




Article

Detecting and Visualizing Outliers in Provider Profiling Using Funnel Plots and Mixed Effects Models—An Example from Prescription Claims Data

Oliver Hirsch ^{1,*} , Norbert Donner-Banzhoff ¹, Maike Schulz ² and Michael Erhart ²

¹ Department of General Practice/Family Medicine, Philipps University Marburg, Karl-von-Frisch-Str.4, 35043 Marburg, Germany; norbert@staff.uni-marburg.de

² Central Research Institute of Ambulatory Health Care in Germany (ZI), Salzufer 8, 10587 Berlin, Germany; mschulz@zi.de (M.S.); merhart@zi.de (M.E.)

* Correspondence: oliver.hirsch@staff.uni-marburg.de; Tel.: +49-6421-2865124

Received: 14 August 2018; Accepted: 13 September 2018; Published: 15 September 2018



Abstract: When prescribing a drug for a patient, a physician also has to consider economic aspects. We were interested in the feasibility and validity of profiling based on funnel plots and mixed effect models for the surveillance of German ambulatory care physicians' prescribing. We analyzed prescriptions issued to patients with a health insurance card attending neurologists' and psychiatrists' ambulatory practices in the German federal state of Saarland. The German National Association of Statutory Health Insurance Physicians developed a prescribing assessment scheme (PAS) which contains a systematic appraisal of the benefit of drugs for so far 12 different indications. The drugs have been classified on the basis of their clinical evidence as "standard", "reserve" or "third level" medication. We had 152,583 prescriptions in 56 practices available for analysis. A total of 38,796 patients received these prescriptions. The funnel plot approach with additive correction for overdispersion was almost equivalent to a mixed effects model which directly took the multilevel structure of the data into account. In the first case three practices were labeled as outliers, the mixed effects model resulted in two outliers. We suggest that both techniques should be routinely applied within a surveillance system of prescription claims data.

Keywords: drug prescriptions; ambulatory care; statistical data interpretation; reference standards

1. Introduction

When prescribing a drug for a patient, a physician not only has to consider the individual needs of a patient, the severity of her/his condition, the characteristics of suitable drugs and potential interactions with simultaneously prescribed treatments, but also economic aspects. For many indications several treatment options exist. Among these, prescribing physicians are expected to choose the less costly ones in order to preserve the economic viability of the health care system. However, the success of treatments should not be compromised. In other words, prescribers are expected to balance quality and costs when prescribing treatments for their patients. In the German health care system this is a requirement based on federal law [1].

In Germany, prescriptions for drugs issued by physicians in the ambulatory care sector are centrally recorded and surveyed regularly. Analytical procedures focus on prescribing costs of individual physicians and/or practices. A certain deviation from the average of a specialty leads to investigations with regard to the appropriateness of prescriptions. If high costs cannot be explained and justified by case-mix of patients, prescribers have to undergo remedial procedures and risk even financial sanctions. These procedures have come under criticisms recently. Patients are being referred

for high-cost prescriptions only which create additional costs and render coordination of care more complex. Primary care physicians providing comprehensive care, including the prescription of costly medications, are being punished, to name but two objections against the established procedure.

The German National Association of Statutory Health Insurance Physicians has developed a prescribing assessment scheme (PAS) which contains a systematic appraisal of the benefit and cost effectiveness of important drugs [2]. It is based on published clinical trials, systematic reviews, guidelines and appraisals by institutions such as the National Institute for Health and Care Excellence (NICE) or the Institute for Quality and Efficiency in Health Care (IQWiG). The following clinical areas have been covered so far: hypertension, coronary heart disease, lipid disorders, osteoporosis, chronic heart failure, atrial fibrillation, Alzheimer dementia, depression, diabetes mellitus type 2, upper and lower respiratory tract infections, urinary tract infection, asthma, and chronic obstructive pulmonary disease (COPD). Drugs for these conditions have been classified on the basis of their clinical evidence as “standard”, “reserve”, or “third level” medication. In the Results section, we present analyses regarding the appropriateness of prescriptions based on the PAS.

The identification of providers issuing a proportion of inappropriate prescriptions regarded as too high is essentially a case of profiling. Profiling is defined as the identification of unusual provider performance (hospitals, departments, practices, physicians) compared to other providers or to a reference standard [3,4]. Previously published profiling analyses were conducted at the patient, individual clinician, and organizational level [3]. Processes, results (clinical outcomes), and costs were investigated. Hospital mortality is perhaps the aspect having gained most attention in the past [5,6].

While explicit policies, e.g., treatment pathways or local guidelines, or implicit policies, e.g., not well-defined thresholds for referral or prescription of antibiotics, are the main focus of profiling procedures, there are other sources of variation in the data analyzed in this way. Apart from random variation, patient characteristics before treatment are particularly relevant [7]. Therefore, it is important to adjust for case-mix which includes factors like age, gender, presence of disease(s) and severity of disease. Within the profiling paradigm, funnel plots have become the preferred graphical means to assess the performance of healthcare providers [8]. Recently, this approach was applied in order to identify inappropriate prescribing patterns [9–12], but few studies explicitly used statistical techniques accounting for the multilevel structure of the data [13,14].

