International Journal of Clinical Pharmacy (2019) 41:414–423
https://doi.org/10.1007/s11096-019-00785-8

RESEARCH ARTICLE

Prospective survey-based study on the categorization quality of hospital pharmacists' interventions using DokuPIK

Angela Ihbe-Heffinger[2,3,11] · Claudia Langebrake[4,5] · Carina Hohmann[6] · Katja Leichenberg[7] · Heike Hilgarth[4] · Mareike Kunkel[8] · Michael Lueb[9] · Tibor Schuster[1,10]

Received: 26 January 2018 / Accepted: 9 January 2019 / Published online: 20 March 2019

Abstract

Background There is a growing need to categorize pharmacists' interventions (PIs) in Germany to document their impact on solving or avoiding drug-related problems. *Objective* To validate the categorization of drug-related problems—one aspect of the categorical internet database DokuPIK, designed for recording routinely PIs. To identify case-specific predictive values. *Setting* German hospitals. *Methods* Within a prospective, nationwide survey-based study, 37 of 498 registered database users volunteered to evaluate 24 standardized case reports independently. Case evaluation was restricted to classify problems, based on 26 given categories with no limit on the number of item choices. Ratings were conducted electronically and anonymously. A gold standard of one or more problems per case was developed by majority consensus of five senior clinical pharmacists. Agreement of raters' case classification with the gold standard was assessed by calculating sensitivity, specificity and positive and negative predictive value and was reported as median and range. *Main outcome* Level of agreement. *Results* Independent assessment yielded a median agreement of 90% [79–94%]. Sensitivity and specificity were 37% [21–57%] and 99% [97–100%], respectively. Median positive and negative predicted value were both 90% [60–100%] and 90% [78–95%]. Mean case-specific agreement was robust ($\geq$ 79%) with respect to a majority and maximum consensus (three and five out of five raters). *Conclusion* DokuPIK seems to have a high level of agreement and a good specificity according to the majority of clinical pharmacists in a panel of assessors. Despite the allowance of multiple choices, predictive values were high and indicated a well-constructed classification by pharmacists.

Keywords Classification system · Clinical pharmacy services · Drug-related problems · Germany · Medication safety · Pharmacist interventions · Pharmaceutical care

✉ Angela Ihbe-Heffinger
a.ihbe-heffinger@klinikum-starnberg.de

[1] Department of Family Medicine, McGill University, Montreal, QC, Canada

[2] Department of Gynecology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

[3] Hospital Pharmacy, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

[4] Hospital Pharmacy, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

[5] Department of Stem Cell Transplantation, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

[6] Hospital Pharmacy Unit, Klinikum Fulda AG, Fulda, Germany

[7] Hospital Pharmacy Unit, Universitätsklinikum der Friedrich-Schiller-Universität Jena, Jena, Germany

[8] Hospital Pharmacy Unit, Klinikum der Johannes Gutenberg-Universität Mainz, Mainz, Germany

[9] Hospital Pharmacy Unit, SLK-Kliniken Heilbronn, Heilbronn, Germany

[10] Institute of Medical Statistics and Epidemiology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

[11] Hospital Pharmacy, Klinikum Starnberg GmbH, Academic Teaching Hospital of the Ludwig-Maximilians-Universität München, Oßwaldstraße 1, 82319 Starnberg, Germany

Impacts on practice

- The utilization of DokuPIK, a German documentation tool for medication errors and pharmaceutical interventions in hospitals, enables a high level of agreement and specificity.
- The online nature of DokuPIK facilitates prospective surveys within defined periods.

Introduction

In recent years, an effort was made to systematically classify and document pharmacists' interventions (PIs) [1], to review their impact on health care and patient-related outcomes [2] and to analyse pooled data for selected settings and pharmaceutical activities [3, 4]. Moreover, several non-comprehensive evidence-based literature web platforms on the impact of pharmacist interventions were developed with the aim to improve pharmaceutical practice and integrate knowledge [5]. The aim and spectrum of PIs have been described in several reports and reviewed recently [1, 6, 7]. Beyond that, a descriptive instrument was developed to characterize the components of the applied clinical pharmacy services [8], with the aim to improve reproducibility and comparability of interventions in future studies. Nevertheless, to our knowledge, there is no internationally recognized PI definition.

With an average number of three pharmacists supplying 1000 beds, staffing level of German hospital pharmacies markedly differs from the European average (11 pharmacists per 1000 beds) and therefore ranks on the second last position within European Countries [9]. Currently, more than 40% of hospital pharmacy departments work routinely as part of multidisciplinary team [10] and more than half document their activities in the patient record and/or in the back office [11]. Therefore, the implementation of routine PIs in German hospitals all over the country and its documentation remains a future challenge.

