Comparing Potentially Inappropriate Prescribing Tools and Their Association With Patient Outcomes

Frank Moriarty, PhD,[*†] *Kathleen Bennett, PhD,*[‡] *Rose Anne Kenny, MD,*[†] *Tom Fahey, MD,*[*] *and Caitriona Cahir, PhD*[‡]

OBJECTIVE: To assess the agreement of several different measures of potentially inappropriate prescribing (PIP) in older people and compare their relationship with patient-reported outcomes.

DESIGN: Prospective cohort study including participants in The Irish Longitudinal Study on Ageing (TILDA).

SETTING: Waves 1 and 2 of TILDA, a nationally representative aging cohort study.

PARTICIPANTS: A total of 1753 community-dwelling TILDA participants with linked administrative pharmacy claims data on medications.

MEASUREMENTS: Potentially inappropriate medications were assessed using the Screening Tool of Older Persons' Potentially Inappropriate Prescriptions (STOPP) v1, American Geriatrics Society (AGS) Beers Criteria® 2012, and relevant Assessing Care of Vulnerable Elders (ACOVE) v3 indicators. Potential prescribing omissions were assessed using the Screening Tool to Alert Doctors to the Right Treatment (START) v1 and ACOVE v3 indicators. Their agreement was assessed via κ statistics, and multivariate regression was used to assess relationships with emergency department visits, general practitioner (GP) visits, quality of life, and functional decline (increased assistance needed for activities of daily living).

RESULTS: There was slight agreement between STOPP and AGS Beers Criteria® (κ = 0.20) and ACOVE indicators (κ = 0.15), while agreement between AGS Beers Criteria® and ACOVE indicators was fair (κ = 0.31). Agreement was fair between START and ACOVE indicators (κ = 0.34). All measures of inappropriate medications were significantly associated with increased GP visits. Only exposure to two or more START indicators was associated with reduced quality of life (adjusted mean difference = −1.12; 95% confidence interval [CI] = −1.92 to −0.33), and only two or more AGS Beers Criteria® were associated with functional decline (adjusted odds ratio = 2.11; 95% CI = 1.37-3.28). For omissions, both measures were associated with functional decline, but only ACOVE indicators were associated with increased GP visits.

CONCLUSION: Prevalence of PIP and relationships with outcomes can differ substantially between tools with little agreement. Choice of PIP measure for research or practice should be considered in light of the circumstances and requirements in each case. **J Am Geriatr Soc 68:526-534, 2020.**

Key words: ACOVE indicators; AGS Beers Criteria®START; STOPP

Potentially inappropriate prescribing (PIP) among older adults has become an important public health issue worldwide and is perceived to have a significant negative clinical and economic impact.[1] It can describe prescription of a medication without a clinical indication, or an indicated medication where the risks outweigh the benefits or where a safer alternative exists.[1] It also includes the omission of a clinically indicated medication. Older people are particularly at risk of adverse consequences of suboptimal prescribing, due to physiological changes in aging and the likelihood they will have multiple conditions and medications, predisposing them to drug interactions.[2-4]

PIP is most frequently assessed using explicit process measures or criteria.[1] Explicit PIP criteria assess whether the prescription accords with accepted standards and can be applied with little or no clinical judgment. They are usually developed from published reviews, expert opinions, and consensus or Delphi techniques. The drugs to avoid or American Geriatrics Society (AGS) Beers Criteria® have been the most frequently used and validated explicit process measure of PIP. The AGS

From the *HRB Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Ireland; †The Irish Longitudinal Study on Ageing, Trinity College Dublin, Dublin, Ireland; and the ‡Division of Population Health Science, Royal College of Surgeons in Ireland, Dublin, Ireland.

Address correspondence to Frank Moriarty, PhD, HRB Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland, Mercer Building, Noel Purcell Walk, Dublin 2, Ireland. E-mail: frankmoriarty@rcsi.ie; Twitter: @FrankMoriarty

DOI: 10.1111/jgs.16239

0002-8614/20/$15.00

Beers Criteria® were developed in the United States in 1991 and updated in 1997, 2003, 2012, 2015, and 2019.[5] In the wake of AGS Beers Criteria®, a significant number of modified or new measures of PIP have followed, including the Screening Tool of Older Persons' Potentially Inappropriate Prescriptions (STOPP) and the Screening Tool to Alert Doctors to the Right Treatment (START), the Assessing Care of Vulnerable Elders (ACOVE), and country-specific measures.[6] They have been used extensively in research, both to assess the prevalence of PIP[7-9] and its relationship with patient and economic outcomes[9-11] and as an outcome measure for interventions aiming to improve prescribing.[12-14]

The choice of PIP measure in studies can be somewhat arbitrary, and often no rationale is provided for why the selected measure was used. Sometimes, country-specific measures are used (ie, in light of medications commonly or uncommonly used).[6] However, ideally, measures of PIP should be generalizable and comparable across countries. While a number of previous studies have used multiple measures of PIP, the agreement between these measures has rarely been assessed formally[15-17] and, to date, different PIP measures of drug omission have not been compared. Also, few studies have compared different PIP measures in terms of their predictive validity or association with adverse outcomes.[18,19] Many of the explicit measures of PIP incorporate a large range of criteria, and a comparative analysis of different measures of PIP may help identify common prevalent inappropriate medication classes or criteria associated with adverse health outcomes that can be applied in clinical practice. For instance, a recent study comparing the 2015 revised version of STOPP and AGS Beers Criteria® 2012 found the five most common instances of PIP were the same for both measures.[20]

The aim of this study is to assess the agreement of different explicit measures of PIP (relating to both errors of commission and omission) and to compare their relationship with patient-reported outcomes in community-dwelling older people in Ireland. We hypothesized that measures of PIP developed in Europe (STOPP/START) may have low agreement with those developed in the United States (AGS Beers Criteria® and ACOVE indicators) and that this may give rise to differing relationships with outcomes.