We were interested in the feasibility and validity of profiling based on funnel plots for the surveillance of ambulatory care physicians’ prescribing. For this analysis we focus on drugs prescribed by psychiatrists and neurologists in the Saarland, one of the 16 German federal states combining both urban and rural characteristics.

2. Materials and Methods

The description of methods follows the recently published Reporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement [15].

2.1. Data

We analyzed prescription data of 2014 according to § 300 Section 2, Social Security Code V (AVD). These data were based on patients with a health insurance card attending neurologists’ and psychiatrists’ ambulatory practices. Virtually all of these had contracts with the statutory health insurance system in the German federal state of Saarland. Since neurologists and psychiatrists could only be identified in the data set if they practiced on their own, multispecialty practices were not included in the analyses presented here. In order to exclude artefacts and practices providing predominantly psychotherapy, we selected only practices with at least five prescriptions per quarter. These were prescriptions covered by an evidence based prescribing assessment scheme (PAS) which were filled by patients of the statutory health insurance system in pharmacies of their choice. The level of analysis was the individual prescription. A total of 56 (67.5%) out of 83 practices fulfilled the inclusion criteria for our analyses.

Patient subjects were constructed by prescription records with the same health insurance card number plus same health insurance institution number and same birthday. This information was stored on the health insurance card and was read every time a patient received a prescription. Before and after patient-construction these variables were pseudonymized by a hashing-function including a secret. Each prescription record contained the prescribing physician's individual identification number as well as his/her practice identification. Both variables were pseudonymized by hashing-function including a secret.

The patient-prescription records were linked with patients' diagnoses taken from the administrative claims data of 2014 according to § 299, Social Security Code V (VDA). A deterministic record-linkage was performed at the patient-level using the patient-pseudonym issued from health insurance card number plus health insurance institution number plus birthday issued from both AVD and VDA data.

Variables under study were patients' age and gender, the total number of different drugs per patient prescribed from all health physicians the patient had contact with. An index of patient need/risk for health services utilization was calculated according to the Patient classification system of the Institut des Bewertungsausschusses (INBA). This index was based on the regression of ambulatory health insurance costs on 34 Age- and gender groups and 72 disease-categories such as diabetes, HIV/AIDS, ophthalmologic-diseases, etc. Resulting from the regression analyses performed by INBA, each patient was assigned his or her expected ambulatory health insurance cost. The latter value constitutes the patient risk score used in this study. This was essentially a measure of morbidity and health service needs [16].

The German National Association of Statutory Health Insurance Physicians developed a prescribing assessment scheme (PAS) which contains a systematic appraisal of the benefit and cost effectiveness of important drugs [2]. The drugs have been classified on the basis of their clinical evidence as "standard", "reserve" or "third level" medication. The main objective of the PAS is to give providers feedback on their prescribing. In the Results section, we present analyses referring to all prescriptions covered by the PAS. We dichotomized the three-level PAS variable to "third level" versus "standard or reserve".

2.2. Statistical Methods

Funnel plots have become the preferred graphical means to assess the performance of healthcare providers within the profiling paradigm. They consist of four components [8]:

- An indicator: this is the outcome of interest, plotted on the y-axis.
- A target: this is the normative expectation for the respective population which is often operationalized by the sample mean or by other measures of central tendency. An alternative would be an external goal (e.g., certain number of infections with multiresistent germs per 1000 patients; certain proportion of prescription of non-evidence-based medication).
- A precision parameter: this parameter determines the accuracy with which the indicator is being measured. Precision is connected to the variance of the outcome and the number of observations, such as total number of patients or total number of prescriptions of non-evidence-based medication.
- Control limits: these are the limits beyond which a provider would be classified as showing unusual performance. They are often set to a probability of 5% and 0.2% for a provider truly within the control limits to be classified as showing unusual behavior and thus constituting a false positive result [8]. Thus, even if a provider truly is showing usual behavior (on average) there is variation in his/her behavior. This variation could be attributable to so called unmeasured heterogeneity and is labelled as variation due to "chance". Picking up hypothetically 1000 providers with truly usual behavior, a 5% limit for example separates the 25 providers with the most positive behavior on the one end and the 25 providers with the most negative behavior on the other end. These are the 5% control limits meaning that 50 out of the 1000 hypothetical providers will be falsely classified as unusual.

However, how to define and measure a deviation from the norm is controversial [17]. The interval of adequate performance should preferably be based on considerations regarding clinical content. Often, however, this poses considerable difficulties and purely statistical considerations prevail [5].

Often an unconvincingly large proportion of providers falls outside the prespecified control limits. This may be due to “overdispersion” which results from not all relevant covariates being taken into account [18]. Procedures have been suggested to adjust for this [19]. Factors contributing to excessive variability causing overdispersion may operate at the individual, e.g., patient, or the aggregate or cluster level, e.g., provider, department or hospital [20].

