European Geriatric Medicine
https://doi.org/10.1007/s41999-025-01386-7

RESEARCH PAPER

Appraisal of the references supporting the STOPP/START.version 3 criteria

B. Boland[1,2] · F. X. Sibille[2,4,5] · A. Mouzon[3] · S. Marien[1,2] · F. Vaillant[6] · A. Spinewine[2,3,5] · O. Dalleur[2,5,6]

Received: 16 July 2025 / Revised: 6 November 2025 / Accepted: 11 December 2025

Key summary points

Aim To describe the strength of evidence of the references provided for the STOPP/START version 3 criteria.

Findings Of the 454 references, 322 are reviews (71%), but 164 are narrative ones (36%: STOPP 41%, START 26%) of which 138 are not based on RCTs. Of the 190 criteria, 107 (56%) are based on RCTs (STOPP 40%, START 84%). Ten of the 133 STOPP criteria are not supported by an appropriate reference.

Message Clinicians should be aware that a substantial proportion of the references are narrative reviews (Fig. 1), most of which not presenting high-level evidence (RCTs).

At the criterion level (Table 2), most START criteria are supported by high-level evidence (RCTs), whereas fewer than half of the STOPP criteria are underpinned by such evidence. The level of evidence supporting a criterion (Appendix 1) shoud be considered by the clinicians when sharing decision-making with older adults.

Abstract

Background The strength of evidence underlying the STOPP/START version 3 (SS.v3) criteria has not been reported and remains controversial, potentially limiting their clinical implementation. This study aimed at appraising the references provided for the SS.v3 criteria.

Methods All references cited for the 190 SS.v3 criteria were evaluated and discussed by a clinical pharmacist–geriatrician pair and a geriatrician expert in literature appraisal who reached agreement. The main endpoint was the evidence level classified by data study design (level I: randomised controlled trials; II: cohort studies; III: case–control; IV: case series; V: expert opinion). References were classified as inappropriate if they showed mismatch, lacked supporting content, or were clinically discordant with the criterion.

Results Of the 454 SS.v3 references, 322 are reviews (STOPP 70%, START 71%), either systematic (STOPP 29%, START 44%) or narrative (STOPP 41%, START 26%). Evidence is level I (STOPP 28%, START 64%), level II (STOPP 13%, START

✉ B. Boland
Benoit.Boland@uclouvain.be

1 Geriatric Medicine, Cliniques Universitaires Saint-Luc, Brussels, Belgium

2 Groupe de Recherche en Gériatrie et Gérontologie, Institut de Recherche Santé Et Société (IRSS), Université Catholique de Louvain, Brussels, Belgium

3 Pharmacy Department, Centre Hospitalier Universitaire UCL-Namur, Yvoir, Belgium

4 Geriatric Medicine, Centre Hospitalier Universitaire UCL-Namur, Yvoir, Belgium

5 Clinical Pharmacy and Pharmacoepidemiology Research Group, Louvain Drug Research Institute (LDRI), UCLouvain, Brussels, Belgium

6 Pharmacy Department, Cliniques Universitaires Saint-Luc, Brussels, Belgium

Published online: 26 December 2025

8%), levels III–IV–V (STOPP 45%, START 19%), or downgraded because of inappropriateness (STOPP 15%, START 8%). At the level of the 190 criteria, the best evidence is level I (STOPP 43%, START 88%), level II (STOPP 13%, START 7%), levels III–IV–V (STOPP 35%, START 4%), while 10 criteria (STOPP 8%, START 0%) lack appropriate supporting reference.
Conclusions Fewer than half of the STOPP criteria are supported by high-level evidence, unlike most START criteria. A reassessment of many STOPP criteria and their references is warranted. Clinicians should be aware of these limitations when involving in shared decision-making with older patients.

Keywords Older adults · Medications · Potential inappropriate prescribing · STOPP/START version 3 · Publication type · Evidencelevel

Introduction

The third version of the STOPP/START criteria (SS.v3) for potentially inappropriate prescribing in older people was eagerly expected by many clinicians in Europe and worldwide. Their publication in 2023 in European Geriatric Medicine [1] was accompanied by positive commentaries from colleagues from North America, Asia and Oceania, who also pointed to some limitations, for example, the lack of a list of the new additions, the difficulty to apply 190 criteria in daily practice, and their non-generalizability to frail older adults [2–4].

