

Reporting of Studies Conducted on Austrian Claims Data

Andrea SUNJIC-ALIC^{a,1}, Karin ZEBENHOLZER^b and Walter GALL^a

^aCenter for Medical Statistics, Informatics and Intelligent Systems, Department of Medical Information Management, Medical University of Vienna, Austria

^bDepartment of Neurology, Medical University of Vienna, Austria

Abstract. In healthcare studies, the analysis of claims data is gaining an increasingly important role. Observational studies should be reported in a manner that promotes internal and external validity assessment, with the exact and standardized description of items. Several international guidelines and checklists for reporting on secondary data are available. The aim of this work was to analyse the applicability of reporting guidelines especially for claims data. The STROSA-2 guidelines recommendations were evaluated by means of a report on a study on triptan medications in Austria. Six items were identified which could be expanded to support complete and transparent report on Austrian claims data. Therefore, we would suggest to add some details in the STROSA-2 guidelines concerning study design, legal foundations, data protection, data flow, descriptive results and risk of bias. The guidelines for reporting on Austrian claims data were successfully compiled with additional items. New guidelines should be further processed and tested with strong recommendations to focus on data limitations and legal aspects.

Keywords. Diagnoses-related groups, Austrian claims data, reporting guidelines, STROSA-2 checklist

1. Introduction

Analysing claims data can help to identify gaps in care, to reduce resource consumption and to evaluate effectiveness and safety of medications and medical procedures. Routinely collected health data (e.g. administrative data or patient records) are frequently used for observational studies and health technology assessment [1], [2]. But the limitations and quality of studies using routinely collected health data are also discussed [3]. Especially claims data have several strong limitations, like incompleteness in terms of clinical diagnoses or inaccuracy [4]. They contain information from different parties in form of bills for the care they provided, where data linkage from different sources can lead to data privacy issues [5]. Despite limitations, the information contained on a bill is still valuable. For example, such data as vital signs or clinical notes are omitted, but records of procedures and diagnoses are kept [6].

Observational studies should be reported in a manner that promotes internal and external validity assessment. The exact and standardized description of studies is very

¹ Corresponding Author: Andrea Sunjic-Alic, Center for Medical Statistics, Informatics and Intelligent Systems, Department of Medical Information Management, Medical University of Vienna, Austria, E-Mail: andrea.sunjic.alic@gmail.com

important. There are several international guidelines for reporting on secondary data available.

The STROBE Statement [7], [8] is a reporting guideline of observational studies in epidemiology, comprising a checklist of 22 items. It was developed in 2007 by the STROBE Initiative, an international collaboration of researchers, epidemiologists, statisticians, and journal editors.

RECORD [9], [10] is an international initiative that developed a STROBE-based reporting guideline for studies conducted using routinely collected health data in 2015.

RECORD-PE [11], [12] is a reporting guideline developed in 2018, based on RECORD for reporting of non-interventional pharmacoepidemiological studies.

Considerations not covered by STROBE and RECORD were described and tested by experts from the German Society for Social Medicine and Prevention, the German Society for Epidemiology and the German Network for Health Services Research. This resulted in a new checklist, STROSA-2 (revised) [13] compiled in 2016 for the reporting of secondary data analysis in Germany.

The aim of this work was to analyse and compare the existing guidelines and adapt them in order to address reporting items, which are specific for studies using the Austrian claims data.

STROSA-2 criteria were used for a preliminary report on a triptan medications study. This study was conducted on Austrian claims data for the year 2007 from the GAP-DRG database. The Austrian GAP-DRG (General Approach for Patient-oriented outpatient-based Diagnosis-Related Groups) database [14] has been used for multiple studies [15], [16], [17] in different clinical areas.

2. Methods

2.1. Reporting Guidelines

We analysed the STROBE, RECORD and STROSA criteria for missing details. A short overview of items from each checklist is presented in Table 1.

2.2. Study on triptan medications

We tested the STROSA-2 recommendations by means of a report on a study on triptan medications. The study was conducted on Austrian claims data from the GAP-DRG database. It explored the health status of the prescription patterns and linkages between triptan use, overuse, and cardiovascular diseases of triptan users over the age of 50.

