ORIGINAL REPORT

Association between drug-specific indicators of prescribing quality and quality of drug treatment: a validation study

Susanna M. Wallerstedt[1]*, Björn Belfrage[2] and Johan Fastbom[3]

[1]*Department of Clinical Pharmacology, Sahlgrenska University Hospital, Sweden*
[2]*Närhälsan Dals-Ed Health Center, Sweden*
[3]*Aging Research Center, Karolinska Institutet and Stockholm University, Sweden*

ABSTRACT

Purpose   To evaluate the concurrent validity of three European sets of drug-specific indicators of prescribing quality
Methods   In 200 hip fracture patients ($\geq$65 years), consecutively recruited to a randomized controlled study in Sahlgrenska University Hospital in 2009, quality of drug treatment at study entry was assessed according to a *gold standard* as well as to three drug-specific indicator sets (Swedish National Board of Health and Welfare, French consensus panel list, and German PRISCUS list). As *gold standard*, two specialist physicians independently assessed and then agreed on the quality for each patient, after initial screening with STOPP (Screening Tool of Older Persons' potentially inappropriate Prescriptions) and START (Screening Tool to Alert to Right Treatment).
Results   According to the Swedish, French, and German indicator sets, 82 (41%), 54 (27%), and 43 (22%) patients had potentially inappropriate drug treatment. A total of 141 (71%) patients had suboptimal drug treatment according to the *gold standard*. The sensitivity for the indicator sets was 0.51 (95% confidence interval: 0.43; 0.59), 0.33 (0.26; 0.41), and 0.29 (0.22; 0.37), respectively. The specificity was 0.83 (0.72; 0.91), 0.88 (0.77; 0.94), and 0.97 (0.88; 0.99). Suboptimal drug treatment was 2.0 (0.8; 5.3), 1.9 (0.7; 5.1), and 6.1 (1.3; 28.6) times as common in patients with potentially inappropriate drug treatment according to the indicator sets, after adjustments for age, sex, cognition, residence, multi-dose drug dispensing, and number of drugs.
Conclusions   In this setting, the indicator sets had high specificity and low sensitivity. This needs to be considered upon use and interpretation. Copyright © 2015 The Authors Pharmacoepidemiology and Drug Safety Published by John Wiley & Sons, Ltd.

INTRODUCTION

Prescribing of drugs is a challenge, particularly in older people. They often suffer from multiple diseases and therefore use many drugs. Further, age-related pharmacokinetic and pharmacodynamic changes increase the sensitivity to drug effects. In fact, it is well-known that suboptimal pharmacotherapy is common in the elderly, such as treatment with inappropriate drugs or dosages, and/or omissions of drugs which the patient would probably benefit from.[1–3]

In order to improve the quality of drug treatment, valid indicators of prescribing quality are essential.

Such an indicator is a measurable element of prescribing performance for which there is evidence or consensus that it can be used to assess quality and, hence, be used in changing the quality of care provided.[4] Indeed, indicators to measure the quality of healthcare are used worldwide as benchmarking and resource allocation according to indicators are believed to promote more efficient healthcare. Further, indicators of prescribing quality are often used in research.[5]

Indicators of prescribing quality can be either drug-specific or diagnosis-specific. The former may be particularly appealing for healthcare providers and researchers as they just require data on the drug treatment. Thus, they can easily be applied on a medication list alone or on drug register data. However, some drug-specific indicators may be harder to extract from register data, for instance those requiring information

Correspondence to: S. M. Wallerstedt, Department of Clinical Pharmacology, Sahlgrenska University Hospital, SE-413 45 Gothenburg, Sweden. E-mail: susanna.wallerstedt@pharm.gu.se

on dose or formulation of the drug. Indeed, in the Swedish Prescribed Drug Register, for example, prescribed daily dose is not a defined variable but has to be interpreted from a text string. Because of the nature of drug-specific indicators, these indicators can identify potentially inappropriate but not potentially missing drugs.

In Europe, three major drug-specific indicator sets, for older patients irrespective of setting, have been established: in Sweden by the National Board of Health and Welfare, in France by the French consensus panel list, and in Germany by the PRISCUS list.[6,7] These sets have been used in research studies,[8–13] and have all been validated by consensus panels using the Delphi technique.[6,14,15] In the United States, the Beers list is the prevailing one.[16] In a recent study, the overlap between different sets of criteria was shown to be small.[17]

Summarized, indicators of prescribing quality have often been evaluated for content and face validity.[4,18] However, little is known on the concurrent validity, that is, the sensitivity (the proportion of patients with suboptimal treatment according to a *gold standard*, also identified by an indicator), the specificity (the proportion of patients with appropriate drug treatment according to a *gold standard*, not captured by an indicator of potentially inappropriate drug treatment), and the predictive value (the proportion of patients correctly characterized by the indicator according to the *gold standard*). As long as we do not know of the ability of the indicators to differentiate between suboptimal and appropriate drug treatment, areas of application may be hard to determine.

