Back pain as reported in survey and claims data – comparison of prevalences and concordance by data linkage

Annemarie Feißel (annemarie.feissel@med.ovgu.de)
Otto von Guericke Universitat Magdeburg
Bernt-Peter Robra
Otto von Guericke Universitat Magdeburg
Christoph Stallmann
Otto von Guericke Universitat Magdeburg
Enno Swart
Otto von Guericke Universitat Magdeburg
Stefanie March
Otto von Guericke Universitat Magdeburg

Research article

Keywords: back pain, survey data, health claims data, data linkage, statutory health insurance funds, validation

Posted Date: August 15th, 2019

DOI: https://doi.org/10.21203/rs.2.12965/v1

Abstract

Objectives Back pain is a burden for those suffering from it as well as for the economy and social insurance system. This explains the great interest in collecting data on back pain in population studies to estimate for instance its prevalence. As part of the lidA study, the agreement between self-reported information on back pain (self-reported prevalence) and health claims data (administrative prevalence) has been examined. Methods In two waves of CAPI, employees (years 1959, 1965) were asked about aspects of health, inter alia, if they experience any symptoms or pain in “upper back or thoracic spine” and “lower back or lumbar region” during the past 12 months. With present informed consent, the CAPI data were individually linked with health claims data (n=1,031). Due to the lack of a gold standard, both data sources were cross-validated. Back pain is logged in claims data across sectors based on two different definitions (2009-2013): Def1) at least one “M54” (ICD-10 coded) entry; Def2) present in at least two quarters within a period of four consecutive quarters. Results The difference between self-reported prevalence (57.8%) and administrative prevalence (58.3%) based on Def1 is minimal in contrast to Def2 (34.6%). Despite almost identical prevalence percentages, Cohen’s Kappa for Def1 indicates a low level of agreement between both data sources (0.23 [95%-CI: 0.17-0.29]). Although prevalence differs significantly based on Def2, Cohen’s Kappa still shows low agreement (0.18 [95%-CI: 0.13-0.24]). Conclusions The low level of agreement between both data sources suggests that each data source identifies a certain group of individuals having few overlaps with one another. Therefore, aiming in identifying the target population for disease-specific health prevention, a combination of both data sources, the prevalence of back pain can be estimated more accurately.

Background

Back pain ranks among the most frequent reasons why people make use of the healthcare system [1]. It is one of the leading causes of work incapacity and years lived with disability (YLDs) [2,3]. It may also lead to early retirement. In addition, back pain creates steep direct costs for treatment and imposes a heavy burden on the economy and the social security system, for instance the pension fund [1,4–6]. Collecting data on back pain in population studies is therefore of great interest in order to estimate prevalence and the need for prevention efforts. However, definitions of back pain show a lack of comparability. Furthermore, a number of different instruments and data pools have been used to collect data on back pain [3,5,7–13]. In some instances and as a consequence, prevalences may differ hugely in public studies and are not comparable. To face this dilemma, it would be helpful to have a standard definition to serve as a basis for uniform operationalization in population-based studies [14].

Health and healthcare research is increasingly using health claims data, including data provided by statutory health insurance funds, in addition to individual data collected in surveys and interviews [15,16]. The advantages of both data sources can be combined to overcome some of the limitations of the individual data sources. The individual linkage of survey data with data provided by statutory health insurers offers new opportunities for analyses conducted by current and future health and healthcare researchers. However, the evaluation and interpretation of linked data sources require special methodological standards. In this respect, the validation of the individual data sources used is of paramount importance [17,18]. The best assessment method is a cross validation, i.e., comparing data from different data sources. Generally, evaluating validity requires a gold standard. There is broad consensus that neither survey nor claims data meets this requirement

[12,19–25]. It has been recommended to use several data sources [20,21,23,26], since the agreement between both data sources depending on the diagnosis examined [24,27]. Self-reported back pain is of limited validity [12], on the other hand there is no need for physicians' contacts if suffering back pain, so claims data also lack on validity.

