Supplemental materials for:

Liew T, Lee C, Goh S, Chang Z. Potentially inappropriate prescribing among older persons: a meta-analysis of observational studies. *Ann Fam Med*. 2019;17(3):257-266.

Supplementary appendix

The adverse outcomes of potentially inappropriate prescribing among older persons in primary care: a meta-analysis of observational studies

Tau Ming Liew, Cia Sin Lee, Shawn Kuan Liang Goh, Zi Ying Chang

Table of content

Section 1. A sample of the search strategy based on PubMed (MeSH, Medical Subject Headings), with no time restriction imposed to the publication year of the studies.

1	"Primary Health Care"[MeSH]
2	“general practice”[MeSH]
3	“general practitioners”[MeSH]
4	"Family Practice"[MeSH]
5	"Physicians, Family"[MeSH]
6	primary[title/abstract] AND care[title/abstract]
7	general[title/abstract] AND practice*[title/abstract]
8	general[title/abstract] AND practitioner*[title/abstract]
9	family[title/abstract] AND practice*[title/abstract]
10	family[title/abstract] AND physician*[title/abstract]
11	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12	"Aged"[MeSH]
13	elder*[title/abstract]
14	older[title/abstract]
15	Geriatric[title/abstract]
16	12 or 13 or 14 or 15
17	"Inappropriate Prescribing"[MeSH]
18	"Medication Errors"[MeSH]
19	"Inappropriate Prescribing"[title/abstract] OR "Inappropriate Prescription"[title/abstract] OR "Inappropriate medication"[title/abstract] OR "Inappropriate drug"[title/abstract] OR "Inappropriate medicine"[title/abstract] OR over-prescribing[title/abstract] OR over-prescription[title/abstract]
20	17 or 18 or 19
21	11 and 16 and 20

Section 2. Detailed results of the risk of bias assessment using the Newcastle-Ottawa Scale

	Newcastle-Ottawa Scale	Studies					
		Barnett (2011)	Cahir (2014)	Hanlon (2002)	Moriarty (2016)	Wallace (2016)	Wauters (2016)
Selection	Representative of the exposed cohort 1. The exposed cohort truly representative of older persons in primary care (whole population registry, or probability sampling with >70% response rate) ❋ 2. The exposed cohort somewhat representative of older persons in primary care (probability sampling with <70% response rate) ❋ 3. The exposed cohort is representative of selected group of users such as nurses, volunteers 4. No description of the derivation of the cohort	1	2	2	1	1	2
	Selection of the non-exposed cohort 1. The non-exposed cohort drawn from the same community as the exposed cohort ❋ 2. The non-exposed cohort drawn from different source 3. No description of the non-exposed cohort	1	1	1	1	1	1
	Ascertainment of exposure 1. From secure record (example, medical records) ❋ 2. From structured interview ❋ 3. From written self-report 4. No description	1	1	2	1	1	1
	Demonstration that outcome of interest was not present at the start of study 1. Yes ❋ 2. No	1	1	1	1	1	1
Comparability	Comparability of cohorts on the basis of the design or analysis: The study control for 1-2 factors (example age and gender) 1. Yes ❋ 2. No	1	1	1	1	1	1
	Comparability of cohorts on the basis of the design or analysis: The study controls for any additional factor(s) 1. Yes ❋ 2. No	1	1	1	1	1	2
Outcome	Assessment of the primary outcome 1. Independent blind assessment ❋ 2. Record linkage ❋ 3. Self-report 4. No description	2	1	2	1	1	2
	Was follow-up long enough for the outcomes to occur 1. Yes (at least 6 months for the follow up period) ❋ 2. No	1	1	1	1	1	1
	Adequacy of follow-up cohorts 1. Complete follow up in which all subjects accounted for ❋ 2. Subjects lost to follow up unlikely to introduce bias such as >70% follow up rate or description provided of those lost ❋ 3. Follow up rate <70% and no description of those lost 4. No statement	1	1	2	2	2	2

Section 3. Assessment of the quality of evidence using GRADE

We used the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) framework to classify the overall certainty of evidence into one of four levels – high, moderate, low or very low. For meta-analytic results based on observational studies, the certainty of evidence was first assumed to be low and then downgraded further in the presence of identified problems in five key domains as described below .[1]

(1) Risk of bias – We downgraded the GRADE assessment if the risk of bias assessment based on Newcastle-Ottawa Scale was <8 in at least one of the studies, suggesting the presence of risk of bias.

(2) Inconsistency – We downgraded the GRADE assessment if the Q test p-value<0.10 or the I^2>75%, indicating significantly high levels of heterogeneity in the results.[2]

(3) Imprecision of effect estimates – For relative risk (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 or more and the lower limit of its confidence interval is below 0.90; or if the RR point estimate is less than 1 and the upper limit of its confidence interval is above 1.10. For standardized mean difference (SMD), we considered a clinically meaningful threshold to be ±0.20 and downgraded the GRADE assessment if the point estimate is 0 or more and the lower limit of its confidence interval is below -0.20; or if the point estimate is less than 0 and the upper limit of its confidence interval is above +0.20.