In December 2008, a nationwide, standardized documentation system was launched to classify and record medication errors (MEs) and routine PIs directed by hospital pharmacists: the database DokuPIK (Documentation of Pharmacists' interventions in Hospital) [12, 13]. The anonymous, self-reported, internet-based, categorical, hierarchical documentation system was designed on the recommendations of van Mil et al. [14] by adjusting several paper-based classification systems for drug-related problems (DRPs) to hospital pharmacists' requirements, e.g.: the PCNE classification for drug-related problems V5.01 [15], APS-Doc [16] or PI-Doc [17]. Moreover, the DokuPIK was designed to cover the special needs of hospital pharmacists, to simplify its use in daily routine and to maximize documentation speed [18]. In summary, it comprises (1) basic patient data: age, gender, renal and liver function; (2) the classification of involved drugs using the WHO ATC coding system [19], the hierarchical classification of DRPs leading to PI, the resulting actions and their acceptance, as well as the severity of an identified ME, where applicable, according to NCC MERP [20]; and (3) a free text option.

After registration, DokuPIK usage is free of charge for members of the German Association of Hospital Pharmacists (ADKA), representing around 85% of German hospital pharmacists. To protect data privacy, all datasets are anonymized automatically. Non-German pharmacists can apply for an extraordinary membership.

DokuPIK was launched to gain nationwide data about the nature of routine PIs in German hospitals to show their impact on minimizing errors and optimizing medication therapy. Database usage therefore includes the permission for data analysis by the administrator—the "PI working group" of the ADKA. Within the DokuPIK database, it was intended to document any communication/action solving and/or avoiding DRPs, defined by the Pharmaceutical Care Network Europe (PCNE) as an "event or circumstance involving drug therapy that actually or potentially interferes with desired health outcomes" [15]. This PI definition therefore includes the management of existing DRPs as well as any proactive approach avoiding potential DRPs within the medication use process.

Some DRP classification systems explicitly separate the potential or manifest type of DRP from its causes—a difficult and time-consuming assessment in daily hospital routine documentation [15]. Within the DokuPIK database, the terminus "cause of PI" summarizes the type(s) of DRP(s) and its (their) cause(s) leading to one PI and consequentially allows multiple item choice. To be more precise, we use the expression "DRP(s) leading to PI" all over this article.

An agreed coding seems to be crucial to establish and document PIs in clinical practice and to facilitate research in this field. Recently, clinical impact grading—which is the subjective assessment of the clinical impact of prescribing error and optimization cases—was shown to vary between health care professionals. Nevertheless, within profession, reliability grading was excellent for pharmacists [21].

Aim of the study

We therefore aimed to validate the categorization quality by measuring the level of agreement with a reference to a SCP majority opinion within a prospective, survey-based

study among DokuPIK users. To improve rating time and to compare the results to the existing literature, the analysis was restricted to the core purpose—the hierarchical classification of DRP types leading to PI. Second objective was to identify case-specific predictive values in order to measure the probability of appropriate item selection.

Ethics approval

We did not evaluate data from actual patients but categorized 24 published, standardized case reports within this study. An ethic approval therefore was not necessary due to the theoretical nature of research.

Methods

Setting and rating procedure

Out of 498 registered ADKA DokuPIK users, 37 (7.4%) survey participants volunteered to evaluate 24 published, standardized case reports between January and March 2012 as independent raters. For case report construction, Ganso et al. detected 24 scenarios with a documentation frequency of at least ten out of more than 2000 case reports documented in a German university hospital over an 18-month period. They represented 92% surgical and 8% internistic patients on four different surgical and one internistic wards. Standardized case reports incorporated all necessary information to describe the underlying DRP, individualized dosing, the description of PI, as well as its result and relevance [22]. These case reports were previously used to validate two German DRP documentation systems: a modified PI-Doc system [22] and the APS-Doc [16].

Raters' case evaluation was restricted to identify and document DRPs and was performed based on 26 items organized into seven main groups (Table 1) with no limit on the number of item choices. Ratings were conducted electronically and anonymously without any preparatory training using the online survey tool Q-SET™ [23]. A handbook with definitions and documentation instructions was available online for all registered users.

DokuPIK database users

The anonymous rating procedure precluded any characterization of the 37 survey participants prospectively recruited out of the ADKA DokuPIK user pool. We alternatively evaluated user characteristics during the first nationwide descriptive analysis of PIs documented between 1st January 2009 and 31st December 2012 [24], to get a broad estimate about the nature of DokuPIK usage. The validation was performed on parallel within this timeframe. Therefore, the total number of registered DokuPIK users at the end of 2012 was somewhat higher than at the beginning of the validation period. The majority of PIs were documented by ward-based clinical pharmacists (82.5%), followed by drug information specialists (8.2%), dispensing pharmacists (3.5%) and hospital pharmacists located at the patient admission (2.9%) or central chemotherapy preparation unit (2.1%).

