Potentially Inappropriate Prescribing Among Older Persons: A Meta-Analysis of Observational Studies

Tau Ming Liew, MRCPsych[1,2]*

Cia Sin Lee, FCFP(S)[3]*

Shawn Kuan Liang Goh, MMed(FM)[3]

Zi Ying Chang, MMed(FM)[3]

[1]Department of Geriatric Psychiatry, Institute of Mental Health, Singapore

[2]Saw Swee Hock School of Public Health, National University of Singapore, Singapore

[3]SingHealth Polyclinics, Singapore

*These authors contributed equally to this work.

MORE ONLINE
www.annfammed.org

CORRESPONDING AUTHOR

Tau Ming Liew
Department of Geriatric Psychiatry
Institute of Mental Health
10 Buangkok View
Singapore 539747
tau_ming_liew@imh.com.sg

ABSTRACT

PURPOSE Potentially inappropriate prescribing (PIP) is a common yet preventable medical error among older persons in primary care. It is uncertain whether PIP produces adverse outcomes in this population, however. We conducted a systematic review with meta-analysis to pool the adverse outcomes of PIP specific to primary care.

METHOD We searched PubMed, Embase, CINAHL, Web of Science, Scopus, PsycINFO, and previous review articles for studies related to "older persons," "primary care," and "inappropriate prescribing." Two reviewers selected eligible articles, extracted data, and evaluated the risk of bias. Meta-analysis was conducted to pool studies with similar PIP criteria and outcome measures.

RESULTS Of the 2,804 articles identified, we included 8 articles with a total of 77,624 participants. All included studies had cohort design and low risk of bias. Although PIP did not affect mortality (risk ratio [RR] 0.98; 95% CI, 0.93-1.05), it was significantly associated with the other available outcomes, including emergency room visits (RR 1.63; 95% CI, 1.32-2.00), adverse drug events (RR 1.34; 95% CI, 1.09-1.66), functional decline (RR 1.53; 95% CI, 1.08-2.18), health-related quality of life (standardized mean difference −0.26; 95% CI, −0.36 to −0.16), and hospitalizations (RR 1.25; 95% CI, 1.09-1.44). A majority of the pooled estimates had negligible heterogeneity.

CONCLUSIONS This meta-analysis provides consolidated evidence on the wide-ranging impact of PIP among older persons in primary care. It highlights the need to identify PIP in primary care, calls for further research on PIP interventions in primary care, and points to the need to consider potential implications when deciding on the operational criteria of PIP.

INTRODUCTION

Potentially inappropriate prescribing (PIP) is the prescribing, or underprescribing, of medications for older persons that may cause significant harm.[1] It can be evaluated by holistic criteria to assess older persons' prescriptions in the context of multiple comorbidities, complex medication regimes, functional and cognitive status, treatment goals, and life expectancy.[1] In the literature, it has been often been operationalized with simpler screening criteria that can be further classified as implicit (judgement based) or explicit (criterion based).[1] Implicit (judgment based) tools are quality indicators that clinicians can apply to a prescription to judge the prescribing appropriateness, and they include the Medication Appropriateness Index. They can often be time consuming to use and rely on each clinician's medical knowledge, which limits their use in clinical practice.[1] Explicit (criterion based) tools comprise lists of drugs or drug classes (developed from literature reviews, expert opinion, and consensus techniques) that when prescribed or underprescribed can cause harm in older persons. They have been more widely adopted in the literature[2] because of the simplicity of their administration, and they include tools such as the Beers Criteria for Potentially Inappropriate Medication Use

in Older Adults (Beers Criteria), Screening Tool of Older Persons' Potentially Inappropriate Prescriptions (STOPP), and Screening Tool to Alert Doctors to Right Treatment (START).[1-3]

In the literature, the pooled prevalence of PIP ranges from 22.6% for community-dwelling older persons[4] to 43.2% for nursing home residents.[5] Previous reviews[3,6,7] have highlighted the associations of PIP with adverse drug events (ADEs), lower quality of life, hospitalizations, and higher health care costs. There has also been 1 meta-analysis[8] that demonstrated the significant effects of PIP on mortality (risk ratio [RR] 1.59; 95% CI, 1.45-1.75). The relationship between PIP and adverse outcomes may be related to the changes in pharmacokinetics and pharmacodynamics in older persons, especially among those who are frail, where there is limited physiologic reserve and lower tolerance of inappropriate medication use.[3]

With rapid aging of populations across the world,[9] we can expect an increasing number of older persons who have chronic diseases and need regular prescriptions from primary care. Although a recent systematic review reports the prevalence of PIP as 1 in 5 older persons attending primary care,[10] it remains unclear whether such PIP produces adverse outcomes in this population. Previous reviews[3,6-8] have included a heterogeneous range of participants, with much of the focus on specialized populations such as those from tertiary health care settings or nursing homes. We are uncertain whether the results from tertiary health care settings or nursing homes can be generalized to primary health care settings, considering that the patients in primary care can differ in their health profiles and comorbidities.[11,12] To this end, we sought to conduct a systematic review with meta-analysis to pool the adverse outcomes of PIP reported in the literature, specifically focusing on older persons in primary care.

METHODS

Search Strategy and Selection Criteria

This systematic review and meta-analysis was developed in accordance with the Preferred Reporting Items for Systematic Review and Meta-analysis statement.[13] The protocol was registered with the International Prospective Register of Systematic Reviews (registration number CRD42016048874) and previously peer reviewed and published in BMJ Open.[14]

We systematically searched PubMed, Embase, CINAHL, Web of Science, Scopus, and PsycINFO from inception to January 7, 2017, using keywords and controlled vocabulary related to "older persons," "primary care," and "inappropriate prescribing." A sample of the search strategy based on PubMed is shown in

the Supplemental Appendix, available at http://www.AnnFamMed.org/content/17/3/257/suppl/DC1/.

