RatSWD
Working Paper Series

Working Paper   No. 122

Administrative Data from Germanys Statutory Health Insurances for Social, Economic and Medical Research

Günther Heller

August 2009

Federal Ministry
of Education
and Research

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Administrative Data from Germanys Statutory Health Insurances for Social, Economic and Medical Research

Günther Heller

Research Institute of the Local Healthcare Funds (WIdO)
(guenther.heller[at]wido.bv.aok.de)

Abstract

This article gives a short description of administrative data sources from Germany's statutory health insurances which are of potential interest for social, economic and medical research.

After an introduction, the underlying legal regulations as well as the structure and contents for the most interesting databases are shortly described.

Current and future accessibility of this data sources are discussed in the context of legal and data privacy protection issues.

Aspects of data validity and completeness are provided for the different data sources.

Additionally an up-to-date example from health care research using administrative data is given.

Finally potentials and limitations of research using administrative data medical data from Germanys Statutory Health Insurances are put into perspective

Keywords: Administrative Data, Claims Data, Health Service Research

1.   Introduction

Recent years have seen a growing importance of administrative medical data for scientific purposes. Clinical research, as well as health service research, and health care economy increasingly demand medical data not only from surveys or clinical trails but also from data from entire populations in order to analyse and evaluate medical innovation or developments in a broader sense based upon real life settings. From this perspective it seems reasonable to address the topic of administrative medical data, its structure, availability and usefulness.

To avoid confusion about the term "administrative medical data" a short description will be given: The term depicts data which is has been established for administrative purposes in the field of health care provision. Typically it has been established for reimbursement of health care providers (e. g. doctors, nurses or hospitals), for usage in official statistics, or other administrative purposes. Thus, administrative medical data are secondary data, when used in scientific research contexts. Other frequently used or similar terms are "administrative data", "claims data", "reimbursement data". Administrative medical data may be contrasted to data especially established in medical surveys, in medical trails, or data derived directly from medical records.

There are numerous examples of administrative medical data from the international field. Some well known examples are the Hospital Episodes Statistics (HES) from Great Britain,[1] the Veterans Affairs data collections, or the Medicare databases from the US.[2]

In the following article I will focus on administrative medical data of the statutory health insurance in Germany, from which health care sectors they derive. Their general structure and availability. An example form the field of health service research will be described, and future developments will be provided.

2.   Administrative Medical Data of the Statutory Health in Germany

Since 125 Years medical care in Germany is financed using a health insurance based system. Currently approximately 86 % of the population are insured within a statutory health insurance (GKV), while approximately 14 % are insured in a private health insurance (PKV) (Jacobs, et al. 2006). Health insurance is mostly regulated in Code of Social Law V (SGB V).

---

1   Url: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=537, zitiert am 27.09.2008.
2   An overview of "coded data from administrative sources" in the US is provided by Iezzoni (2003).

For smaller parts, also in the Code of Social Law IX (Rehabilitation) (SGB IX) and XI (Nursing) (SGB XI). Establishment of routine medical data as well as its transfer to the statutory health insurances[3] is also regulated here. Provision of health care in Germany is divided into different sectors. Accordingly, regulations refer to these different sectors of health care provision. For the sectors listed in table 1 health care providers are obliged to provide individualized case specific medical data together with their invoice to statutory health insurances in order to get reimbursed.

Although, it is unclear whether all statutory health insurances have established databases linking these data from different health related contacts and from different health sectors to an (anonymized) individualized level.

Table 1 provides an overview of the different laws regulating different health care sectors (column 1 and 2) according to SGB V as well as examples of the main specific medical contents to be transferred to statutory health insurances. Additionally the availability of Databases on a anonymized individualized level within the Research Institute of the Local Healthcare Funds (WIdO) is provided. The WIdO runs the databases for the regional healthcare funds (AOK). Currently approximately 24 Million people are in the AOK, making it – compared with other health insurances - the largest database of this kind in Germany.

