BMJ Open

Evaluation of clarity of the STOPP/START criteria for clinical applicability in prescribing for older people: a quality appraisal study

Bastiaan Theodoor Gerard Marie Sallevelt ![ORCID],[1] Corlina Johanna Alida Huibers,[2] Wilma Knol,[3] Eugene van Puijenbroek,[4,5] Toine Egberts,[1,6] Ingeborg Wilting[1]

For numbered affiliations see end of article.

Correspondence to
Mr Bastiaan Theodoor Gerard Marie Sallevelt;
B.T.G.Sallevelt@umcutrecht.nl

ABSTRACT

Objectives Appropriate prescribing in older people continues to be challenging. Studies still report a high prevalence of inappropriate prescribing in older people. To reduce the problem of underprescribing and overprescribing in this population, explicit drug optimisation tools like Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert to Right Treatment (STOPP/START) have been developed. The aim of this study was to evaluate the clinical applicability of STOPP/START criteria in daily patient care by assessing the clarity of singular criteria.

Design Quality appraisal study.

Methods For each of the 114 STOPP/START criteria V.2, elements describing the action (*what/how* to do), condition (*when* to do) and explanation (*why* to do) were identified. Next, the clarity of these three elements was quantified on a 7-point Likert scale using tools provided by the Appraisal of Guidelines for Research and Evaluation (AGREE) Consortium.

Primary and secondary outcomes The primary outcome measure was the clarity rating per element, categorised into high (>67.7%), moderate (33.3%–67.7%) or low (<33.3%). Secondary, factors that positively or negatively affected clarity most were identified. Additionally, the nature of the conditions was further classified into five descriptive components: disease, sign, symptom, laboratory finding and medication.

Results STOPP recommendations had an average clarity rating of 64%, 60% and 69% for actions, conditions and explanations, respectively. The average clarity rating in START recommendations was 60% and 57% for actions and conditions, respectively. There were no statements present to substantiate the prescription of potential omissions for the 34 START criteria.

Conclusions Our results show that the clarity of the STOPP/START criteria can be improved. For future development of explicit drug optimisation tools, such as STOPP/START, our findings identified facilitators (high clarity) and barriers (low clarity) that can be used to improve the clarity of clinical practice guidelines on a language level and therefore enhance clinical applicability.

Strengths and limitations of this study

► To the best of our knowledge, this is the first study that explores the clarity of Screening Tool of Older Persons' potentially inappropriate Prescriptions/ Screening Tool to Alert to Right Treatment (STOPP/ START) criteria.

► Clarity ratings were scored independently by appraisers who were experienced in applying STOPP/ START criteria in clinical practice.

► The scoring process remains partly subjective, however, consensus ratings show high inter-rater agreement.

► By evaluating the *what*, *when* and *why* of recommendations, element-specific strategies were formulated to improve their clarity.

INTRODUCTION

Clinical practice guidelines (CPGs) are instruments intended to provide guidance to healthcare professionals in patient care. Translation of healthcare knowledge, evidence and experience into clear recommendations for patient care, however, is challenging. Studies in the USA and the Netherlands suggest that about 30%–40% of patients do not receive care according to evidence based guidelines. A clear description of the desired behaviour has been associated with better compliance with guideline recommendations.[1 2]

Recommendations about safe and effective pharmacotherapy are an important part of CPGs. However, it is often unclear whether recommendations also apply to older people.[3–5] A complicating factor is that older people experience more concomitant morbidities, while CPGs often focus on best treatment for a single disease. Ambiguity among prescribers about pharmacotherapy in older people results in inappropriate prescribing, which causes adverse drug

reactions, drug-related hospitalisations, decreased quality of life and even death.[6 7]

Due to the lack of clear statements in CPGs about (in) appropriate prescribing in older people with multimorbidity, several explicit screening tools have been developed.[8 9] The most widely used are the Beers criteria[10] and the Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert to Right Treatment (STOPP/START) criteria.[11] CPG recommendations are rarely specified in precise behavioural terms such as what, how, when and why to stop or start a drug, while explicit screening tools are designed to make clear statements and therefore ease clinical implementation.[2] However, studies continue to report a high prevalence of inappropriate prescribing in older people.[12–14] This suggests that implementation can still be improved.