Importantly, criticisms were also formulated related to the lack of information on the underlying literature data and the evidence levels of the provided references [5, 6]. These latter aspects are important to support the clinician's understanding of the SS.v3 criteria strength and the shared decision making of (de)prescribing with the patients and their relatives as it is based on the communication of the most accurate benefit–harm estimation [7]. Despite the answers provided [8, 9], these aspects remain insufficiently documented, which may hamper the proper implementation of SS.v3 criteria in clinical practice.

This study is a critical appraisal of the references supporting the SS.v3 criteria, primarily in terms of evidence level and relevance to older adults, to better inform the clinicians involved in medication review for older adults.

Methods

This study is a comprehensive critical appraisal of the references supporting the 133 STOPP and 57 START version 3 criteria. The 308 STOPP and 160 START references were evaluated and discussed by a pair of clinician experts in pharmacotherapy (a geriatrician and a clinical pharmacist involved in geriatric care at two teaching hospitals from UCLouvain and trained in evidence-based practice, each pair appraising one-third of the references) and by the main investigator (geriatrician with expertise in evaluation of the medical literature), who reached agreement. Although the original plan was for independent assessments followed by end-of-process discussions, pilot testing revealed the process's complexity. Consequently, iterative discussions between reviewer pairs and the main investigator were held throughout the appraisal.

An Excel file was used to collect the following pieces of information for each reference: the criterion message (A1 to M1 for STOPP.v3 and A1 to L4 for START.v3), the first author, Journal, publication year, the publication type (guideline, systematic review, narrative review, original article), a focus on older adults (patient's mean/median age ≥ 65 years or report on a subgroup ≥ 65 years) and the number of PubMed-indexed citations (on May 31st, 2024). Due to unavailability of these pieces of information, the 13 references from the British National Formulary (STOPP 13, START 0) were removed, as well as 1 duplicated reference (START B1 iv and v), leaving 454 references (STOPP 295, START 159) for analysis. The retracted publication (START H2 iii) was kept in the analysis.

The evidence level (main endpoint) of each reference was determined according to the study design of the original studies (randomised controlled trials, cohort studies, case–control studies, cases series, register) underlying the data related to the criterion message. For references that were guidelines or reviews, we extracted information on the study design of the original articles (cited within these sources) that were relevant to the criterion. The evidence level was reported as follows: Ia (systematic review of randomised controlled trials ± meta-analysis, or ≥ 3 concordant randomised controlled trials), Ib (1–2 randomised controlled trial(s)), IIa (systematic review of cohorts ± meta-analysis, or ≥ 3 concordant cohorts), IIb (1–2 cohort(s)), IIIa (systematic review of case–control studies), IIIb (1–2 case–control studies), IV (case series, registers, surveys), V (expert opinion, that is no evidence-based data found in the article). References were downgraded to an inappropriate status under three circumstances: a major mismatch between the article and its associated criterion (in terms of patient population, medication, or clinical event); no relevant message found in the reference to support the criterion; or a clinically discrepant message between the reference and the criterion.

Analyses were conducted at both the reference level and the criterion level. At the level of the 454 references, we

Table 1 Appraisal of the references provided for the STOPP/START.v3 criteria

	All references $n=454$	STOPP references $n=295$	START references $n=159$
Publication type			
Guidelines	30 (7%)	14 (5%)	16 (9%)
Systematic reviews/MA	158 (35%)	87 (29%)	71 (44%)
Narrative reviews	164 (36%)	122 (41%)	42 (26%)
Original articles	102 (23%)	72 (25%)	30 (19%)
Randomized trials (RCTs)	37 (8%)	19 (6%)	18 (11%)
Observational studies	65 (14%)	53 (18%)	12 (8%)
Older patients			
Focus on ≥ 65 years	202 (44%)	136 (46%)	66 (42%)
Word in the title	110 (24%)	86 (29%)	24 (15%)
PubMed citations, per year			
Median	2.8 [0.9; 8.3]	2.1 [0.8; 6.5]	4.4 [1.5; 14.5]
< 2 citations/year	184 (41%)	137 (46%)	46 (29%)
Publication year			
Median [P_{25}; P_{75}]	2009 [2005,2014]	2009 [2005,2013]	2011 [2006,2016]
Range	1968–2023	1968–2023	1987–2022
Evidence Level *			
High (level I)	185 (41%)	82 (28%)	103 (64%)
Intermediate (level II)	49 (11%)	37 (13%)	12 (8%)
Low (levels III–IV–V)	163 (36%)	132 (45%)	31 (19%)
III	17 (4%)	17 (5%)	0 (0%)
IV	44 (10%)	41 (14%)	3 (2%)
V	102 (22%)	74 (25%)	28 (18%)
Inappropriate reference °	57 (13%)	44 (15%)	13 (8%)
Mismatch in design	23 (5%)	16 (5%)	7 (4%)
No such message	20 (4%)	16 (5%)	4 (3%)
Discordant message	14 (3%)	12 (4%)	2 (1%)