Table 1:   Health Care Provision according to Code of Social Law V (SGB V) in different medical sectors and data transfer to statutory health insurances in Germany

	Medical Sector	Main Medical Contents (Examples)	WIdO Database since (2)
SGB V § 295	Ambulatory care	Diagnosis, type of medical care provided	2004
SGB V § 295	Incapable of work	Diagnosis, time being incapable of working	2006
SGB V § 300	Pharmaceutical prescription	type of pharmaceutical product, price, quantity of prescribed pharmaceutical	1998
SGB V § 301	Hospital care	Admitted hospital, diagnoses, operations, procedures, length of stay	1998
SGB V § 302	Prescription of remedies and medical aids (1)	Diagnosis, type and quantity of medical care provided, price	2004
SGB XI § 93 – 98	Nursing Care	Diagnoses, Type of care	-
SGB IX	Rehabilitation	Admitted hospital, diagnoses, operations, procedures, length of stay	2004

(1) e. g. massages, ergo therapy, physical therapy, prosthesis
(2) case specific, linkable on a anonymized individual level

---

3   Or maybe different insurances when referring to SGB IX and XI.

In general the described administrative medical data contains the following information, which are usually stored in SQL-databases:

- personal identifier
- date(s) of medical care provision (episode)
- type of disease (e. g. ICD)
- type of treatment (e. g. procedural classification)
- invoice
- other information

These data may be linked to additionally administrative data related to the health insured. Therefore information about e. g. survival of the individual, place of residence, status of insurance, or end of insurance may be used for further scientific purposes.

As a consequence it seems possible and tempting to perform individual longitudinal analyses using data from (nearly) all medical sectors and considering all contacts to the medical system, addressing long time follow up endpoints. It is thus of high interest for research analyses.

Accessibility of Data: It may be seen as a serious drawback that the described data is currently only available to health insurances, as well as to some researchers performing research in cooperation with these health insurances (e. g. AOK-Bundesverband et al. 2007, Bramesfeld et al. 2007, Geyer 2008, Grobe et al. 2008, Heller et al. 2004, Heller 2006, Heller 2007, Heller et al. 2007, Ihle et al. 2005, Müller and Braun 2006, Schubert et al. 2007, Swart and Heller 2007, Swart and Ihle 2005).

This limited data access is due to the fact of data privacy protection of individual insured as well as of the involved institutions (e. g. hospitals). However, there are (at least) two regulations which could make these data available to the scientific community in the future:

§ 303 SGB V Data Transparency: According to this act administrative medical data from all health insurances should be pooled in a data trust centre. These data should be made available to health insurers, health authorities, and several other defined user groups, e. g. independent scientific organisations. However, the level of aggregation in which the data may be available to the public is not fixed. Thus, data may be distributed on the individual level, or on some aggregated level. Additionally it is still unclear how expensive using these data would be for external researches. How the data will be compiled, distributed, and to which expense it will be provided is still discussed within the bodies of Germany's joint self administrated health system.

§ 137 SGB V Quality of Medical Care: In the latest health care reform this act was renewed in March 2007. According to this reform a quality agency shall be assigned to develop and provide „inter sectoral" comprehensive quality assurance. This agency is now allowed to use administrative medical data. As a consequence it might be interesting to focus on this agency when thinking about using administrative medical data in future years. Which institution will be the quality agency is still unclear, as the pan-European tendering procedure for this agency is still in progress.

Aspects of Data Quality

Is the needed information present: Administrative medical data has been established primarily for other reasons. Thus, interesting information might not be, or might be insufficiently present. As an example consider social status of an individual or individual quality of life measures. To illustrate this further: there has been a long going discussion, whether administrative medical data provides sufficient information to perform risk adjusted analyses when comparing hospitals performance measurements, e. g. 30-day survival after admission of acute myocardial infarction. While older analyses have questioned sufficient risk adjusted analyses using administrative medical data, recent works form Great Britain and the US showed similar performance of administrative medical data compared to clinical data or clinical register data when predicting survival after hospital admission due to several tracer-diagnoses or procedures (Aylin et al. 2007, Pine et al. 2007). Although it should be noted that good prediction is only a poor indicator for good risk adjustment (Heller and Schnell 2007).