METHODS

Study Design and Setting

This is a prospective cohort study including participants in The Irish Longitudinal Study on Ageing (TILDA),[21] and is reported in line with the Strengthening Reporting of Observational Studies in Epidemiology statement.[22] Previously, the prevalence of PIP in this cohort and the association between STOPP/START and participant outcomes have been reported.[23,24] In brief, these older people, aged 65 years and older, completed baseline data collection between 2009 and 2011 and were followed up after approximately 2 years (2012-2013). They completed an in-depth computer-assisted personal interview and self-completion questionnaire at both waves, and a subset also underwent a baseline health assessment. A selection of TILDA participants eligible for the General Medical Services (GMS) scheme, a form of public health cover providing most health services free of charge, provided their GMS identifier. They consented for the use of this to examine medicines dispensed under the scheme, which was obtained from the Health Service Executive Primary Care Reimbursement Service.[25] These data provide information relating to medications (product name, World Health Organization Anatomical Therapeutic Classification code and defined daily dose, and product strength) and the dispensing (date of dispensing and quantity dispensed).

Data Collection

Measures of PIP

PIP exposure was assessed separately for both commissions and omissions, using a subset of applicable AGS Beers Criteria® 2012, STOPP version 1, and ACOVE version 3 indicators in the former case,[26-28] and START version 1 and ACOVE version 3 indicators in the latter case.[27,28] These represented the current versions of each PIP measure during the study period. Information was available to apply 69% of STOPP criteria (45/65), 81% of AGS Beers Criteria® (42/52), and 77% of ACOVE indicators relating to inappropriate medicines (17/22), while 68% of START criteria (15/22) and 34% of ACOVE indicators relating to prescribing indicated medicines (21/65) could be applied.

Outcomes

We considered the following patient-reported outcomes collected in TILDA (Table 1). Healthcare utilization was based on participants reporting how many times they attended an emergency department (ED) and went to their general practitioner (GP) as a patient in the 12 months preceding their follow-up interview, capturing both secondary and primary care utilization. Functional decline was defined as an increase between baseline and follow-up in the number of activities of daily living that a participant reported needing assistance with. Six basic activities were asked about: dressing, walking, bathing, eating, transferring in/out of bed, and toileting. Last, quality of life was assessed using the Control Autonomy Self-Realization Pleasure (CASP) measure included in the self-completion questionnaire, specifically the 12-item revised version, which has improved validity in this population.[29] The CASP measure, based on Maslow's theory of needs satisfaction, was developed for use at middle and older ages and is included in a number of national aging cohort studies.[29]

Covariates

A range of other participant characteristics were assessed at wave 1 (Supplementary Table S1). These included demographics (age, sex, level of education, and living arrangements), number of medicines, number of physician-diagnosed health conditions reported by the participant, private health insurance status, reported physical activity, reported hospital admissions, screened level of depressive symptoms (measured using the Center for Epidemiologic Studies Depression scale), cognition (measured using the animal naming task), and reported social participation. These were considered as confounders in the analysis of PIP and patient-reported outcomes.

Analysis

Agreement

The agreement of measures of PIP was assessed using the κ statistic, considering the number of criteria (ie, 0, 1, or ≥2) as

Table 1. Cross Tabulation of PIP Prevalence, According to Different Measures of Inappropriate Medications

Measures	STOPP, No. (%) 0 PIPs	1 PIP	≥2 PIPs	Total	κ (95% CI) Categorical	Binary
AGS Beers Criteria®						
0 PIPs	625 (35.7)	334 (19.1)	196 (11.2)	1155 (65.9)		
1 PIP	86 (4.9)	105 (6.0)	116 (6.6)	307 (17.5)		
≥2 PIPs	43 (2.5)	88 (5.0)	160 (9.1)	291 (16.6)	0.20 (0.17-0.23)	0.28 (0.24-0.32)
ACOVE indicators						
0 PIPs	691 (39.4)	418 (23.8)	251 (14.3)	1360 (77.6)		
1 PIP	63 (3.6)	101 (5.8)	156 (8.9)	320 (18.3)		
≥2 PIPs	0 (0.0)	8 (0.5)	65 (3.7)	73 (4.2)	0.15 (0.12-0.17)	0.22 (0.19-0.26)
Total	754 (43.0)	527 (30.1)	472 (26.9)			
	AGS Beers Criteria®, No. (%)					
	0 PIPs	1 PIP	≥2 PIPs			
ACOVE indicators						
0 PIPs	1052 (60)	177 (10.1)	131 (7.5)			
1 PIP	98 (5.6)	111 (6.3)	111 (6.3)			
≥2 PIPs	5 (0.3)	19 (1.1)	49 (2.8)		0.31 (0.27-0.35)	0.43 (0.39-0.48)

Abbreviations: ACOVE, Assessing Care of Vulnerable Elders; AGS, American Geriatrics Society; CI, confidence interval; PIP, potentially inappropriate prescription; STOPP, Screening Tool for Older Persons' Prescriptions.

well as the binary classification of any PIP (vs none) per each measure. To calculate 95% confidence intervals (CIs), bootstrapping with 100 replications was used for categorical classification of PIP. Agreement was classified as poor (κ statistic = <0), slight (κ = 0.00-0.20), fair (κ = 0.21-0.40), moderate (κ = 0.41-0.60), substantial (κ = 0.61-0.80), and almost perfect (κ = 0.81-1.00).[30]

Relationship With Outcomes

The relationship between participant-reported outcomes and PIP exposure (which was assessed in the 12 months preceding outcome measurement) was determined separately for each measure of PIP. The primary outcome of interest was healthcare utilization, including both ED and GP visits, with functional decline and quality of life as secondary outcomes. Details of assessment and analysis of these outcomes are provided in Supplementary Table S1. Multivariate regression models, as detailed in Supplementary Table S1, were fitted separately for each measure of PIP and each outcome; and the threshold for significance was adjusted using the false discovery rate approach at an α of 0.05 to account for multiple testing.[31]

Prevalence of Common Criteria

Last, among participants classified as having two or more criteria, according to each measure of PIP, the prevalence of individual criteria across all measures was compared graphically for commissions and omissions. For example, among participants exposed to two or more criteria according to STOPP, we assessed the prevalence of individual STOPP criteria, AGS Beers Criteria®, and ACOVE criteria for potentially inappropriate medications. The intention of this was to assess the predominant types of medication use that may give rise to differing relationships with outcomes across different measures. Statistical analysis was conducted using Stata 14 (StataCorp), and significance was assumed at $P < .05$.