Abbreviations MA: Meta-analysis. RCT: Randomised controlled trial. SR: Systematic review

* Evidence levels of the original data I: randomised controlled trial(s): Ia: SR ± MA, or ≥ 3 RCTs; Ib: 1–2 RCTs II: Cohort(s): IIa SR ± MA, or ≥ 3 concording cohorts; IIb: 1–2 cohort(s) III: Case–control studies IV: Case series, registers, surveys V: Expert opinion, with no provided evidence-based data

° Inappropriateness: downgraded reference (due to major mismatch between the reference and its criterion, no message found in the reference to support the criterion, or a clinically discrepant message between the reference and its criterion)

reported the scientific evidence distinguishing a) the STOPP and the START criteria, b) the publication types (guidelines, systematic reviews/meta-analyses, narrative reviews, original articles), and c) the new and the previously existing criteria as identified in our recent publication [10]. At the level of each of the 190 criteria, we identified the best reference as the one with the highest evidence level (randomised controlled trial > cohort > case–control > case series > no data). When several references provided for the criterion had the same highest level, the reference with the more citations per year in PubMed was selected. For criteria lacking an appropriate reference, we conducted a literature search to test the feasibility of identification a relevant article available at the time of SS.v3 development (that is, up to 2022).

Continuous data were expressed as means ± standard deviation or medians [interquartile range]. Categorical data were expressed as number of participants and percentages. As this study is a qualitative assessment, we chose not to report the *p* values of the observed differences.

Results

Analysis at the reference level

Table 1 presents the appraisal of the 454 references provided for the STOPP/START criteria. The most frequent publication types are narrative reviews (36%) and systematic reviews (35%), followed by original studies (22%) and