Although STOPP/START criteria have shown good inter-rater reliability in studies involving physicians and (hospital)pharmacists working in geriatric units, data on how physicians less familiar with medication optimisation would interpret STOPP/START criteria are lacking.[15 16] The question then arises whether the recommended actions are formulated clearly enough to guide prescribers less experienced in geriatric patient care.

The aim of this study was to evaluate the clinical applicability of STOPP/START criteria in daily patient care by assessing the clarity of singular criteria with the purpose of improving future clinical guideline recommendations for appropriate prescribing in older people.

METHODS
STOPP/START criteria
The STOPP/START criteria were first published in 2008 and have been updated in 2015 to STOPP/START V.2.[17] STOPP/START is a product of two Delphi rounds by 19 experts from 13 European countries.

For this study, the supplementary data of the corrigendum of the STOPP/START criteria V.2 as published in November 2017 were used.[18] STOPP/START V.2 consists of a list of 80 potentially inappropriate medications (STOPP criteria) and 34 potential prescribing omissions (START criteria).

Clarity assessment
The Appraisal of Guidelines for Research & Evaluation (AGREE) II Instrument and Guideline Implementability Decision Excellence Model (GUIDE-M) were used to develop a framework to assess the clarity of language used in STOPP/START. AGREE II Instrument is an internationally validated tool to rate the quality of CPGs, developed by the AGREE Consortium.[19] In addition to the AGREE II Instrument, AGREE developed a GUIDE-M.[20] This model identifies 'communicating content' as a core tactic for CPG implementability. Obviously, language is an important domain of this tactic. The language subdomain promotes a clear, simple and persuasive message.

The relevant part of the AGREE II Instrument ('clarity of presentation', domain 4, item 15) states that recommendations should be 'specific and unambiguous', which is defined as 'a concrete and precise description of which option is appropriate for which situation and for what population group'. In line with this statement and the corresponding section of the AGREE II Instrument, three elements were identified that influence the clarity of recommendations:

► Action: description of the recommended action, i.e. what to do and how to act?
► Condition: identification of the relevant target population and statements about patients or conditions for whom the recommendations would apply or not apply, i.e. when?
► Explanation: identification of the intent or purpose of the recommended action, i.e. why?

In order to quantify the clarity of STOPP/START criteria, the three elements of each recommendation were rated independently on a 7-point Likert scale by a panel of two appraisers, consisting of a geriatric resident (CJAH) and a hospital pharmacy resident (BTGMS), both experienced with the application of STOPP/START criteria in daily practice. The clarity for each of these three elements was rated from the perspective of a 'junior' physician or pharmacist with a basic level of knowledge (≤5 years of clinical postgraduate experience). The appraisers were trained with a rating guidance, developed and approved by senior clinicians (TE/EvP/IW/WK) prior to rating the elements independently. If ratings differed more than 1 point, a senior hospital pharmacist/clinical pharmacologist (IW) or a senior geriatrician/clinical pharmacologist (WK) was consulted as a third appraiser until consensus was reached.

Descriptive components of conditions
In addition to the calculation of clarity ratings for the action, condition and explanation, the nature of the conditions was further explored. The condition identifies the target population and is the most heterogeneous element. By stratifying the conditions into descriptive components, the nature of the components in relation to their clarity could be assessed. These components could lead to different strategies to optimise 'specific and unambiguous' wording in describing conditions.

The conditions were subdivided into five components that were considered essential for identification of the target population: disease, sign, symptom, laboratory finding and medication. Definitions of four components were based on the ontology as described by Scheuermann et al.[21] Signs are defined as bodily features observed in a physical examination including measurements (e.g. blood pressure), while symptoms are bodily features experienced by a patient (e.g. restless legs). Since optimisation of polypharmacy is the main focus of the STOPP/START, the target population can also be described by (co)medication. Medication is not defined by Scheuermann et al. Therefore, medication was added as a fifth component

using the definition for medicinal products by the European Medicines Agency as 'a substance or combination of substances that is intended to treat, prevent or diagnose a disease or to restore, correct or modify physiological functions by exerting a pharmacological, immunological or metabolic action'.[22]

Data analysis

Clarity ratings for each of the three elements (action, condition, explanation) were calculated as a percentage of the obtained scores given by appraiser 1 and 2 divided by the maximum score.

$$Clarity\ rating\ (\%) = \frac{obtained\ score\ (sum\ of\ 2\ appraisers) - minimum\ possible\ score\ (2)}{maximum\ possible\ score\ (14) - minimum\ possible\ score\ (2)}$$

This calculation method is in accordance with the approach provided by AGREE II Instrument. The scores of appraisers 1 and 2 were both replaced by the consensus score if a third appraiser was consulted. After scoring the elements, clarity ratings were categorised into low (<33.3%), moderate (33.3%–67.7%) and high (>67.7%).