To evaluate the extent of routine database usage, anonymized login data were retrospectively analysed between April 2012 and September 2012. By 31st December 2012, 536 DokuPIK users were registered. Almost one quarter (24.3%) worked at a university and 68.5% at a non-university hospital. Between April and September 2012, 6386 database logins were registered: 187 anonymized user IDs logged in at least once, 115 at least six times and 66 at least twenty times during the half-year period.

Selection of study method

To evaluate the level of concordance between independent users of categorical, hierarchical documentation systems, interrater agreement (IRA) and interrater reliability (IRR) measures are common concepts. The latter offer a correction by chance and assume that categories are independent, as well as mutually exclusive and exhaustive [25]. For DRPs or PIs, *Cohen's kappa* coefficient is a widespread index, although the likelihood of random agreement decreases with increasing number of categories. Moreover, exclusive- and exhaustiveness—a prerequisite of *kappa* usage—is ambitious and often not practicable [1]. As the DokuPIK allows multiple ratings, the measurement of IRR was not feasible. Thus, we decided to evaluate the level of concordance to a gold standard of SCPs' majority opinion using classical metrics of a binary classifier: sensitivity, specificity, PPV and NPV.

Measures and statistics

Three out of five senior clinical pharmacists (SCPs) determined a gold standard of one or more DRPs per case report by majority opinion ($\geq 60\%$ consensus). SCPs were instructed not only to vote for the most appropriate DRP but also to select all acceptable DRP types for each case report. Gold standard was created before statistical analysis. SCPs were blinded to each other and to survey participants' results. 60% consensus level was selected to respect the diversity of item selection among SCPs and altered to 100% within sensitivity analysis. Gold standard generation process is shown in Table 2 for one selected case report.

The case-specific rater agreement is presented as median and 95% confidence interval against the gold standard. Sensitivity, specificity, as well as positive (PPV) and negative

Table 1 Documentation of drug-related problems (DRPs) leading to a pharmaceutical intervention (PI) within the DokuPIK documentation tool—hierarchical classification based on $n = 26$ categories (English translation)

Cause of PI	
Drug	
DR 1	(Clear) indication not (or no longer) given
DR 2	(Clear) indication, but no drug prescribed
DR 3	Drug allergy or medical history not considered
DR 4	Double prescription
DR 5	Dispensing error on the ward
DR 6	Generic/therapeutic substitution
DR 7	Transcription error
DR 8	Inappropriate or not most suitable drug formulation in terms of indication
DR 9	Inappropriately or not most suitable drug in terms of costs
DR 10	Inappropriate or not most suitable drug in terms of indication
DR 11	Prescription/documentation incomplete/incorrect
Administration	
ADM 1	Request/query concerning administration/compatibility
ADM 2	Administration (route)
ADM 3	Administration (duration)
ADM 4	Incompatibility or incorrect preparation or reconstitution
Dosage	
D 1	Failure to adjust dose for organ dysfunction
D 2	(Inappropriate) dose
D 3	(Inappropriate) administration interval
D 4	TDM not performed or not considered
Contraindication	
CI 1	Contraindication
Adverse drug reaction	
ADR 1	Adverse drug reaction
Other	
O 1	Advisory service/drug choice
O 2	Procurement/costs
O 3	Failure to discontinue relevant drugs pre-/perioperatively
O 4	Patient counselling or education
Interaction	
I 1	Interaction

The original database uses the expression "cause of PI" instead of DRP

TDM therapeutic drug monitoring

predictive values (NPV) were calculated per case report and are reported as median and range. Measures are defined as follows:

- *Proportion of rater agreement*: [positive and negative votes concordant with the gold standard/total votes] × 100%.
- *Sensitivity*: [true positive votes/(true positive votes + false negative votes)] × 100%—therefore, raters' ability to select (all) case-specific appropriate categories defined by SCPs.
- *Specificity*: [true negative votes/(true negative votes + false positive votes)] × 100%—thus, raters' ability to perceive inappropriate items defined by SCPs.
- *PPV*: [true positive votes/(true positive votes + false positive votes)] × 100%—the probability that raters' positive item selection is appropriate according to the gold standard.
- *NPV*: [true negative votes/(true negative votes + false negative votes)] × 100%—the probability that raters' negative item selection is appropriate according to the gold standard.