Similar search strategies were used for the other databases. We also hand searched the references of 6 review articles related to PIP[3,4,6-8,10] to retrieve relevant articles that were not captured through our search of the electronic databases. Included studies met the following criteria: recruited participants from primary care settings; either had ≥90% of the participants who were aged ≥65 years or reported subgroup analyses based on participants who were aged ≥65 years; were conducted via observational study designs, such as cross-sectional, case-control, or cohort studies; reported the adverse outcomes related to PIP, such as accident and emergency department (A&E) visits, ADEs, functional decline, health-related quality of life (HRQoL), hospitalizations, and mortality; and were reported in the English language. Studies were excluded if they recruited participants from non–primary care settings, such as tertiary hospitals or nursing homes; did not assess PIP based on published criteria; or focused only on PIP related to a single class of drug, such as analgesics or antibiotics.

After retrieving potential articles, 2 reviewers (S.K.L.G., Z.Y.C.) independently selected eligible articles, extracted the relevant data, and assessed the risk of bias. Any discordance between the 2 reviewers was resolved by discussion with a third independent reviewer (C.S.L. or T.M.L.). The extracted data included information on participants, study characteristics, criteria of PIP, measurement of adverse outcomes, effect estimates, and their 95% CI. The risk of bias was assessed with the original 8-item Newcastle-Ottawa scale,[15] which focuses on 3 key areas of potential bias: selection of participants, comparability of groups, and measurement of outcome. The exact items in the Newcastle-Ottawa scale are shown in the Supplemental Appendix.

Data Analysis

We conducted meta-analyses to pool the results for available outcomes that had >1 included study. We used the fixed-effect model (Mantel-Haenszel method)[16] in the meta-analysis because the random effects model would have been imprecise in its estimations when <5 studies were included in the analysis.[17,18] For studies that reported >1 model of statistical adjustment, we included only the result with the largest number of confounders adjusted for in the statistical model. For continuous outcomes, we pooled the results based on Cohen's standardized mean difference (SMD). For binary outcomes, we log-transformed the effect estimates before including them in the meta-analytic models. Although the RR and odds ratio have been used in different studies, we chose RR as a measure of

risk estimates in our meta-analyses because the odds ratio would have approximated the RR in the presence of low incidence rate of the outcomes.[19] We used the Q test and the I^2 statistic to assess the extent of heterogeneity among the pooled studies, with Q test P value <.10 or I^2 >75% indicating significantly high heterogeneity.[20] To minimize heterogeneity, we pooled only studies with similar PIP criteria and outcome measures. Although originally planned, we were not able to conduct meta-regression analysis or evaluate for publication bias because these analyses would not have been appropriate when there are <10 included studies.[21] All analyses were conducted in Stata version 14 (StataCorp LP).

We used the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) framework to classify the overall certainty of evidence into 1 of 4 levels (high, moderate, low, or very low) based on 7 key domains that focused on risk of bias, inconsistency, imprecision of effect estimates, risk of publication bias, indirectness of evidence, large magnitude of effect, and dose-response gradient.[22] Detailed descriptions of the other domains in GRADE are shown in the Supplemental Appendix.

RESULTS

We identified 2,804 articles from the 6 databases. After the selection process, we included 8 articles that are related to 6 unique studies (2 pairs of articles[23-26] reported different outcome measures originating from the same studies). The flowchart with details of the selection process is shown in Figure 1. The 6 included studies had a total of 77,624 par-

ticipants. All of them were cohort studies, with a mean follow-up duration of 2.0 years. The key characteristics of the included studies are presented in Table 1. All the included studies had low risk of bias and achieved

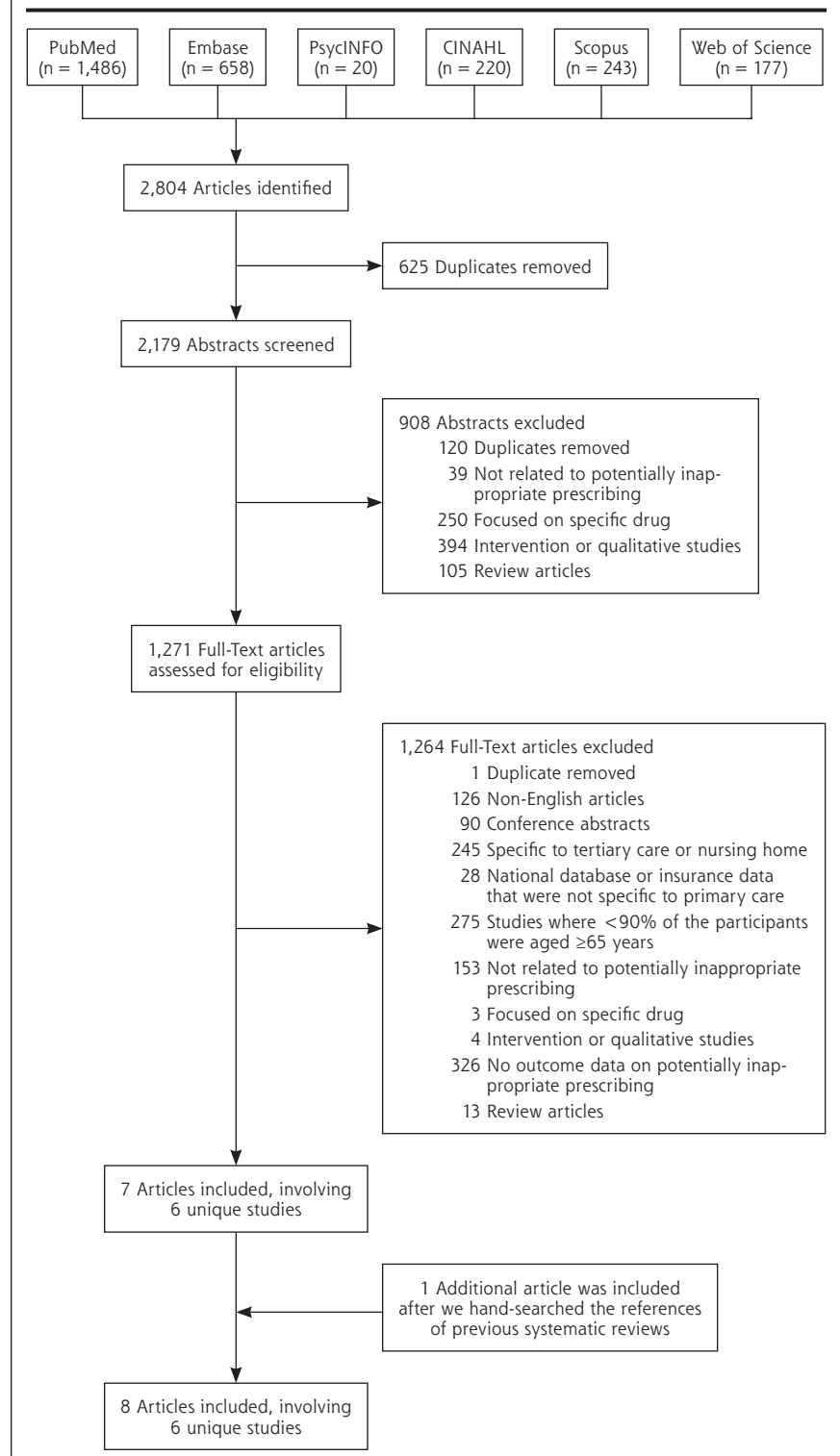**Figure 1. Flowchart of the study selection.**