Patient and public involvement

Since this is an appraisal study of clinical guideline recommendations intended to be used by clinicians, this research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

RESULTS

The elements 'action' and 'condition' in STOPP and START recommendations were rated on their clarity, resulting in 80 and 34 scores per element, respectively. The element 'explanation' was present in all but three (A1, A2, B11) STOPP recommendations, resulting in 77 scores. None of the START criteria contained an explanation to substantiate the prescription of potential omissions. Therefore, Likert scores for explanations were only assessed in STOPP recommendations.

The agreement among the two appraisers for Likert scores was high and ranged from 76.3% (STOPP—condition) to 91.3% (STOPP—action). Forty-four out of 305 (14.4%) scores were replaced after consensus meetings with a third appraiser. Replacements did not alter average Likert scores per element with more than 0.2 points compared with the average scores prior to consensus.

Average clarity ratings for STOPP recommendations were 64%, 60% and 69% for actions, conditions and explanations, respectively. Average clarity ratings for START recommendations were 60% and 57% for actions and conditions, respectively (figure 1).

In 80 STOPP and 34 START recommendations, the clarity ratings of 35 actions were categorised as high (30.7%), 65 as moderate (57.0%) and 14 as low (12.3%). Thirty-eight (33.3%), 67 (58.8%) and 9 (7.9%) conditions
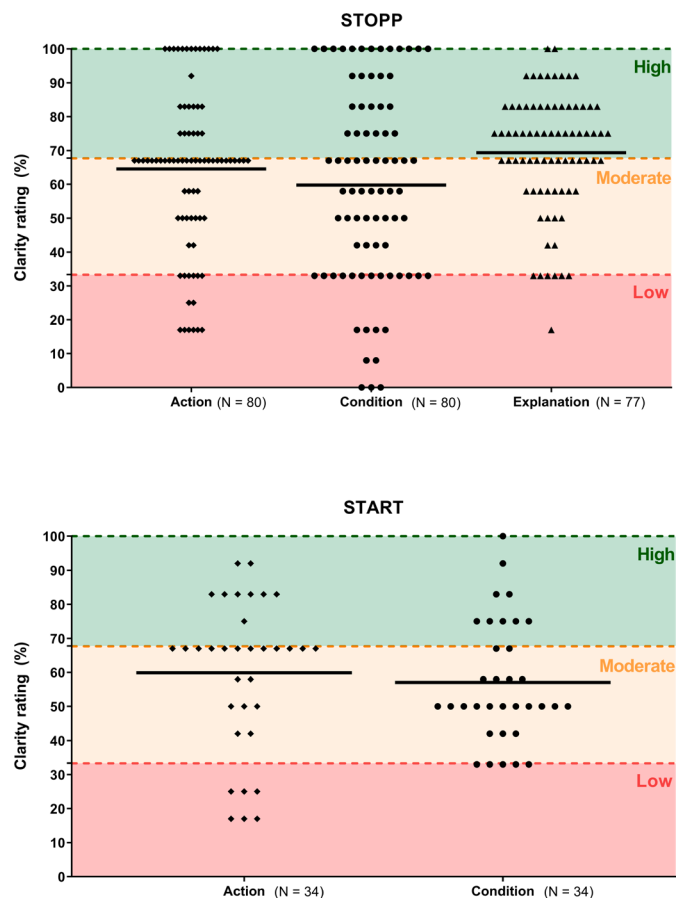

Figure 1 Distribution of clarity ratings for STOPP and START recommendations per element. Average clarity ratings for STOPP recommendations were 64%, 60% and 69% for actions, conditions and explanations, respectively. Average clarity ratings for start recommendations were 60% and 57% for actions and conditions, respectively. STOPP/START, Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert to Right Treatment.

had a high, moderate or low clarity rating, respectively. In 77 STOPP criteria, the clarity ratings of 41 (53.2%) explanations were categorised as high, 35 (45.5%) as moderate and 1 (1.3%) as low.