RESULTS

A total of 1753 TILDA respondents, aged 65 years and older, were included in this analysis (Supplementary Figure S1). At follow-up interview, their median age was 76 years (interquartile range = 72-80 years), 54.4% were female, 50.2% had no or primary education only, and 32.2% had secondary education. A majority lived with a spouse (47.6%), while 34.4% lived alone and 18% lived with others. A substantial majority (79.7%) reported having multimorbidity (ie, ≥2 physician-diagnosed chronic conditions), and participants took a median of six regular medications (interquartile range = 4-9).

Agreement

Considering categorical PIP exposure (0, 1, or ≥2 criteria), the κ statistics for STOPP compared to the AGS Beers Criteria® and the ACOVE indicators were 0.20 (95% CI = 0.17-0.23) and 0.15 (95% CI = 0.12-0.17), respectively, indicating slight agreement. In both cases, STOPP classified a higher proportion as exposed to PIP. Agreement between AGS Beers Criteria® and ACOVE indicators was fair (κ = 0.31; 95% CI = 0.27-0.35), and AGS Beers Criteria® identified a higher proportion of individuals as exposed. Considering binary PIP exposure (Table 1), the κ statistics were 0.28 (95% CI = 0.24-0.32) for STOPP vs AGS Beers Criteria® and 0.22 (95% CI = 0.19-0.26) vs ACOVE indicators (fair agreement), while agreement was moderate for AGS Beers Criteria® vs ACOVE (κ = 0.43; 95% CI = 0.39-0.48). There was fair agreement between START and the ACOVE indicators (κ statistic = 0.34; 95% CI = 0.31-0.38), and the ACOVE indicators identified a higher proportion of participants as exposed. The κ statistic for binary PIP exposure was 0.51 (95% CI = 0.47-0.55) for START and the ACOVE indicators, indicating moderate agreement (Table 2).

Relationship With Outcomes

The results of the multivariate regression analysis are presented in Table 3. Comparing measures of inappropriate

Table 2. Cross Tabulation of PIP Prevalence, According to Different Measures of Omissions

Measures	START, No. (%) 0 PIPs	1 PIP	≥2 PIPs	Total	κ (95% CI) Categorical	Binary
ACOVE indicators						
0 PIPs	724 (41.3)	122 (7.0)	14 (0.8)	860 (49.1)		
1 PIP	235 (13.4)	163 (9.3)	46 (2.6)	444 (25.3)		
≥2 PIPs	62 (3.5)	226 (12.9)	161 (9.2)	449 (25.6)	0.34 (0.31-0.38)	0.51 (0.47-0.55)
Total	1021 (58.2)	511 (29.2)	221 (12.6)			

Abbreviations: ACOVE, Assessing Care of Vulnerable Elders; CI, confidence interval; PIP, potentially inappropriate prescription; START, Screening Tool to Alert Doctors to Right Treatment.

medications, all three (AGS Beers Criteria®, STOPP, and ACOVE) showed a statistically significant association with increased rate of GP visits. Only exposure to two or more ACOVE indicators was associated with reduced quality of life (adjusted mean difference = −1.68; 95% CI = −3.06 to −0.30), and only two or more AGS Beers Criteria® were associated with functional decline (adjusted odds ratio [OR] = 2.27; 95% CI = 1.47-3.52). For measures of medication omissions, Table 3 shows the likelihood of functional decline increase with either two or more START criteria (adjusted OR = 2.04; 95% CI = 1.24-3.35) or two or more ACOVE indicators (adjusted OR = 1.89; 95% CI = 1.21-2.96). Only two or more ACOVE indicators were associated with increased GP visits (adjusted incident rate ratio = 1.13; 95% CI = 1.04-1.24).

Prevalence of Common Criteria

For potentially inappropriate medications, the prevalence of criteria relating to benzodiazepines, anticholinergics, tricyclic antidepressants, and coxibs/nonsteroidal anti-inflammatory drugs with hypertension appeared higher in the two or more ACOVE indicators group relative to the two or more STOPP and two or more AGS Beers Criteria® groups (Figure 1). The prevalence of long-term maximal dose PPI use appeared lower among those with two or more AGS Beers Criteria® compared to those with two or more ACOVE or STOPP criteria (Figure 1). The prevalence of prescribing omissions relating to statins, antiarrhythmics, and regular inhaled bronchodilators appeared lower among those with two or more ACOVE omissions compared to those with two or more START omissions (Figure 2).

DISCUSSION

Principal Findings

This study found slight or fair levels of agreement between different measures of PIP applied to an older community-dwelling cohort. Generally, there was higher agreement between the measures relating to inappropriate omissions of medications (eg, START and ACOVE indicators) compared to measures relating to inappropriate medications (eg, STOPP, AGS Beers Criteria®, and ACOVE indicators). Additionally, there was evidence that the association between PIP and outcomes did differ depending on which measure of PIP was used to define exposure. All measures of inappropriate medication were associated with an increased rate of GP visits, while only the AGS Beers Criteria® were associated with functional decline, and only STOPP was significantly associated with ED visits. Conversely, prescribing omissions were more commonly associated with reduced quality of life and functional decline and less so with increased healthcare use.