ANNALS OF FAMILY MEDICINE • WWW.ANNFAMMED.ORG • VOL. 17, NO. 3 • MAY/JUNE 2019

Table 1. Key Characteristics of the Studies Included in the Systematic Review

Author (Year)	Sampling Method	Sample Size	Age, Years	% Female	Follow-Up Duration in Years	Data Source of PIP
Barnett et al,[27] 2011	Whole population registry	70,299	Mean 75.2 SD 6.8	57.0	2.0	Dispensed prescribing database
Cahir et al,[23] 2014 Cahir et al,[24] 2014	Stratified random sampling	931	Mean 78.0 SD 5.4	54.0	0.5	Pharmacy claim database
Hanlon et al,[25] 2002 Fillenbaum et al,[26] 2004	Stratified probability sampling	3,234	65-74 years: 49.1% 75-84 years: 41.1% ≥85 years: 9.8%	64.8	3.0	Structured interview
Moriarty et al,[28] 2016	Stratified random sampling	1,753	Mean 76.5 SD 6.0	54.4	1.0	Pharmacy claim database
Wallace et al,[29] 2017	Stratified random sampling	904	Median 77 IQR 74-81	54.0	2.0	Pharmacy claim databases
Wauters et al,[30] 2016	Whole GP registry and consecutive sampling	503	Mean 84.4	61.2	1.5	Secured record

A&E = accident and emergency department; ADE = adverse drug event; ADL = activity of daily living; CASP-R12 = control, autonomy, self-realization, and pleasure revised 12-item quality of life scale; DUR = drug utilization review; EQ-5D = health-related quality of life states consisting of 5 dimensions; GP = general practitioner; HRQoL = health-related quality of life; IQR = interquartile range; OARS = Older American Resources and Service; PIP = potentially inappropriate prescribing; START = Screening Tool to Alert Right Treatment; STOPP = Screening Tool of Older Persons' Potentially Inappropriate Prescriptions; VES = Vulnerable Elders Survey.

Table 2. Covariates That Were Adjusted for in the Statistical Models of the Included Studies

Author (Year)	Adjusted Covariate						
	Age	Sex	Socioeconomic status	Comorbidities	Number of Medications	Medication Adherence	Education
Barnett et al,[27] 2011	✓	✓	✓	—	✓	—	—
Cahir et al,[23] 2014	✓	✓	✓	✓	✓	✓	—
Cahir et al,[24] 2014	✓	✓	✓	✓	✓	✓	—
Hanlon et al,[25] 2002	✓	✓	✓	✓	✓	—	✓
Fillenbaum et al,[26] 2004	✓	✓	✓	✓	—	—	✓
Moriarty et al,[28] 2016	✓	✓	—	✓	✓	—	✓
Wallace et al,[29] 2017	✓	✓	✓	✓	✓	✓	✓
Wauters et al,[30] 2016	—	—	—	—	✓	—	—

maximum or near-maximum scores on the Newcastle-Ottawa Scale (the summary results are shown in Table 1, and the detailed results are shown in the Supplemental Appendix. All included studies reported results based on adjusted estimates. The covariates that were adjusted for in the statistical models are presented in Table 2. They include key confounders such as age, socioeconomic status, disease comorbidities, and number of medications.

We conducted meta-analyses only when there were ≥2 studies with results based on similar PIP criteria and outcome measures; results that could not be pooled in the meta-analysis are separately shown in the Supple-

mental Appendix. The PIP criteria commonly used in the included studies were the Beers Criteria and the STOPP criteria. Six outcome measures were consistently reported in ≥2 studies and could be subjected to meta-analysis, including A&E visits, ADEs, functional decline, HRQoL, hospitalizations, and mortality. Figure 2 shows the forest plots for the adverse outcomes of PIP based on the Beers Criteria, with PIP significantly associated with functional decline (pooled RR 1.38; 95% CI, 1.06-1.80) and hospitalizations (pooled RR 1.14; 95% CI, 1.01-1.29) but not with mortality (pooled RR 0.98; 95% CI, 0.93-1.05). Figure 3 shows the forest plots for the adverse outcomes of PIP based