Data-Validity: Administrative medical data is secondary data. But validation of this data usually refers to it's primary purpose. E. g. reimbursement data from the hospital sector is usually checked by health insurances addressing the accuracy of the invoice using plausibility checks (internal validity) as well as external validity by audits comparing transferred administrative medical data with clinical data from medical records. Several contents are usually not important for reimbursement. Time of coded procedures, or admission diagnoses and are therefore neither checked nor corrected. Nevertheless such information may be of great importance in other analytic perspectives. Performing medical administrative data analyses for different purposes should therefore be aware of data validity of every data field used.

To my knowledge there has been no comprehensive validity study from either health sector in Germany examining the validity of administrative medical data from a medical /

scientific perspective. It seems reasonable to perform such an external validation study in the near future before using administrative medical data on a wider basis.

In general, data validity in administrative medical data, may be expected to increase with longer duration of administrative establishment of data transfer. Thus administrative medical data from hospital care or pharmaceutical prescriptions are usually considered to be valid and reliable (regular data transfer since 1998). Data from the ambulatory sector, where data transfer started in 2004, has been questioned recently (Gerste and Gutschmidt 2006, Giersiepen et al. 2007, Trautner et al. 2005). Another aspect might be, whether the interesting information is linked to reimbursement. E. g. Type and amount of prescribed pharmaceuticals is directly linked to reimbursement and is thus considered to be valid. Documentation of (primary and secondary) diagnoses in the hospital sector has dramatically increased during the implementation of DRG-related reimbursement of hospitals, whereas diagnosis from the ambulatory sector are not directly related to reimbursement and have been shown to be invalid to a considerable extent (Gerste and Gutschmidt 2006, Giersiepen et al. 2007, Trautner et al. 2005). However, related to the morbidity orientated risk structure compensation (Morbititäts-Orientierter-Risikostrukturausgleich) this might change dramatically within the nearest future.

Completeness of data: Usually this kind of data is considered to be highly complete in the sense that virtually all cases a present in the data. Several other systems rely on administrative medical data by using them a data source (e. g. DRG-based statistics of Hospital diagnoses from the federal statistical office (Spindler 2008)), or using it as a gold standard (e. g. completeness of documented quality data by the federal office for quality assurance (Veith et al. 2008)).

For several years there is a working group addressing administrative medical data's contents and possible usage (Working Group Secondary Data Analyses, Arbeitsgruppe Sekundärdatenanalyse, AGENS) in Germany. Some years ago a handbook using administrative medical data has been published which provides a good and detailed overview of administrative medical data in Germany (Swart and Ihle 2005). A shorter more up-to-date description may be find in (Swart and Heller 2007). Additionally, guidelines addressing "Good Practice Secondary Data Analysis" have been developed by this group (Swart et al. 2005). Meanwhile they have been adapted in cooperation with several scientific societies, and have been adopted by funding Agencies recently.[4]

---

[4]   url: http://www.gesundheitsforschung-bmbf.de/_media/GPS.pdf, cited 02/10/2008.

3. A Health service research example using administrative medical data

The volume outcome relationship of very low birth weight infants (VLBWs) might serve as an example of health service research using administrative medical data. In Germany, the Federal Joint Committee (Gemeinsamer Bundesausschuss) has addressed the issue whether a minimum provider volume should be introduced for hospitals treating VLBW-infants. Before doing so the Federal Joint Committee commissioned the Institute for Quality and Efficiency in Health Care (Institut für Qualität und Wirschaftlichkeit im Gesundheitswesen, IQWiG) to evaluate whether volume outcome relationship for these patient group exists. In its final report the IQWiG used analyses by Heller based upon administrative Hospital claims data (Heller 2007, IQWiG 2008).