Comparison With Previous Literature

The present study found STOPP had lower agreement with the other measures of inappropriate medications. The agreement between AGS Beers Criteria® and STOPP in the present study (κ = 0.28) was broadly in line with previous studies where the κ statistics ranged from 0.22 to 0.35 to 0.58.[15-17] In contrast, previous studies comparing successive versions of the AGS Beers Criteria® (2003 and 2012) reported κ statistics of 0.42 and 0.80, respectively, although higher agreement may be expected in the case of a measure being updated and revised.[15,16] Agreement between measures can depend both on the types of medication being used as well as how they are prescribed, with definitions of inappropriateness varying in both respects. Differences in agreement between studies may reflect differences in the profile of medicines and prescribing in the study populations, as well as differences in how the criteria are operationalized. While the agreement between explicit measures was fair at best, implicit or expert judgement–based measures of appropriateness (ie, the Medicines Appropriateness Index [MAI]) also had low agreement (κ statistic = 0.14) when compared to the AGS Beers Criteria®.[32] However, the MAI identified a higher proportion of participants as having inappropriate prescribing compared to AGS Beers Criteria® (82% vs 37%).

There is an extensive literature on adverse effects of PIP; however, relatively few studies have previously compared PIP measures with regard to their relationship with patient outcomes.[18,19,33-37] There has been little consistency in findings across studies, although generally the STOPP criteria have been associated with adverse outcomes. Two studies evaluating a community-dwelling cohort of adults aged 70 years and older found that, although AGS Beers Criteria® and STOPP had similar relationships with vulnerability, AGS Beers Criteria® were not associated with other adverse outcomes, such as hospitalization, adverse drug events, poorer quality of life, and ED visits.[18,19] A study of older hospital inpatients found that inappropriate medicines identified by STOPP contributed to a significantly higher proportion of admissions than those identified by AGS

Table 3. Association of Measures of PIP With ED and GP Visits, Functional Decline, and Quality of Life

Measure	ED Visits			GP Visits			Functional Decline			CASP-R12		
	Any Visit, %	IRR(95% CI)	*P* Value	Mean (SD)	IRR(95% CI)	*P* Value	Any Decline, %	OR (95% CI)	*P* Value	Score, Mean (SD)	Mean Difference (95% CI)	*P* Value
STOPP criteria												
0	17.0			4.5 (4.9)			4.6			27.0 (5.0)		
1	22.4	1.24 (0.94 to 1.62)	.127	5.9 (6.3)	1.13 (1.04 to 1.23)	.004[a]	8.5	1.15 (0.70 to 1.89)	.577	26.2 (5.1)	−0.23 (−0.84 to 0.38)	.462
≥2	29.6	1.39 (1.04 to 1.87)	.028[a]	6.8 (6.6)	1.15 (1.05 to 1.27)	.003[a]	13.8	1.22 (0.74 to 2.01)	.439	24.7 (5.5)	−0.51 (−1.24 to 0.22)	.168
AGS Beers Criteria®												
0	18.8			4.8 (4.5)			5.2			26.8 (4.9)		
1	29.3	1.24 (0.93 to 1.65)	.145	6.1 (5.4)	1.05 (0.96 to 1.15)	.28	8.8	1.16 (0.70 to 1.93)	.556	25.1 (5.5)	−0.45 (−1.12 to 0.23)	.193
≥2	27.2	1.09 (0.80 to 1.49)	.595	7.4 (7.3)	1.14 (1.03 to 1.26)	.009[a]	19.9	2.11 (1.36 to 3.28)	.001[a]	24.5 (5.7)	−0.80 (−1.58 to −0.02)	.044
ACOVE PIMs												
0	20.4			5.0 (4.7)			7.2			26.4 (5.2)		
1	26.4	1.00 (0.76 to 1.31)	.987	6.6 (6.5)	1.11 (1.02 to 1.21)	.017[a]	10.6	0.80 (0.51 to 1.26)	.335	25.7 (5.2)	−0.16 (−0.83 to 0.51)	.647
≥2	27.9	1.02 (0.52 to 1.97)	.96	7.9 (7.8)	1.12 (0.91 to 1.38)	.288	17.8	1.10 (0.54 to 2.24)	.792	23.6 (4.8)	−1.41 (−2.80 to −0.02)	.047
START criteria												
0	19.0			5.2 (5.7)			6.0			26.7 (5.1)		
1	24.3	1.16 (0.90 to 1.49)	.242	5.6 (5.8)	1.00 (0.92 to 1.08)	.939	9.8	1.32 (0.87 to 2.00)	.195	26.0 (5.2)	−0.05 (−0.62 to 0.52)	.867
≥2	30.9	1.51 (1.07 to 2.13)	.019[a]	6.9 (6.6)	1.09 (0.97 to 1.23)	.135	15.4	1.98 (1.20 to 3.26)	.008[a]	24.2 (5.7)	−1.12 (−1.92 to −0.33)	.006[a]
ACOVE PPOs												
0	16.9			4.6 (4.3)			5.0			26.9 (5.1)		
1	24.0	1.08 (0.82 to 1.42)	.574	6.0 (6.2)	1.08 (1.00 to 1.18)	.064	9.2	1.15 (0.71 to 1.86)	.568	25.8 (5.3)	−0.24 (−0.87 to 0.39)	.451
≥2	30.6	1.34 (1.01 to 1.77)	.039	6.5 (5.8)	1.11 (1.01 to 1.21)	.027[a]	13.6	1.82 (1.16 to 2.86)	.009[a]	25.1 (5.3)	−0.63 (−1.27 to 0.01)	.055

Abbreviations: ACOVE, Assessing Care of Vulnerable Elders; AGS, American Geriatrics Society; CASP-R12, Control Autonomy Self-Realization Pleasure Revised 12-item scale; CI, confidence interval; ED, emergency department; GP, general practitioner; IRR, incidence rate ratio; OR, odds ratio; PIM, potentially inappropriate medication; PIP, potentially inappropriate prescription; PPO, potential prescribing omission; START, Screening Tool to Alert Doctors to Right Treatment; STOPP, Screening Tool for Older Persons' Prescriptions.

