharmacoepidemiological research with large health databases. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2008; *51:* 1135-1144 (in German).

Ärztestatistik der Bundesärztekammer. (Physicians' statistics of the German Medical Association) 2007, http://www.baek.de.

Benson K, Hartz AJ. A comparison of observation studies and randomised, controlled trials. N Engl J Med. 2000; *342:* 1878-1886.

BKK. BKK Gesundheitsreport 2005, (accessed August 16, 2007). www.bkk.de/htdocs/bkk/psfile/downloaddatei/64/Gesundheit438ae1c8de09a.pdf.

Burkowitz J, Brüggenjürgen B. Resource utilisation of patients with chronic pain conditions before and during treatment with long-acting opioids in Germany. Value in Health. 2004; *7:* 725.

Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observation studies, and the hierarchy of research design. N Engl J Med. 2000; *342:* 1887-1892.

Dietlein G, Schröder-Bernhardi D. Use of the mediplus patient database in healthcare research. Int J Clin Pharmacol Therap. 2002; *40:* 130-133.

Dietlein G, Schröder-Bernhardi D. Doctor's prescription behavior regarding dosage recommendations for preparations of kava extracts. Pharmacoepidemiol Drug Saf. 2003; *12:*.417-421.

Farmer RDT, Todd J-C, MacRae KD, Williams TJ, Lewis MA. Oral contraception was not associated with venous thromboembolic disease in recent study. BMJ. 1998a; *316:* 1090-1091.

Farmer RD, Todd JC, Lewis MA, MacRae KD, Williams TJ. The risks of venous thromboembolic disease among German women using oral contraceptives: a database study. Contraception. 1998b; *57:* 67-70.

Feige A, Rempen A, Würfel W, Jawny J, Caffier H. Frauenheilkunde, 2nd edition, München-Jena: Urban & Fischer Verlag; 2001.

Gaus W, Westendorf J, Diebow R, Kieser M. Identification of adverse drug reactions by evaluation of a prescription database. Methods Inf Med. 2005; *44:* 697-703.

Glaeske G, Janhsen K. GEK Arzneimittelreport 2005, Schriftenreihe zur Gesundheitsanalyse, Band 36. Sankt Augustin: Asgard; 2005.

Glaeske G, Janhsen K. GEK Arzneimittelreport 2007, Schriftenreihe zur Gesundheitsanalyse, Band 55, Bremen-Schwäbisch Gmünd; June 2007.

Grimmsmann T, Schwabe U, Himmel W. The influence of hospitalisation on drug prescription in primary care-a large-scale follow-up study. Eur J Clin Pharmacol. 2007; *63:* 783-790.

Hamann J, Ruppert A, Auby P, Pugner K, Kissling W. Antipsychotic prescribing patterns in Germany: a retrospective analysis using a large outpatient prescription database. Int Clin Psychopharmacol. 2003; *18:* 237-242.

Hasford J, Mimran A, Simons WR. A population-based European cohort study of persistence in newly diagnosed hypertensive patients. Journal of Human Hypertension. 2002; *16:* 569-575.

Hasford J, Schubert I, Garbe E, Dietlein G, Glaeske G. Memorandum zu Nutzen und Notwendigkeit Pharmakoepidemiologischer Datenbanken in Deutschland (GEK-Edition; Band 33). Hippe: Asgard-Verlag; 2004.

Hasford J, Schröder-Bernhardi D, Rottenkolber M, Kostev K, Dietlein G. Persistence in Hypertension. Eur J Clin Pharmacol. 2007; *63:* 1055-1061.

Hauner H, Köster I, von Ferber L. Ambulante Versorgung von Patienten mit Diabetes mellitus, Deutsche Medizinische Wochenschrift. 2003; *128:* 2638-2643.

Icks A, Haastert B, Giani G, Rathmann W. Incremental prescription and drug costs during the years preceding diabetes diagnosis in primary care practices in Germany. Exp Clin Endocrinol Diabetes. 2006; *114:* 348-355.

Jick SS, Kaye JA, Vasilakis-Scaramozza C, Garcia Rodriguez LA, Ruigómez A, Meier CR, Schlienger RG, Black C, Jick H. Validity of the general practice research database. Pharmacotherapy. 2003; *23:* 686-689.