Profiling and other statistical methods trying to identify divergent providers face a fundamental paradox. On the basis of available data, a statistical model (distribution) has to be defined. Deviating units, which the procedure is meant to detect, are part of the sample providing the data. The more these divergent entities are incorporated into the statistical model the less likely investigators or administrators will be able to detect them [21]. As a consequence, one should select robust models that are less affected by idiosyncrasies contained in the data available for analysis [22].

We followed the approach of detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models demonstrated by Ieva and Paganoni [23]. This approach implies analyses on an individual and an aggregate level. The statistical unit of the individual level 1 is the prescription of a drug. The statistical level on the aggregate level is the practice.

In a first step we plotted the number of prescriptions covered by the PAS against the proportion of medication with “third-level”. We then calculated a common mean model with the R procedure `glm` and level 1 predictors patient age, -gender, -individual morbidity without age and gender effects, and individual polypharmacy ≥ 2 quarters to predict the occurrence of a third level medication (yes/no). These predictors were selected on the basis of procedures `glmer` of R package `lme4` and `glmmPQL` of R package `nlme`.

Next, we calculated the observed versus the expected ratio of third level medication for each practice and plotted it against the expected number of third level medication in an unadjusted funnel plot [23]. Practices above the upper control limits of the funnel plot can be regarded as outliers, i.e., in need of measures to have their prescribing habits explored and possibly improved. In a next step, the $\hat{\phi}W$ was calculated according to the method of Spiegelhalter [8] to verify if overdispersion was present. We then applied the additive adjustment factor for the band limits of the funnel plot described in [23] and demonstrated the change in the number of outliers in an adjusted funnel plot. To validate the funnel plot procedure, we fitted a mixed effect model which accounted for the cluster structure of the data with the R procedure `glmer` and predictors age, gender, individual morbidity as described above, and individual polypharmacy ≥ 2 quarters as fixed effects and a random intercept.

A particular patient could receive several prescriptions. Thus, the individual prescriptions were not statistically independent but were clustered at the patient level. Furthermore, one patient could receive prescriptions from several physicians. Thus, from a strict statistical perspective a 3-level hierarchical model accounting for cross-classification should be employed. However, approximately 150 thousand prescriptions nested in approximately 39 thousand patients nested in 57 practices would demand a sizeable computational capacity. For the sake of presenting a feasible and already tested analytic strategy, we therefore decided to follow the approach of detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models demonstrated by Ieva and Paganoni [23] while at the same time acknowledging that we might have underestimated complexity and variation of the data.

A boxplot and a Q-Q plot of the point estimates of the random intercept of the multilevel model were drawn to assess the agreement between the two approaches. We performed all calculations with R, Version 3.2.2 (<https://www.r-project.org/>).

We obtained a waiver by the local ethics committee of the Department of Medicine at Phillips University Marburg for this study of fully anonymous prescription claims data, dated 9 December 2015, project identification code: “Az.: Prof. Ri./ra”.

3. Results

Our data set contained 155,234 prescriptions issued by 56 practices. After elimination of missing data in age, gender, prescription or diagnosis variables 152,583 prescriptions remained. A total of 38,796 patients received these prescriptions. Mean age was 59.2 years, 25% were younger than 49 years, 25% were older than 73 years. A proportion of 63.6% were female. On average, every patient received 3.9 prescriptions from a psychiatrist/neurologist. About 46.1% of the patients received 6 or more agents on at least two quarters and thus were classified as polypharmacy patients. On average the patients had a risk score of 1417. This means that their expected use of ambulatory health care-services was about 41.7% higher than the average of patients. We first plotted the total number of prescriptions against the proportion of medication with third level which is displayed in Figure 1.

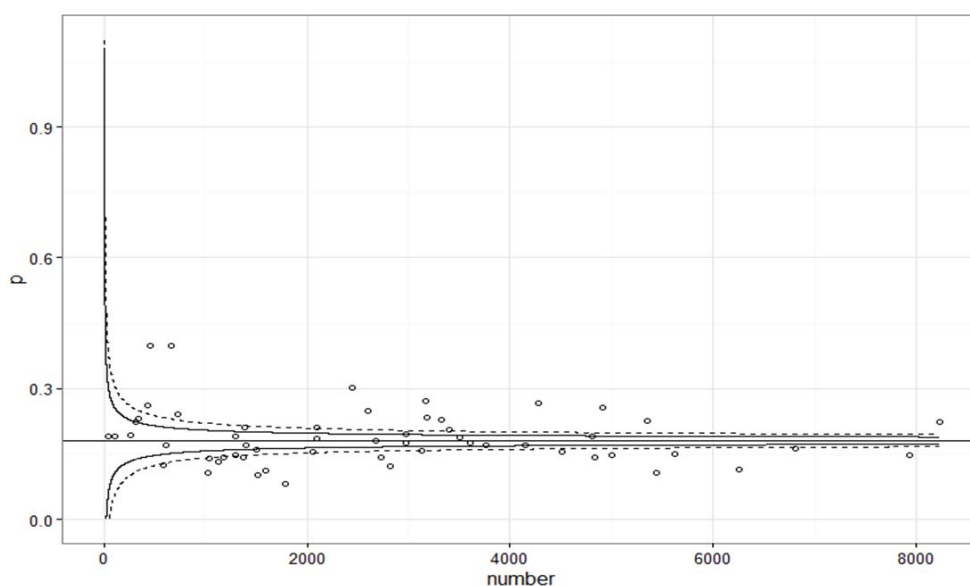


Figure 1. Number of prescriptions by neurologists and psychiatrists in the German federal state of Saarland plotted against the proportion of medication with third level. Practices are units of observation. Control limits were calculated according to Ieva and Paganoni and Spiegelhalter [8,23]. The solid horizontal line represents Huber's M estimator, a robust substitute for the mean.