Due to the missing gold standard, the lack of comparability of definitions of back pain and the limited validity of back pain in different data sources we use several individually linked data sources, to analyze the prevalence of back pain. We have examined – as part of the lidA study – the degree of agreement between subjective information on back pain reported in a computer-assisted personal interview compared to diagnoses documented by doctors (claims data). The recorded diagnoses pertain to claims data provided by statutory health insurers. For cross-validation purposes, both data sources are considered equal after the individual data linkage in order to show the prevalence of back pain more accurately for other content analyses.

Methods

The lidA study is a cohort study on work, age, health, and work participation. It aims at examining the long-term effect of work on the health and work participation of older working employees based on a sample from two age cohorts born in 1959 and 1965, respectively [28]. They present the beginning and the end of Germany's so-called baby boomers who will make up a significant share of the potential older work force in the coming years [29]. The sample was drawn from registry data pertaining to the integrated employment biographies (IEB) of the German Federal Labour Office (Bundesagentur für Arbeit (BA)). It is based on all gainfully employed people covered by social insurance in Germany on December 31, 2009. Public servants and self-employed persons are not part of the study population. The utilization rate is 27.3% [28]. During the computer-assisted interviews (CAPI) of the first two waves in 2011 and 2014, participants were asked for their informed consent to have their interview data linked individual with their claims data. During the second survey in 2014, those participants who had changed statutory health insurance funds were asked again for their consent, as well as those who had not given their consent during the first survey or if other mismatches had occurred [30,31]. A total of 63 percent of all respondents who participated in both survey waves agreed to have their survey and claims data linked [31]. Cooperation agreements with eleven statutory health insurance funds were signed [32]. Finally, data on outpatient and inpatient treatment as well as sick leave and outpatient drug prescription provided by a total of ten statutory health insurance funds[1] were used for the linkage [33-35].

A total of 4,244 persons participated in the follow-up [31]. Of these only those participants were included in the validation study for whom linked survey and claims data was available (n=1,031) (Fig. 1).

During both waves of the primary survey (first wave: 2011, second wave: 2014), study participants were asked exhaustively about their work situation. As regards back pain, we used the questionnaire for the analysis of musculoskeletal symptoms (Nordic questionnaire) [36]. To this end, participants had to answer the following question: "Did you experience any symptoms or pain in the following parts of your body during the past 12 months?" Possible answers included "upper back or thoracic spine" and "lower back or lumbar region". The

analysis included all respondents, who affirmed the question of back pain during the past twelve months (self-reported prevalence) in both waves.

The claims data comprised both in- and outpatient data as well as data on sick leave for the period from 2009-2013. Due to unspecific medication and a lack of means to assign it to specific diagnoses, we decided not to use outpatient drug prescription data. Besides, the data would have comprised subscription pain medication only. Data on self-medication using over-the-counter (OTC-)analgetics is not documented in claims data. Back pain is coded using the ICD-10-Code M54 ("back pain") when it is diagnosed either as a principal or secondary diagnosis made by hospitals, as an outpatient diagnosis (classified as 'confirmed' or 'condition after recovery'), and in medical certificates of sick leave. Two different definitions are used to define administrative prevalence in claims data:

- Definition 1 (Def1): A person is considered to suffer from back pain if "M54" was stated at least once in one of the three sectors between 2009 and 2013.
- Definition 2 (Def2): A person is considered to suffer from back pain if the person received two "M54" diagnoses in at least two quarters within four consecutive quarters (M2Q criterion) [37] across (all three) sectors between 2009 and 2013.

There seems to be a gap of three years, because claims data were used for the whole years 2009 – 2013 but the two questionnaires conducted in 2011 and 2014 respectively. Since we use all respondents, who reported back pain in both waves, the period of claims data is almost covered. With the definition of back pain in the CAPI data (back pain in both waves) and the definitions of back pain in claims data, especially in Def2, the study looks less at actually back pain than at chronic back pain.

Statistical analyses

First, self-reported and administrative prevalence of back pain was determined descriptively. The agreement of survey and claims data was determined using Cohen's Kappa. A Cohen's Kappa < 0.40 indicates a low level of agreement, a Kappa between 0.41 – 0.60 means a moderate level of agreement and a Kappa between 0.61 – 1 means a high level of agreement [38]. The overall prevalence of back pain was calculated as the sum of self-reported and administrative prevalence. In order to examine possible differences between the two cohorts as well as between women and men, Cohen's Kappa was then analyzed carefully according to cohort affiliation and gender. Sensitivity and specificity as well as the positive predictive value and the negative predictive value were not calculated because of the missing gold standard and deviations in both directions.