[a]Statistically significant after correcting for false discovery rate.

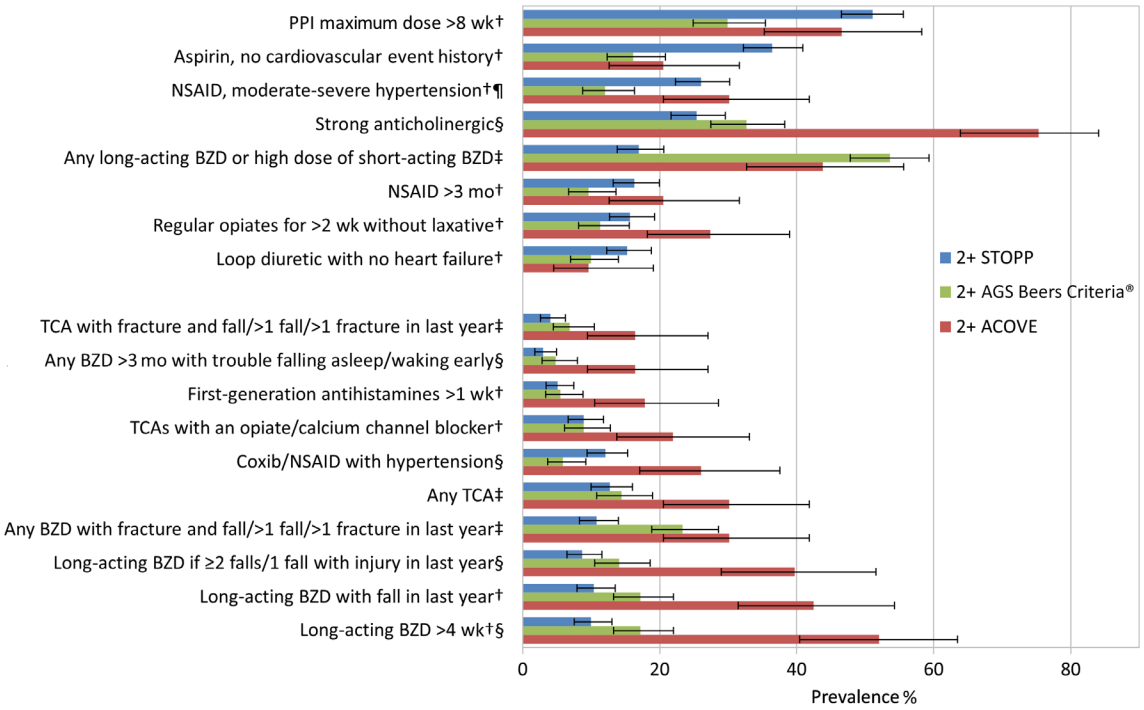

Figure 1. Prevalence of most common criteria (prevalence ≥10%) among participants with two or more Screening Tool of Older Persons' Potentially Inappropriate Prescriptions (STOPP) criteria, two or more American Geriatrics Society (AGS) Beers Criteria®, and two or more Assessing Care of Vulnerable Elders (ACOVE) indicators. Indicator included in †STOPP criteria, ‡AGS Beers Criteria®, or §ACOVE indicators. ¶Only assessable in 1192 participants (68%) who underwent health assessment. BDZ indicates benzodiazepine; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor, TCA, tricyclic antidepressant.

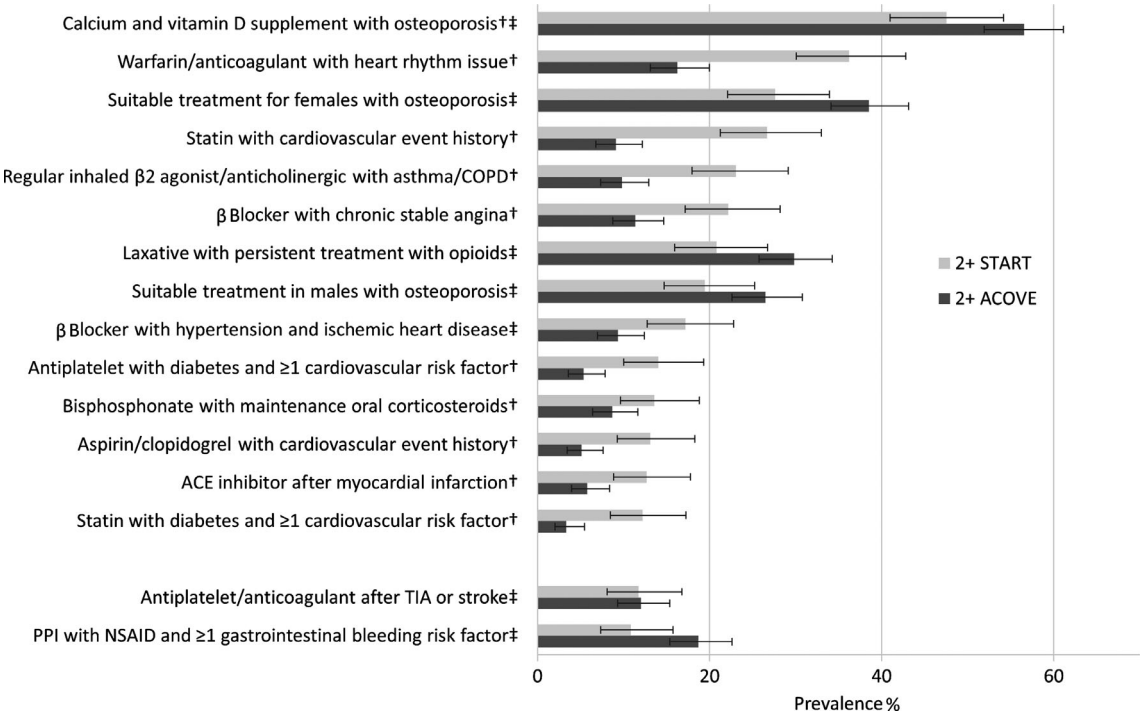

Figure 2. Prevalence of most common criteria (prevalence ≥10%) among participants with two or more Screening Tool to Alert Doctors to the Right Treatment (START) criteria and two or more Assessing Care of Vulnerable Elders (ACOVE) indicators. Indicator included in †START criteria or ‡ACOVE indicators. ACE indicates angiotensin-converting enzyme; COPD, chronic obstructive pulmonary disease; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor; TIA, transient ischemic attack.