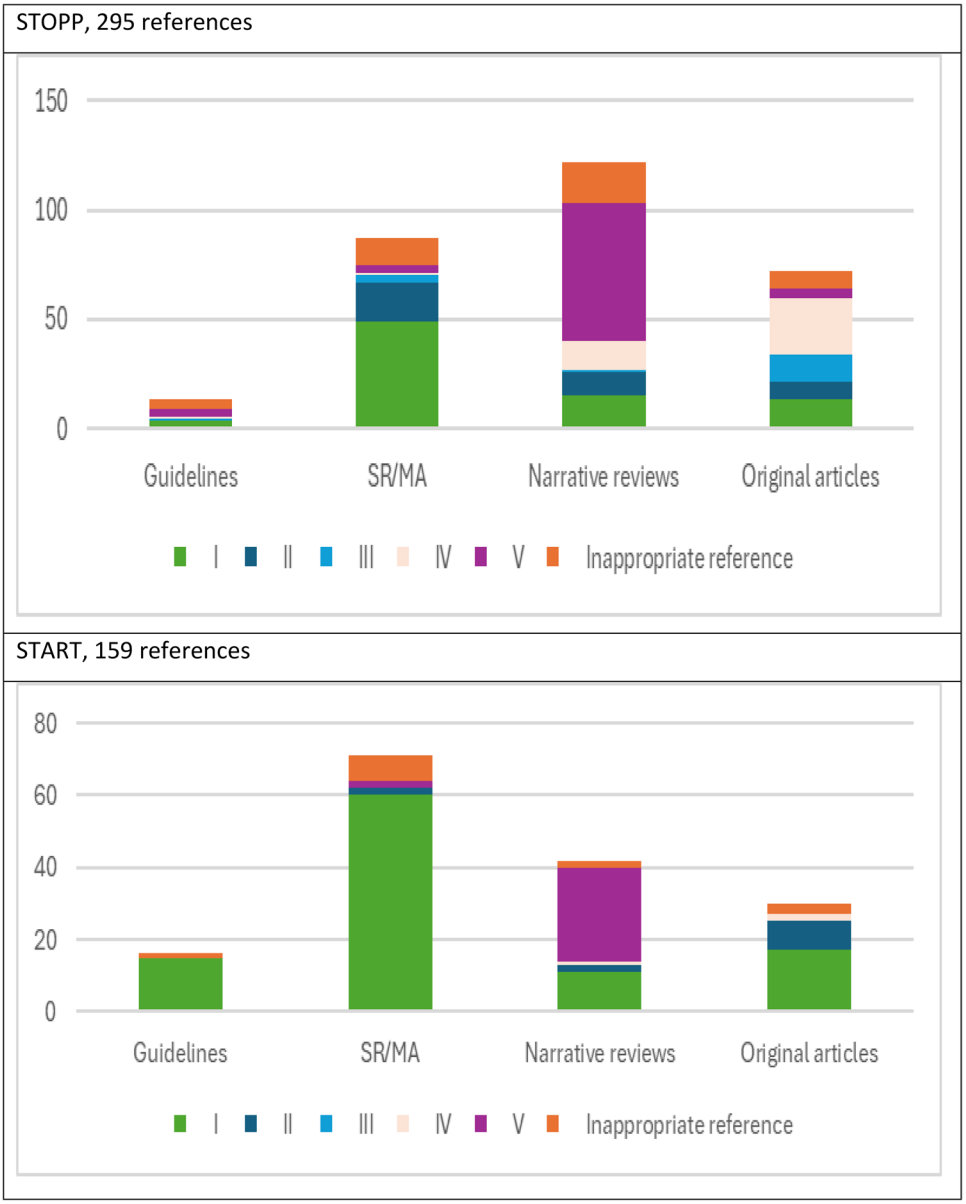

Fig. 1 Evidence levels of the STOPP and START references according to publication types. Abbreviations: SR: Systematic review. MA: Meta-analysis Evidence levels of the original data: I: Randomised controlled trial(s), in green II: Cohort(s), in dark blue III: Case–control, in light blue IV: Cases (series, registers, surveys), in pink V: Expert opinion, with no evidence-based data found in the reference, in purple Inappropriate reference (see methods), in orange

guidelines (7%). A focus on older adults (≥ 65 years) is present 202 references (44%). Comparing STOPP and START references, we found that reviews are of similar frequency (STOPP 70%, START 71%) but different design (systematic reviews: STOPP 29%, START 44%; narrative reviews: STOPP 41%, START 26%). The proportion of references with a < 2 PubMed citations per year was also markedly different (STOPP 46%, START 29%).

Regarding evidence level (main endpoint) of the data, 185 references are randomised controlled trials (STOPP 28%, START 64%), 49 are cohorts (STOPP 13%, START 8%), and 163 are case–control, case series or expert opinion (STOPP 45%, START 19%). The 57 remaining references (13%) were considered as inappropriate (STOPP 15%, START 8%), due to a major mismatch between the reference and its criterion, no message found in the reference to support the criterion, or a clinically discrepant message between the reference and its criterion (Table 1, bottom).

We further analysed the evidence level according to the publication type. High evidence (level I) was found in 19 guidelines (63%), 109 systematic reviews/meta-analyses (69%), 26 narrative reviews (16%) and 31 original RCTs (84%). Low evidence (levels III–IV–V) was detected in 5 guidelines (17%), 10 systematic reviews/meta-analyses (6%), 104 narrative reviews (63%), 1 original RCT (3%) and 43 original observational studies (66%). Inappropriateness ($n = 57$) concerned 6 guidelines (20%), 19 systematic reviews/meta-analyses (12%), 21 narrative reviews (13%), 4 original RCTs (11%) and 7 original observational studies (11%). Figure 1 illustrates the evidence level distribution

Table 2 Best evidence level * and its focus on older adults + of the 190 STOPP/START v 3 criteria (The criteria are listed in the Appendix 1 of the publication: STOPP/START criteria for potentially inappropriate prescribing in older people: version 3 | European Geriatric Medicine