Table 2 Process of gold standard generation reached by five senior clinical pharmacists (SCPs): majority consensus (≥60%) and maximum consensus (100%) for a positive vote

Cause of pharmaceutical intervention case report 11 according [22]	SCP 1	SCP 2	SCP 3	SCP 4	SCP 5	60% consensus	100% consensus
(Clear) indication not (or no longer) given	0	0	0	0	0	0	0
(Clear) indication, but no drug prescribed	0	0	0	0	0	0	0
Drug allergy or medical history not considered	0	0	0	0	0	0	0
Double prescription	0	0	0	0	0	0	0
Dispensing error on the ward	0	0	0	0	**1**	0	0
Generic/therapeutic substitution	0	0	0	0	0	0	0
Transcription error	0	0	0	0	0	0	0
Inappropriate or not most suitable drug formulation in terms of indication	0	0	0	0	0	0	0
Inappropriately or not most suitable drug in terms of costs	0	0	0	0	0	0	0
Inappropriate or not most suitable drug in terms of indication	0	0	0	0	0	0	0
Prescription/documentation incomplete/incorrect	**1**	**1**	0	0	**1**	**1**	0
Request/query concerning administration/compatibility	0	0	0	0	0	0	0
Administration (route)	0	0	0	0	**1**	0	0
Administration (duration)	0	0	0	0	0	0	0
Incompatibility or incorrect preparation or reconstitution	0	0	0	0	0	0	0
Failure to adjust dose for organ dysfunction	0	0	0	0	0	0	0
(Inappropriate) dose	0	0	0	0	**1**	0	0
(Inappropriate) administration interval	**1**	**1**	**1**	**1**	**1**	**1**	**1**
TDM not performed or not considered	0	0	0	0	0	0	0
Contraindication	0	0	0	0	0	0	0
Adverse drug reaction	0	0	0	0	0	0	0
Advisory service/drug choice	**1**	**1**	**1**	0	**1**	**1**	0
Procurement/costs	0	0	0	0	0	0	0
Failure to discontinue relevant drugs pre-/perioperatively	0	0	0	0	0	0	0
Patient counselling or education	0	0	0	0	**1**	0	0
Interaction	0	0	0	0	0	0	0

Case report 11: Mr. S, age 49 years, 189 cm, 93 kg, was hospitalized at a general surgical ward because of a suspected gastric cancer. Due to a severe respiratory infection, Tarivid 200 mg coated tablets 1-1-1-1 (=9.36 Euro) were applicated orally. For pharmacokinetic/-dynamic reasons, you recommend the administration of 2-0-2-0

We altered the degree of SCPs' consensus to create a positive vote from a majority to a maximal 100% consensus (five out of five SCPs) within a sensitivity analysis. As both the PPV and NPV depend on the actual prevalence of DRP types provided by SCPs for consensus-based gold standard, the sensitivity analysis also varied the possible range of prevalence values in different case scenarios (0% to 100%) [26].

Data were analysed using the statistical software package R [27] or Microsoft Excel™ 2010.

Results

Case report assessment

The 24 standardized case reports yielded a median case-specific rater agreement of 90% to the majority consensus and varied between case reports with a range from 79 to 94% (Table 3). Sensitivity and specificity were 37%

Table 3 Median case-specific agreement [range] for n = 24 case classifications and other measures in dependence of consensus reached by five senior clinical pharmacists (SCPs): 60% and 100% consensus for a positive vote

Case-specific measures (%) median [range]	Majority consensus 60% (≥ 3 out of 5 SCP)	Maximal consensus 100% (5 out of 5 SCP)
Rater agreement	90 [79–94]	92 [92–99]
Sensitivity	37 [21–57]	41 [0–100]
Specificity	99 [97–100]	95 [92–99]
PPV	90 [60–100]	30 [0–70]
NPV	90 [78–95]	97 [88–100]

PPV positive predictive value, *NPV* negative predictive value

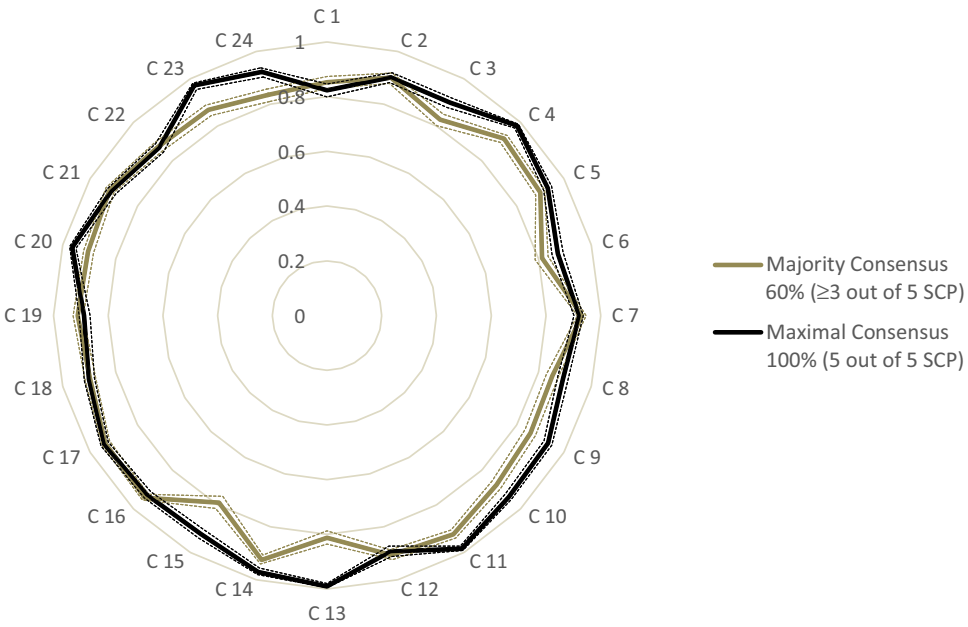

Fig. 1 Agreement of the raters' DRP classification with the gold standard for *n* = 24 case reports (C) and variation of gold standard consensus reached by five senior clinical pharmacists (SCP): line graphs represent mean case-specific rater agreements with their 95% confidence interval