As is apparent in Figure 1, about one fourth of practices above the upper control limits of the funnel plot would be categorized as outliers. Given the size of this group, overdispersion is likely to contribute to these findings [8,23].

We then calculated a common mean model with the R procedure `glm` and level 1 predictors age, gender, individual morbidity, and individual polypharmacy ≥ 2 quarters to predict the occurrence of a third level medication (yes/no). Age ($p < 0.001$), gender ($p < 0.001$), individual morbidity and individual polypharmacy ≥ 2 quarters ($p < 0.001$), were significant predictors. Next, the observed versus the expected ratio of third level medication was calculated for each practice and plotted against the expected number of third level medication (see Figure 2) which explains the difference to the numbers on the x-axis of Figure 1.

Again, about one fourth of practices would be classified as outliers. So far, we did not consider overdispersion. In our case, the $\hat{\phi}W = 32.01$ which signals overdispersion as it should not be higher than $\hat{\phi}W = 1 + 2\sqrt{2/56} = 1.38$ [23].

Consequently, we applied the additive adjustment factor for the band limits of the funnel plot. The resulting funnel plot is displayed in Figure 3.

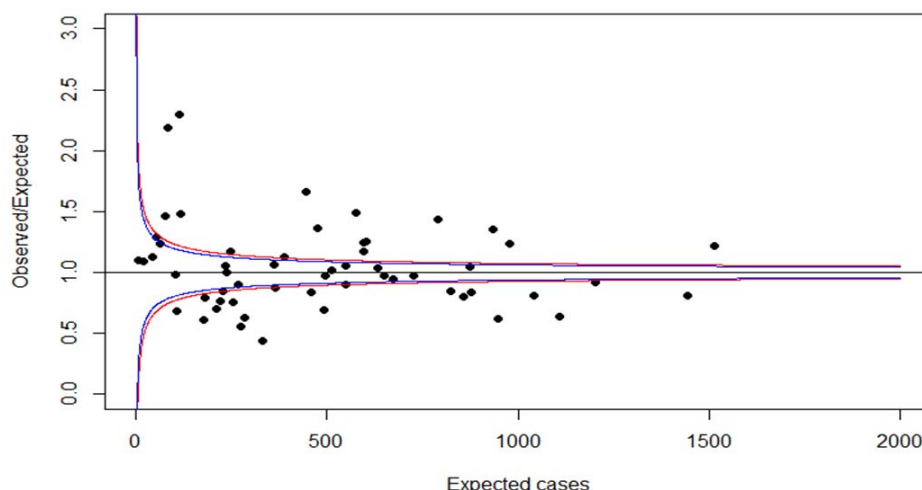


Figure 2. Standardized ratio of third level medication against expected number of third level medication in neurologists and psychiatrists in the German federal state of Saarland. Adjusted funnel plot with band limits at 95% (blue lines) and 98% (red lines).

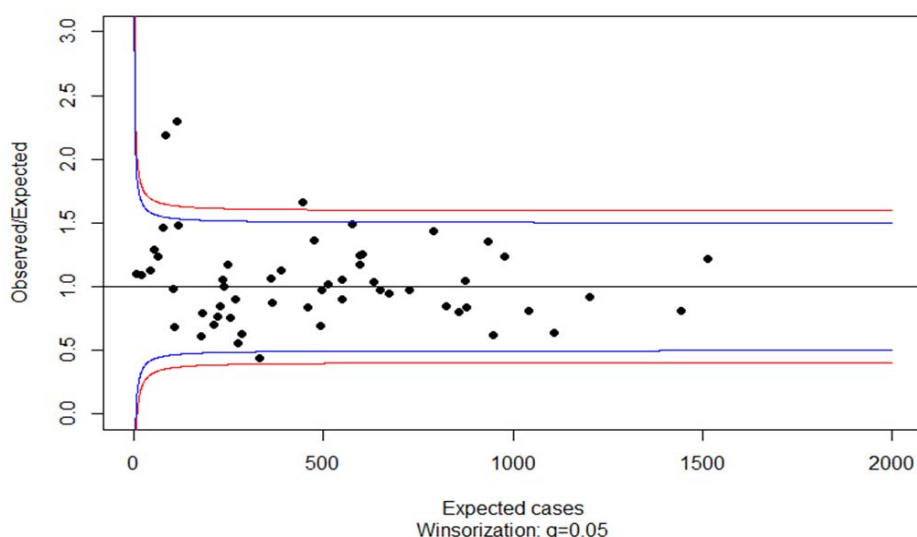


Figure 3. Standardized ratio of third level medication in neurologists and psychiatrists in the German federal state of Saarland. Additive adjusted funnel plot with band limits at 95% (blue lines) and 98% (red lines).