Criteria of PIP	Data Source of Outcomes	Risk of Bias (Newcastle-Ottawa Scale)[a]		
		Selection (Max 4✶)	Comparability (Max 2✶)	Outcome (Max 3✶)
Beers Criteria 2003	Mortality: Death certification database	✶✶✶✶	✶✶	✶✶✶
Beers Criteria 2012 STOPP	ADEs, hospitalizations, and A&E visits: Structured interview and medical records	✶✶✶✶	✶✶	✶✶✶
	Functional decline: VES scale			
	HRQoL: EQ-5D scale			
Beers Criteria 1997 DUR	Mortality: National Death Index	✶✶✶✶	✶✶	✶✶✶
	Functional decline: Combination of Katz ADL, OARS Instrumental ADLs, and abbreviated Rosow-Breslau scale			
	Hospitalizations: Medicare and Medicaid database			
	Outpatient visits and nursing home entry: self- or proxy report			
STOPP START	A&E visits, GP visits: Structured interview	✶✶✶✶	✶✶	✶✶✶
	Functional decline: Difficulty in doing 6 named ADLs			
	HRQoL: CASP-R12 scale			
Beers Criteria 2012 STOPP	ADEs: Patient interview and medical records	✶✶✶✶	✶✶	✶✶✶
	A&E visits and hospitalizations: Medical records			
	HRQoL: EQ-5D scale			
STOPP START	Mortality: Secured record	✶✶✶✶	✶	✶✶✶
	Hospitalizations: Secured record			

[a] A star (✶) is awarded if a specific criterion in the Newcastle-Ottawa Scale was met, indicating low risk of bias in that criterion. Further details on the Newcastle-Ottawa Scale are available in the supplemental appendix, available at http://www.AnnFamMed.org/content/17/3/257/suppl/DC1/.

		Adjusted Covariate					
Race	Functional Status	Cognitive Impairment	Social Support	Prior Healthcare Utilization	Mental Health Condition	Other
—	—	—	—	—	—	—
—	—	—	✓	—	—	Sex of the general practitioner
—	—	—	✓	—	—	Sex of the general practitioner
✓	✓	✓	—	—	—	Body mass index
✓	✓	✓	—	✓	—	Marital status, insurance coverage
—	—	—	✓	✓	✓	Insurance coverage
—	—	—	—	✓	✓	—
—	—	—	—	—	—	—

on the STOPP criteria, with PIP significantly associated with A&E visits (pooled RR 1.63; 95% CI, 1.32-2.00), ADEs (pooled RR 1.34; 95% CI, 1.09-1.66), functional decline (pooled RR 1.53; 95% CI, 1.08-2.18), HRQoL (pooled SMD −0.26; 95% CI, −0.36 to −0.16), and hospitalizations (pooled RR 1.25; 95% CI, 1.09-1.44).

A majority of the pooled estimates had negligible heterogeneity. Notwithstanding that, 1 of the outcomes, HRQoL, showed significantly high heterogeneity (I^2 82.3% and $P = 0.003$) because of the different scales used in measuring HRQoL (2 studies[24,29] used the EQ-5D, and 1 study[28] used the Control, Autonomy, Self-realization, and Pleasure Revised 12-item Qual-

ity of Life scale). The GRADE assessment of overall certainty of evidence is shown in Table 3. Apart from HRQoL, the rest of the outcomes maintained "low" certainty of evidence consistent with meta-analytic results based on observational studies, with no further downgrades or upgrades in the GRADE assessment. HRQoL was downgraded to "very low" certainty of evidence because of the high heterogeneity.
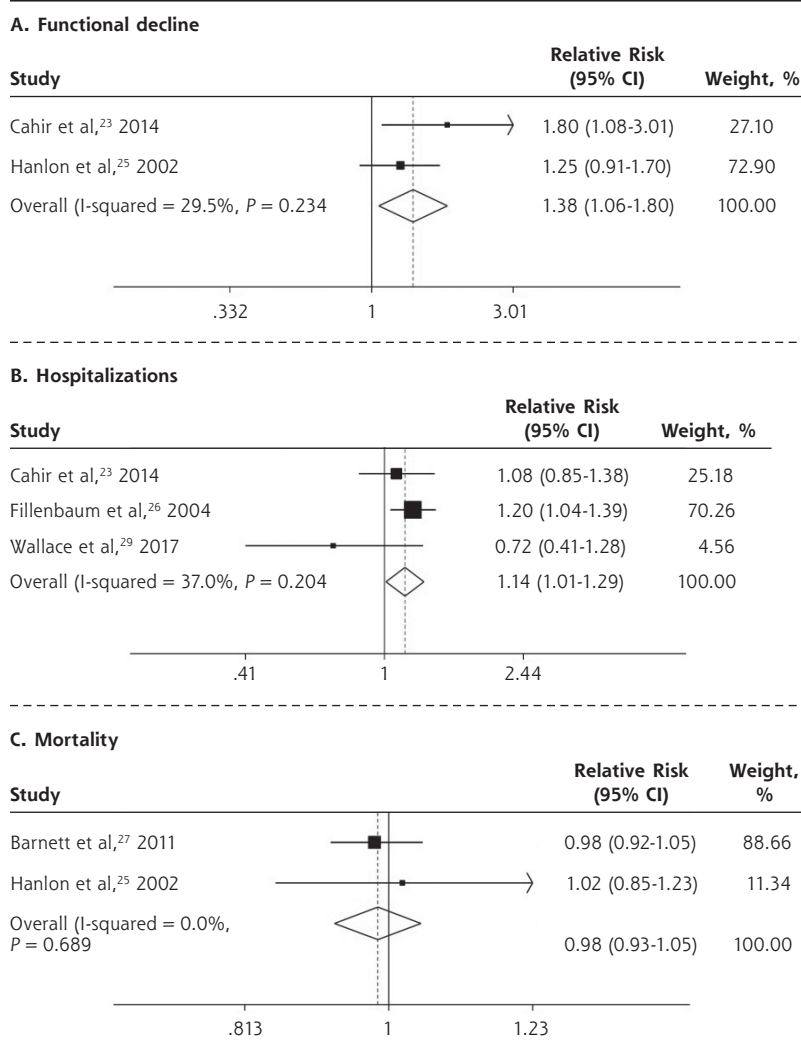
DISCUSSION