2.3. GAP-DRG

The Austrian healthcare system includes a compulsory healthcare insurance and covers three major areas: inpatient care, outpatient care and the drug supply. Inpatient care is carried out mainly in hospitals. Outpatient care includes all the treatments by practitioners, specialists and outpatient clinics. The supply of prescription drugs takes place in pharmacies.

Table 1. Comparison of STROBE, STROSA 1, STROSA 2 and RECORD reporting guidelines (adapted from [13]); N = no specific recommendations, E = suggested extensions.

STROBE	STROSA-1	STROSA-2	RECORD
Title, Abstract	Title, Abstract	Title, Abstract, Keywords	Title, Abstract, Keywords
Title, Abstract N	Title, Abstract N	Title, Abstract Keywords	Title, Abstract N
Introduction	Introduction	Introduction	Introduction
Background/rationale Objectives	Background/rationale Objectives	Background/rationale Objectives	Background/rationale Objectives
Methods	Methods	Methods	Methods
Study design N	Study design Framework	Study design Data origin	EStudy design N
N	Legal foundation	Legal foundation	EN
N	N	Data protection	EN
N	Data flow	Data flow	EN
N	Study curriculum	Selection criteria	N
N	Unit of analysis	Unit of analysis	N
Participants N	Participants Internal validations	N N	Participants N
Variables	Variables	Variables	Variables
Data sources/ measurement	Classification Systems	N	Data sources/ measurement
Bias	Bias	N	Bias
Study size	Study size	Study population size	Study size
Quantitative variables	Quantitative variables	N	Quantitative variables
Statistical methods N	Statistical methods N	Statistical methods N	Statistical methods Data access and processing
N	N	N	Data linkage
Results	Results	Results	Results
Participants	Selection of study population	Selection of study population	Participants
Descriptive data	Description of participants	Descriptive results	EDescriptive data
Outcome data	Statistical measures	N	Outcome data
Main results	Main results	Main results	Main results
Other analyses	Other analyses	Other results	Other analyses
Discussion	Discussion	Discussion	Discussion
Key results	Key results	Key results	Key results
Limitations N	Limitations	Internal validity and risk of bias	ELimitations
Interpretation	Strengthen Interpretation	Strengthen and weaknesses	N
Generalisability	Generalisability	Interpretation Generalisability	Interpretation Generalisability
N	N	Conclusion	Conclusion
N	N	Conclusion	Conclusion
N	N	Conflicts of interest	Further Information
Funding N	Funding Role of data owners	Funding Role of data owners	Funding N
N	N	Other conflicts of interest	N
N	N	N	Accessibility of proto- col, raw data, and programming code

The research database GAP-DRG of the Main Association of Austrian Social Security Institutions (HBV) contains routinely collected data from various data sources from 2006 to 2011. The data available in GAP-DRG include hospital stays, prescribed

and dispensed medications, the medical field of the physicians, social security providers, sick leaves (including their duration) and others.

The diagnosis-related model (DRG) is the regulatory framework for standardized grouping and scoring of inpatient hospital stays [18]. The assignment of a DRG depends on a patient's age and sex, their main and secondary diagnoses, procedures, comorbidities, complications and discharge status [19].

The billing data of Austrian health insurances (2006–2011) were transmitted to the HVB, integrating data from the Folgekosten-Datenbank (FoKo database), the Minimal Basic Data Set (MBDS) of the Federal Ministry of Health (BMG), billing data of public hospitals and the Private Hospital Financing funds.

All data collected from the individual health insurances and BMG were merged into one database. Double entries were removed and regional characteristics of both, the insurers and the insured parties were aggregated. The review of the data quality focused mainly on the completeness and consistency of the data.

2.4. Legal Aspects

Data protection requirements are governed by the European Data Protection Regulation (GDPR) [20]. Article 9 thereof provides an exemption for processing of personal data for scientific purposes [21]. The Austrian data protection law (DSG) [22] is compliant with European data protection regulations.