Thirteen STOPP criteria (C1, C2, C4, C7, D6, D12, D13, E5, E6, F1, G1, H1 and H9) had high clarity ratings for all three elements. Four START criteria (B3, G3, I1 and I2) had high clarity ratings for both action and condition. Detailed information of clarity ratings per element for all individual STOPP/START criteria can be found in online supplementary data S1.

Elements with high (>67.7%) and moderate or low (≤67.7%) clarity ratings were analysed in more detail to identify factors that either positively or negatively affected 'specific and unambiguous' language most. These findings for actions, conditions and explanations with illustrative examples for STOPP and START recommendations are presented in table 1.

The results of stratifying the element 'condition' into the five descriptive components medication, disease, sign, symptom and laboratory finding are shown per STOPP/

Table 1 Main barriers and facilitators that affected clarity of the elements action, condition and explanation of STOPP/START recommendations

Barriers	Example* (clarity rating, %)
Action	
Lack of explicit drug (class)	STOPP D7/8. **Anticholinergics/antimuscarinics**… (17%)
► 'For example' represents a non-limitative list and is therefore inconclusive	STOPP B10. Centrally acting antihypertensives (**e.g. methyldopa, clonidine, moxonidine, rilmenidine, guanfacine**)… (33%)
► Use of adjectives that need further investigation to allow use	STOPP D14. **First-generation** antihistamines (17%) START H1. **High potency** opioids… (17%)
Lack of drug deprescribing schedules while considered necessary	STOPP K2. **Neuroleptic drugs** (17%)
Starting dose and target dose not mentioned	START A6. **ACE inhibitor** with systolic heart failure… (67%)
Lack of directions how and what to monitor after starting a drug	START E1. **Disease-modifying antirheumatic drug**(DMARD)… (25%)
Condition	
General—patient population for whom recommendations would not apply was not (clearly/unambiguously) defined ► In patients with a strong indication for a potentially inappropriate drug, it may be harmful to stop it ► In patients with potential omissions, warnings for important contra indications are lacking/not clearly defined	STOPP B5. **…as first-line antiarrhythmic therapy in supraventricular tachyarrhythmias** (33%) START A2. **…where vitamin K antagonists or direct thrombin inhibitors or factor Xa inhibitors are contraindicated** (33%)
Medication—see also action ► Ambiguous adjectives were used ► Description of drug therapy (substance/dosage) not specific enough	STOPP D2. **…as first-line antidepressant treatment** (33%) START E7. …in patients taking **methotrexate** (33%)
Disease—clinical interpretation of 'disease (state)' for defining population needed	STOPP D1. …with **dementia**, narrow angle glaucoma, **cardiac conduction abnormalities**, prostatism or **prior history of urinary retention** (33%) START A5. …with a documented **history of coronary, cerebral or peripheral vascular disease** (33%)
Sign—measurement or scores were not described unambiguously	STOPP H2. …with **severe hypertension** or **severe heart failure** (33%) START E1. …with **active, disabling rheumatoid disease** (42%)
Symptom—symptoms were not described unambiguously	STOPP K-section. Not clear whether the occurrence of **'falls'**—as mentioned only in the title of section K—is a prerequisite for the applicability of the recommendation or only used to address the increased risk of falls. If 'falls' is considered a condition, the frequency of 'falls' is not specified. (0%) STOPP D10. …unless **sleep disorder** is due to… (33%) START C2. …with **persistent major depressive symptoms** (33%)
Laboratory finding—parameters lack clear cut-off levels with reference ranges	START C6. …once **iron deficiency** and **severe renal failure** have been excluded (33%)
Explanation	
Risk of continuing therapy not clearly described: explanation does not cover clinical relevance of benefit/harm balance (specific adverse drug reactions, toxicity).	STOPP D7. …(**risk of anticholinergic toxicity**) (17%) START N/A