	I a	I b	II a	II b	III	IV	V	Inapp°
The 133 STOPP.v3 criteria								
Section	$n=46$	$n=11$	$n=7$	$n=10$	$n=3$	$n=21$	$n=23$	$n=10$
A. Indication	–	–	–	–	–	**3**	–	–
B. Cardiovascular	3,**5**,6,16, 17,19,21	11,**12**, 13,14	**15**,18	**9**	–	**3**,**10**,**20**	2,4,7,**8**	1
C. Coagulation	**3**,**7**,**11**, 15,16	–	12	**2**,**4**,**13**	**10, 14**	–	5,6	1,8,9
D. Central nervous	1**,2**,3,**5**,**8**, 9,10,**11, 12**, **15**,**20**	–	7	**6**	13	4,**17**,**18**, 19,23,24	14,**16**, **21**,22,25	–
E. Renal	–	**1**	–	–	–	2,3,8,10	4,**5**,**6**,**7**	9
F. Gastrointestinal	2,3,7	**4**,**8**	–	–	**1**	5,**6**	–	–
G. Respiratory	1,2	–	–	–	–	3,4	–	–
H. Musculoskeletal	1,2,3,4,**6**,9	–	–	8	–	**7**	–	5
I. Urogenital	**4**,**8**	5	–	–	–	**1**	**3**	2,6,7
J. Endorcrine	1,**2**,**4**,7,**8**	6,9,10		5			**3**	–
K. Falls risk drugs	2**,4**,**7**,**9**	–	**1**,**8**	**5**,**6**,**10**	–	–	**3**,**11**,**12**	–
L. Analgesic drugs	5	–	2	–	–	**6**	1,4	**3**
M. Anti-cholinergic	–	–	–	–	–	–	**1**	–
	I a	I b	II a	II b	III	IV	V	Inapp°
The 57 START.v3 criteria								
Section	$n=42$	$n=8$	$n=1$	$n=3$	$n=0$	$n=1$	$n=1$	$n=0$
A. Indicated								
B. Cardiovascular	2,3,**4**,5, 6,8,10	**1**,**7**,**9, 11**						
C. Coagulation	**1**,**2**							
D. Central nervous	**2**,**3**,**4**,**5**,6,7						**1**	
E. Renal	2,4	3		1				
F. Gastrointestinal	2,3,**5**,6,7	**1**,4						
G. Respiratory	1,2,3							
H.Musculo-skeletal	1,2,**3**,**4**,**5**, 6,8,**9**			**7**				
I. Urogenital	**1**,**2**,3,**4**,5							
J. Endorcrine	1							
K. Analgesics	**1**	3				**2**		
L. Vaccines	**3**,**4**		**1**	**2**				

* Evidence levels of the original data(SR: Systematic review; M-A: Meta-analysis) I: RCT(s): Ia: SR ± MA, or ≥ 3 concordant RCTs; I b: 1–2 RCT(s) II: cohort(s): IIa SR ± MA, or ≥ 3 concordant cohorts; IIb: 1–2 cohort(s) III: case–control studies; IV: cases (series, registers, and surveys) V: expert opinion, with no evidence-based data found in the reference Inapp: inappropriateness (see methods)

+ Focus on adults ≥ 65 years (reference in **bold** characters)

according to the publication type of the STOPP and the START references. Supplementary Table 1 shows the numbers of this distribution.

Of the 454 references, 174 are provided for the 78 new criteria and the other 280 ones for the pre-existing 112 criteria. For new and pre-existing criteria, evidence was, respectively, level I (44% and 34%), level II (19% and 5%), level III–IV–V (23% and 42%), while inappropriateness concerned 11 and 21 references (14% and 19%). Both in the new and pre-existing criteria, respectively, the inappropriate references related more to STOPP (18 and 23%) than to START (5 and 13%).

The Appendix 1 provides detailed information for each of the 454 references, namely, its publication type, focus on older adults, number of PubMed-indexed citations (overall and per year), and evidence level. The best reference for each of the 190 criteria appears in bold characters. The 57 references (13%) considered as inappropriate appear in italic characters, and the reason for inappropriateness is provided.

Analysis at the criterion level

Table 2 shows the best reference of each of the 190 STOPP/START version 3 criteria. The best reference is level I (randomised controlled trials) in 107 criteria (56%: STOPP 43%,