[21–57%] and 99% [97–100%], respectively. Median PPV and NPV were 90% [60–100%] and 90% [78–95%] (Table 3).

Sensitivity analysis

Figure 1 shows the variability of rater agreement among the 24 case reports by presenting arithmetic means and 95% confidence intervals for the two scenarios: majority and maximum SCPs consensus. For all case reports, rater agreements were robust with respect to the majority and maximum consensus definitions (average agreement ≥ 79%). Within the two scenarios, specificity and NPV showed higher consistency among raters than sensitivity and PPV (Table 3).

Due to a statistical relationship, the actual case report-specific prevalence of DRP types provided by SCPs for consensus-based gold standard influences PPV and NPV. Figure 2 illustrates the relationship between a potentially given DRP type prevalence per case report (ranging from 0 to 100%) and the associated predictive values. The actual mean prevalence of DRP types per case report for two different consensus levels was 16.4 or 5.1% for a majority or maximal consensus. Applying the majority consensus definition, the PPV takes values of at least 80% in scenarios, whereas the DRP type prevalence exceeds 10%. The corresponding NPV is 93%, declining approximately linearly with increasing DRP type prevalence.

Discussion

This was a prospective nationwide survey-based study to evaluate coding accuracy of DRPs—one aspect of the categorical internet database DokuPIK, designed for recording routinely PIs. A comprehensive literature review recently stated that prior validation does not prevent the need of subsequent modification of DRP or PI classification systems and seems not necessarily be a study prerequisite [1]. Nevertheless, an agreed subjective coding seems to be crucial to establish and document PIs in clinical practice and to facilitate further research in this field [21]. The reported high