To assess quality of drug treatment is a delicate matter, and there is no established *gold standard* for such assessments. However, a medical assessment is the key step, as all prescribing has to be adapted to the characteristics of the individual patient. Further, it is important to include both inappropriate and missing drugs in the concept of suboptimal drug treatment as undertreatment is also a prevailing problem.[19] Thus, an approach towards a *gold standard* for quality of drug treatment may be to let specialist physicians assess the medication list in relation to the patient's medical history. Moreover, to ascertain that the assessments are made systematically, it may be useful to start from validated screening tools. To minimize the impact of the assessors' professional skills and clinical judgement in this first step of the *gold standard* assessment, explicit criteria may be preferable. Summarized, the Screening Tool of Older Persons' potentially inappropriate Prescriptions (STOPP) and the Screening Tool to Alert to Right Treatment (START) may be useful for the purpose.[20] These tools are comprehensive, take into account the clinical situation, and cover both over- and undertreatment.

The aim of this study was to investigate the concurrent validity of three European sets of drug-specific indicators of prescribing quality, that is, how well the indicators correlate with a *gold standard* for suboptimal drug treatment.

METHODS

Setting and participants

The study cohort consisted of 200 hip fracture patients, consecutively recruited in 2009 to a randomized controlled study.[21] Inclusion criteria in the original study were patients, ≥65 years of age, who had undergone surgery for a hip fracture at the Sahlgrenska University Hospital, were residing in the Gothenburg area, and provided informed consent. In all, 200 out of 253 patients undergoing hip fracture surgery during the inclusion period were included in the original study; 23 declined participation, 14 did not fulfil the other inclusion criteria, and 10 were deceased before inclusion.

Quality of drug treatment at study entry (admission to the hospital) was assessed according to three sets of drug-specific indicators of prescribing quality as well as to a *gold standard*. These assessments were performed in 2012–2013.

The medication list was determined in the original study, and included products used regularly and as needed. Drugs for external use were included only if having potential systemic effects. In order to include prescribed drug treatment not captured in the medication reconciliation by the attending physician, the Swedish Register of Dispensed Drugs (Läkemedelsförteckningen) was consulted. This register is used in clinical practice when a patient cannot report satisfactorily on his/her medications, and holds information on prescribed drugs purchased from any Swedish pharmacy during the preceding 15 months.

Indicators of prescribing quality

We evaluated indicators of prescribing quality from three European countries: the Swedish National Board of Health and Welfare,[22] the French consensus panel list,[14] and the German PRISCUS list.[15] We included indicators which could be applied on a medication list alone, that is, drug-specific indicators. These could either comprise individual drugs deemed inappropriate in the elderly, or combinations of drugs deemed inappropriate for concomitant use

(polypharmacy indicators). We excluded indicators where the dose or the formulation of the drug had to be considered.

In Table 1 and Appendix 1 in Supporting Information, characteristics of the three sets of indicators are described. From the Swedish National Board of Health and Welfare, we included 22 potentially inappropriate drugs or drug groups and 4 polypharmacy indicators (six dose-dependent indicators were excluded, as were indicators concerning 11 conditions). Out of 34 criteria in the French consensus panel list, we included 103 potentially inappropriate drugs and four polypharmacy criteria, while two diagnosis-dependent and five diagnose-dependent indicators were excluded. Out of 83 potentially inappropriate drugs in the PRISCUS list, we included 71 drugs, whereas nine dose-dependent, one diagnosis-dependent, and two formulation-dependent indicators were excluded.

Gold standard

Gold standard for quality of drug treatment was systematically assessed in two steps aiming to identify inappropriate and missing drugs. Suboptimal drug treatment was defined as ≥1 inappropriate drugs or ≥1 missing drugs. For a patient without inappropriate/missing drugs, the treatment was considered appropriate.

First, we identified *potentially* suboptimal drug treatment by the use of STOPP and START, which provide 65 criteria for *potentially* inappropriate drugs and 22 criteria for *potentially* missing drugs, respectively.[19] Then the clinical relevance of identified STOPP and/or START outcomes was assessed at the individual level. An inappropriate drug was defined as a *clinically relevant* STOPP outcome. Thus, if the expected benefit of a particular medication was judged to outweigh the potential harm, such as an antipsychotic drug in a patient with schizophrenia, the STOPP outcome was assessed as not clinically relevant, i.e. not representing an inappropriate drug. Similarly, a missing drug was defined as a *clinically relevant* START outcome. Thus, if there was a clinical reason not to treat the patient with the drug, such as an adverse drug reaction or a contraindication, the START outcome was assessed as not clinically relevant, i.e. not representing a missing drug. In order to keep a conservative approach to categorizing drugs as inappropriate or missing, we chose to categorize STOPP and START outcomes not possible to assess regarding clinical relevance (e.g. because of missing information) as not clinically relevant.