The data from the GAP-DRG are pseudonymised and stored on a multi-layered encrypted server. They can be restored to their original state, allowing for individuals to be re-identified. Direct access to the GAP-DRG servers is exclusively granted to the so-called custodians, which are specifically trained in data protection. Only data necessary for a specific study are transferred to the research server.

Access to the research server has to be granted by the HVB. Authorized persons (e.g. statisticians) can access it via a VPN (virtual private network) connection.

3. Results

The STROBE and RECORD criteria meet most of the requirements for reporting of secondary data analysis. Although RECORD was initiated as a reporting standard for routine data analysis, following the introduction of the GDPR in the EU, neither RECORD nor STROBE were fully suitable for reporting of claims data.

We identified six items not covered by STROSA-2, which focus on reporting of Austrian claims data, concerning study design, legal foundations, data protection, data flow, descriptive results and risk of bias.

The report should clearly state the original purpose of the data collection, underlining that the study is based on secondary data. We recommend to capture this under item 5 ('Study design').

The research question should be formulated precisely and included in the study design. This is important for the analysis, study design, data extraction, time framework and costs. A written protocol that determines the study characteristics is essential for the analysis of the secondary data. All studies conducted on Austrian claims data must be approved by the Ethics committee. We recommend to expand item 7 ('Legal foundations') to provide the information on the Ethics committee decision, project name, protocol, enrollees, institution and sponsors.

The analysis of billing data must comply with Austrian and European data protection law. It must not be possible to identify an individual based on their healthcare data used in the research. Safety measures, such as involving data custodians in the data selection process, anonymization or pseudonymisation of personal references, and data encryption, must be applied. Our recommendation is to expand item 8 ('Data protection') to indicate how the data from the claims database are pseudonymised and encrypted.

However, even though the data are merged into a single database, many data preparation steps are necessary for each new project. Data extraction from the database requires both expertise in database management and replication, and deep knowledge of the Austrian healthcare system. We recommend to expand the item 9 ('Data flow') to describe if and how data sources in the database were linked at a macro and micro level.

The biggest disadvantage of the data from the claims database is their incompleteness in terms of patients' outpatient diagnoses, clinical information and socio-demographic characteristics. The BMG provides the MBDS data without the personal reference (anonymized data) [14]. Since MBDS contains only information on stays and services provided but no patient reference, patients' paths cannot be completely retraced throughout the healthcare system. Thus, we recommend to include the characteristics of the study population in item 16 ('Descriptive results') and describe selection criteria based on data quality and linkage.

The STROSA-2 reporting guidelines were developed following the GPS (Good Practice in Secondary Data Analysis) recommendations [23]. Internal validity and risk of bias are aspects covered in the 'Discussion' section of the STROSA-2 guidelines. Any potential selection bias or confounders should be considered and documented. It is a fact that Austrian claims data do not contain the same information value as clinical data. The database covers most of the general population, but for example does not include the information on diagnoses for outpatient treatments. Therefore, some illnesses can only be inferred if the patient was prescribed and dispensed certain medications for the treatment of those diseases. Prescribed medications dispensed to patients who are exempt from payment due to low income or severe chronic diseases are also not included in the database. ICD codes documented for billing purposes often differ from diagnoses used in everyday clinical practice. Our recommendation is to expand item 20 ('Internal validity and risk of bias') to discuss the limitations of the study such as the missing data, and methods and quality of the data linkage.

Items resulting in the report on the study being incomplete, are marked with 'E' in Table 1. Additional recommendations with explanations are presented in Table 2.

Table 2. Recommendations for reporting on Austrian Claims Data

5. Study design (METHODS)
STROSA-2 recommendations
Demonstrate that the study is based on secondary data, what primary purpose it served, and whether you have selected a cross-sectional, cohort, case-control, or other study design within your secondary data analysis
Additional recommendations
Demonstrate that the study is based on secondary data and state the original purpose of the data collection, such as if data were administrative or claims data. Study design should include the methods of the study populations selection and elaboration on used classifications or developed algorithms. <i>Example: One of the data sources integrated in the GAP-DRG database is MBDS which consists of administrative data (e.g. personal data), patient's medical records (e.g. diagnosis), and LKF data (claims).</i>
7. Legal foundation (METHODS)
STROSA-2 recommendations