The funnel plot in Figure 3 classifies only three practices as having an excessive proportion of third level prescriptions. In order to validate the funnel plot procedure for detecting outliers we applied the method of identifying extreme values of the random effect point estimates from a mixed effect model. We fitted a mixed effect model with the R procedure `glmer` with the former covariates as fixed effects and a random intercept. This time accounting for the cluster structure of the data, age, gender, and individual morbidity without age and gender effects (all $p < 0.001$) and individual polypharmacy ($p = 0.003$) were significant predictors of a third level prescription at the patient level, with odds ratios of age 0.9912, female gender 0.7938, morbidity 1.2362, and polypharmacy 1.0487. Results hinting at males and patients suffering from severe illness are more likely to receive third level medication. However, the actual effects are rather small indicating that the appropriateness of prescription depends only to a very small degree on the variables explored in our data set.

The two points over the limit of the whiskers in the boxplot of the point estimates of the random intercept (Figure 4) correspond to two of the three practices that were out of the 95% and 98% band limits of the funnel plot in Figure 3.

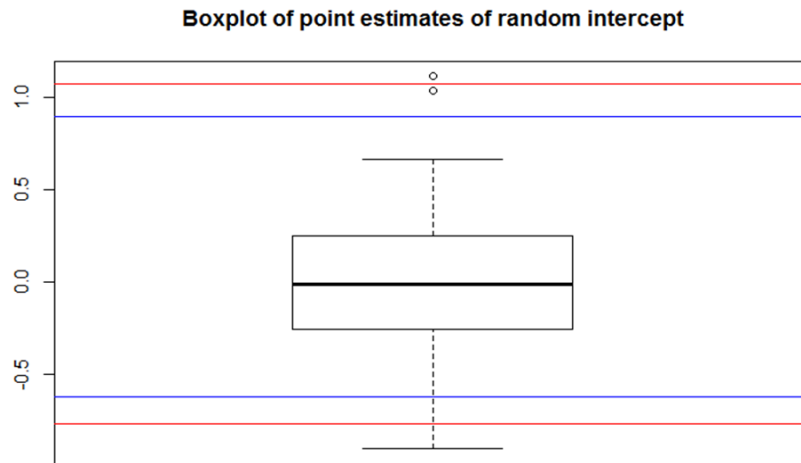


Figure 4. Boxplot of the point estimates of the random intercept of the multilevel model. The horizontal blue lines correspond to the quantiles of 2.5% and 97.5%. The horizontal red lines correspond to the quantiles of 1% and 99%.

These two practices can further be located in the upper left part of the Normal Q-Q plot in Figure 5. We conclude that funnel plots and mixed effect models reached almost the same conclusions in detecting outliers in German prescription claims data.

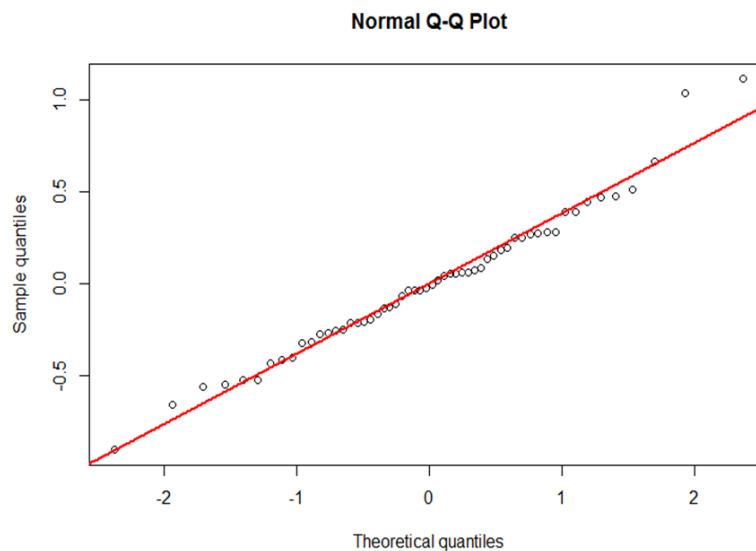


Figure 5. Normal Q-Q plot of the point estimates of the random intercept of the multilevel model to visualize outliers.

4. Discussion

To our knowledge this is the first time the method for detecting and visualizing outliers in provider profiling using funnel plots and mixed effects models [23] has been applied to prescription claims data in Germany. Prescriptions issued by neurologists and psychiatrists in the German federal state of Saarland were evaluated by an appropriateness of prescribing index. The funnel plot approach with additive correction for overdispersion was almost equivalent to a mixed effects model which directly took the multilevel structure of the data into account. In the first case three practices were

labeled as outliers, the mixed effects model resulted in two, which were both among the practices identified by the first model. In our study, morbidity and polypharmacy as significant predictors corroborate earlier findings showing considerable associations to high risk prescribing [14].