The analyses were performed using IBM SPSS 24 ©.

[1] The data oft he following health insurance funds are part oft the cumulative data set: AOK Bremen/Bremerhaven – Die Gesundheitskasse, AOK – Die Gesundheitskasse für Niedersachsen, AOK Nordost – Die Gesundheitskasse, AOK NORDWEST – Die Gesundheitskasse, AOK Rheinland/Hamburg – Die

Gesundheitskasse, AOK Rheinland-Pfalz/Saarland – Die Gesundheitskasse, AOK Sachsen-Anhalt – Die Gesundheitskasse, BARMER GEK, IKK gesund plus, Techniker Krankenkasse [33].

Results

Out of 1,031 participants, 596 persons reported back pain in both waves, which equals a self-reported prevalence of 57.8%. Claims data showed an administrative prevalence of 58.3% (n=601) based on Definition 1, and of 34.6% (n=357) according to Definition 2. A comparison of both data sources based on Def1 results in a Cohen's Kappa of 0.23 [95% CI: 0.17-0.29], which indicates a low level of agreement [38]. Based on Def2, the Kappa value is 0.18 [95% CI: 0.13-0.24], which also indicates a low level of agreement.

Table 1: Comparison of back pain in claims data and CAPI data

		CAPI data against Def1				CAPI data against Def2			
					Cohen's Kappa [95%-CI]				Cohen's Kappa [95%-CI]
Claims data	**Total**		yes	no	0.23 [0.17-0.29]		yes	nein	0.18 [0.13-0.24]
		yes	406	190		yes	256	340	
		no	195	240		no	101	334	
	Men		0	1	0.20 [0.11-0.29]		0	1	0.14 [0.07-0.22]
		yes	158	89		yes	89	158	
		no	86	111		no	41	156	
	Women		0	1	0.25 [0.17-0.33]		0	1	0.21 [0.14-0.28]
		yes	248	101		yes	167	182	
		no	109	129		no	60	178	
	1959		0	1	0.24 [0.15-0.33]		0	1	0.18 [0.10-0.26]
		yes	190	78		yes	118	150	
		no	94	106		no	49	151	
	1965		0	1	0.23 [0.15-0.31]		0	1	0.18 [0.11-0.25]
		yes	216	112		yes	138	190	
		no	101	134		no	52	183	

The individual data linkage made it possible to identify another 195 persons (Def1) and 101 persons (Def2), respectively, who received an M54 diagnosis in claims data in addition to the self-reported prevalence of back pain. On the other hand, the analysis identified 190 persons (Def1) and 340 persons (Def2), respectively, who reported back pain but were not diagnosed with it. Consequently, the calculation showed an overall prevalence of 77.2% (n=791) based on Def1 and of 67.6% based on Def2 (n= 697).

If analyzed according to cohort affiliation, Cohen's Kappa shows a low level of agreement of 0.24 [95% CI: 0.15-0.33] (persons born in 1959) and 0.23 [95% CI: 0.15-0.31] (persons born in 1965) based on Def1. Def2

also indicates low agreement (persons born in 1959: 0.18 [95% CI: 0.10-0.26], persons born in 1965: 0.18 [95% CI: 0.11-0.25]) for both age groups.

If analyzed according to gender, Cohen's Kappa for men is 0.20 [95% CI: 0.11-0.29] and 0.25 [95% CI: 0.17-0.33] for women based on Def1, which presents a low level of agreement. Likewise, Def2 also shows low agreement.