Describe the legal basis for data disposal and analysis.
Additional recommendations
Include the information on: Vote (Votum) from the Ethics committee, project name and protocol, enrolees, institution and sponsors.
<i>Example: All studies conducted on Austrian health care records must be approved by the Ethics committee.</i>
8. Data protection (METHODS)
STROSA-2 recommendations
Indicate how personal and/or institutional data protection was ensured
Additional recommendations
Indicate how the health data from the research database are pseudonymised and stored.
<i>Example: Pseudonymised data can be restored to their original state, with the addition of information by a data custodian, allowing for individuals to be re-identified. Removing of personal references guarantees the anonymity of the study population, but the possibility of retracing individuals has to be disabled as well. Therefore, for study on triptans, the exact date of birth and full postal codes are removed, keeping only the birth year and the district.</i>
Indicate if and how data from different sources in the database were linked.
<i>Example: Linkage of the claims data from healthcare providers with the data reported through the Regional Health Funds to the Federal Ministry of Health. Data from insurance carriers include social security numbers but personal data are pseudonymised due to privacy rules and regulations. On the other hand, the social security number is completely removed (anonymized) from the records by the Regional Health Funds.</i>
9. Data flow (METHODS)
STROSA-2 recommendations
Represent the data flow and indicate who carried out the data provision and analysis. If necessary, describe how data from different data sources was linked
Additional recommendations
Indicate if and how the data sources in the claims database were linked at a macro level.
<i>Example: Healthcare providers report to insurance carriers, where hospitals additionally report through the Regional Health Funds to the BMG. The data from insurance carriers include pseudonymised personal data. The data reported to BMG are anonymized and it is not possible to determine if two registered hospitalizations belong to the same patient. A record linkage was developed by the DEXHELPP team (Decision Support for Health Policy and Planning: Methods, Models and Technologies based on Existing Health Care Data) to find a unique person identifier for each event recorded in MDBS and hospital reports to the insurance carriers.</i>
Indicate if and which data sources were linked at a micro level, such as data not included in the database, but necessary for the study analysis. Describe the data architecture and all layers of the ETL process. Describe additional materials imported into the data staging layer.
<i>Example: NUTS region codes can be used to link patients district of residence to the ÖSG Versorgungscode (care-supply region codes). The linkage can be performed on the research server.</i>
16. Descriptive results (RESULTS)
STROSA-2 recommendations
Describe the characteristics of the study population as well as exposures and possible confounders. Take into account whether there is a case or personal reference or the level of aggregation of the data
Additional recommendations
Describe in detail the characteristics of the study population and include selection criteria based on the data quality and linkage.
<i>Example: In study on triptan medication, a cardiovascular disease was presumed if participants were dispensed certain medications. The majority of those participants didn't have the diagnosis recorded in the claims database. Characteristics of the study population can be described in the form of a diagram.</i>
20. Internal validity and risk of bias (DISCUSSION)
STROSA-2 recommendations
Discuss the risk of bias (selection bias, information bias, confounding, etc.) and the measures you have taken to determine its presence and extent
Additional recommendations
Discuss the limitations of the study such as missing data. If data linkage was performed, methods and quality of the outcome should be specified.
<i>Example: The GAP-DRG claims database does not contain data on outpatient diagnoses or over-the-counter medications (OTC).</i>

4. Discussion

Austrian claims data can be reused in studies to serve as a foundation for better understanding and decision support in healthcare. For conducting a secondary data analyses we suggest to follow the GPS recommendations [23]. The aim of this work was to analyse the applicability of reporting guidelines specifically to claims data. Several international guidelines and checklists for reporting on secondary data are available. The STROBE and RECORD criteria meet most of the requirements for reporting of secondary data analysis, but none were fully suitable for reporting of Austrian claims data. The STROSA-2 guidelines recommendations were evaluated by means of a report on a study on triptan medications in Austria. We found more details could be provided in six items, in particular study design, legal foundations, data protection, data flow, descriptive results and risk of bias. Other items allowed complete and transparent reporting.