The assessments were independently performed by one general practitioner and one geriatrician. They were based on (i) electronic medical records from the hospital (introduced early in the first decade of 2000) and faxed medical records from the primary care for the two years preceding the hip fracture, and (ii) previously collected data including information on risk of falls, cognition, residence, and glomerular filtration rate. The latter, estimated with the Cockcroft–Gault equation, was dichotomized as either ≥50 or <50 ml/min to fit the STOPP and START criteria. In a final consensus discussion, the two specialist physicians reached agreement on all STOPP/START outcomes, and the clinical relevance of these.

Table 1. Overall description of drug-specific indicators of prescribing quality included in this study. The content of each indicator set is described in Appendix 1 in Supporting Information

		Swedish National Board of Health and Welfare	French consensus panel list	German PRISCUS list
Potentially inappropriate drugs	Analgesics	X	X	X
	Antibiotics		X	X
	Antidepressants			X
	Antiepileptic drugs			X
	Cardiovascular drugs: antihypertensives, antiarrhythmics, antiplatelet drugs, and cerebral vasodilators		X	X
	Drugs with anticholinergic effects	X	X	X
	Ergotamine and derivates			X
	Gastrointestinal drugs		X	X
	Hypoglycaemic drugs		X	
	Long-acting benzodiazepines	X	X	X
	Muscle relaxants		X	X
	Neuroleptic drugs			X
	Sedatives and hypnotics (excl long-acting benzodiazepines)	X		X
Potentially inappropriate polypharmacy	Excessive polypharmacy (≥10 drugs)	X		
	Three or more psychotropic drugs	X		
	Two or more drugs from the same therapeutic class	X	X	
	Two drugs inappropriate to combine	X	X	

Data analysis

All statistical analyses were performed with SPSS (IBM SPSS Statistics for Windows, Version 17.0, Armonk, NY). The Mann Whitney and the Chi-square tests were used for comparisons of characteristics between patients. Kappa statistics was used to assess inter-rater agreement for STOPP/START outcomes. As for the association between the indicators of prescribing quality and suboptimal drug treatment, we calculated sensitivity and specificity as well as positive and negative predictive values including 95% confidence intervals. The concurrent validity of indicators of potentially inappropriate drugs and potentially inappropriate polypharmacy was determined combined and separately. Logistic regression was performed to obtain odds ratios (and 95% confidence intervals) for suboptimal drug treatment according to outcomes on the indicators. Adjustments were made for age, sex, cognition (defined as impaired or not), residence (defined as nursing home or not), multi-dose drug dispensing (a system which has been associated with an extensive medication list and quality of drug treatment),[9,10,23] and number of drugs (a proxy for burden of disease).[24] As drug-specific indicators can only detect potentially inappropriate drugs and not potentially missing drugs, we also performed sensitivity analyses where the *gold standard* included clinically relevant STOPP, but not START, outcomes. Values are presented as mean ± standard deviation if not stated otherwise.

RESULTS

Study population

Characteristics of patients by drug treatment quality—according to the three sets of drug-specific indicators of prescribing quality and according to the *gold standard*—are presented in Table 2. Summarized, the patients had a mean age of 84.5 years, ranging from 65 to 98 years, and 133 (67%) were women. The mean number of drugs in the medication list was 7.2 ± 3.9 (range 0–21). Multi-dose drug dispensing was consistently more common in patients with potentially suboptimal drug treatment according to the indicator sets as well as in patients with suboptimal drug treatment according to the *gold standard*. These patients also had more drugs in their medication list.

Quality of drug treatment

According to the Swedish, French, and German indicators, 82 (41%), 54 (27%), and 43 (22%) patients had potentially inappropriate drug treatment. These

Table 2. Characteristics of patients with and without potentially inappropriate drug treatment according to three European sets of drug-specific indicators of prescribing quality, and patients with and without suboptimal drug treatment according to the *gold standard*. Values are presented as mean ± standard deviation or n (percent of those identified by the indicator sets or the *gold standard* as either having/potentially/suboptimal drug treatment ("Yes") or not ("No")). Comparisons with a P-value below 0.05 are underlined.