Discussion

Due to its design – an individual data linkage of survey and claims data bases in informed consent – the lidA study offered an opportunity to compare data from both data sources in the scope of a cross validation to analyze the prevalence of back pain. To this day, we are not aware of any comparable analysis of back pain based on data from German cohort studies [32]. Since there is no applicable gold standard, both data sources were considered equal. Self-reported prevalence hardly differed from administrative prevalence based on Def1, however, prevalences differed significantly based on Def2. Despite an almost identical prevalence of back pain, Kappa values based on Def1 show only a low level of agreement of both data sources (Tab. 1). Although prevalences based on Definition 2 differ significantly, Cohen's Kappa indicates yet again only little agreement. Irrespective of the definition chosen, both data sources indicate only a small overlap of individuals and thus largely identify different groups of individuals suffering from back pain.

Tisnado et al. [12] also compared self-reported data with medical records regarding back pain and clinical services. At a Kappa of 0.3 and 0.1, respectively, they also arrived at only low levels of agreement. This low value may be explained by the fact that two different groups were logged in the two data sources, back pain in whole versus back pain diagnosed by a physician. On the one hand, respondents subjectively report back pain without having received a diagnosis from a physician. On the other hand, respondents might only receive a M54 diagnosis if doctors see an indication to treat them. Since self-reported information may be incorrect [12], for instance by exhibiting a recall bias, we consider the use of self-reported information due to its limited validity as a limitation for healthcare research. The question regarding pain within the past 12 months may also have been cause for misunderstanding, which may have led to an over-coverage of back pain in the CAPI data. It is also conceivable that respondents mention pain sooner and more frequently if asked about the diseases diagnosed by a doctor. Some people often experience pain but do not automatically go to see a doctor. According to Green et al. [39], only 25% of individuals having health problems will go to the doctor. In this light, a question about a disease diagnosed by a physician would be more specific in order to specify chronic back pain or pain compromising a person's health. Schmidt et al. [9] also assume that the phrasing of a question about back pain within the past twelve months may have contributed to the high prevalence. Then again, the Nordic questionnaire is an internationally recognized instrument that has been tested on a wide spectrum of occupational groups and validated using clinical data [36,40]. Another factor to keep in mind is the temporal dimension of the question with regard to the period of the diagnosis. During each of the two waves, respondents were asked about pain within the past twelve months, which means that two years (2011 and 2014) of experiencing back pain were measured. The claims data considers the M54 diagnosis in the period from 2009 to 2013, which means back pain is recorded for a period of five years.

Admittedly, these reasons may also have led to an under-coverage in the claims data because having complete data depends, among other things, on a doctor's accuracy and thoroughness in terms of documentation [12]. What is more, claims data only comprises claim-relevant diagnoses. If a person experiences back pain but does not mention it or does not get treatment, it will not appear in the data. Consequently, the data only comprises individuals who actually went to the doctor. That means there is a utilization bias. It is also possible that respondents experienced other reportable diseases as more relevant and subjectively more impairing and stressful.

Strengths and limitations of the lidA study

Analyses of the representativeness of the lidA study indicate a largely unbiased sample [28,31,41]. Despite the fact that survey data were linked with claims data provided by only ten out of currently 109 statutory health insurers [42], March et al. [32] and Stallmann et al. [43] found selectivity effects to be negligible.

Conclusions

In conclusion, the low level of agreement between both data sources suggests that a data linkage may present a benefit since each data source identifies a certain group of individuals having few overlaps with one another. There are limitations in each of the used data sources which can be minimized by data linkage [44]. We are aware that the period in both data sources does not completely overlap and different things in both data sources were measured. Nevertheless, within a combination of both data sources, the prevalence of back pain can be estimated more accurately. The data linkage captures both persons with back pain who may not have gotten treatment but should have been treated (self-reported) and persons whose back pain was diagnosed by a physician. For the lidA study we interviewed middle-aged employees covered by social insurance. Regarding back pain the cohort itself is a limitation of the study as the prevalence during employment increases with age [3,14]. Especially with regard to health problems such as back pain where self-reported and claims data may differ, it is worth linking several data sources in order to arrive at the true prevalence. Which definition to use for further analyses of back pain in claims data depends on the question to be asked. For analyses of chronic back pain, we recommend to use a more conservative approach such as Def2 while Def1 is more suitable for studies to estimate the need for prevention measures.