The adaptation of STROSA-2 reporting guidelines can provide reviewers and readers with reliable information, promote the quality of research design and help to assess the validity of the results. The report should include the information on the original purpose of the data collection, vote from the Ethics committee and data protection methods. The characteristics of the study population and selection criteria should be provided. The data architecture, all layers of the ETL process, and how the data sources in the claims database were linked both at a macro and micro level should be described. One of the most important aspects of the report is to discuss the limitations of the study in question, such as missing data (e.g., OTC medications) or inaccurate data (e.g., discrepancies caused by delays in billing dispensed drugs). If data linkage was performed, methods and quality of the outcome should be specified.

Our suggested guidelines extensions must be further analysed and discussed, especially the international applicability. E.g., the main and secondary diagnoses are the most important usable data from claims databases. However, in different countries they may be recorded for different purposes and could as well differ in meaning. It is therefore crucial to describe their interpretation precisely in order to secure the comparability of studies.

New recommendations expand the STROSA-2 criteria to reporting on observational studies conducted on Austrian claims data. The present work is an evaluation of the STROSA-2 guidelines based on a clinical study. It provides suggestions for a refinement on limitations and legal aspects. However, these guidelines could be further developed to support complete and transparent reporting of studies conducted on claims data.

References

- [1] A. Makady, A. van Veelen, P. Jonsson, O. Moseley, A. D'Andon, A. de Boer, H. Hillege, O. Klungel, and W. Goettsch, Using Real-World Data in Health Technology Assessment (HTA) Practice: A Comparative Study of Five HTA Agencies, *PharmacoEconomics*. **36** (2018). doi:10.1007/s40273-017-0596-z.
- [2] B. Blanch, N.A. Buckley, L. Mellish, A.H. Dawson, P.S. Haber, and S.-A. Pearson, Harmonizing post-market surveillance of prescription drug misuse: a systematic review of observational studies using routinely collected data (2000-2013), *Drug Saf*. **38** (2015) 553–564. doi:10.1007/s40264-015-0294-8.
- [3] L.G. Hemkens, E.I. Benchimol, S.M. Langan, M. Briel, B. Kasenda, J.-M. Januel, E. Herrett, and E. von Elm, The reporting of studies using routinely collected health data was often insufficient, *J Clin Epidemiol*. **79** (2016) 104–111. doi:10.1016/j.jclinepi.2016.06.005.