Beers Criteria®.[33] Among inpatients aged 80 years and older, the MAI was found to have better discrimination for rehospitalization or death than STOPP or START[34]; however, STOPP was associated with a greater risk of drug-related readmission than the MAI.[35] In a study of 90-day readmission, although not statistically significant, exposure to STOPP criteria appeared to be positively associated with readmission, whereas exposure to AGS Beers Criteria® was not.[36] In an Australian study of community dwellers, STOPP identified more medication problems than AGS Beers Criteria® but fewer than Australia-specific prescribing indicators; however, STOPP was most closely aligned with pharmacist implicit review for medication problems.[37] STOPP considers clinical effectiveness as well as other aspects of PIP, such as drug-drug interactions and the dosage and duration of drug treatment, which may explain the more frequent association with outcomes.

Strengths and Limitations

This is the first study to undertake a comparative investigation of both the agreement and association with outcomes of three well-known measures of both inappropriate medication use and prescribing omissions. Strengths of this study include the large sample size and the use of pharmacy claims to assess medication exposure; this may be more reliable than self-reported drug use,[38] but may not reflect actual medication use in cases of nonadherence, and use of over-the-counter drugs was not considered. Participants in the study were community dwelling and, thus, the findings are likely to be representative of a large proportion of the older population, although participants may have a lower socioeconomic status due to the use of the GMS eligible subpopulation. While it is possible more socially disadvantaged individuals may have poorer outcomes and more PIP exposure, it is unlikely that the observed associations differ by socioeconomic status. This was a prospective study and so baseline characteristics could be accounted for, and this also allowed for the temporal relationship between medication exposure and outcomes measured subsequently to be assessed. We controlled for a large number of factors that could confound the relationship between inappropriate prescribing and the outcomes considered.

A limitation is that not all criteria within each measure could be applied to the data used in this study. However, this is unlikely to substantially impact the results as those criteria that could be applied from each measure were applicable to the same types of data (eg, medications, doses, durations, and diagnosed conditions) because the reason for being not applicable was the same across each measure. The outcomes relating to healthcare utilization were based on participant self-report over a period of 12 months, which may lead to potential inaccuracies in recall. However, we controlled for multiple participant characteristics, reducing the chance of systematic reporting bias among subgroups.[39,40] Given the measures of PIP incorporate prescribing across a range of therapeutic areas, we adjusted for participants' medications and conditions using counts rather than specific comorbidities; therefore, there is potential for residual confounding. While we accounted for multiple testing in our interpretation of results to reduce the false-positive rate, this may be overly conservative and a lack of statistical power could also contribute to nonsignificant associations.

Implications

Those who are implementing measures of PIP in research or practice should carefully consider which to use as agreement between different measures is low. Therefore, different measures will identify different levels of PIP among different patients. Depending on the purposes, it may be necessary to implement other criteria in addition to those in the measure of choice. Although no single measure had a greater association with all the outcomes considered, if one end point was particularly of interest, this may guide the choice of measure (eg, using STOPP if concerned with ED visits). Given the AGS Beers Criteria® identified fewer participants as having PIP, and were significantly associated with functional decline, this may support use of this measure over STOPP or the ACOVE indicators. Further research should investigate which individual indicators are most important (particularly from a patient perspective, as well as in relation to prevalence and likelihood and magnitude of potential harm), which could make use of large routine data sources to conduct appropriately powered analysis.

An ongoing systematic review of all explicit individual patient indicators of PIP will provide a comprehensive bank of existing indicators, and this may reduce the need to choose a single measure of PIP for use in practice or research.[41] The development and sharing of common standards for operationalizing indicators in electronic health records may also improve the comparability and consistency of findings from different research teams or settings.[42]

CONCLUSIONS

This study found that depending on how PIP is measured, the prevalence and relationships with outcomes can differ substantially, with little agreement between different tools. Different relationships with patient outcomes may be driven by varying types of inappropriate medicines or omissions included in each measure. No single PIP measure appears to be superior to another and, therefore, the choice of how to measure PIP in research or practice should be considered in light of the circumstances and requirements in each case.

ACKNOWLEDGMENTS

The authors wish to acknowledge all those who participated in The Irish Longitudinal Study on Ageing and the Primary Care Reimbursement Service for provision of the medication dispensing data used in this study.

Financial Disclosure: This study was funded by the Health Research Board in Ireland (HRB) through the HRB PhD Scholars Programme in Health Services Research (grant PHD/2007/16) and the HRB Centre for Primary Care Research (grant HRC/2007/1). The Irish Longitudinal Study on Ageing was supported by the Department of Health and Children, The Atlantic Philanthropies, and Irish Life.

Conflict of Interest: The authors have no conflicts.

Author Contributions: F.M. and C.C. conceived the study, and all authors were involved in study design.

K.B. and R.A.K. acquired study data, F.M. completed analysis, and all authors interpreted the data. F.M. prepared the initial draft, and all authors were involved in the critical revision of this and approval of the final manuscript.

Sponsor's Role: The sponsor had no role in the design, methods, subject recruitment/data collection, or analysis of this study, or in the preparation of this article.