Abbreviations

BA: Bundesagentur für Arbeit

CAPI: computer-assisted personal interviews

CI: confidence interval

Def1: Definition 1

Def2: Definition 2

ICD-10: **I**nternational **C**lassification of **D**iseases, 10. Revision

IEB: integrated employment biographies

lidA: living at work

OTC: over-the-couner

YLD's: years lived with disability

Declarations

Ethics approval and consent to participate

The lidA Study was approved by the Ethics Committees of the University of Wuppertal (Study Coordination) and the University of Magdeburg.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable. For reasons of data protection and deletion deadlines the data is no longer available after the end of the project.

Competing Interests

None declared.

Funding

The lidA study was supported by the German Federal Ministry of Education and Research [grant numbers 01ER0825, 01ER0826, 01ER0827 and 01ER0806].

Authors' contributions

All authors contributed to the design and concept of the study. AF drafted main parts of the manuscript and performed the statistical analysis. SM was responsible for the claims data and data linkage of survey and claims data. All authors contributed to the interpretation of the results. All authors revised the manuscript critically. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Melloh M, Röder C, Elfering A, et al. Differences across health care systems in outcome and cost-utility of surgical and conservative treatment of chronic low back pain: A study protocol. BMC Musculoskelet

Disord. 2008;9:81. doi: 10.1186/1471-2474-9-81.

2. Vos T, Abajobir AA, Abbafati C, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1211-1259. doi: 10.1016/S0140-6736(17)32154-2.
3. Hoy D, March L, Brooks P, Blyth F, Woolf Aea. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis. 2014;(73):968-974. doi: 10.1136/annrheumdis-2013-204428.
4. Wynne-Jones G, Cowen J, Jordan JL, et al. Absence from work and return to work in people with back pain: A systematic review and meta-analysis. Occup Environ Med. 2014;71(6):448-456. doi: 10.1136/oemed-2013-101571.
5. Liebers, Claudia Brendler, Ute Latza. Berufsspezifisches Risiko für das Auftreten von Arbeitsunfähigkeit durch Muskel-Skelett-Erkrankungen und Krankheiten des Herz-Kreislauf-Systems. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Berlin, 2016. doi: 10.21934/baua:bericht20160629.https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2255.pdf?__blob=publicationFile&v=10. Updated 2016. [accessed April 20 2017].
6. Kent PM, Keating JL. The epidemiology of low back pain in primary care. Chiropr Osteopat. 2005;13:13. doi: 10.1186/1746-1340-13-13.
7. Schmidt CO, Raspe H, Pfingsten M, et al. Does attrition bias longitudinal population-based studies on back pain? Eur J Pain. 2011;15(1):84-91. doi: 10.1016/j.ejpain.2010.05.007.
8. Schmidt CO, Raspe H, Kohlmann T. Graded back pain revisited - do latent variable models change our understanding of severe back pain in the general population? Pain. 2010;149(1):50-56. doi: 10.1016/j.pain.2010.01.025.
9. Schmidt CO, Raspe H, Pfingsten M, et al. Back pain in the German adult population: Prevalence, severity, and sociodemographic correlates in a multiregional survey. Spine. 2007;32(18):2005-2011. doi: 10.1097/BRS.0b013e318133fad8.
10. Hüppe A, Brockow T, Raspe H. Chronische ausgebreitete Schmerzen und Tender Points bei Rückenschmerzen in der Bevölkerung. Z Rheumatol. 2004;63(1):76-83. doi: 10.1007/s00393-004-0531-5.
11. Latza U, Kohlmann T, Deck R, Raspe H. Influence of Occupational Factors on the Relation Between Socioeconomic Status and Self-Reported Back Pain in a Population-Based Sample of German Adults With Back Pain. Spine. 2000;25(11):1390-1397. doi: 10.1097/00007632-200006010-00011.
12. Tisnado DM, Adams JL, Liu H, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? Med Care. 2006;44(2):132-140.
13. Kohlmann T, Raspe H. Hanover Functional Questionnaire in ambulatory diagnosis of functional disability caused by backache. Rehabilitation 1996;35(1):I-VIII.
14. Meucci RD, Fassa AG, Faria NMX. Prevalence of chronic low back pain: Systematic review. Rev Saude Publica. 2015;49. doi: 10.1590/S0034-8910.2015049005874.
15. Ohlmeier C, Frick J, Prütz F, et al. Nutzungsmöglichkeiten von Routinedaten der Gesetzlichen Krankenversicherung in der Gesundheitsberichterstattung des Bundes. Bundesgesundheitsblatt

Gesundheitsforschung Gesundheitsschutz. 2014;57(4):464-472. doi: 10.1007/s00103-013-1912-1.