Facilitators	Example* (clarity rating, %)
Action	
Drugs were specified on individual drug level and—if necessary—route/dosage was specified	STOPP C7. **Ticlopidine**… (100%) START A2. **Aspirin (75–160 mg once daily)**… (92%)
Condition	
Medication—see also action Specific description of drug therapy (substance/dosage) to clearly identify the target population (i.e. patients using a certain drug regimen).	STOPP B3. …in combination with **verapamil** or **diltiazem** (92%) START I2. **…at least once after age 65 according to national guidelines** (83%)
Disease—diseases clearly described, the target population could be easily identified	STOPP H9. …in patients **with a current or recent history of upper gastrointestinal disease that is, dysphagia, oesophagitis, gastritis, duodenitis or peptic ulcer disease, or upper gastrointestinal bleeding** (92%) START C4. …for **primary open-angle glaucoma** (100%)
Signs—signs clearly described as scores or measurements and therefore unambiguous	START B3. …with documented chronic hypoxaemia (**i.e. pO_2 <8.0 kPa or 60 mm Hg or SaO_2 <89%**) (92%)
Symptom—symptoms clearly and unambiguous described	STOPP F1. …with **parkinsonism** (92%)

Continued

Table 1 Continued

Facilitators	Example* (clarity rating, %)
Laboratory findings—clear cut-off levels with reference ranges present	STOPP E6. …if **eGFR <30 mL/min/1.73 m²** (100%)
Explanation	
Risk of discontinuing clearly described	STOPP D5. …(**no indication for longer treatment; risk of prolonged sedation, confusion, impaired balance, falls, road traffic accidents; all benzodiazepines should be withdrawn gradually if taken for >2 weeks as there is a risk of causing a benzodiazepine withdrawal syndrome if stopped abruptly**) (100%) START N/A

*The examples shown are selected from elements with low and moderate (≤67.7%) clarity ratings for barriers and from high (>67.7%) clarity ratings for facilitators to substantiate the main findings. An overview of all clarity ratings can be found in the online supplementary data S1.
eGFR, estimated glomerular filtration rate; N/A, not applicable; pO_2, partial pressure of oxygen; SaO_2, arterial oxygen saturation; STOPP/START, Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert to Right Treatment.

START recommendation in figure 2. Clarity ratings were scored on the level of condition as an element and not on the sublevel of the five descriptive components. Therefore, all components of one condition share the same colouring for their clarity.

In 33 (41%) STOPP criteria and 17 (50%) START criteria, the condition consisted of more than one component. No strong association was found between the clarity of conditions and the nature of the descriptive components, as the clarity ratings of the condition section varied regardless of the nature of the component. However, laboratory findings used to identify the target population were discovered to have the highest clarity rating compared with other descriptive components in STOPP recommendations; 9 out of 13 laboratory-based conditions had a high clarity rating (>67.7%).
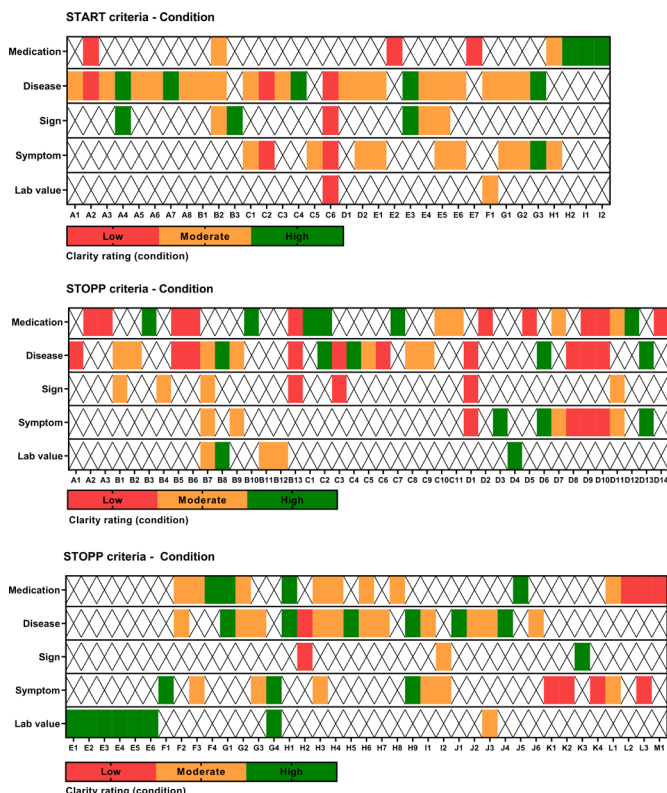

Figure 2 Clarity ratings of conditions for STOPP and start criteria related to five descriptive components. Green, orange and red colours correspond with high (>67.7%), moderate (33.3%–67.7%) or low (<33.3%) clarity ratings of conditions. STOPP/START, Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert to Right Treatment.