	Potentially inappropriate drugs or polypharmacy according to indicator sets									Suboptimal drug treatment according to *gold standard*		
	Swedish National Board of Health and Welfare			French consensus panel list			German PRISCUS list					
	Yes	No	P-value	Yes	No	P-value	Yes	No	P-value	Yes	No	P-value
	n = 82	n = 118		n = 54	n = 146		n = 43	n = 157		n = 141	n = 59	
Age, years	84.4 ± 6.8	84.6 ± 7.3	0.82	84.9 ± 6.7	84.4 ± 7.2	0.77	84.6 ± 7.0	84.5 ± 7.1	0.86	84.8 ± 6.8	83.7 ± 7.6	0.37
Female sex	57 (70)	76 (64)	0.45	39 (72)	94 (64)	0.30	34 (79)	99 (63)	0.049	92 (65)	41 (69)	0.56
Impaired cognition, including dementia	39 (48)	51 (43)	0.54	30 (56)	60 (41)	0.068	20 (47)	70 (46)	0.82	69 (49)	21 (36)	0.084
Nursing home resident	27 (33)	33 (28)	0.45	25 (46)	35 (24)	0.002	14 (33)	46 (29)	0.68	47 (33)	13 (22)	0.11
Multi-dose drug dispensing	49 (60)	51 (43)	0.021	36 (67)	64 (44)	0.004	28 (65)	72 (46)	0.025	86 (61)	14 (24)	<0.0001
Number of drugs	10.2 ± 3.4	5.1 ± 2.6	<0.0001	9.7 ± 3.5	6.2 ± 3.6	<0.0001	9.7 ± 3.2	6.5 ± 3.8	<0.0001	8.2 ± 3.5	4.8 ± 3.6	<0.0001

patients had a mean of 0.5±0.7 (range: 0–3), 1.0±0.5 (0–3), and 1.1±0.4 (1–3) potentially inappropriate drugs per person, and, according to the Swedish and the French indicators, 67 (82%) and 18 (33%) had potentially inappropriate polypharmacy.

A total of 141 patients (71%) had suboptimal drug treatment according to the *gold standard* (kappa for inter-rater agreement: 0.52). These patients had a mean of 1.5±1.3 (range: 0–6) inappropriate drugs and 0.6±0.7 (0–3) missing drugs. In the sensitivity analysis, 118 patients (59%) were treated with inappropriate drugs.

Concurrent validity

The concurrent validity of the indicators of prescribing quality is presented in Table 3. The sensitivity was higher for the Swedish indicator set: 0.51 than for the French and the German ones: 0.33 and 0.29, respectively. The specificity and the positive predictive value was >0.80 for all indicator sets. The negative predictive value was <0.5 for all sets.

The indicators focusing on potentially inappropriate drugs had a sensitivity of 0.24 (Swedish) and 0.29 (French, German), with overlapping confidence intervals. Concerning the polypharmacy indicators, the Swedish one had higher sensitivity than the French one: 0.42 vs. 0.11, with a maintained high specificity: 0.86 vs. 0.97.

Within the Swedish indicator set, the sensitivity was greater for the polypharmacy indicator than for the inappropriate drugs indicator: 0.42 vs. 0.24. The opposite was found within the French indicator set: 0.11 vs. 0.29.

In the sensitivity analysis, where the *gold standard* included treatment with inappropriate drugs only (not missing drugs), the sensitivity and the negative predictive value increased numerically, whereas the sensitivity and the positive predictive value decreased. The changes were generally small, and the confidence intervals overlapped the main analysis.

For patients with potentially inappropriate drugs or drug combinations according to the Swedish, French, and German indicator sets, the crude odds ratios for

Table 3. Concurrent validity of three European sets of drug-specific indicators of prescribing quality, i.e. how well the indicators correlate with a *gold standard* for suboptimal drug treatment. Figures in italics represent results when the *gold standard* includes treatment with inappropriate drugs only, i.e. not missing drugs