This is the first systematic review and meta-analysis of observational studies on the adverse outcomes of PIP

among older persons in primary care. Although PIP is not associated with mortality in this study, it is significantly associated with the other available outcomes, including A&E visits, ADEs, functional decline, HRQoL, and hospitalizations. The meta-analytic results were based on cohort studies with large numbers of participants, samples representative of the primary health care settings, and low risk of bias. Moreover, most of the pooled estimates had low heterogeneity, which suggests consistency of results across the included studies.

Comparison With Other Studies

To date, only 1 other meta-analysis has reported on the adverse outcomes of PIP.[8] This meta-analysis by Muhlack et al[8] included a heterogeneous range of populations (with a large proportion of participants from tertiary health care settings or nursing homes) and focused mainly on the outcome of mortality, however. Our result on mortality showed both similarity and dissimilarity to that of Muhlack et al.[8] In the current meta-analysis, we found that PIP (assessed at a single time point, also known as the prevalent-user design) is not associated with mortality longitudinally (pooled RR 0.98; 95% CI, 0.93-1.05). This is similar to the result by Muhlack et al,[8] where the pooled RR based on prevalent-user design of PIP is 1.01 (95% CI, 0.97-1.04). However, Muhlack et al[8] also identified a subset of studies that evaluated PIP based on the new-user design (which also captures those with new onset of PIP) and showed that the risk of mortality is higher (RR 1.59; 95% CI, 1.45-1.75) when PIP is based on evaluations at multiple time points. The discrepancy in findings is consistent with our understanding of the prevalent-user design and its association with healthy-user/sick-stopper bias, which can underestimate the true effect.[8,31,32] At any specific time point, those found to have PIP according to the prevalent-user design are more likely to be healthier people who have better physiologic reserve and can adhere

Figure 2. Forest plots for the adverse outcomes of potentially inappropriate prescribing based on the Beers Criteria.

A. Functional decline

Study		Relative Risk (95% CI)	Weight, %
Cahir et al,[23] 2014		1.80 (1.08-3.01)	27.10
Hanlon et al,[25] 2002		1.25 (0.91-1.70)	72.90
Overall (I-squared = 29.5%, P = 0.234		1.38 (1.06-1.80)	100.00

.332　　1　　3.01

B. Hospitalizations

Study		Relative Risk (95% CI)	Weight, %
Cahir et al,[23] 2014		1.08 (0.85-1.38)	25.18
Fillenbaum et al,[26] 2004		1.20 (1.04-1.39)	70.26
Wallace et al,[29] 2017		0.72 (0.41-1.28)	4.56
Overall (I-squared = 37.0%, P = 0.204		1.14 (1.01-1.29)	100.00

.41　　1　　2.44

C. Mortality

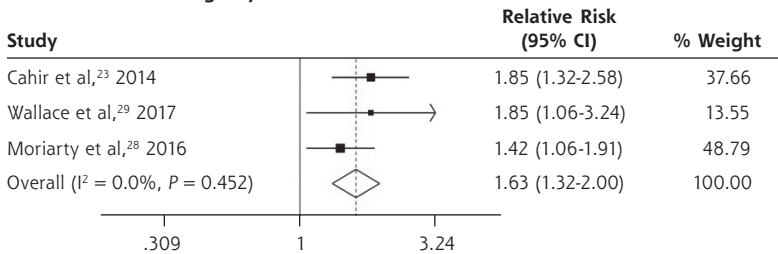| Study		Relative Risk (95% CI)	Weight, %
Barnett et al,[27] 2011		0.98 (0.92-1.05)	88.66
Hanlon et al,[25] 2002		1.02 (0.85-1.23)	11.34
Overall (I-squared = 0.0%, P = 0.689		0.98 (0.93-1.05)	100.00

.813　　1　　1.23

to the treatment regimen even in the presence of PIP. They may be less affected by PIP and are less likely to die from PIP. In contrast, those who have poorer health would be more likely to have discontinued PIP because of poor tolerance and may not be captured by the prevalent-user design at a specific time point. Although these people may have a higher risk of dying from PIP, the exclusion of them in a prevalent-user design can bias the result toward null effect, which may possibly explain the nonsignificant finding on mortality in the current meta-analysis.
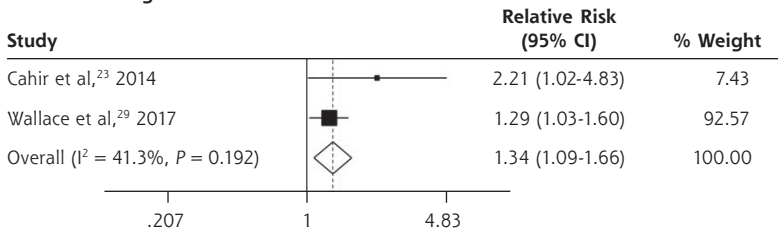
Limitations

Several limitations should be noted before the implications of the current findings are considered. First, we included only studies reported in the English language and may have missed other relevant evidence in the