DISCUSSION
Main findings
In this study, we evaluated the clinical applicability of STOPP/START criteria in daily patient care by assessing the clarity of singular criteria. We found that 13 out of 80 STOPP and 4 out of 34 START criteria had a high clarity rating for the three elements action, condition and explanation. To improve clarity of recommendations, element-specific strategies can be formulated (table 1).

Actions were considered unclear if recommendations included non-explicitly specified drug classes (e.g. 'anticholinergics'). To improve clear description of the action (what and how) we advise to specify drugs at an individual substance level. The addition of how to start or stop a drug (immediately vs gradually, including monitoring guidelines and deprescribing schedules), route of administration and dosage were considered necessary for some actions to further improve clarity.

The definition of the condition (the when) had the lowest average clarity rating in both START and STOPP. Low clarity ratings for conditions resulted from insufficient distinctiveness in the identification of patients for whom recommendations do or do not apply. Conditions were described by medication, diseases, signs, symptoms and laboratory findings. To increase the clarity of the conditions, laboratory findings and signs have the highest potential to be optimised by adding statements about clear cut-off levels (e.g. 'potassium >5.0 mmol/L' instead of 'hyperkalaemia') and measurements (e.g. 'systolic blood pressure >160 mm Hg' instead of 'uncontrolled severe hypertension'). For conditions defined by medication use, the same improvements as suggested for

actions apply. In some cases even a description on a drug substance level was not specific enough. For instance, folic acid for patients on methotrexate therapy (START E7) only applies to patients using a low dose, weekly methotrexate schedule and not for patients on high dose methotrexate. In such cases, a more detailed description of a drug dosage, route or indication was deemed necessary. Conditions described by diseases—like 'heart failure'— might seem clear at first, but often need further specification (reduced vs preserved ejection fraction) to avoid ambiguity. Moreover, international cardiology guidelines distinguish between these subtypes of heart failure, subsequently affecting treatment recommendations. Adherence to terminology of internationally used dictionaries to describe diseases, such as International Classification of Primary Care (ICPC) and International Classification of Diseases (ICD), could be a solution.

Furthermore, no explanations were present for START criteria to substantiate why a potential omitted drug should be initiated. Even though the reason to start a drug might seem obvious in most cases, the risk–benefit balance should always be addressed to assist a physician's decision-making process whether or not to expose a patient to additional drug therapies.

Other remarks

STOPP/START criteria provide best evidence-based practices for the overtreatment and undertreatment of single conditions. However, it should be noted that STOPP/START criteria provide conflicting recommendations. For example, if a patient has a clear indication for a beta blocker to treat ischaemic heart disease (START A7), this is contradicted if a patient is already using verapamil or diltiazem (STOPP B3). Merging such recommendations could increase implementation and prevent potential patient harm by overlooking relevant contraindications.

Besides making the what, how, when and why as clear as possible, guideline developers should consider whether recommendations are tailored for its intended end users (i.e. the who). Explicit screening tools to detect inappropriate prescribing in older people, such as Beers criteria and STOPP/START, are likely to be developed to reach all professionals involved in prescribing, as all prescribers encounter the problem of underprescribing and overprescribing in older people. Clinicians with high affinity for geriatric medicine may not need explicit treatment recommendation to provide best patient care, whereas some clinicians—such as surgical specialists—who treat older people but may be less experienced with (in)appropriate prescribing in older people, probably require more clear guidance. Clear recommendations are therefore important to reach all prescribers, because the success of STOPP/START criteria as an intervention depends on its integration and implementation in clinical practice.[23] Some recommendations may be best applied by physicians with a certain expertise, such as to start an 'acetylcholinesterase inhibitor for mild-to-moderate Alzheimer's dementia or Lewy body dementia (START C3)'. In such cases, the focus for all clinicians should probably be the recognition and detection of a potential omission, rather than to actually start drug treatment. An explicit action could be to refer such patients to a geriatrician or neurologist, thus separating the trigger for potential undertreatment from the actual prescriber.