Table 3 Ten STOPP criteria with no appropriate reference

Medication, condition	References	Reason for inappropriateness
Digoxin in HFpEF (B1)	(i) Circulation 2009; 119: 1977–2016 (ii) Am J Ger Pharm 2009; 7(5): 233–49	No message on digoxin No data on HFpEF
Aspirin > 100 mg/day, long term (C1)	(i) Am J Med 2006; 119(3):198–202 (ii) Curr Med Res Opin 2006; 22:1239–48	Discordance: *160 mg/day* Discordance: *75–325 mg/day*
ACO > 6 months for first DVT without continuing RF (C8)	(i) Circulation 2001; 103(20): 2453–60 (ii) Chest 2012; 141(2 Suppl): e419S-94S	Discordance: *3 months* Discordance: *3 months*
ACO > 12 months for first PE without continuing RF (C9)	(i) Chest 2012; 141(2 Suppl): e419S-94S	Discordance: *3 months*
Biphosphonate if eGFR < 30 ml/min/m^2 (E9)	(i) J Bone Miner Res. 2013; 28: 2049–59 (ii) British Nat. Formulary, 2021; 81: 768	Discordance: *no effect on renal function* (Not evaluable)
Corticosteroid (systemic) for osteoarthritis (H5)	(i) British Nat. Formulary 2021; 81: 1140 (ii) Arthritis Rheum 2000; 43: 1905–15	(Not evaluable) No message on corticosteroid
Antimuscarinic drugs with narrow-angle glaucoma (I2)	(i) Arch Ophthalmol. 1970; 84:719–723 (ii) British Nat. Form. 2021; 81: 1223–24	Mismatch: open angle glaucoma (Not evaluable)
Mirabegron in labile or severe hypertension (I6)	(i) Ther Adv Urol. 2012; 4: 315–24 (ii) Int J Clin Pract. 2013; 67: 619–32	Discordance: *similar SE incidence* Discordance: *efficacy and safety*
Duloxetine for urge incontinence (I7)	(i) BJOG. 2004; 111: 249–57 (ii) Am J Obstet Gyn. 2002; 187: 40–8	Discordance: *potential treatment* Mismatch: stress, not urge
Long-acting opioid without short-acting opioid for break-through pain (L3) [Should be in START]	(i) Drug Saf 1993; 8: 99–127 (ii) Drugs Aging 2007; 24:815–28 (iii) Exp Opin Drug Saf 2005; 4:157–69 (iv) Am J Med 2004; 117 S 5A:63S-71S	Mismatch for all four references : NSAIDs, not opioids

Abbreviations ACO: Anticoagulant. DVT: Deep venous thrombosis. eGFR: Estimated glomerular filtration rate. HFpEF: Heart failure with preserved ejection fraction. NSAIDs: Non-steroid anti-inflammatory drugs. PE: Pulmonary embolism. PPI: Proton pomp inhibitor. RF: Risk factor. SE: Side effects

START 88%), level II (cohorts) in 21 criteria (11%: STOPP 13%, START 7%) and levels III–IV–V (case–control, case series or expert opinion) in 49 criteria (26%: STOPP 35%, START 4%) criteria. In the remaining 10 criteria (5%: STOPP 8%, START none), no reference was considered as appropriate.

Table 3 presents these latter 10 STOPP criteria and provides the reason for inappropriateness of their references, i.e., no message found in the reference to support the criterion, or a clinically discordant message between the reference and its criterion, or a major mismatch between the reference and its criterion. Appendix 2 provides a proposal for a valid reference for each of these STOPP criteria.

Discussion

To our knowledge, this is the first study reporting on the scientific evidence provided for the criteria of STOPP/START, an important issue regarding medication management in older adults. Surprisingly, one-third of all SS.v3 references (36%) were rated as low evidence level, with a notably higher proportion for STOPP criteria compared to START criteria (45% vs. 19%). This difference was not unexpected. Most medication studies primarily focus on evaluating efficacy—addressing whether a medication should be initiated—whereas evidence on safety and deprescribing is comparatively sparse. More robust evidence is needed to support decisions related to discontinuing medications. Generating such evidence through randomized controlled trials is challenging in older and frail populations. However, cohort studies controlling for confounders can offer valuable insights into medication efficacy and safety in these groups.

Most of the low-level ratings originated from narrative reviews used by the SS.v3 authors to support the criteria. We believe that the inclusion of narrative reviews with low evidence level in such consensus-based work is questionable. We suggest that such information should be given to the experts involved in the consensus process and to the clinicians applying the STOPP/START criteria in their practice. Consensus-based evidence with no data provided was classified as level V, the lowest category in our framework, consistent with international grading systems that do not consider narrative review data as valid evidence. For this reason, we sought to identify and appraise the data provided by the original studies cited within these reviews—an approach that could be considered for future revisions of the STOPP/START criteria.

Another important finding is the absence of appropriate provided reference for 10 STOPP criteria. These 10 criteria (Table 3) relate to common medications, such as antiplatelet drugs, anticoagulant, corticosteroids, bisphosphonates and opiates, which are used for prevalent medical conditions, such as cardiovascular prevention, venous thrombo-embolic disease, arthritis, osteoporosis and pain. We searched a valid reference for each of the 10 STOPP criteria. The Appendix 2 lists a proposal for such a reference for each of these criteria [11–19].

This study has several strengths. This is, to the best of our knowledge, the first detailed analysis and appraisal of the references that were used for developing and supporting the STOPP/START criteria. Second, our thorough assessment of the 454 references provided in SS.v3 addressed the study design of the underlying data, the focus on adults aged 65 years or more, and the alignment between each reference and its corresponding criterion. Third, the analysis was performed both at level of the 454 references and the level of the 190 criteria. Finally, for the ten STOPP criteria with no appropriate reference, we searched the medical literature for available evidence and we provided a valid reference for nine of these ten criteria.