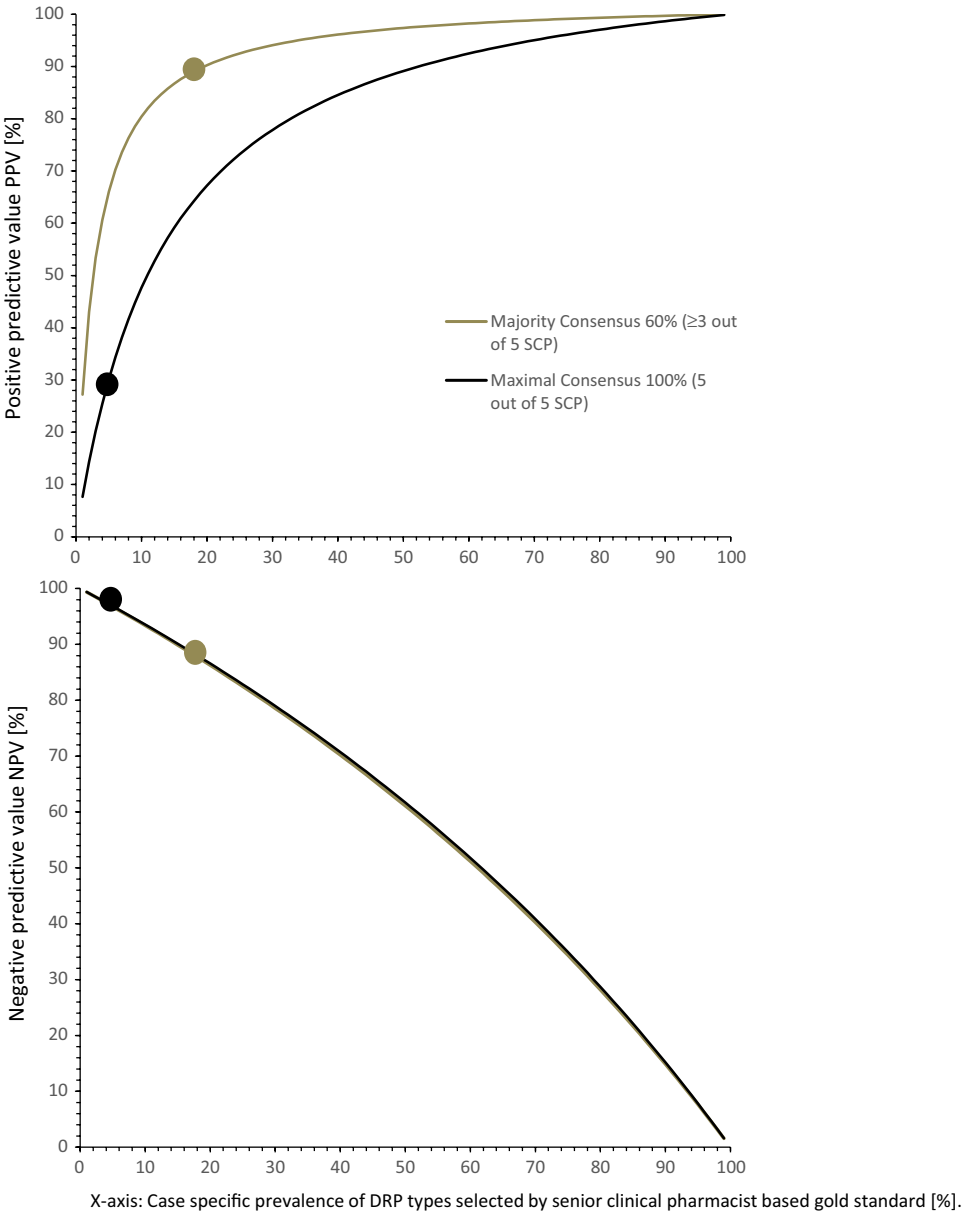

Fig. 2 Positive (PPV) and negative predicted values (NPV) and their relationship to the case-specific prevalence of DRP types (drug-related problems), potentially defined by expert consensus gold standard. Dots represent actual mean DRP type prevalence per case report for two different consensus levels among five senior clinical pharmacists (SCP), which was 16.4% or 5.1% for a majority or maximal consensus definition

intra-professional level of agreement to a SCPs' majority vote as gold standard therefore should add further value to the international literature and studies based on the DokuPIK [24].

Methodological issues

Common concepts of interrater agreement or interrater reliability measure the coding variability among a group of raters. However, even if coding is homogenous, item selection itself could be erroneous. The level of concordance to a gold standard of SCPs' majority opinion offers the chance to compare coding agreement to clinical pharmacists with a higher experience level, who seem to reduce medication errors more effectively than their less experienced colleagues do within the PROTECTED-UK cohort [28]. Our method therefore should add further value to this research field.

The level of agreement was assessed by using classical metrics of a binary classifier: sensitivity, specificity, PPV and NPV. While sensitivity and specificity are measures of the diagnostic accuracy of a test, they are of no practical use when it comes to estimate the probability of presence or absence of a certain condition for an individual patient. Positive and negative predictive values provide such estimates. However, they strongly depend on the prevalence of the condition in the population under investigation [29]. In this study, prevalence of DRP types selected for gold standard definition depended on the selected case reports and the consensus level of the five SCPs. We therefore advise caution when generalizing the observed PPV and NPV values

reported in this article. However, the provided sensitivity analyses demonstrate robustness of results for a wide range of DRP type prevalence values potentially defined by SCP consensus gold standard.

Comparison with the literature

Many authors used *Cohen's kappa* to evaluate IRR of their DRP or PI classification system. With κ values around 0.6 over all tested subitems, these highly diversifying systems demonstrate an agreement at the threshold between moderate to substantial according to Landis and Koch's widely used benchmarks [30]. The German APS-Doc, for example, showed a κ value of 0.58 for 48 subcategories [16], and the adapted PI-Doc version of Ganso et al. [22] showed a κ value of 0.60 for 57 DRP subitems. Similar values were published for the internationally widely used PCNE V6.2 with $\kappa = 0.614$ for 11 problem codes and $\kappa = 0.601$ for $n = 35$ cause codes [31]. A higher kappa was measured by Allenet et al. [32] for a short non-hierarchical questionnaire with ten DRP codes ($\kappa = 0.76$). The interpretation of κ values is difficult, and the observed broad subitem range itself (variation: $n = 10$–57) should lessens comparability due to its intrinsic effect on the kappa methodology [33]. Nevertheless, data show that reliability of existing classification systems is good but not ideal.

Around 70% of raters were able to identify one single appropriate main category provided by research pharmacists for around twenty scenarios during the validation of a Norwegian and an Australian DRP documentation system [34, 35]. In case of multiple item selection, the value 1.0 for a single positive vote was split into decimals, for example 0.5 per positive item for a classification into two categories. As the provided ratio only includes right positive votes, the concept is not comparable to our rater agreement definition, which comprises all votes—negative or positive—concordant with a SCP majority consensus. We therefore found a substantial higher median case-specific agreement of 90%, range 79–94% (Table 3).