(4) Risk of publication bias – We could not assess for publication bias as there were <10 studies for each of the outcomes. Hence, we did not downgrade any of the GRADE assessment due to publication bias.

(5) Indirectness of evidence – We downgraded the GRADE assessment if the recruited participants were not representative of older persons in the primary care setting.

Additionally, the certainty of evidence was upgraded in the presence of the following two criteria:

(1) Large magnitude of effect – We upgraded the GRADE assessment if the RR was >2 or <0.5.

(2) Dose-response gradient – We upgraded the GRADE assessment in the presence of a dose-response gradient, which provides stronger evidence of a cause-effect relationship.

References

1. Schunemann H BJ, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendation 2013. http://gdt.guidelinedevelopment.org/app/ (assessed 1 Mar 2018).

2. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed)* 2003; **327**(7414): 557-60.

Section 4. Results that had only been reported in one study and were not eligible for meta-analysis

Author (Year)	Outcome	Criteria of PIP	Effect estimate (95% CI)	Reason for not being included in the meta-analysis
SIGNIFICANT FINDINGS				
Hanlon (2002) & Fillenbaum (2004)	Functional decline	DUR	OR 1.36 (1.00, 1.83)	No other studies had defined PIP using the DUR criteria
Hanlon (2002) & Fillenbaum (2004)	Outpatient visits	DUR	Beta coefficient 0.82 (0.27, 1.37)	No other studies had reported on this outcome measure
Moriarty (2016)	GP Visits	STOPP	RR 1.16 (1.06, 1.28)	No other studies had reported on this outcome measure based on the STOPP criteria of PIP
Moriarty (2016)	A&E visits	START	RR 1.45 (1.03, 2.04)	No other studies had reported on this outcome measure based on the START criteria of PIP
Moriarty (2016)	GP Visits	START	RR 1.13 (1.01, 1.27)	No other studies had reported on this outcome measure based on the START criteria of PIP
Moriarty (2016)	Functional decline	START	OR 2.06 (1.25, 3.39)	No other studies had reported on this outcome measure based on the START criteria of PIP
Moriarty (2016)	HRQoL	START	Regression Coefficient -1.06 (-1.84, -0.27)	No other studies had reported on this outcome measure based on the START criteria of PIP
Wauters (2016)	Mortality	START	HR 2.91 (1.28, 6.61)	No other studies had reported on this outcome measure based on the START criteria of PIP
Wauters (2016)	Hospitalizations	START	HR 2.08 (1.29, 3.36)	No other studies had reported on this outcome measure based on the START criteria of PIP
NON-SIGNIFICANT FINDINGS				
Hanlon (2002) & Fillenbaum (2004)	Mortality	DUR	RR 0.85 (0.69, 1.24)	No other studies had defined PIP using the DUR criteria
Hanlon (2002) & Fillenbaum (2004)	Hospitalizations	DUR	RR 1.06 (0.90, 1.25)	No other studies had defined PIP using the DUR criteria
Hanlon (2002) & Fillenbaum (2004)	Nursing home entry	DUR	RR 1.06 (0.76, 1.47)	No other studies had reported on this outcome measure
Hanlon (2002) & Fillenbaum (2004)	Outpatient visits	Beers 1997	Beta coefficient 0.48 (-0.01, 0.97)	No other studies had reported on this outcome measure
Hanlon (2002) & Fillenbaum (2004)	Nursing home entry	Beers 1997	RR 0.93 (0.69, 1.08)	No other studies had reported on this outcome measure
Wallace (2016)	A&E visits	Beers 2012	OR 1.54 (0.88, 2.71)	No other studies had reported on this outcome measure based on the Beers criteria of PIP
Wallace (2016)	ADE	Beers 2012	RR 1.00 (0.78, 1.29)	No other studies had reported on this outcome measure based on the Beers criteria of PIP
Wallace (2016)	HRQoL	Beers 2012	MD -0.05 (-0.11, 0.003)	No other studies had reported on this outcome measure based on the Beers criteria of PIP
Wauters (2016)	Mortality	STOPP	HR 1.07 (0.36, 3.17)	No other studies had reported on this outcome measure based on the STOPP criteria of PIP

A&E, accident and emergency department; ADE, adverse drug events ; CI, confidence interval; DUR, drug utilization review; GP, General practitioner; HR, hazard ratio; HRQoL, health-related quality of life; MD, mean difference; OR, odds ratio; PIP, potentially inappropriate prescribing; START, Screening tool to alert right treatment; STOPP, screening tool of older persons' potentially inappropriate prescriptions; RR, relative risk.