		Suboptimal drug treatment ≥1 inappropriate drugs		Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
		According to indicator, n	According to gold standard				
			Indicator result confirmed, n (%)	Not identified by indicator, n				
Swedish National Board of Health and Welfare	Potentially inappropriate drug treatment	82	72 (88) / *64 (78)*	69 / *54*	0.51 (0.43; 0.59) / *0.54 (0.45; 0.63)*	0.83 (0.72; 0.91) / *0.78 (0.68; 0.86)*	0.88 (0.79; 0.93) / *0.78 (0.68; 0.86)*	0.42 (0.33; 0.51) / *0.54 (0.45; 0.63)*
	Potentially inappropriate drugs	36	34 (94) / *33 (92)*	107 / *85*	0.24 (0.18; 0.32) / *0.28 (0.21; 0.37)*	0.97 (0.88; 0.99) / *0.96 (0.90; 0.99)*	0.94 (0.82; 0.98) / *0.92 (0.78; 0.97)*	0.35 (0.28; 0.42) / *0.48 (0.41; 0.56)*
	Potentially inappropriate polypharmacy	67	59 (88) / *52 (78)*	82 / *66*	0.42 (0.34; 0.50) / *0.44 (0.35; 0.53)*	0.86 (0.75; 0.93) / *0.81 (0.72; 0.89)*	0.88 (0.78; 0.94) / *0.78 (0.66; 0.86)*	0.38 (0.31; 0.47) / *0.50 (0.42; 0.59)*
French consensus panel list	Potentially inappropriate drug treatment	54	47 (87) / *45 (83)*	94 / *73*	0.33 (0.26; 0.41) / *0.38 (0.30; 0.47)*	0.88 (0.77; 0.94) / *0.89 (0.80; 0.94)*	0.87 (0.76; 0.94) / *0.83 (0.71; 0.91)*	0.36 (0.28; 0.44) / *0.50 (0.42; 0.58)*
	Potentially inappropriate drugs	47	41 (87) / *40 (85)*	100 / *78*	0.29 (0.22; 0.37) / *0.34 (0.26; 0.43)*	0.90 (0.80; 0.95) / *0.91 (0.83; 0.96)*	0.87 (0.75; 0.94) / *0.85 (0.72; 0.93)*	0.35 (0.28; 0.42) / *0.49 (0.41; 0.57)*
	Potentially inappropriate polypharmacy	18	16 (89) / *15 (83)*	125 / *103*	0.11 (0.07; 0.18) / *0.13 (0.08; 0.20)*	0.97 (0.88; 0.99) / *0.96 (0.90; 0.99)*	0.89 (0.67; 0.97) / *0.83 (0.61; 0.94)*	0.31 (0.25; 0.38) / *0.43 (0.36; 0.51)*
German PRISCUS list	Potentially inappropriate drug treatment (drugs)	43	41 (95) / *37 (86)*	100 / *81*	0.29 (0.22; 0.37) / *0.31 (0.24; 0.40)*	0.97 (0.88; 0.99) / *0.93 (0.85; 0.97)*	0.95 (0.85; 0.99) / *0.86 (0.75; 0.93)*	0.36 (0.29; 0.44) / *0.48 (0.41; 0.56)*

NPV, negative predictive value; PPV, positive predictive value.

suboptimal drug treatment according to the *gold standard* were 5.1 (2.4; 10.9), 3.7 (1.6; 8.8), and 11.7 (2.7; 50.1). After age, sex, cognition, residence, multi-dose drug dispensing, and number of drugs had been considered, the adjusted odds were 2.0 (0.8; 5.3), 1.9 (0.7; 5.1), and 6.1 (1.3; 28.6). In the sensitivity analysis, the corresponding adjusted odds were 1.7 (0.7; 3.9), 2.6 (1.1; 6.4), and 3.1 (1.1; 8.5).

A total of 60 patients had suboptimal drug treatment according to the *gold standard* and were not identified by any of the indicator sets. For these patients, the most frequent inappropriate drugs were aspirin at too high a dose (n=11) or without indication (n=7), benzodiazepines in those prone to falls (n=13), and loop diuretics without clinical signs of heart failure (n=8). The most frequent missing drugs were statin therapy in the presence of vascular disease and a life expectancy of >5 years (n=6), and calcium and vitamin D supplement in patients with known osteoporosis (n=6).

DISCUSSION

Our study demonstrates that the Swedish, French, and German sets of drug-specific indicators, applied on the drug lists of the study participants, had acceptable specificity and positive predictive value. Indeed, more than eight out of ten patients with appropriate drug treatment according to the *gold standard* were not identified by any of the three indicator sets to have suboptimal drug treatment. Further, about nine out of ten patients characterized as having suboptimal drug treatment by these indicator sets were confirmed to have such treatment according to the *gold standard*.

The sensitivity and the negative predictive value, on the other hand, were generally poor. Indeed, sensitivity figures ranging from 0.51 to 0.29 indicate that five to seven out of ten patients with suboptimal drug treatment were not captured by the drug-specific indicator sets. Thus, the indicators may be inappropriate to use to identify patients with suboptimal drug treatment. Further, the indicators may be of limited value to characterize the drug treatment of patients, as more than every other patient with appropriate drug treatment according to the indicators, actually had suboptimal drug treatment.

In the sensitivity analyses, the sensitivity and the specificity were only marginally affected when inappropriate drugs only (not missing drugs) were included in the *gold standard* for suboptimal drug treatment. This finding may, at least partly, be explained by previous findings that about two thirds of patients with potential prescribing omissions have potentially inappropriate drugs at the same time.[25]

Interestingly, the Swedish indicator set had higher sensitivity than the other ones. This result was driven by the relatively high sensitivity of the polypharmacy indicators, as the indicators of potentially inappropriate drugs had low sensitivity in all sets. The fact that the Swedish polypharmacy indicators are quite extensive compared to the French ones may contribute to this result.