Although results of funnel plot and mixed effect model profiling were similar in our case, we suggest that both techniques should be routinely applied in profiling of prescription behavior. The prescription behavior of individual physicians is usually highly consistent across their respective patients but different from other physicians of their discipline. Physician-level clustering should therefore be accommodated for [24].

For decades it has been recognized that claims-based profiling methods can be applied to improve the quality of care [25]. The focus, however, should be more on quality than on the cost of care as evidence-based treatments are more likely to reduce symptom load in patients and should therefore have priority.

Profiling can be applied in a formative, non-punitive sense. Providers can thus identify areas of problematic actions and draw conclusions for improvements. Information from profiling analyses can be integrated into clinical pathways and treatment protocols in order to improve the provision of healthcare [3]. One of the primary goals of provider profiling in the area of prescription data should be to provide physicians with meaningful feedback regarding their behavior. If this is not in accordance with common guidelines or a certain target behavior, quality-oriented audits may help to elucidate the processes behind possible unusual results [18]. It is important to stress that the profiling approach discussed in this article primarily focuses on quality of prescriptions [25]. We are in line with other studies showing that practices with high prescribing costs per patient do not necessarily have prescription patterns which have to be modified. This might be due to special patient characteristics [12].

Some issues should be considered in connection with profiling. For example, statistical models seek to reduce information inherent in data which are extracted from observations of a group of entities. The aim is to represent the data as good as possible (“goodness of fit”). Such models are not suitable to draw conclusions about cases that were not in the calibration sample (“bias-variance-dilemma”). Therefore, more simple robust models should be preferred.

Funnel plots are an attractive tool for the surveillance physicians’ prescriptions. Within such a system there could be a continuous, systematic collection, analysis, and interpretation of prescription data used to inform health policy and planning [18] based on sound statistical methods and incorporating special provider characteristics. They are superior to simple rank-based methods like league tables as they are able to adjust for confounding variables. They can easily be created by common statistical software like R but trained personnel is needed to adequately interpret the results.

Past studies have shown equivocal results of feedback on prescribing behavior [26–28]. However, feedback was given delayed and irregularly. Immediate electronic feedback while prescriptions are issued should be the preferred method. This might be done best by providing electronic support during the prescription process. Primary care physicians who received feedback by a computerized decision support system (CDSS) considerably changed their prescription behavior compared to a control group [29]. This was confirmed in another study in which simple data feedback led to a significant reduction in high risk prescribing in primary care [30].

As a method to identify third-level-drugs by psychiatrists and neurologists in Saarland, the funnel method was feasible and largely concordant with a mixed effect model.

Limitations

Prescription data was available for drugs that were completely reimbursed by the statutory health insurances in Germany. No data were available for over-the-counter (OTC) drugs, drugs prescribed for hospital in-patients or drugs reimbursed by private health insurances (<10% market share in the area covered). We were not able to evaluate the appropriateness of prescriptions in the individual clinical case. We had only limited information on patient characteristics and almost none on provider

characteristics. This was due to logistical as well as formal reasons, such as privacy regulations. As a result, possibilities to adjust for confounders at these levels were limited [31,32]. Since neurologists and psychiatrists could be only identified in the data set if they practiced on their own, multispecialty practices were not included in the analyses presented here.

We chose the Saarland as the smallest German territorial federal state for practical and logistical reasons. Moreover, its proportion of rural and urban areas is largely representative for the whole country. This is confirmed by data from comparisons with other German federal states. The success index of Saarland which consists of parameters like economic growth, social security, unemployment rate, per capita gross national product, inner security, and gainful employment was with 5.86 close to the average in Germany of 5.56. The activity index of Saarland which contains factors that are under the influence of the German federal states and have interactions with the above-mentioned parameters was also close to the German average (5.18 versus 5.70, respectively) [33,34].

Although we could show the concordance between the funnel plot method and a mixed effects model, a true validity study would have required the individual investigation of practices and their prescribing behaviors which would have been beyond the scope of this project.

5. Conclusions

The prescribing assessment scheme (PAS) of the German National Association of Statutory Health Insurance Physicians is an instrument with which quality and cost-effectiveness of prescriptions can be assessed on the basis of evidence-based criteria. In this it is superior to costs, which have been used to evaluate physicians' prescriptions in the past. The results of our study support the feasibility and validity of a funnel-plot approach together with mixed effect models.

Author Contributions: M.E. conceived of the study, participated in the design and in the interpretation of the data, performed part of the statistical analyses, and helped to draft the manuscript. M.S. conceived of the study, prepared the data sets, participated in the design and in the interpretation of the data. N.D.-B. conceived of the study, participated in the design and in the interpretation of the data, and helped to draft the manuscript. O.H. conceived of the study, participated in the design and in the interpretation of the data, drafted the manuscript and performed the statistical analyses. All authors read and approved the final version of the manuscript.

Funding: The study was funded by the Central Research Institute of Ambulatory Health Care in Germany (ZI), Berlin.