Figure 3. Forest plots for the adverse outcomes of potentially inappropriate prescribing based on the STOPP (Screening Tool of Older Persons' Potentially Inappropriate Prescriptions) criteria.
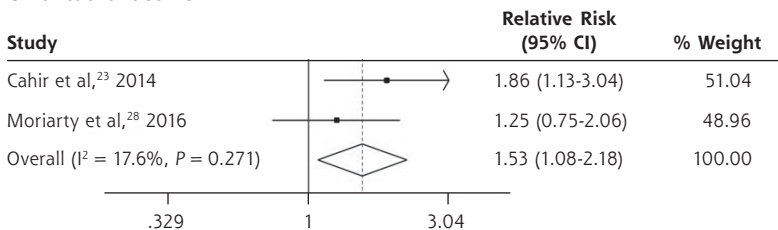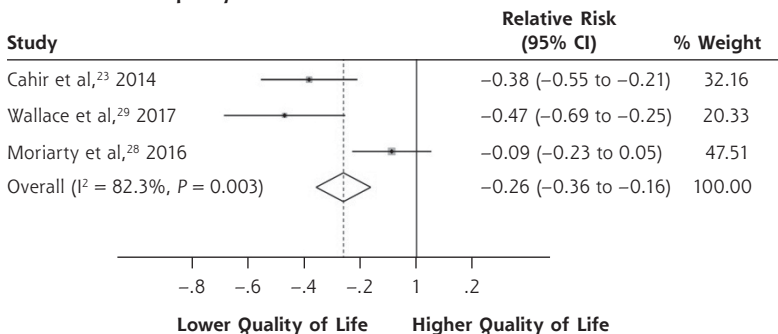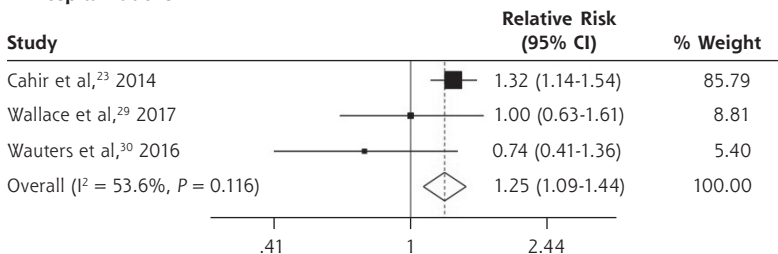
A. Accident and emergency visits

Study		Relative Risk (95% CI)	% Weight
Cahir et al,[23] 2014		1.85 (1.32-2.58)	37.66
Wallace et al,[29] 2017		1.85 (1.06-3.24)	13.55
Moriarty et al,[28] 2016		1.42 (1.06-1.91)	48.79
Overall ($I^2 = 0.0\%$, $P = 0.452$)		1.63 (1.32-2.00)	100.00

.309 1 3.24

B. Adverse drug events

Study		Relative Risk (95% CI)	% Weight
Cahir et al,[23] 2014		2.21 (1.02-4.83)	7.43
Wallace et al,[29] 2017		1.29 (1.03-1.60)	92.57
Overall ($I^2 = 41.3\%$, $P = 0.192$)		1.34 (1.09-1.66)	100.00

.207 1 4.83

C. Functional decline

Study		Relative Risk (95% CI)	% Weight
Cahir et al,[23] 2014		1.86 (1.13-3.04)	51.04
Moriarty et al,[28] 2016		1.25 (0.75-2.06)	48.96
Overall ($I^2 = 17.6\%$, $P = 0.271$)		1.53 (1.08-2.18)	100.00

.329 1 3.04

D. Health-related quality of life

Study		Relative Risk (95% CI)	% Weight
Cahir et al,[23] 2014		−0.38 (−0.55 to −0.21)	32.16
Wallace et al,[29] 2017		−0.47 (−0.69 to −0.25)	20.33
Moriarty et al,[28] 2016		−0.09 (−0.23 to 0.05)	47.51
Overall ($I^2 = 82.3\%$, $P = 0.003$)		−0.26 (−0.36 to −0.16)	100.00

−.8 −.6 −.4 −.2 1 .2

Lower Quality of Life   **Higher Quality of Life**

E. Hospitalizations

Study		Relative Risk (95% CI)	% Weight
Cahir et al,[23] 2014		1.32 (1.14-1.54)	85.79
Wallace et al,[29] 2017		1.00 (0.63-1.61)	8.81
Wauters et al,[30] 2016		0.74 (0.41-1.36)	5.40
Overall ($I^2 = 53.6\%$, $P = 0.116$)		1.25 (1.09-1.44)	100.00

.41 1 2.44

A&E = accident and emergency department; HRQoL = health-related quality of life.

non-English literature. Second, the limited number of included studies did not allow further investigations of individual- or study-level associations. Third, we did not include studies of PIP that are specific to certain drug classes (eg, analgesics or antibiotics) and therefore may not draw conclusions about the outcomes associated with specific classes of drugs. Fourth, although we also identified studies of PIP related to underprescribing (eg, those using the START criteria), we could not pool results in the meta-analysis because of the limited number of available studies related to underprescribing (Supplementary Appendix), which remains an area for further research in future. Fifth, there were some variations in the magnitude of pooled estimates between the Beers and STOPP criteria, which were probably related to the different drugs included in the respective criteria (some of which do not necessarily apply to the different populations).[1] Sixth, the meta-analytic results were based on observational studies, which may not have sufficiently controlled for all relevant confounders. Therefore, the findings may not allow definitive conclusions about the causal role between PIP and the adverse outcomes.