Strengths and weaknesses

The most important strength of this study is that it provides scientific knowledge on the concurrent validity of three European sets of drug-specific indicators of prescribing quality, that is, how well they correlate with a *gold standard* for suboptimal drug treatment. As far as we are aware, such information is lacking in the scientific literature. An additional strength of our study is the choice of *gold standard* to characterize the quality of drug treatment. Indeed, this *gold standard* includes assessments of the quality of drug treatment at the individual level based on quite extensive data both from the original randomized controlled study on medication reviews[21] and from hospital and primary care. Another advantage is that all assessments were performed by two specialist physicians with expertise in the relevant area.

A limitation of this study is the limited number of patients included. However, the confidence limits were sufficiently narrow to provide valuable information. Another limitation is that the STOPP/START tools, which were used to systemize the specialist assessments, may not capture all kinds of suboptimal drug treatment. In addition, our kappa value for the STOPP/START assessments was lower than previously reported.[19] Apart from the fact that the assessors may have differed in performance, a potential explanation for the moderate inter-rater agreement may be that the sources of information for applying the criteria were limited to medical records and previously collected study data. Indeed, the quality of the documentation in the medical records may vary, allowing for divergent interpretations. In fact, a systematic review showed that multiple sources may be needed when to apply the criteria, such as direct contact with patients, caregivers, and healthcare professionals.[26]

The inclusion criteria in the original study may have implications for the generalizability of the results. However, few patients were excluded, and thus, the external validity of the results should be acceptable. Further, hip fracture patients may represent a relevant subgroup of older patients because hip fracture is a common diagnosis in Sweden where every fourth

middle-aged woman will sustain a hip fracture during her lifetime, and one out of three hip fracture patients is a man.[27] In addition, suboptimal drug treatment is common in this patient group.[28] Nevertheless, the prevalence of suboptimal drug treatment, especially inappropriate drugs related to fall risk, may differ from that found in a general population of older people. Thus, generalizing the results to other settings and applications of the indicators need to be done with the above mentioned precautions in mind.

Comparison to prior research

Although several indicators have been used to evaluate the quality of drug treatment, scientific knowledge on the concurrent validity is limited. Indeed, several articles have described the prevalence of inappropriate drug treatment according to indicators,[5] and some have compared different explicit criteria with each other.[17,26] However, this study is the first one to evaluate how well drug-specific indicators intended for a general older population correlate to a *gold standard* for suboptimal drug treatment, providing figures on, e.g. sensitivity and specificity.

Our study confirms that suboptimal drug treatment is common in the elderly. Interestingly, such treatment was about six times as common in patients with potentially inappropriate drug treatment according to the PRISCUS list compared with patients without such treatment. These results are driven by the high specificity of this indicator set. However, the sensitivity was low as less than one third of those with suboptimal drug treatment were identified. Nevertheless, the confidence interval was >1, and thus, the German indicator set can predict suboptimal drug treatment. Correspondingly, suboptimal drug treatment was twice as common in patients with potentially inappropriate drug treatment according to the Swedish and the French indicator sets, after relevant covariates had been considered. The confidence limits passed the line of unity, and thus, these indicator sets were not significant predictors for suboptimal drug treatment. Indeed, the sample size may have been too small to detect predictor properties of this magnitude. However, in the sensitivity analysis where inappropriate, but not missing, drugs were included in the *gold standard*, both the French and the German indicator sets predicted inappropriate treatment.

The large difference between the crude and the adjusted odds ratios indicate that other factors than the drug list *per se*, for example the presence of multi-dose drug dispensing,[23] may be more important when to identify patients with suboptimal drug treatment. This is further supported by recent results on the concurrent validity of indicators of prescribing quality based on the number of drugs in the medication list, where no specific cut-off could serve as a general indicator of prescribing quality.[29]

The results of this study may contribute to the understanding of the lack of effects on relevant patient outcomes for third party medication reviews.[30–32] In case these are based on indicators applied to medication lists alone, they apparently capture only a minor part of patients with suboptimal drug treatment, and thus the effects cannot be expected to be great. In addition, the fact that only a limited proportion of alerts to drug treatment changes upon such reviews are acted upon,[33] may, at least partly, be explained by varying positive predictive value. On the other hand, the results are encouraging as drug-specific indicators are the easiest ones to incorporate into computerized decision support systems. Because of the high positive predictive value, such guidance for the physician at the moment of prescribing may improve prescribing performance.

Implications and future research

Because of the low sensitivity of the drug-specific indicators, these alone may not be sufficient to identify patients with potentially suboptimal drug treatment, an important area of application for an indicator. Further, the indicators have limitations within the area of application to monitor prescriber performance, even though face and content validity is appropriate.[6,14,15] Indeed, to assess the performance of physicians, an indicator needs to identify as many suboptimally treated patients as possible (acceptable sensitivity), and those identified as suboptimally treated should actually have suboptimal drug treatment (acceptable positive predictive value). Nevertheless, the high positive predictive values of the indicators suggest that an area of application may be the monitoring of changes in drug utilization and effects of interventions. Indeed, in these cases, it may be acceptable to include only a subgroup of the patients with the actual kind of suboptimal drug treatment captured by the indicator.