16. Hoffmann F, Andersohn F, Giersiepen K, Scharnetzky E, Garbe E. Validierung von Sekundärdaten. Grenzen und Möglichkeiten. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2008;51(10):1118-1126. doi: 10.1007/s00103-008-0646-y.
17. Hartmann J, Weidmann C, Biehle R. Validierung von GKV-Routinedaten am Beispiel von geschlechtsspezifischen Gesundheitswesen. 2016;78(10):e53-58. doi: 10.1055/s-0035-1565072.
18. Hoffmann W, Bobrowski C, Fendrich K. Sekundärdatenanalyse in der Versorgungsepidemiologie: Potenzial und Limitationen. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2008;51(10):1193-1201. doi: 10.1007/s00103-008-0654-y.
19. Hure AJ, Chojenta CL, Powers JR, Byles JE, Loxton D. Validity and reliability of stillbirth data using linked self-reported and administrative datasets. J Epidemiol. 2015;25(1):30-37. doi: 10.2188/jea.JE20140032.
20. Koller KR, Wilson AS, Asay ED, Metzger JS, Neal DE. Agreement Between Self-Report and Medical Record Prevalence of 16 Chronic Conditions in the Alaska EARTH Study. J Prim Care Community Health. 2014;5(3):160-165. doi: 10.1177/2150131913517902.
21. Carter K, Barber PA, Shaw C. How does self-reported history of stroke compare to hospitalization data in a population-based survey in New Zealand? Stroke. 2010;41(11):2678-2680. doi: 10.1161/STROKEAHA.110.598268.
22. Schubert I, Ihle P, Köster I. Interne Validierung von Diagnosen in GKV-Routinedaten: Konzeption mit Beispielen und Falldefinition. Gesundheitswesen. 2010;72(6):316-322. doi: 10.1055/s-0030-1249688.
23. Corser W, Sikorskii A, Olomu A, Stommel M, Proden C, Holmes-Rovner M. "Concordance between comorbidity data from patient self-report interviews and medical record documentation". BMC Health Serv Res. 2008;8:85. doi: 10.1186/1472-6963-8-85.
24. Hall HI, van den Eeden SK, Tolsma DD, et al. Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. Prev Med. 2004;39(1):27-35. doi: 10.1016/j.ypmed.2004.02.024.
25. Newell SA, Girgis A, Sanson-Fisher RW, Savolainen NJ. The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: a critical review. Am J Prev Med. 1999;17(3):211-229.
26. Swart E, Bitzer EM, Gothe H, et al. A Consensus German Reporting Standard for Secondary Data Analyses, Version 2 (STROSA-STandardisierte BerichtsROutine für SekundärdatenAnalysen). Gesundheitswesen. 2016;78(S 01):e145-e160. doi: 10.1055/s-0042-108647.
27. Barber J, Muller S, Whitehurst T, Hay E. Measuring morbidity: self-report or health care records? Fam Pract. 2010;27(1):25-30. doi: 10.1093/fampra/cmp098.
28. Hasselhorn HM, Peter R, Rauch A, et al. Cohort profile: the lidA Cohort Study-a German Cohort Study on Work, Age, Health and Work Participation. Int J Epidemiol. 2014;43(6):1736-1749. doi: 10.1093/ije/dyu021.
29. Tisch A, Tophoven S. Erwerbseinstieg und bisheriges Erwerbsleben der deutschen Babyboomerkohorten 1959 und 1965: Vorarbeiten zu einer Kohortenstudie [Entry into employment and previous working life of the German baby boomer cohorts 1959 and 1965: preparatory work for a cohort study]. Nürnberg: IAB 2011: IAB-Forschungsbericht, No. 8/2011.