Further implications

Although comparable data of similar documentation systems are missing, DokuPIK seems to have a good specificity with reference to a gold standard. With both 90%, case report-specific PPV (range 60–100%) and NPV (range 78–95%) are good, despite the allowance of multiple choices. Predictive values indicate, for the given case reports (prevalence based on SCP consensuses), a well-considered decision-making by the hospital pharmacist based on the DokuPIK. With 37% (range 21–57%), median case-specific sensitivity is rather low, but acceptable for an exclusive informative tool. These results are explained by a general conservative rater attitude to vote for the most appropriate case-specific DRP, whereas gold standard definition comprised all acceptable DRP types selected by majority opinion. Variation of gold standard consensus for a positive vote has an impact on case-specific measures (Table 2) and underlines the challenge to develop an exclusive and exhaustive hierarchical DRP documentation system.

Overall, to the best of our knowledge, similar data do not exist for established DRP/PI documentation systems, and therefore, our study adds to the available literature.

Limitations and future challenges

The anonymous survey precludes the description of survey participants' spread across German hospitals and their professional experience. Similar to the first nationwide database evaluation, they mainly could have been ward-based pharmacists at university and non-university hospitals [24]. Time-consuming rating was performed voluntarily without any preparatory training in a given period. Thus, rating results and number of participants may be improved, e.g. by well-trained coders incentivized for their effort. Otherwise, the sampling method by volunteers may have a positive impact on rating quality as they may generally have more experience and be more careful with documenting their interventions.

The difference between login characteristic and total number of registered users partially explains the difficulties in recruiting survey participants. Referring to 187 active users the rater proportion was markedly higher ($n = 37$ out of 187; 19.8%).

On creating majority opinions for positive votes, SCP detected the following problems during case evaluation: (1) missing case-specific information, e.g. the result of PI, (2) incomplete definitions for DokuPIK usage, e.g. for the term double prescription, and (3) one missing subcategory on advisory service concerning drug dosage. Consequently, we improved category prescriptions complemented by clinical examples and adequate instructions. Additionally, we slightly modified the DokuPIK within the rubric "others" by adding the proactive subcategory "advisory service/drug dose" and translated them to English for international usage.

DokuPIK is a well-accepted coding system for PIs performed by hospital pharmacists. Nevertheless, a future challenge should be to minimize documentation time in daily routine and to implement quality assurance measures to quantify, guarantee and improve CPs' professional performance within their interdisciplinary teams. Moreover, the online nature of the DokuPIK documentation tool should enable us to conduct prospective surveys within defined periods, to describe not only the diversity of day-to-day PIs directed by German hospital pharmacists [24] but also their incidence.

Conclusion

Although comparable data of similar documentation systems are missing, the use of DokuPIK shows a high level of agreement and a good specificity, PPV and NPV when using SCPs' majority vote as gold standard.

Despite allowing multiple choices, predictive values are of practical relevance and indicate a well-considered DRP categorization by the hospital pharmacist using DokuPIK. The low, but for an exclusive informative tool acceptable, sensitivity is explained by a general conservative attitude to vote for the most appropriate case-specific DRP by a single pharmacist, whereas gold standard definition comprised all acceptable DRP types selected by majority decision.

Acknowledgements We would like to thank the survey participants. Angela Ihbe-Heffinger, Claudia Langebrake, Carina Hohmann, Katja Leichenberg, Heike Hilgarth, Mareike Kunkel, and Michael Lueb for the Pharmacists' interventions working group of the German Association of Hospital Pharmacists, ADKA.

Funding This study did not receive funding.

Conflicts of interest All authors declared no conflicts of interest.