These analyses have meanwhile been extended to spatial simulation analyses for different minimum provider volumes. In a first step distances from patients places of residence to hospitals providing care were calculated. In a second step it was simulated how these distances changed after introducing minimum provider values. Additionally, changes in survival after introduction different minimum provider volumes were estimated (Heller 2009). These and similar analyses are currently being used for the counselling referring to this topic at the Federal Joint Committee.

4. Discussion and Conclusion

- Administrative medical data provides detailed information about medical care provision for large samples or even entire populations.
- Existing diseases, medical therapy as well as health outcomes, are available in these data.
- Especially the possibility to perform individual longitudinal analyses using individual identifiers seems to draw special interest from researchers with clinical economic and social research interests.

In the current article I have aimed to described or at least mention the most important administrative medical data collections in Germany, addressed issues like data validity, completeness and (future) accessibility of administrative medical data. Additionally I have provided an up-to-date example referring to health service research based referring to very low birth weight infants based upon administrative administrative medical data.

- It should be kept in mind, however, that health and illness also exists outside the official medical system within the lay system (Borgetto and Trojan 2007). The onset of a disease might occur long before the contact to a doctor and self administered treatments, e. g. using ubiquitous available over the counter drugs, are known to cure numerous diseases without any contact to the official medical system. Incidence or prevalence studies may be difficult to justify under this restriction.

From this perspective administrative medical data of Statutory Health Insurances tells only part of the story. Thus, depending on the research question, it may seem wise to enrich administrative medical data with other data sources, when addressing research questions.

References:

AOK-Bundesverband, Forschungs- und Entwicklungsinstitut für das Sozial- und Gesundheitswesen in Sachsen Anhalt (FEISA), HELIOS Kliniken, Wissenschaftliches Institut der AOK (WIdO), Editor. Qualitätssicherung der stationären Versorgung mit Routinedaten (QSR).

Aylin, P./Bottle, A. and Majeed, A. (2007): Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. In: BMJ; 334 (7602), 1044.

Borgetto, B./Trojan, A. (2007): Versorgungsforschung und Laiensystem. In: Janßen, C./Borgetto, B. and Heller, G. (Eds.): Medizinsoziologische Versorgungsforschung. Theoretische Ansätze, Methoden, Instrumente und empirische Befunde. Weinheim, 25-47.

Bramesfeld, A./Grobe, T. and Schwartz, F.W. (2007): Who is treated, and how, for depression? An analysis of statutory health insurance data in Germany. Soc Psychiatry Psychiatr Epidemiol 42(9), 740-746.

Gerste, B. and Gutschmidt, S. (2006): Datenqualität von Diagnosen aus dem ambulanten Bereich. Gesundheits- und Sozialpolitik 2006, (3-4), 29-43.

Geyer, S. (2008): Social inequalities in the incidence and case fatality of cancers of the lung, the stomach, the bowels, and the breast. Cancer Causes Control 19(9), 965-974.

Giersiepen, K./Pohlabeln, H./Egidi, G. and Pigeot, I. (2007): Die ICD-Kodierqualität für Diagnosen in der ambulanten Versorgung. Bundesgesundheitsbl - Gesundheitsforsch – Gesundheitsschutz (8), 1028-1038.

Grobe, T.G./Gerhardus, A./A'Walelu, O./Meisinger, C. and Krauth, C. (2008): [Hospitalisations for acute myocardial infarction--comparing data from three different sources]. In: Gesundheitswesen;70 (8-9), e37-46.

Heller, G./Swart, E. and Mansky, T. (2004): Qualitätsanalysen mit Routinedaten. Ansatz und erste Analysen aus dem Gemeinschaftsprojekt "Qualitätssicherung mit Routinedaten" (QSR). In: Klauber, J./Robra, B.P. and Schellschmdit, H. (Eds.): Krankenhaus-Report 2003. Stuttgart/New York, 271-288.

Heller, G. (2006): Sind risikoadjustierte Analysen mit administrativen Routinedaten möglich? In: Hey, M. and Maschewski-Schneider, U. (Eds.): Kursbuch Versorgungsforschung. Berlin, 252-256.