30. Schröder H, Kleudgen M, Steinwede J, March S, Swart E, Stallmann C. Zustimmung von Befragten zur Verknüpfung von Daten - selektionsfrei? [Data linkage - respondents consent without selectivity?]. Gesundheitswesen 2015; 77:e57-62. doi: 10.1055/s-0034-1398594.
31. Steinwede J, Kleudgen M, Häring A, Schröder H. Methodenbericht zur Haupterhebung lidA – leben in der Arbeit, 2. Welle: FDZ Methodenreport 07/2015. [Bundesagentur für Arbeit, Nürnberg]. http://doku.iab.de/fdz/reporte/2015/MR_07-15.pdf; 2015 [accessed November 26, 2016].
32. Individual Data Linkage of Survey Data with Claims Data in Germany—An Overview Based on a Cohort Study. Int J Environ Res Public Health 2017, 14(12), 1543. doi.org/10.3390/ijerph14121543.
33. Kooperierende Krankenkassen. http://www.arbeit.uni-wuppertal.de/fileadmin/arbeit/vor2015/index.php%3Fkooperierende-krankenkassen.html; 2013 [accessed 24 October, 2017].
34. March S, Swart E, Robra B-P. Können Krankenkassendaten Primärdaten verzerrungsfrei ergänzen?: – Selektivitätsanalysen im Rahmen der lidA-Studie. Gesundh ökon Qual manag. 2017;22(02):104-115. doi: 10.1055/s-0042-117963.
35. March S, Powietzka J, Stallmann C, Swart E. Viele Krankenkassen, Fusionen und deren Bedeutung für die Versorgungsforschung mit Daten der Gesetzlichen Krankenversicherung in Deutschland - Erfahrungen aus der lidA-(leben in der Arbeit)-Studie. Gesundheitswesen. 2015;77(2):e32-36. doi: 10.1055/s-0034-1390443.
36. Kuorinka I, Jonsson B, Kilbom A, et al. Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. Appl Ergon. 1987;18(3):233-237. doi: 10.1016/0268-0033(88)90149-0.
37. Busse R, Drösler S, Glaeske G, Greiner W, Schäfer T, Schrappe M. Gutachten zur Auswahl von 50 bis 80 Krankheiten zur Berücksichtigung im morbiditätsorientierten Risikostrukturausgleich. http://www.der-gesundheitsfonds.de/fileadmin/redaktion/Dokumente/Gutachten_Beirat_Krankheitsauswahl_gesamt.pdf. Updated 2007; 2007 [accessed 13 November 2017].
38. Grouven U, Bender R, Ziegler A, Lange S. Der Kappa-Koeffizient. Dtsch Med Wochenschr. 2007;132 Suppl 1:e65-8. doi: 10.1055/s-2007-959046.
39. Green, L.A.; Fryer, G.E.; Yawn, B.P.; Lanier, D.; Dovey, S.M. The ecology of medical care revisited. N Engl J Med 2001; 344:2021–2025. doi: 10.1056/NEJM200106283442611.
40. Crawford JO. The Nordic Musculoskeletal Questionnaire. Occupational Medicine. 2007;57(4):300-301. doi: 10.1093/occmed/kqm036.
41. Schröder H, Kersting A, Gilberg R, Steinwede J. Methodenbericht zur Haupterhebung lidA – leben in der Arbeit; 2013. http://doku.iab.de/fdz/reporte/2013/MR_01-13.pdf, 2013 [accessed 1 March 2015].
42. GKV Spitzenverband. Krankenkassenliste [Register of German statutory health insurance funds]; 2017. https://www.gkv-spitzenverband.de/service/versicherten_service/krankenkassenliste/krankenkassen.jsp, 2017 [accessed 25 October 2017].
43. Stallmann C, Swart E, Robra BP, March S. Linking primary study data with administrative and claims data in a German cohort study on work, age, health and work participation: is there a consent bias? Public health 2017;150:9–16. doi: 10.1016/j.puhe.2017.05.001.