Clinical Implications

This study demonstrated the associations between PIP and a wide range of adverse outcomes and highlighted the relevance of PIP among older persons in primary care. Essentially, the findings showed that the construct of PIP is more than just a consensus of good clinical practice, and they underscored the need to focus on PIP in primary care to improve patient outcomes. A variety of interventions have been evaluated to address PIP among community-dwelling older persons in previous

Table 3. Summary of Key Findings and GRADE Assessment

Outcomes	Number of Participants	Pooled Effect Size (95% CI)	Certainty of Evidence (GRADE)	GRADE Assessment			
				Risk of Bias[a]	Inconsistency[b]	Imprecision[c]/ Publication Bias[d]/ Indirectness[e]	Large Effect[f]/ Dose Response[g]
PIP based on the Beers Criteria							
Functional decline	4,165	RR 1.38 (1.06-1.80)	●●○○○ Low	No downgrade (NOS = 9)	No downgrade (I^2 = 29.5%, P = .234)	No downgrade	No upgrade
Hospitalizations	5,069	RR 1.14 (1.01-1.29)	●●○○○ Low	No downgrade (NOS = 9)	No downgrade (I^2 = 37.0%, P = .204)	No downgrade	No upgrade
Mortality	73,533	RR 0.98 (0.93-1.05)	●●○○○ Low	No downgrade (NOS = 9)	No downgrade (I^2 = 0.0%, P = .689)	No downgrade	No upgrade
PIP based on the STOPP criteria							
A&E visits	3,588	RR 1.63 (1.32-2.00)	●●○○○ Low	No downgrade (NOS = 9)	No downgrade (I^2 = 0.0%, P = .452)	No downgrade	No upgrade
ADEs	1,835	RR 1.34 (1.09-1.66)	●●○○○ Low	No downgrade (NOS = 9)	No downgrade (I^2 = 41.3%, P = .192)	No downgrade	No upgrade
Functional decline	2,684	RR 1.53 (1.08-2.18)	●●○○○ Low	No downgrade (NOS = 9)	No downgrade (I^2 = 17.6%, P = .271)	No downgrade	No upgrade
HRQoL	3,588	SMD −0.26 (−0.36 to −0.16)	●○○○○ Very low	No downgrade (NOS = 9)	Downgrade (I^2 = 82.3%, P = .003)	No downgrade	No upgrade
Hospitalizations	2,338	RR 1.25 (1.09-1.44)	●●○○○ Low	No downgrade (NOS = 8)	No downgrade (I^2 = 53.6%, P = .116)	No downgrade	No upgrade

A&E = accident and emergency department; ADE = adverse drug event; GRADE = Grading of Recommendations, Assessment, Development and Evaluations; HRQoL = health-related quality of life; NOS = Newcastle-Ottawa Scale; PIP = potentially inappropriate prescribing; RR = relative risk; SMD = standardized mean difference; STOPP = Screening Tool of Older Persons' Potentially Inappropriate Prescriptions.

[a] We downgraded the GRADE assessment if the risk of bias assessment based on the NOS is <8 in at least one of the studies, suggesting the presence of risk of bias.

[b] We downgraded the GRADE assessment if the Q test P < 0.10 or the I^2 > 75%, indicating significantly high levels of heterogeneity in the results.

[c] For RR, we considered a clinically meaningful threshold to be 0.90 or 1.10 and downgraded the GRADE assessment if the RR point estimate is ≥1 and the lower limit of its CI is <0.90, or if the RR point estimate is <1 and the upper limit of its CI is >1.10. For SMD, we considered a clinically meaningful threshold to be ±0.20 and downgraded the GRADE assessment if the point estimate is ≥0 and the lower limit of its CI is <−0.20, or if the point estimate is <0 and the upper limit of its CI is >0.20.

[d] We could not assess for publication bias because there were <10 studies for each of the outcomes. Therefore, we did not downgrade any of the GRADE assessments due to publication bias.

[e] We downgraded the GRADE assessment if the recruited participants were not representative of older persons in the primary care settings.

[f] We upgraded the GRADE assessment if the RR is >2 or <0.5.

[g] We upgraded the GRADE assessment in the presence of dose-response gradient, which provides stronger evidence of the cause-effect relationship.

randomized controlled trials (RCTs), some of which may be adopted in routine practice. These interventions can be broadly classified into organizational, professional, or multifaceted interventions.[33] Organizational interventions focus on changing the delivery of health care services (eg, by including medication reviews by pharmacists), professional interventions aim to improve the practice of health care professionals (eg, with the use of computerized clinical decision support systems [CCDSSs]), and multifaceted interventions may involve combinations of organizational or profes-

sional interventions. Most of the interventions have been shown to be useful in the literature, with some of them (eg, medication reviews, CCDSSs, and multifaceted interventions) having had consistent evidence of effectiveness in reducing PIP.[33]

Implications for Future Research

Notwithstanding the promising benefits of interventions, more research is needed to evaluate the effects of interventions on clinically relevant outcomes. Prior RCTs have infrequently reported on the clinical out-

comes of interventions, and the few that have done so had mixed results for HRQoL, hospitalizations, A&E visits, and medication costs.[33] Such research is especially relevant in light of the current meta-analytic findings, to further confirm that the reduction of PIP can have a direct effect in improving outcomes and that PIP has a direct causal relationship with the adverse outcomes. Ideally, future RCTs should be adequately powered and have a sufficiently long follow-up period so that any difference in the outcomes can be captured sensitively.[33] Inasmuch as intervention studies are relevant, further research is also needed in the area of implementation science, to evaluate how PIP interventions can be optimally integrated into routine clinical practice, especially in primary health care settings where the patient load can be high and changes can be difficult to implement or sustain.

In conducting future studies, researchers should be mindful of how PIP is operationalized and of the potential implications associated with the choice of PIP criteria. Researchers may consider capturing PIP based on the new-user design (which includes those with new onset of PIP)—instead of the conventional approach of PIP based on prevalent-user design (which identifies those with PIP only at a single time point)—to improve the predictive validity of PIP and to identify patients who can receive a greater benefit from PIP interventions. When selecting a screening tool for PIP, researchers may need to consider the applicability of individual drugs within the tool and possibly adapt the tool to fit the local prescribing practice. Researchers may also need to be aware that different tools may capture different aspects of PIP. Although most available tools focus on overprescribing or misprescribing of medications, it may be equally important to include tools related to underprescribing of medications (such as the START) to capture the whole spectrum of PIP.