For future research, it may valuable to further explore how to define/assess a *gold standard* for appropriate/suboptimal drug treatment. Future research could also focus on improving the available indicator sets for enhanced sensitivity. Indeed, to increase the ability of the indicators to capture suboptimally treated patients, our results suggest that it may be preferable to include extensive polypharmacy criteria in the set. In fact, this approach seems not to decrease the specificity too much.

In conclusion, this study shows that the three European sets of drug-specific indicators of prescribing quality, easy to apply on register data, have limitations when it comes to the concurrent validity. Indeed, the indicators need to be used with caution, and decision-makers, healthcare providers, and researchers may find the results of the present study valuable when to use indicators and to interpret results.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

KEY POINTS

- Drug-specific indicators of prescribing quality are appealing as they are easy to apply on register data. Examples of applications are to measure the performance of healthcare in order to achieve improvements and to constitute outcome measures in scientific evaluations.
- For drug-specific indicators from Sweden, France, and Germany, evidence on content and face validity is available, but evidence on concurrent validity, i.e. how well the indicators correlate with a *gold standard* for suboptimal drug treatment, is lacking.
- For the indicator sets evaluated in this study, the specificity and positive predictive values were acceptable, whereas the sensitivity and the negative predictive values were not. Results on the indicators need to be interpreted with this in mind.

ETHICS STATEMENT

The study complies with the Declaration of Helsinki, and ethics approval was obtained from the Regional Ethical Review Board in Gothenburg.