We recognize some limitations. First, our assessment of the references did not follow an international standard, such as GRADE (Grading of Recommendations Assessment, Development and Evaluation), nor did we apply formal risk-of-bias tools, such as RoB 2 (for randomized trials), ROBINS-I (for non-randomized studies), or AMSTAR-2 (for systematic reviews). While these methods would have allowed a more rigorous appraisal, they were beyond the scope and resources available for 454 references. Consequently, our finding that 46 STOPP criteria (35%) and 42 START criteria (74%) are supported by level Ia evidence (systematic reviews or ≥ 3 concordant RCTs) may be overestimated due to the presence of low-quality systematic reviews. Such bias is also likely among the 19 criteria supported by a single RCT (level Ib). Nonetheless, our approach effectively highlights key evidence gaps and directions for future refinement and application of the STOPP/START criteria. Second, we were not able to conduct independent assessments, as planned, between the pair of clinician experts in pharmacotherapy and the main investigator. Indeed, during this appraisal process, a majority of the references kept leading to discussions between the pair and the main investigator mainly regarding 1) the evidence level (based on the study design of the original studies underlying the data related to the criterion message): these data were often difficult to identify, particularly in the many narrative reviews ($n = 164$); 2) the identification in these narrative reviews of data (provided or cited; evidence levels I–IV) or their absence (evidence level V, $n = 89$, supplementary table); and 3) an agreement on the inappropriateness of a reference ($n = 57$). Third, this study involved investigators from one university. However, all are physicians and pharmacists with expertise in pharmacotherapy and a long experience in the use of STOPP/START criteria, including the OPERAM trial [20–25]. Fourth, a few references could not be accessed, and their appraisal was done based on their abstract, preventing us to be confident that the criterion message is not mentioned somewhere in the full text. As the difference between evidence levels IV (cases) and V (no data provided) is sometimes difficult to draw and clinically not important, we gathered levels IV and V along with level III into the low evidence class.

There is both a need and an opportunity to strengthen scientific basis of many criteria of the STOPP version 3. Particular attention should also be given to STOPP criteria that currently lack a suitable reference. We suggest that any inappropriate reference be removed in the next version of STOPP/START, for the sake of validity and homogeneity. When high-level evidence supports a criterion, these sources should be clearly identified and a strong recommendation might be given. Such improvements would increase the uptake in clinical practice of STOPP criteria and thereby help decrease the use of potentially inappropriate medications in older adults.

In conclusion, one-third of the references underlying the criteria are narrative reviews, with the large majority not based on high-level evidence (RCTs). At the criterion level, high evidence level is provided for most START criteria but less than half of the STOPP criteria. Awareness of the criteria with low levels of evidence is important for clinicians involved in shared decision-making with older adults.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41999-025-01386-7.

Data availability Our data are available to other colleagues, including the Excel table on the 454 STOPP/START references.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent For this type of study formal consent is not required.

References

1. O'Mahony D, Cherubini A, Guiteras AR, Denkinger M, Beuscart JB, Onder G et al (2023) Stopp/START criteria for potentially inappropriate prescribing in older people: version 3. Eur Geriatr Med 14(4):625–632