Strengths and limitations

To the best of our knowledge, this is the first study that explores the clarity of STOPP/START criteria. By systematically reviewing the clarity of the given action, condition and explanation, we identified facilitators (high clarity) and barriers (low clarity) that may be used to improve the content on a language level. As a result, element-specific strategies can be extracted to improve items requiring refinement. Although no previous studies have reviewed the clarity of singular recommendations of explicit drug screening tools, comparable research has been conducted concerning clarity of monitoring instructions in CPGs and drug labels. Their conclusions to improve ambiguous instructions concerning the monitoring of laboratory values are in line with our suggestions to add clear statements about the what, why, when and how of recommendations.[24 25]

Moreover, studies to refine the methodology of developing deprescribing guidelines to facilitate the deprescribing process were conducted.[26 27] A good example are the tools provided by the Bruyère Research Institute, based on their research about developing deprescribing guidelines. The Bruyère research group has published evidence-based CPGs (for instance how to deprescribe benzodiazepines), accompanied by clear algorithms including well-described populations (including for which patients the recommendation does not apply), a list of available drugs and dosages, monitoring recommendations and tapering regimes, thereby complementing the clarity some STOPP-recommendations are lacking.[28]

Tools that have been developed to review the quality of entire CPGs underline the importance of clear and unambiguous recommendations,[29] but no validated tool exists to date to rate singular clinical recommendations. As clarity of presentation is both part of the AGREE II Instrument and described by GUIDE-M, we used tools from the AGREE Consortium to develop a review method. Moreover, the AGREE II Instrument is internationally formally endorsed for guideline assessment and provides a Likert scale that allowed us to quantify clarity.

Clarity ratings were scored by appraisers who are experienced in applying STOPP/START criteria in clinical practice, as they contributed to a large multicentre, randomised controlled trial that evaluated the impact of a STOPP/START-based medication review in older people with polypharmacy. We believe that these experiences allowed clear identification of difficulties prescribers not familiar with STOPP/START may encounter. Although the scoring process remains partly subjective, the consensus ratings show high inter-rater agreement. Differences (>1 point) were discussed with a third appraiser

and consensus was reached for all items. Therefore, the final clarity ratings were considered reliable.

One concern of further specifying recommendations might be that they 'replace' important clinical considerations made by physicians. However, guideline recommendations are never meant to fully substitute clinical judgement to treat individual patients. This is why the explanation of a recommendation—next to the action and condition sections—is important for facilitating translation to an individual patient level.

A lack of strong evidence to support the recommended actions could impede formulating clear explanations. For example, clear statements on numbers needed to treat or numbers needed to harm might be difficult to extract from currently available evidence. In such cases, the addition of the strength of recommendations and supporting evidence could further direct clinicians. This is also endorsed by internationally renowned CPG quality assessment tools from AGREE and Grading of Recommendations Assessment, Development and Evaluation (GRADE).[30]

Furthermore, our study only highlights barriers that could be optimised to prevent unintentional deviations from STOPP/START due to unclear language. Apart from the clarity of presentation, many other factors attribute to clinical implementation of evidence-based recommendations.[27 31]

Implications

To clarify the action, condition and explanation sections of a recommendation, a more detailed statement is often required. This may directly affect choices regarding the presentation of recommendations. In addition to improvements in 'language', the 'format' of a guideline could have a high impact on applicability as well. In a time where almost all evidence-based knowledge is electronically requested, a dynamic, electronic format could be used to integrate information that will improve clarity of presentation without making recommendations too extensive. Integrating clinical rules within electronic healthcare systems—with an option to request more detailed information—could contribute to a continuing learning cycle as part of (but without slowing down) the usual care process. For example, a drug class (stop benzodiazepines) may be provided with a hyperlink including information on drug substance levels (ATC5-codes) and a deprescribing tool, accessible on request. Once a prescriber has become familiar with all the details of a certain recommendation, such information is no longer required. However, converting recommendations into effective software assistance starts with a clear message of the initial statements.