Heller, G./Günster, C./Misselwitz, B./Feller, A. and Schmidt, S. (2007): Jährliche Fallzahl pro Klinik und Überlebensrate sehr untergewichtiger Frühgeborener (VLBW) in Deutschland. Eine bundesweite Analyse mit Routinedaten. Z Geburtshilfe Neonatol; 211 (3), 123-131.

Heller, G. and Schnell, R. (2007): Hospital Mortality Risk Adjustment Using Claims Data. In: JAMA; 297 (18), 1983.

Heller, G. (2009): Auswirkungen der Einführung von Mindestmengen in der Behandlung von sehr untergewichtigen Früh- und Neugeborenen (VLBWs). Eine Simulation mit Echtdaten. In: Klauber, J./Robra, B. and Schellschmidt, H. (Eds.): Krankenhaus-Report 2008. Stuttgart, 183-199.

Ihle, P./Köster, I./Herholz, H./Rambow-Bertram, P./Schardt, T. and Schubert, I. (2005): [Sample survey of persons insured in statutory health insurance institutions in Hessen--concept and realisation of person-related data base]. In: Gesundheitswesen 67 (8-9), 638-645.

Iezzoni, L.I. (2003): Coded data from administrative soruces. In: Iezzoni, L.I. (Ed.): Risk adjustment for measuring health care outcomes. 3rd ed. Chicago.

IQWiG (2008): http://www.iqwig.de/index.681.html.

Jacobs, K./Klauber, J. and Leinert, J. (2006): Fairer Wettbewerb oder Risikoselektion. Analysen zur gesetzlichen und privaten Krankenversicherung. Bonn.

Müller, R. and Braun, B. (Eds.): Vom Quer zum Längsschnitt. Möglichkeiten der Analysen mit GKV-Daten. St. Augustin.

Pine, M./Jordan, H.S./Elixhauser, A./Fry, D.E./Hoaglin, D.C./Jones, B. et al. (2007): Enhancement of Claims Data to Improve Risk Adjustment of Hospital Mortality. JAMA; 297 (1), 71-76.

Schubert, I./Kupper-Nybelen, J./Ihle, P. and Krappweis, J. (2007): [Utilization patterns of dementia patients in the light of statutory health insurance data]. In: Z Arztl Fortbild Qualitatssich, 101(1), 7-13.

Spindler, J. (2008): Fallpauschalenbezogene Krankenhausstatistik: Diagnosen und Prozeduren der Patienten auf Basis der Daten nach § 21 Krankenhausentgeltgesetz. In: Klauber, J./Robra, B. and Schellschmidt, H. (Eds.): Krankenhaus-Report 2007. Stuttgart, 343-370.

Swart, E. and Heller, G. (2007): Nutzung und Bedeutung von (GKV-)Routinedaten für die Versorgungsforschung. In: Janßen, C./Borgetto, B. and Heller, G. (Eds.): Medizinsoziologische Versorgungsforschung. Theoretische Ansätze, Methoden, Instrumente und empirische Befunde. Weinheim, 93-112.

Swart, E. and Ihle, P. (Eds.): Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven. Bern.

Swart, E./Ihle, P./Geyer, S./Grobe, T. and Hofmann, W. (2005): GPS - Gute Praxis Sekundärdatenanalyse Arbeitsgruppe Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) [GPS--good practice secondary data analysis. Working Group for the Survey and Utilization of Secondary Data (AGENS) of the German Society for Social Medicine and Prevention (DGSMP)]. In: Gesundheitswesen 67(6), 416-421.

Trautner, C./Dong, Y./Ryll, A. and Stillfried, D.G. (2005): Verlässlichkeit von Diagnosen niedergelassener Ärzte in Niedersachsen. Gesundheits- und Sozialpolitik (1-2), 36-42.

Veith, C./Bauer, J./Döbler, K./Eckert, O./Fischer, B. and Woldenga,C. (2008): Qualität sichtbar machen. BQS-Qualitätsreport 2007. Düsseldorf.