2. Rochon PA, Stall NM, Reppas-Rindlisbacher C, Gurwitz JH (2023) Stopp/START version 3: even better with age. Eur Geriatr Med 14(4):635–637
3. Kojima T, Akishita M (2023) STOPP/START version 3: overhauled resources to support older people. What will we do next? Eur Geriatr Med 14(4):643–644
4. Hilmer SN (2023) Stopp/START version 3: looks great, fits well, but itches. Eur Geriatr Med 14(4):639–641
5. Boland B, Marien S, Sibille FX, Locke S, Mouzon A, Spinewine A, Dalleur O (2023) Clinical appraisal of STOPP/START version 3 criteria. Eur Geriatr Med 14(5):1149–1150
6. Massé O, Flamand Villeneuve J, Lahaie A, Marcoux C, Hill J, Papillon-Ferland L, Desforges K (2024) Stopp/START version 3: clinical pharmacists are raising concerns. Eur Geriatr Med 15(2):589–591
7. Bomhof-Roordink H, Gärtner FR, Stiggelbout AM, Pieterse AH (2019) Key components of shared decision making models: a systematic review. BMJ Open 9(12):e031763
8. O'Mahony D, Cherubini A, Guiteras AR, Denkinger M, Beuscart JB, Onder G et al (2023) Authors' response to points raised by Boland et al. regarding STOPP/START version 3 criteria. Eur Geriatr Med 14(5):1151–1154
9. O'Mahony D, Cherubini A, Guiteras AR, Denkinger M, Beuscart JB, Onder G et al (2024) Response to letter by Masse O et al. STOPP/START version 3: clinical pharmacists are raising alarms. Eur Geriatr Med 15(2):593–596
10. Dalleur O, Sibille FX, Mouzon A, Vaillant F, Marien S, Spinewine A, Boland B (2025) Improving usability of the STOPP/START version 3 criteria: development of a practice tool for clinicians, students and researchers. Eur Geriatr Med 16(4):1403–1413
11. Feinberg M (1993) The problems of anticholinergic adverse effects in older patients. Drugs Aging 3(4):335–348
12. Takigawa C, Goto F, Tanda S, Shima Y, Yomiya K, Matoba M et al (2015) Breakthrough pain management using fentanyl buccal tablet (FBT) in combination with around-the-clock (ATC) opioids based on the efficacy and safety of FBT, and its relationship with ATC opioids. Jpn J Clin Oncol 45(1):67–74
13. Ahmed A, Rich MW, Fleg JL, Zile MR, Young JB, Kitzman DW et al (2006) Effects of digoxin on morbidity and mortality in diastolic heart failure: the ancillary digitalis investigation group trial. Circulation 114(5):397–403
14. Jones WS, Mulder H, Wruck LM, Pencina MJ, Kripalani S, Muñoz D et al (2021) Comparative effectiveness of aspirin dosing in cardiovascular disease. N Engl J Med 384(21):1981–1990
15. Stevens SM, Woller SC, Kreuziger LB, Bounameaux H, Doerschug K, Geersing GJ et al (2021) Antithrombotic therapy for VTE disease: second update of the CHEST guideline and expert panel report. Chest 160(6):e545–e608
16. Ortel TL, Neumann I, Ageno W, Beyth R, Clark NP, Cuker A et al (2020) American Society of Hematology 2020 guidelines for management of venous thromboembolism: treatment of deep vein thrombosis and pulmonary embolism. Blood Adv 4(19):4693–4738
17. Miller PD, Roux C, Boonen S, Barton IP, Dunlap LE, Burgio DE (2005) Safety and efficacy of risedronate in patients with age-related reduced renal function as estimated by the Cockcroft and Gault method: a pooled analysis of nine clinical trials. J Bone Miner Res 20(12):2105–2115
18. Yi W, Yang Y, Yang J (2021) Monotherapy with mirabegron had a better tolerance than the anticholinergic agents on overactive bladder. Medicine (Baltimore) 100(41):e27469
19. Mirzaei M, Daneshpajooh A, Anvari SO, Dozchizadeh S, Teimorian M (2021) Evaluation of the clinical efficacy and complications of duloxetine in comparison to solifenacin in the treatment of overactive bladder disease in women: a randomized clinical trial. Urol J 18(5):543–548
20. Dalleur O, Spinewine A, Henrard S, Losseau C, Speybroeck N, Boland B (2012) Inappropriate prescribing and related hospital admissions in frail older persons according to the STOPP and START criteria. Drugs Aging 29(10):829–837
21. Dalleur O, Boland B, Losseau C, Henrard S, Wouters D, Speybroeck N, Degryse JM, Spinewine A (2014) Reduction of potentially inappropriate medications using the STOPP criteria in frail older inpatients: a randomised controlled study. Drugs Aging 31(4):291–298
22. Lang PO, Dalleur O, Petrovic M, Benetos A, Ferahta N, Boland B (2016) The exercise in applying STOPP/START.v2 in vulnerable very old patients. Eur Geriatr Med 7(2):176–179
23. Boland B, Guignard B, Dalleur O, Lang P-O (2016) Application of STOPP/START and Beers criteria: compared analysis on identification and relevance of potentially inappropriate prescriptions. Eur Geriatr Med 7(5):416–423
24. Thevelin S, Mounaouar LE, Marien S, Boland B, Henrard S, Dalleur O (2019) Potentially inappropriate prescribing and related hospital admissions in geriatric patients: a comparative analysis between the STOPP and START criteria versions 1 and 2. Drugs Aging 36(5):453–459
25. Blum MR, Sallevelt BTGM, Spinewine A, O'Mahony D, Moutzouri E, Feller M et al (2021) Optimizing therapy to prevent avoidable hospital admissions in multimorbid older adults (OPERAM): cluster randomised controlled trial. BMJ 374:1585