44. Swart E. Health Care Utilization Research using Secondary Data. In: Janßen C, Swart E, von Lengerke T, editors. Health care utilization in Germany: Theory, methodology, and results. New York: Springer; 2014. p.63-86.

Figures

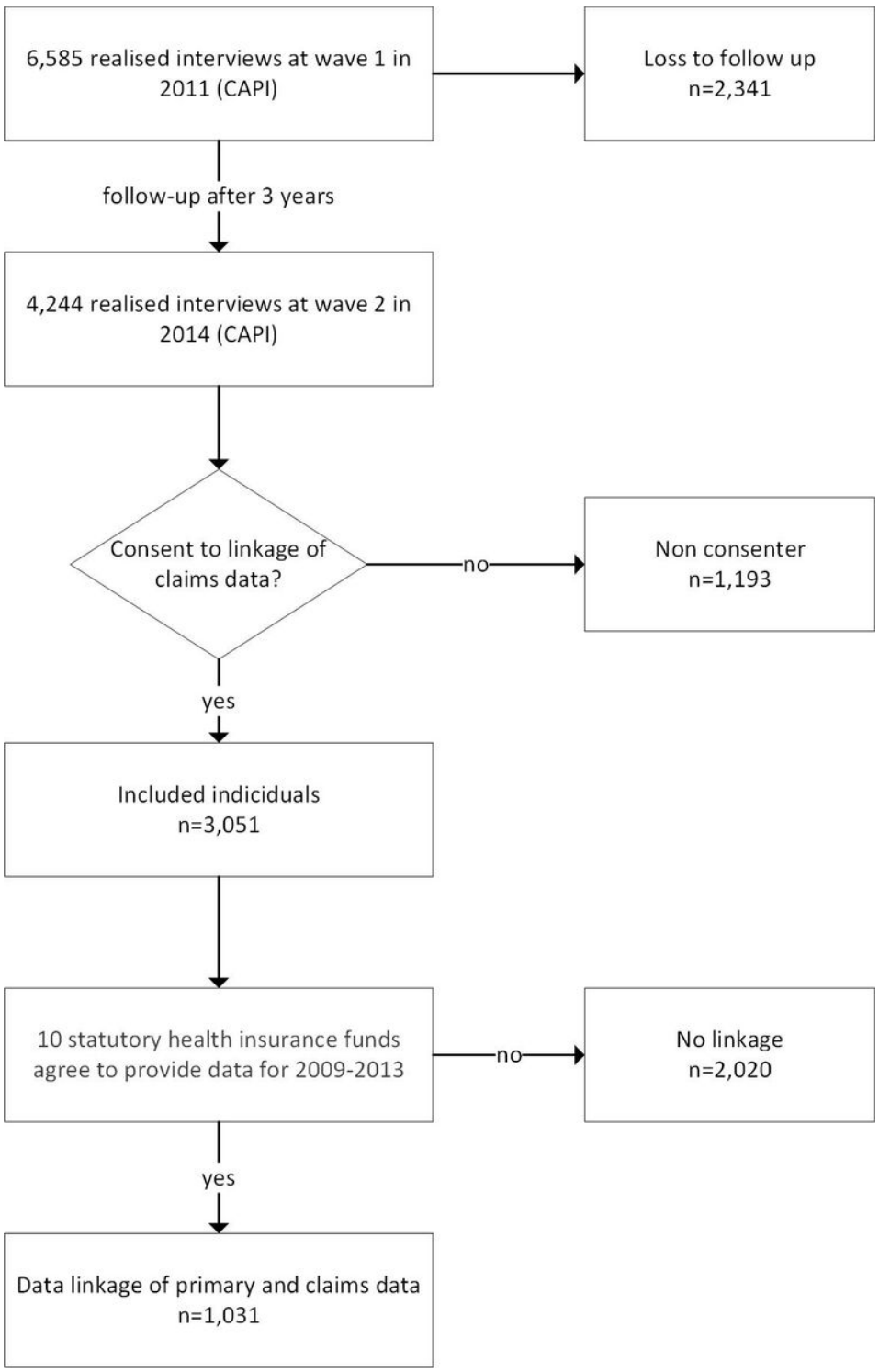

Figure 1

Data base and methodology of the data linkage