Krobot KJ, Schröder-Bernhardi D, Pfaffenrath V. Migraine consultation patterns in primary care. Cephalalgia. 1999a; *19:* 831-840.

Krobot KJ, Steinberg HW, Pfaffenrath V. Migraine prescription density and recommendations. Results of the PCAOM study. Cephalalgia. 1999b; *19:* 511-519.

Krobot KJ et al. Quantifying delay in access to new medical treatment: an application of risk advancement period methodology. Epidemiology. 2004; *15:* 202-207.

Linsell L, Dawson J, Zondervan K, Randall T. Prospective study of elderly people comparing trearments following first primary care consultation for a symptomatic hip or knee. Fam Pract. 2005; *22:* 118-125.

Mockenhaupt M, Schlingmann J, Schröder W, Rathmann W, Schöpf E, Martinez C. Severe Cutaneous Adverse Reactions (SCAR): The Risk for Diabetes Mellitus and Anti-infective Sulphonamides. B J Dermatol. 1998; *138:* 374-377.

Mockenhaupt M, Messenheimer J, Tennis P. Risk of Stevens-Johnson syndrome and toxic epidermal necrolysis in new users of antiepileptics. Neurology. 2005; *64:* 1134-1138.

Perez E, Schröder-Bernhardi D, Dietlein G. Treatment behavior of doctors regarding Helicobacter pylori infections. Int J Clin Pharmacol Ther. 2002; *40:* 126-129.

Rathmann W, Haastert B, Delling B, Gries FA, Giani G. Postmarketing surveillance of adverse drug reactions: a correlational study approach using multiple data sources. Pharmacoepidemiol Drug Safety. 1998a; *7:* 51-57.

Rathmann W, Haastert B, Roseman J, Gries FA, Giani G. Prescription drug use and costs among diabetic patients in primary health care practices in Germany. Diabetes Care. 1998b; *21:* 389-397.

Rathmann W, Haastert B, Roseman J, Gries FA, Giani G. Cardiovascular drug prescriptions and risk of depression in diabetic patients. J Clin Epidemiol. 1999; *52:* 1103-1109.

Robert Koch Institut. Krebs in Deutschland, Schätzung des Robert Koch-Instituts. (accessed August 16, 2008.) http://www.rki.de/cln_091/nn_204078/DE/Content/GBE/DachdokKrebs/Datenbankabfragen/Neuerkrankungen/neuerkrankungen__node.html?__nnn = true.

Rodriguez LA, Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. Br J Clin Pharmocol. 1998; *45:* 419-426.

Sauer H, Artmann A, de Waal JC, Eiermann W, Gutschow K, Harbeck N, Heinemann V, Kessler M, Lutz L, Mahl G, Prechtl A, Rauthe G, Schmid L, Sittek H, Sommer H, Stieber P, Untch M, Wuttge-Hannig A. Nachsorge, Manual Mammakarzinome, Tumorzentrum München (ed.). München: W. Zuckschwerdt Verlag; 2003.

Schröder-Bernhardi D, Dietlein G. Compliance with prescription recommendations by physicians in practices. Int J Clin Pharmacol Ther. 2001; *39:* 477-479.

Schröder-Bernhardi D, Dietlein G. Lipid lowering therapy: do hospitals influence the prescription behavior of general practitioners? Int J Clin Pharmacol Ther. 2002; *40:* 317-321.

Schröder-Bernhardi D, Roth K, Dietlein G. Off-label-use of PPI andβ-blockers in general practices. Int J Clin Pharmacol Ther. 2004; *42:* 581-588.

Schwabe U, Pfaffrath D. Arzneiverordnungsreport 2005, Heidelberg: Springer Verlag; 2005.

Schwabe U, Pfaffrath D. Arzneiverordnungsreport 2006. Heidelberg: Springer Verlag; 2006.

Sittl R, Nuitjen M, Nautrup BP. Changes in the prescribed daily doses of transdermal fentanyl and transdermal buprenorphine during treatment of patients with cancer and noncancer pain in Germany. Clin Ther. 2005; *27:* 1022-1031.

Statistisches Bundesamt. www.eds-destatis.de

von Zepelin H-H, Meden H, Kostev K, Schröder-Bernhardi D, Stammwitz U, Becher H. Isopropanolic black cohosh extract and recurrence-free survival after breast cancer, Int J Clin Pharmacol Ther. 2007, *45:* 143-154.