- [4] H. Gothe, P. Ihle, D. Matusiewicz, and E. Swart, Routinedaten im Gesundheitswesen, Handbuch Sekundärdatenanalyse: Grundlagen, Methoden, und Perspektiven, 2., vollständig überarbeitete und erweiterte Auflage., Verlag Hans Huber, Hogrefe AG, Bern, 2014.
- [5] S. Mueller, F. Gottschalk, A. Groth, W. Meeraus, M. Driessen, T. Kohlmann, and T. Wilke, Primary data, claims data, and linked data in observational research: the case of COPD in Germany, *Respir Res.* **19** (2018) 161. doi:10.1186/s12931-018-0865-1.
- [6] Claims Data - The Good, The Bad and The Ugly, *Health IT Answers.* (2014). <https://www.healthitanswers.net/claims-data-the-good-the-bad-and-the-ugly/> (accessed January 10, 2021).
- [7] STROBE Group, STROBE Statement, (2019). <https://www.strobe-statement.org/index.php?id=strobe-home> (accessed January 10, 2021).
- [8] E. von Elm, D.G. Altman, M. Egger, S.J. Pocock, P.C. Gøtzsche, J.P. Vandenbroucke, and STROBE Initiative, The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies, *Int J Surg.* **12** (2014) 1495–1499. doi:10.1016/j.ijso.2014.07.013.
- [9] RECORD Group, RECORD Reporting Guidelines, (2020). <https://www.record-statement.org/> (accessed January 10, 2021).
- [10] E.I. Benchimol, L. Smeeth, A. Guttman, K. Harron, D. Moher, I. Petersen, H.T. Sørensen, E. von Elm, S.M. Langan, and RECORD Working Committee, The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement, *PLoS Med.* **12** (2015) e1001885. doi:10.1371/journal.pmed.1001885.
- [11] S. Langan, S. Schmidt, K.J. Wing, V. Ehrenstein, S. Nicholls, K. Filion, O. Klungel, I. Petersen, H. Sørensen, W. Dixon, A. Guttman, K. Harron, L. Hemkens, D. Moher, S. Schneeweiss, L. Smeeth, M. Sturkenboom, E. von Elm, S. Wang, and E. Benchimol, The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE), *British Medical Journal.* (2018). doi:10.1136/bmj.k3532.
- [12] RECORD-PE Reporting Checklist, (n.d.). <http://www.record-statement.org/checklist-pe.php> (accessed January 12, 2021).
- [13] E. Swart, E.M. Bitzer, H. Gothe, M. Harling, F. Hoffmann, D. Horenkamp-Sonntag, B. Maier, S. March, T. Petzold, R. Röhrig, A. Rommel, T. Schink, C. Wagner, S. Wobbe, and J. Schmitt, SStandardisierte BerichtsROutine für Sekundärdaten Analysen (STROSA) – ein konsentierter Berichtsstandard für Deutschland, Version 2, *Gesundheitswesen.* **78** (2016) e145–e160. doi:10.1055/s-0042-108647.
- [14] E. Florian, GAP-DRG, *GAP-DRG Dokumentation.* (2013). <http://gapdrg.endel.at/dokuwiki/doku.php/gapdrg:start> (accessed January 10, 2021).
- [15] A. Geroldinger, S.K. Sauter, G. Heinze, G. Endel, W. Dorda, and G. Duftschmid, Mortality and continuity of care - Definitions matter! A cohort study in diabetics, *PLoS One.* **13** (2018) e0191386. doi:10.1371/journal.pone.0191386.
- [16] H. Katschnig, C. Straßmayr, F. Endel, M. Berger, G. Zauner, J. Kalseth, R. Sfetcu, K. Wahlbeck, F. Tedeschi, L. Šprah, and CEPHOS-LINK study group, Using national electronic health care registries for comparing the risk of psychiatric re-hospitalisation in six European countries: Opportunities and limitations, *Health Policy.* **123** (2019) 1028–1035. doi:10.1016/j.healthpol.2019.07.006.
- [17] C. Rinner, S.K. Sauter, L.M. Neuhofer, D. Edlinger, W. Grossmann, M. Wolzt, G. Endel, and W. Gall, Estimation of severe drug-drug interaction warnings by medical specialist groups for Austrian nationwide eMedication, *Appl Clin Inform.* **5** (2014) 603–611. doi:10.4338/ACI-2014-04-RA-0030.
- [18] E. Hagenbichler, The Austrian DRG system, Federal Ministry of Health, Sector I/B, 2010. <https://broschuerenservice.sozialministerium.at/Home/Download?publicationId=576> (accessed January 10, 2021).
- [19] Diagnosis Related Group (DRG), *HMSA Provider Resource Center.* (2018). https://hmsa.com/portal/provider/zav_pel.fh.DIA.650.htm (accessed January 10, 2021).
- [20] Protecting Patients' Medical Records under the GDPR, (2019). <https://idpc.org/mt/en/articles/Pages/synapse-article.aspx> (accessed January 10, 2021).
- [21] General Data Protection Regulation (GDPR) Compliance Guidelines, *GDPR.Eu.* (n.d.). <https://gdpr.eu/> (accessed January 10, 2021).
- [22] Datenschutzrecht in Österreich - Datenschutzbehörde, (2020). <https://www.dsb.gv.at/gesetze-in-osterreich> (accessed January 10, 2021).
- [23] P. Ihle, E. Swart, T. Lampert, and S. Klug, GPS – Good Practice in Secondary Data Analysis: Revision after Fundamental Reworking, *Deutsche Gesellschaft Für Epidemiologie.* (n.d.). <https://www.dgepi.de/assets/Leitlinien-und-Empfehlungen/Practice-in-Secondary-Data-Analysis.pdf> (accessed January 10, 2021).