References

1. Basger BJ, Moles RJ, Chen TF. Application of drug-related problem (DRP) classification systems: a review of the literature. Eur J Clin Pharmacol. 2014;70(7):799–815.
2. Melchiors AC, Correr CJ, Venson R, Pontarolo R. An analysis of quality of systematic reviews on pharmacist health interventions. Int J Clin Pharm. 2012;34(1):32–42.
3. Patterson SM, Cadogan CA, Kerse N, Cardwell CR, Bradley MC, Ryan C, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Database Syst Rev. 2014;10:Cd008165.
4. Christensen M, Lundh A. Medication review in hospitalised patients to reduce morbidity and mortality. Cochrane Database Syst Rev. 2016;2:Cd008986.
5. Guerin A, Tanguay C, Lebel D, Bussieres J-F. Website on the impact of pharmacists. Eur J Hosp Pharm. 2014;21:192.
6. Batty R BE. Prescription monitoring interventions—is recording justified? Hospital Pharmacy. 1994;1.
7. Brady DFB. An evaluation of the contribution of the medical admissions pharmacist at a London Teaching Hospital. Int J Pharm Pract. 2004;1:1–6.
8. Correr CJ, Melchiors AC, de Souza TT, Rotta I, Salgado TM, Fernandez-Llimos F. A tool to characterize the components of pharmacist interventions in clinical pharmacy services: the DEPICT project. Ann Pharmacother. 2013;47(7–8):946–52.
9. Frontini R, Miharija-Gala T, Sykora J. EAHP Survey 2010 on hospital pharmacy in Europe: Part 1. General frame and staffing. Eur J Hosp Pharm. 2012;19:385–7.
10. EAHP. European Statements of Hospital Pharmacy. Survey Results 2014–15 http://www.eahp.eu/publications/survey/content/2015-eahp-survey-results2016.
11. Frontini R, Miharija-Gala T, Sykora J. EAHP Survey 2010 on hospital pharmacy in Europe: arts 4 and 5. Clinical services and patient safety. Eur J Hosp Pharm. 2013;20:69–73.
12. Kantelhardt P, Langebrake C. DokuPik—a new database for the documentation of medication errors and pharmaceutical interventions (part I medication errors). Krankenhauspharmazie. 2009;30(2):63–8.
13. Langebrake C, Kantelhardt P. DokuPik—a new database for the documentation of medication errors and pharmaceutical interventions (part II interventions). Krankenhauspharmazie. 2009;30(9):149–55.
14. van Mil JW, Westerlund LO, Hersberger KE, Schaefer MA. Drug-related problem classification systems. Ann Pharmacother. 2004;38(5):859–67.
15. Europe PCN. PCNE Classification for Drug-Related Problems V5.01. http://www.pcne.org/sig/drp/documents/PCNE%20classification%20V5.01.pdf2006 [cited 2014 2014-08-25].
16. Hohmann C, Eickhoff C, Klotz JM, Schulz M, Radziwill R. Development of a classification system for drug-related problems in the hospital setting (APS-Doc) and assessment of the inter-rater reliability. J Clin Pharm Ther. 2012;37(3):276–81.
17. Schaefer M. Discussing basic principles for a coding system of drug-related problems: the case of PI-Doc. Pharm World Sci. 2002;24(4):120–7.
18. Wolf C, Pauly A, Nutt S, Mangartz L, Friedland K, Langebrake C. Comparison of two documentation systems for drug-related problems: DokuPIK and PIE-System. Krankenhauspharmazie. 2015;36(2):69–78.
19. Methodology WCCfDS. ATC/DDD Index 2016 http://www.whocc.no/atc_ddd_index/[updated 2015-04-23. Last updated: 2013-12-19.
20. Reporting NCCoME, Prevention. a. NCC MERP Taxonomy of Medication Errors. http://www.nccmerp.org/sites/default/files/taxonomy2001-07-31.pdf2016.
21. Bourne RS, Shulman R, Tomlin M, Borthwick M, Berry W, Mills GH. Reliability of clinical impact grading by healthcare professionals of common prescribing error and optimisation cases in critical care patients. Int J Qual Health Care. 2017;29(2):250–5.
22. Ganso M, Areschin S, Lange P, Emser A, Rössler J, Krämer I. Verlässlichkeit eines Klassifikationssystems für pharmazeutische Interventionen. Krankenhauspharmazie. 2007;28(7):273–83.
23. Q-SET, Design of on-line surveys http://www.q-set.de/: Goldecker GmbH.
24. Langebrake C, Ihbe-Heffinger A, Leichenberg K, Kaden S, Kunkel M, Lueb M, et al. Nationwide evaluation of day-to-day clinical pharmacists' interventions in German hospitals. Pharmacotherapy. 2015;35(4):370–9.
25. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm. 2013;9(3):330–8.
26. Rehling M. Visualizing the impact of prevalence on a diagnostic test. Scand J Clin Lab Invest. 2010;70(6):458–61.
27. Team RC. R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing; 2015 [Available from: http://www.R-project.org/.
28. Rudall N, McKenzie C, Landa J, Bourne RS, Bates I, Shulman R. PROTECTED-UK—Clinical pharmacist interventions in the UK critical care unit: exploration of relationship between intervention, service characteristics and experience level. Int J Pharm Pract. 2017;25(4):311–9.
29. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. Acta Paediatrica (Oslo, Norway: 1992). 2007;96(3):338–41.
30. Landis JRKG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
31. Nielsen TR, Andersen SE, Rasmussen M, Honore PH. Clinical pharmacist service in the acute ward. Int J Clin Pharm. 2013;35(6):1137–51.

32. Allenet B, Bedouch P, Rose FX, Escofier L, Roubille R, Charpiat B, et al. Validation of an instrument for the documentation of clinical pharmacists' interventions. Pharm World Sci. 2006;28(4):181–8.
33. Wirtz M, Kutschmann M. Analyzing interrater agreement for categorical data using Cohen's kappa and alternative coefficients. Die Rehabil. 2007;46(6):370–7.
34. Ruths S, Viktil KK, Blix HS. Classification of drug-related problems. Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke. 2007;127(23):3073–6.
35. Williams M, Peterson GM, Tenni PC, Bindoff IK, Stafford AC. DOCUMENT: a system for classifying drug-related problems in community pharmacy. Int J Clin Pharm. 2012;34(1):43–52.