ACKNOWLEDGEMENTS

REFERENCES

1. Gillaizeau F, Chan E, Trinquart L, *et al.* Computerized advice on drug dosage to improve prescribing practice. *Cochrane Database Syst Rev.* 2013; **11** CD002894.
2. Patterson SM, Hughes C, Kerse N, Cardwell CR, Bradley MC. Interventions to improve the appropriate use of polypharmacy for older people. *Cochrane Database Syst Rev.* 2012; **5** CD008165.
3. Alldred DP, Raynor DK, Hughes C, Barber N, Chen TF, Spoor P. Interventions to optimise prescribing for older people in care homes. *Cochrane Database Syst Rev.* 2013; **2** CD009095.
4. Hoven JL, Haaijer-Ruskamp FM, Vander Stichele RH. Indicators of prescribing quality in drug utilisation research: report of a European meeting (DURQUIM, 13–15 May 2004). *Eur J Clin Pharmacol.* 2005; **60**: 831–4.
5. Santos AP, Silva DT, Alves-Conceicao V, Antoniolli AR, Lyra DP, Jr. Conceptualizing and measuring potentially inappropriate drug therapy. *J Clin Pharm Ther.* 2015; **40**: 167–76.
6. Fastbom J, Johnell K. National indicators for quality of drug therapy in older persons: the Swedish experience from the first 10 years. *Drugs Aging.* 2015; **32**: 189–99.
7. Kaufmann CP, Tremp R, Hersberger KE, Lampert ML. Inappropriate prescribing: a systematic overview of published assessment tools. *Eur J Clin Pharmacol.* 2014; **70**: 1–11.
8. Bladh L, Ottosson E, Karlsson J, Klintberg L, Wallerstedt SM. Effects of a clinical pharmacist service on health-related quality of life and prescribing of drugs: a randomised controlled trial. *BMJ Qual Saf.* 2011; **20**: 738–46.
9. Sjöberg C, Edward C, Fastbom J, *et al.* Association between multi-dose drug dispensing and quality of drug treatment—a register-based study. *PLoS One.* 2011; **6** e26574.
10. Wallerstedt SM, Fastbom J, Johnell K, Sjöberg C, Landahl S, Sundström A. Drug treatment in older people before and after the transition to a multi-dose drug dispensing system—a longitudinal analysis. *PLoS One.* 2013; **8** e67088.
11. Johnell K, Fastbom J, Rosen M, Leimanis A. Inappropriate drug use in the elderly: a nationwide register-based study. *Ann Pharmacother.* 2007; **41**: 1243–8.
12. Bongue B, Laroche ML, Gutton S, *et al.* Potentially inappropriate drug prescription in the elderly in France: a population-based study from the French National Insurance Healthcare system. *Eur J Clin Pharmacol.* 2011; **67**: 1291–9.
13. Schubert I, Kupper-Nybelen J, Ihle P, Thurmann P. Prescribing potentially inappropriate medication (PIM) in Germany's elderly as indicated by the PRISCUS list. An analysis based on regional claims data. *Pharmacoepidemiol Drug Saf* 2013; **22**: 719–27.
14. Laroche ML, Charmes JP, Merle L. Potentially inappropriate medications in the elderly: a French consensus panel list. *Eur J Clin Pharmacol.* 2007; **63**: 725–31.
15. Holt S, Schmiedl S, Thurmann PA. Potentially inappropriate medications in the elderly: the PRISCUS list. *Dtsch Arztebl Int.* 2010; **107**: 543–51.
16. American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society updated Beers Criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc.* 2012; **60**: 616–31.
17. Morin L, Fastbom J, Laroche ML, Johnell K. Potentially inappropriate drug use in older people: a nationwide comparison of different explicit criteria for population-based estimates. *Br J Clin Pharmacol.* 2015; Epub ahead of print.
18. Andersen M. Is it possible to measure prescribing quality using only prescription data? *Basic Clin Pharmacol Toxicol.* 2006; **98**: 314–9.
19. Gallagher PF, O'Connor MN, O'Mahony D. Prevention of potentially inappropriate prescribing for elderly patients: a randomized controlled trial using STOPP/START criteria. *Clin Pharmacol Ther.* 2011; **89**: 845–54.
20. Gallagher P, Ryan C, Byrne S, Kennedy J, O'Mahony D. STOPP (Screening Tool of Older Person's Prescriptions) and START (Screening Tool to Alert doctors to Right Treatment). *Consensus validation. Int J Clin Pharmacol Ther.* 2008; **46**: 72–83.
21. Sjöberg C, Wallerstedt SM. Effects of medication reviews performed by a physician on treatment with fracture-preventing and fall-risk-increasing drugs in older adults with hip fracture-a randomized controlled study. *J Am Geriatr Soc.* 2013; **61**: 1464–72.
22. National Board of Health and Welfare. Indikatorer för god läkemedelsterapi hos äldre [Indicators for drug treatment in older people]. 2010.
23. Belfrage B, Koldestam A, Sjöberg C, Wallerstedt SM. Prevalence of suboptimal drug treatment in patients with and without multidose drug dispensing—a cross-sectional study. *Eur J Clin Pharmacol.* 2014; **70**: 867–72.
24. Brilleman SL, Salisbury C. Comparing measures of multimorbidity to predict outcomes in primary care: a cross sectional study. *Fam Pract.* 2013; **30**: 172–8.
25. Steinman MA, Landefeld CS, Rosenthal GE, Berthenthal D, Sen S, Kaboli PJ. Polypharmacy and prescribing quality in older people. *J Am Geriatr Soc.* 2006; **54**: 1516–23.
26. Hill-Taylor B, Sketris I, Hayden J, Byrne S, O'Sullivan D, Christie R. Application of the STOPP/START criteria: a systematic review of the prevalence of potentially inappropriate prescribing in older adults, and evidence of clinical, humanistic and economic impact. *J Clin Pharm Ther.* 2013; **38**: 360–72.
27. Kanis JA, Johnell O, De Laet C, Jonsson B, Oden A, Ogelsby AK. International variations in hip fracture probabilities: implications for risk assessment. *J Bone Miner Res.* 2002; **17**: 1237–44.

28. Sjöberg C, Bladh L, Klintberg L, Mellström D, Ohlsson C, Wallerstedt SM. Treatment with fall-risk-increasing and fracture-preventing drugs before and after a hip fracture: an observational study. *Drugs Aging.* 2010; **27**: 653–61.

29. Belfrage B, Koldestam A, Sjöberg C, Wallerstedt SM. Number of drugs in the medication list as an indicator of prescribing quality: a validation study of polypharmacy indicators in older hip fracture patients. *Eur J Clin Pharmacol.* 2015; **71**: 363–8.

30. Christensen M, Lundh A. Medication review in hospitalised patients to reduce morbidity and mortality. *Cochrane Database Syst Rev.* 2013; **2** CD008986.

31. Wallerstedt SM, Kindblom JM, Nylén K, Samuelsson O, Strandell A. Medication reviews for nursing home residents to reduce mortality and hospitalization: systematic review and meta-analysis. *Br J Clin Pharmacol.* 2014; **78**: 488–97.

32. Hohl CM, Wickham ME, Sobolev B, *et al.* The effect of early in-hospital medication review on health outcomes: a systematic review. *Br J Clin Pharmacol.* 2015; **80**: 51–61.

33. Kwint HF, Bermingham L, Faber A, Gussekloo J, Bouvy ML. The relationship between the extent of collaboration of general practitioners and pharmacists and the implementation of recommendations arising from medication review: a systematic review. *Drugs Aging.* 2013; **30**: 91–102.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.