Acknowledgments: We gratefully acknowledge the most valuable support of Anna Maria Paganoni and Francesca Ieva who provided us with their R code which was the basis for our analyses and who gave us an inspiring lecture about this topic.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Gemeinsamer Bundesausschuss (Ed.) *Richtlinie des Gemeinsamen Bundesausschusses über die Verordnung von Arzneimitteln in der Vertragsärztlichen Versorgung [Guideline of the Common Federal Committee on the Prescription of Drugs in Ambulatory Care]*; Gemeinsamer Bundesausschuss: Berlin, Germany, 2016.
2. Schwenzer, S. ARMIN in Sachsen und Thüringen—Mehr Arzneimitteltherapiesicherheit durch rationale und evidenzbasierte Arzneimitteltherapie und patientenindividuelles Medikationsmanagement. *Arzneiverordnung Praxis* **2015**, *42*, 130–137.
3. Tucker, J.L. The theory and methodology of provider profiling. *Int. J. Health Care Qual. Assur.* **2000**, *13*, 316–323. [[CrossRef](#)]
4. Gomes, M.; Gutacker, N.; Bojke, C.; Street, A. Addressing Missing Data in Patient-Reported Outcome Measures (PROMS): Implications for the Use of PROMS for Comparing Provider Performance. *Health Econ.* **2016**, *25*, 515–528. [[CrossRef](#)] [[PubMed](#)]
5. Burgess, J.F.; Christiansen, C.L.; Michalak, S.E.; Morris, C.N. Medical profiling: Improving standards and risk adjustments using hierarchical models. *J. Health Econ.* **2000**, *19*, 291–309. [[CrossRef](#)]

6. Bragg, F.; Cromwell, D.A.; Edozien, L.C.; Gurol-Urganci, I.; Mahmood, T.A.; Templeton, A.; van der Meulen, J.H. Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: Cross sectional study. *BMJ Clin. Res.* **2010**, *341*, c5065. [[CrossRef](#)] [[PubMed](#)]
7. Marshall, E.C.; Spiegelhalter, D.J. Institutional Performance. In *Multilevel Modelling of Health Statistics*; Leyland, A.H., Goldstein, H., Eds.; Wiley: Chichester, NY, USA, 2001; pp. 127–142.
8. Spiegelhalter, D.J. Funnel plots for comparing institutional performance. *Stat. Med.* **2005**, *24*, 1185–1202. [[CrossRef](#)] [[PubMed](#)]
9. Pouwels, K.B.; Dolk, F.C.K.; Smith, D.R.M.; Robotham, J.V.; Smieszek, T. Actual versus ‘ideal’ antibiotic prescribing for common conditions in English primary care. *J. Antimicrob. Chemother.* **2018**, *73*, 19–26. [[CrossRef](#)] [[PubMed](#)]
10. Hawker, J.I.; Smith, S.; Smith, G.E.; Morbey, R.; Johnson, A.P.; Fleming, D.M.; Shallcross, L.; Hayward, A.C. Trends in antibiotic prescribing in primary care for clinical syndromes subject to national recommendations to reduce antibiotic resistance, UK 1995–2011: Analysis of a large database of primary care consultations. *J. Antimicrob. Chemother.* **2014**, *69*, 3423–3430. [[CrossRef](#)] [[PubMed](#)]
11. Gharbi, M.; Doerholt, K.; Vergnano, S.; Bielicki, J.A.; Paulus, S.; Menson, E.; Riordan, A.; Lyall, H.; Patel, S.V.; Bernatoniene, J.; et al. Using a simple point-prevalence survey to define appropriate antibiotic prescribing in hospitalised children across the UK. *BMJ Open* **2016**, *6*, e012675. [[CrossRef](#)] [[PubMed](#)]
12. Tomlin, A.M.; Gillies, T.D.; Tilyard, M.W.; Dovey, S.M. Variation in the pharmaceutical costs of New Zealand general practices: A national database linkage study. *J. Public Health* **2016**, *38*, 138–146. [[CrossRef](#)] [[PubMed](#)]
13. Cahir, C.; Fahey, T.; Teljeur, C.; Bennett, K. Prescriber variation in potentially inappropriate prescribing in older populations in Ireland. *BMC Fam. Pract.* **2014**, *15*, 59. [[CrossRef](#)] [[PubMed](#)]
14. Byrne, C.J.; Cahir, C.; Curran, C.; Bennett, K. High-risk prescribing in an Irish primary care population: Trends and variation. *Br. J. Clin. Pharmacol.* **2017**, *83*, 2821–2830. [[CrossRef](#)] [[PubMed](#)]
15. Benchimol, E.I.; Smeeth, L.; Guttman, A.; Harron, K.; Moher, D.; Petersen, I.; Sorensen, H.T.; von Elm, E.; Langan, S.M. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* **2015**, *12*, e1001885. [[CrossRef](#)] [[PubMed](#)]
16. Berger, I.; Hinz, A.; Naftaliev, B.; Schwan, M.; Spillner, A. Bericht des Instituts des Bewertungsausschusses zur Weiterentwicklung des Klassifikationssystems sowie zur Ermittlung der Veränderungsdaten für das Jahr 2015 Gemäß § 87a Abs. 5 SGB V. 2016. Available online: https://institut-ba.de/publikationen/InBA_Bericht_KM87a2014.pdf (accessed on 13 September 2018).
17. Christiansen, C.L.; Morris, C.N. Improving the Statistical Approach to Health Care Provider Profiling. *Ann. Intern. Med.* **1997**, *127*, 764–768. [[CrossRef](#)] [[PubMed](#)]
18. Dover, D.C.; Schopflicher, D.P. Using funnel plots in public health surveillance. *Popul. Health Metr.* **2011**, *9*, 58. [[CrossRef](#)] [[PubMed](#)]
19. Cousineau, D.; Chartier, S. Outliers detection and treatment: A review. *Int. J. Psychol. Res.* **2010**, *3*, 58–67. [[CrossRef](#)]
20. Spiegelhalter, D.J. Handling over-dispersion of performance indicators. *Qual. Saf. Health Care* **2005**, *14*, 347–351. [[CrossRef](#)] [[PubMed](#)]
21. Ohlssen, D.I.; Sharples, L.D.; Spiegelhalter, D.J. A hierarchical modelling framework for identifying unusual performance in health care providers. *J. R. Stat. Soc. Ser. A* **2007**, *170*, 865–890. [[CrossRef](#)]
22. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*; Wiley-Interscience: Hoboken, NJ, USA, 2003.
23. Ieva, F.; Paganoni, A.M. Detecting and visualizing outliers in provider profiling via funnel plots and mixed effect models. *Health Care Manag. Sci.* **2015**, *18*, 166–172. [[CrossRef](#)] [[PubMed](#)]
24. Greenfield, S.; Kaplan, S.H.; Kahn, R.; Ninomiya, J.; Griffith, J.L. Profiling care provided by different groups of physicians: Effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann. Intern. Med.* **2002**, *136*, 111–121. [[CrossRef](#)] [[PubMed](#)]
25. Garnick, D.W.; Fowles, J.; Lawthers, A.G.; Weiner, J.P.; Parente, S.T.; Palmer, R.H. Focus on quality: Profiling physicians’ practice patterns. *J. Ambul. Care Manag.* **1994**, *17*, 44–75. [[CrossRef](#)]
26. Schectman, J.M.; Kanwal, N.K.; Schroth, W.S.; Elinsky, E.G. The effect of an education and feedback intervention on group-model and network-model health maintenance organization physician prescribing behavior. *Med. Care* **1995**, *33*, 139–144. [[CrossRef](#)] [[PubMed](#)]
27. Mainous, A.G.; Hueston, W.J.; Love, M.M.; Evans, M.E.; Finger, R. An evaluation of statewide strategies to reduce antibiotic overuse. *Fam. Med.* **2000**, *32*, 22–29. [[PubMed](#)]