CONCLUSION

This is the first meta-analysis to consolidate the quantitative evidence on the wide-ranging impact of PIP among older persons in primary care. The outcome of mortality was the only nonsignificant finding, possibly explained by the recruitment of healthier participants, which may have biased the results toward null effect. The findings highlight the need to address PIP in primary care and call for further research on PIP interventions in primary care. They also point to the need for researchers to consider the potential implications of how PIP is operationalized when designing future research on PIP.

To read or post commentaries in response to this article, see it online at http://www.AnnFamMed.org/content/17/3/257.

Key words: primary health care; general practice; general practitioners; family practice; family physicians; inappropriate prescribing; medication errors; aged; adverse outcomes; systematic review; meta-analysis

References

1. O'Connor MN, Gallagher P, O'Mahony D. Inappropriate prescribing: criteria, detection and prevention. *Drugs Aging.* 2012;29(6): 437-452.

2. Kaufmann CP, Tremp R, Hersberger KE, Lampert ML. Inappropriate prescribing: a systematic overview of published assessment tools. *Eur J Clin Pharmacol.* 2014;70(1):1-11.

3. 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.

4. Tommelein E, Mehuys E, Petrovic M, Somers A, Colin P, Boussery K. Potentially inappropriate prescribing in community-dwelling older people across Europe: a systematic literature review. *Eur J Clin Pharmacol.* 2015;71(12):1415-1427.

5. Morin L, Laroche ML, Texier G, Johnell K. Prevalence of potentially inappropriate medication use in older adults living in nursing homes: a systematic review. *J Am Med Dir Assoc.* 2016;17(9):862. e1-862.e9.

6. Hyttinen V, Jyrkkä J, Valtonen H. A systematic review of the impact of potentially inappropriate medication on health care utilization and costs among older adults. *Med Care.* 2016;54(10):950-964.

7. Tesfaye WH, Castelino RL, Wimmer BC, Zaidi STR. Inappropriate prescribing in chronic kidney disease: a systematic review of prevalence, associated clinical outcomes and impact of interventions. *Int J Clin Pract.* 2017;71(7).

8. 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.

9. World Health Organization. *Global Health and Aging.* Geneva, Switzerland: World Health Organization; 2011.

10. Opondo D, Eslami S, Visscher S, et al. Inappropriateness of medication prescriptions to elderly patients in the primary care setting: a systematic review. *PLoS One.* 2012;7(8):e43617.

11. World Health Organization. *Medication Errors: Technical Series on Safer Primary Care.* Geneva, Switzerland: World Health Organization; 2016.

12. Matthews FE, Bennett H, Wittenberg R, Jagger C, Dening T, Brayne C; Cognitive Function, Ageing Studies (CFAS) Collaboration. Who lives where and does it matter? Changes in the health profiles of older people living in long term care and the community over two decades in a high income country. *PLoS One.* 2016;11(9):e0161705.

13. Moher D, Shamseer L, Clarke M, et al.; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4:1.

14. Lee CS, Liew TM. Inappropriate prescribing among older persons in primary care: protocol for systematic review and meta-analysis of observational studies. *BMJ Open.* 2017;7(2):e015395.

15. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm. Accessed Oct 6, 2016.

16. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22(4):719-748.

17. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc.* 2009;172(1):137-159.

18. Guyatt G. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice.* 3rd ed. New York, NY: McGraw-Hill Education; 2014.

19. Thomas DC. *Statistical Methods in Environmental Epidemiology.* Oxford, UK: Oxford University Press; 2009.

20. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* 2003;327(7414):557-560.

21. Higgins JPT. *Cochrane Handbook for Systematic Reviews of Interventions.* Chichester, UK: Wiley; 2008.

22. Schunemann HBJ, Guyatt G, Oxman A. Handbook for grading quality of evidence and strength of recommendations using the GRADE approach. 2013. http://gdt.guidelinedevelopment.org/app/. Accessed Mar 1, 2018.

23. Cahir C, Moriarty F, Teljeur C, Fahey T, Bennett K. Potentially inappropriate prescribing and vulnerability and hospitalization in older community-dwelling patients. *Ann Pharmacother.* 2014;48(12):1546-1554.

24. Cahir C, Bennett K, Teljeur C, Fahey T. Potentially inappropriate prescribing and adverse health outcomes in community dwelling older patients. *Br J Clin Pharmacol.* 2014;77(1):201-210.

25. Hanlon JT, Fillenbaum GG, Kuchibhatla M, et al. Impact of inappropriate drug use on mortality and functional status in representative community dwelling elders. *Med Care.* 2002;40(2):166-176.

26. Fillenbaum GG, Hanlon JT, Landerman LR, et al. Impact of inappropriate drug use on health services utilization among representative older community-dwelling residents. *Am J Geriatr Pharmacother.* 2004;2(2):92-101.

27. Barnett K, McCowan C, Evans JM, Gillespie ND, Davey PG, Fahey T. Prevalence and outcomes of use of potentially inappropriate medicines in older people: cohort study stratified by residence in nursing home or in the community. *BMJ Qual Saf.* 2011;20(3):275-281.

28. 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;82(3):849-857.

29. 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 A Biol Sci Med Sci.* 2017;72(2):271-277.

30. Wauters M, Elseviers M, Vaes B, et al. Too many, too few, or too unsafe? Impact of inappropriate prescribing on mortality, and hospitalization in a cohort of community-dwelling oldest old. *Br J Clin Pharmacol.* 2016;82(5):1382-1392.

31. Wang SV, Gagne JJ, Glynn RJ, Schneeweiss S. Case-crossover studies of therapeutics: design approaches to addressing time-varying prognosis in elderly populations. *Epidemiology.* 2013;24(3):375-378.

32. Glynn RJ, Knight EL, Levin R, Avorn J. Paradoxical relations of drug treatment with mortality in older persons. *Epidemiology.* 2001;12(6):682-689.

33. 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.