REFERENCES

1. Spinewine A, Schmader KE, Barber N, et al. Appropriate prescribing in elderly people: how well can it be measured and optimised? Lancet. 2007; 370(9582):173-184. https://doi.org/10.1016/S0140-6736(07)61091-5.
2. Mangoni A, Jackson S. Age-related changes in pharmacokinetics and pharmacodynamics: basic principles and practical applications. Br J Clin Pharmacol. 2003;57(1):6-14. https://doi.org/10.1046/j.1365-2125.2003.02007.x.
3. Corsonello A, Pedone C, Incalzi R. Age-related pharmacokinetic and pharmacodynamic changes and related risk of adverse drug reactions. Curr Med Chem. 2010;17(6):571-584. https://doi.org/10.2174/092986710790416326.
4. Mallet L, Spinewine A, Huang A. The challenge of managing drug interactions in elderly people. Lancet. 2007;370(9582):185-191. https://doi.org/10.1016/S0140-6736(07)61092-7.
5. Beers M, Ouslander J, Rollingher I, Reuben D, Brooks J, Beck J. Explicit criteria for determining inappropriate medication use in nursing home residents. Arch Intern Med. 1991;151(9):1825-1832. http://www.ncbi.nlm.nih.gov/pubmed/1888249. Accessed May 6, 2014.
6. Levy HB, Marcus E-L, Christen C. Beyond the Beers criteria: a comparative overview of explicit criteria. Ann Pharmacother. 2010;44(12):1968-1975. https://doi.org/10.1345/aph.1P426.
7. Dalleur O, Boland B, De Groot A, et al. Detection of potentially inappropriate prescribing in the very old: cross-sectional analysis of the data from the BELFRAIL observational cohort study. BMC Geriatr. 2015;15(1):156. https://doi.org/10.1186/s12877-015-0149-2.
8. Bahat G, Bay I, Tufan A, Tufan F, Kilic C, Karan MA. Prevalence of potentially inappropriate prescribing among older adults: a comparison of the beers 2012 and screening tool of older person's prescriptions criteria version 2. Geriatr Gerontol Int. 2017;17(9):1245-1251. https://doi.org/10.1111/ggi.12850.
9. Hill-Taylor B, Walsh KA, Stewart S, Hayden J, Byrne S, Sketris IS. Effectiveness of the STOPP/START (screening tool of older persons' potentially inappropriate prescriptions/screening tool to alert doctors to the right treatment) criteria: systematic review and meta-analysis of randomized controlled studies. J Clin Pharm Ther. 2016;41(2):158-169. https://doi.org/10.1111/jcpt.12372.
10. Muhlack DC, Hoppe LK, Weberpals J, Brenner H, Schöttker B. The association of potentially inappropriate medication at older age with cardiovascular events and overall mortality: a systematic review and meta-analysis of cohort studies. J Am Med Dir Assoc. 2017;18(3):211-220. https://doi.org/10.1016/j.jamda.2016.11.025.
11. Moriarty F, Cahir C, Bennett K, Fahey T. Economic impact of potentially inappropriate prescribing and related adverse events in older people: a cost-utility analysis using Markov models. BMJ Open. 2019;9(1):e021832. https://doi.org/10.1136/bmjopen-2018-021832.
12. Clyne B, Smith SM, Hughes CM, et al. Effectiveness of a multifaceted intervention on potentially inappropriate prescribing in older patients in primary care: a cluster randomised controlled trial (the OPTI-SCRIPT study). Ann Fam Med. 2015;13:545-553. https://doi.org/10.1370/afm.1838.
13. Clyne B, Fitzgerald C, Quinlan A, et al. Interventions to address potentially inappropriate prescribing in community-dwelling older adults: a systematic review of randomized controlled trials. J Am Geriatr Soc. 2016;64(6):1210-1222. https://doi.org/10.1111/jgs.14133.
14. Cardwell K, Clyne B, Moriarty F, et al. Supporting prescribing in Irish primary care: protocol for a non-randomised pilot study of a general practice pharmacist (GPP) intervention to optimise prescribing in primary care. Pilot Feasibility Stud. 2018;4:122. https://doi.org/10.1186/s40814-018-0311-7.
15. Brown JD, Hutchison LC, Li C, Painter JT, Martin BC. Predictive validity of the beers and screening tool of older persons' potentially inappropriate prescriptions (STOPP) criteria to detect adverse drug events, hospitalizations, and emergency department visits in the United States. J Am Geriatr Soc. 2016;64(1):22-30. https://doi.org/10.1111/jgs.13884.
16. Blanco-Reina E, Ariza-Zafra G, Ocaña-Riola R, León-Ortiz M. 2012 American Geriatrics Society Beers Criteria: enhanced applicability for detecting potentially inappropriate medications in European older adults? a comparison with the screening tool of older person's potentially inappropriate prescriptions. J Am Geriatr Soc. 2014;62(7):1217-1223. https://doi.org/10.1111/jgs.12891.
17. Hudhra K, García-Caballos M, Jucja B, Casado-Fernández E, Espigares-Rodriguez E, Bueno-Cavanillas A. Frequency of potentially inappropriate prescriptions in older people at discharge according to Beers and STOPP criteria. Int J Clin Pharmacol. 2014;36(3):596-603. https://doi.org/10.1007/s11096-014-9943-8.
18. Cahir C, Moriarty F, Teljeur C. Potentially inappropriate prescribing and vulnerability and hospitalization in older community-dwelling patients. Ann Pharmacother. 2014;48(12):1546-1554. https://doi.org/10.1177/1060028014552821.
19. Wallace E, McDowell R, Bennett K, Fahey T, Smith SM. Impact of potentially inappropriate prescribing on adverse drug events, health related quality of life and emergency hospital attendance in older people attending general practice: a prospective cohort study. J Gerontol Ser A Biol Sci Med Sci. 2017;72(2):271-277. https://doi.org/10.1093/gerona/glw140.
20. Novaes PH, da Cruz DT, Lucchetti ALG, Leite ICG, Lucchetti G. The "iatrogenic triad": polypharmacy, drug–drug interactions, and potentially inappropriate medications in older adults. Int J Clin Pharmacol. 2017;39(4):818-825. https://doi.org/10.1007/s11096-017-0470-2.
21. Donoghue OA, McGarrigle CA, Foley M, Fagan A, Meaney J, Kenny RA. Cohort profile update: the Irish longitudinal study on ageing (TILDA). Int J Epidemiol. 2018;47:1398-1398l. https://doi.org/10.1093/ije/dyy163.
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007;4(10):e296. https://doi.org/10.1371/journal.pmed.0040296.
23. Moriarty F, Bennett K, Fahey T, Kenny RA, Cahir C. Longitudinal prevalence of potentially inappropriate medicines and potential prescribing omissions in a cohort of community-dwelling older people. Eur J Clin Pharmacol. 2015;71(4):473-482. https://doi.org/10.1007/s00228-015-1815-1.
24. Moriarty F, Bennett K, Cahir C, Kenny RA, Fahey T. Potentially inappropriate prescribing according to STOPP and START and adverse outcomes in community-dwelling older people: a prospective cohort study. Br J Clin Pharmacol. 2016:[Epub ahead of print];82:849-857. https://doi.org/10.1111/bcp.12995.
25. Sinnott S-J, Bennett K, Cahir C. Pharmacoepidemiology resources in Ireland—an introduction to pharmacy claims data. Eur J Clin Pharmacol. 2017;73(11):1449-1455. https://doi.org/10.1007/s00228-017-2310-7.
26. 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(4):616-631. https://doi.org/10.1111/j.1532-5415.2012.03923.x.
27. 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(2):72-83. http://cat.inist.fr/?aModele=afficheN&cpsidt=20054663. Accessed January 20, 2014.
28. San-José A, Agustí A, Vidal X, et al. An inter-rater reliability study of the prescribing indicated medications quality indicators of the Assessing Care Of Vulnerable Elders (ACOVE) 3 criteria as a potentially inappropriate prescribing tool. Arch Gerontol Geriatr. 2014;58(3):460-464. https://doi.org/10.1016/j.archger.2013.12.006.
29. Sexton E, King-Kallimanis BL, Conroy RM, Hickey A. Psychometric evaluation of the CASP-19 quality of life scale in an older Irish cohort. Qual Life Res. 2013;22(9):2549-2559. https://doi.org/10.1007/s11136-013-0388-7.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174. https://doi.org/10.2307/2529310.
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1):289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
32. Steinman MA, Rosenthal GE, Landefeld CS, Bertenthal D, Sen S, Kaboli PJ. Conflicts and concordance between measures of medication prescribing quality. Med Care. 2007;45(1):95-99. https://doi.org/10.1097/01.mlr.0000241111.11991.62.
33. Gallagher P, O'Mahony D. STOPP (Screening Tool of Older Persons' potentially inappropriate Prescriptions): application to acutely ill elderly patients and comparison with Beers' criteria. Age Ageing. 2008;37(6):673-679. https://doi.org/10.1093/ageing/afn197.
34. Alassaad A, Melhus H, Hammarlund-Udenaes M, Bertilsson M, Gillespie U, Sundstrom J. A tool for prediction of risk of rehospitalisation and mortality in the hospitalised elderly: secondary analysis of clinical trial data. BMJ Open. 2015;5(2):e007259. https://doi.org/10.1136/bmjopen-2014-007259.
35. Gillespie U, Alassaad A, Hammarlund-Udenaes M, et al. Effects of pharmacists' interventions on appropriateness of prescribing and evaluation of the instruments' (MAI, STOPP and STARTs') ability to predict hospitalization:

analyses from a randomized controlled trial. PLoS One. 2013;8(5):e62401. https://doi.org/10.1371/journal.pone.0062401.
36. Fabbietti P, Di Stefano G, Moresi R, et al. Impact of potentially inappropriate medications and polypharmacy on 3-month readmission among older patients discharged from acute care hospital: a prospective study. Aging Clin Exp Res. 2018;30(8):977-984. https://doi.org/10.1007/s40520-017-0856-y.
37. Curtain CM, Bindoff IK, Westbury JL, Peterson GM. A comparison of prescribing criteria when applied to older community-based patients. Drugs Aging. 2013;30(11):935-943. https://doi.org/10.1007/s40266-013-0116-6.
38. Richardson K, Kenny RA, Peklar J, Bennett K. Agreement between patient interview data on prescription medication use and pharmacy records in those aged older than 50 years varied by therapeutic group and reporting of indicated health conditions. J Clin Epidemiol. 2013;66(11):1308-1316. https://doi.org/10.1016/j.jclinepi.2013.02.016.
39. Reijneveld S, Stronks K. The validity of self-reported use of health care across socioeconomic strata: a comparison of survey and registration data. Int J Epidemiol. 2001;30(6):1407-1414. https://doi.org/10.1093/ije/30.6.1407.
40. Short M, Goetzel R, Pei X, et al. How accurate are self-reports? an analysis of self-reported healthcare utilization and absence when compared to administrative data. J Occup Environ Med. 2009;51(7):786-796. https://doi.org/10.1097/JOM.0b013e3181a86671.How.
41. Persaud N, Bali A, Rajakulasingam Y, et al. PROSPERO International prospective register of systematic reviews explicit individual patient indicators of potentially inappropriate prescribing: a systematic review. PROSPERO. 2016; CRD42016051037 http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42016051037. Accessed January 16, 2019.
42. Huibers CJ, Sallevelt BT, de Groot DA, et al. Conversion of STOPP/START version 2 into coded algorithms for software implementation: a multidisciplinary consensus procedure. Int J Med Inform. 2018;125:110-117.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Supplementary Figure S1: Inclusion of participants from The Irish Longitudinal Study on Ageing (TILDA) in the present analysis.

Supplementary Table S1: Methods of Measurement and Analysis of Included Outcomes.