28. Balas, E.A.; Boren, S.A.; Brown, G.D.; Ewigman, B.G.; Mitchell, J.A.; Perkoff, G.T. Effect of physician profiling on utilization. Meta-analysis of randomized clinical trials. *J. Gen. Intern. Med.* **1996**, *11*, 584–590. [[CrossRef](#)] [[PubMed](#)]
29. McMullin, S.T.; Loneragan, T.P.; Rynearson, C.S.; Doerr, T.D.; Veregge, P.A.; Scanlan, E.S. Impact of an evidence-based computerized decision support system on primary care prescription costs. *Ann. Fam. Med.* **2004**, *2*, 494–498. [[CrossRef](#)] [[PubMed](#)]
30. Guthrie, B.; Kavanagh, K.; Robertson, C.; Barnett, K.; Treweek, S.; Petrie, D.; Ritchie, L.; Bennie, M. Data feedback and behavioural change intervention to improve primary care prescribing safety (EFIPPS): Multicentre, three arm, cluster randomised controlled trial. *BMJ Clin. Res.* **2016**, *354*, i4079. [[CrossRef](#)] [[PubMed](#)]
31. Mazzaglia, G.; Caputi, A.P.; Rossi, A.; Bettoncelli, G.; Stefanini, G.; Ventriglia, G.; Nardi, R.; Brignoli, O.; Cricelli, C. Exploring patient- and doctor-related variables associated with antibiotic prescribing for respiratory infections in primary care. *Eur. J. Clin. Pharmacol.* **2003**, *59*, 651–657. [[CrossRef](#)] [[PubMed](#)]
32. Orzella, L.; Chini, F.; Giorgi Rossi, P.; Borgia, P. Physician and patient characteristics associated with prescriptions and costs of drugs in the Lazio region of Italy. *Health Policy* **2010**, *95*, 236–244. [[CrossRef](#)] [[PubMed](#)]
33. Berthold, N.; Kosturkova, N.; Müller, A. Die Bundesländer im Standortwettbewerb-gestern, heute und morgen [Federal states in competition-yesterday, today, and tomorrow]. *ifo Schnelldienst* **2010**, *63*, 15–20.
34. Hessisches Ministerium für Wirtschaft, Energie, Verkehr und Landesentwicklung. *Bundesländer-, Regional- und Städterankings im Vergleich [Rankings of Federal States, Regions, and Cities—A Comparison]*; HA Hessen Agentur GmbH: Wiesbaden, Germany, 2014.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).