To make the current version of STOPP/START criteria suitable for software engines, multiple multidisciplinary expert rounds turned out to be necessary to reach consensus on how to interpret ambiguous wordings.[32] For instance, due to different lists of anticholinergic drugs in current literature, expert opinion is needed to translate this drug class to clinically relevant, individual drugs with high anticholinergic burden. Furthermore, it was found that some recommendations, such as to 'stop any drug beyond the recommended duration (STOPP A3)' were too general or unspecific to convert into an algorithm. Selecting specific recommendations concerning potentially inappropriate long-term use of medication, such as long-term corticosteroids (>3 months) as monotherapy for rheumatoid arthritis (STOPP H4) or continuing bisphosphonates >5 years without evaluating efficacy (not a criterion), will probably result in a better uptake among clinicians and can be easily integrated into clinical decision support systems. Consequently, the lack of clear statements may impede software implementation.[32 33]

Another advantage to present clear recommendations in an electronic, dynamic format is that content could be easily modified based on updates in evidence, country-specific guidelines, available drugs and local expertise. Collaboration of guideline developers with experts in medical informatics for considering content formatting could, therefore, be of great value to facilitate future implementation of recommendations in clinical practice.

Conclusion

In conclusion, for future development of CPGs, our findings provide direction to assure the clarity of recommendations. We believe in the opportunity to transform STOPP/START from a tool to detect inappropriate prescribing to a guideline that provides clear statements on how to act after detection. The use of specific and unambiguous language in CPG recommendations is likely to assist physicians in prescribing the right drug to the right patient at the right time.

Author affiliations
[1]Clinical Pharmacy, University Medical Center Utrecht, Utrecht, Utrecht, The Netherlands
[2]Geriatrics, Department of Geriatric Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[3]Geriatrics, Department of Geriatric Medicine and Expertise Centre Pharmacotherapy in Old Persons, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands
[4]The Netherlands Pharmacovigilance Centre Lareb, Lareb, 's-Hertogenbosch, The Netherlands
[5]PharmacoTherapy, Epidemiology & Economics, University of Groningen, Groningen, Groningen, Netherlands
[6]Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, Utrecht, The Netherlands

ORCID iD
Bastiaan Theodoor Gerard Marie Sallevelt http://orcid.org/0000-0003-4687-4048

REFERENCES

1 Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *The Lancet* 2003;362:1225–30.
2 Michie S, Johnston M. Changing clinical behaviour by making guidelines specific. *BMJ* 2004;328:343–5.
3 Munster BC, Portielje JEA, Maier AB, et al. Methodology for senior-proof guidelines: a practice example from the Netherlands. *J Eval Clin Pract* 2018;24:254–7.
4 Mutasingwa DR, Ge H, Upshur REG. How applicable are clinical practice guidelines to elderly patients with comorbidities? *Can Fam Physician* 2011;57:253–62.
5 Beers E, Egberts TCG, Leufkens HGM, et al. Information for adequate prescribing to older patients : an evaluation of the product information of 53 recently approved medicines. *Drugs Aging* 2013;30:255–62.
6 Howard RL, Avery AJ, Slavenburg S, et al. Which drugs cause preventable admissions to hospital? A systematic review. *Br J Clin Pharmacol* 2007;63:136–47.
7 Leendertse AJ, Egberts ACG, Stoker LJ, et al. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. *Arch Intern Med* 2008;168:1890–6.
8 Lavan AH, Gallagher PF, O'Mahony D. Methods to reduce prescribing errors in elderly patients with multimorbidity. *Clin Interv Aging* 2016;11:857–66.
9 Page II RL, Linnebur SA, Bryant LL. Inappropriate prescribing in the hospitalized elderly patient: defining the problem, evaluation tools, and possible solutions. *Clin Interv Aging* 2010;5:75–87.
10 American Geriatrics Society 2019 Expert Panel. American geriatrics Society 2019 updated AGS beers criteria for potentially inappropriate medication use in older adults. *J Am Geriatr Soc* 2019;67:674–94.
11 Mangin D, Bahat G, Golomb BA, et al. International Group for Reducing Inappropriate Medication Use & Polypharmacy (IGRIMUP): Position Statement and 10 Recommendations for Action. *Drugs Aging* 2018;35:575–87.
12 Gallagher P, Lang PO, Cherubini A, et al. Prevalence of potentially inappropriate prescribing in an acutely ill population of older patients admitted to six European hospitals. *Eur J Clin Pharmacol* 2011;67:1175–88.
13 Galvin R, Moriarty F, Cousins G, et al. Prevalence of potentially inappropriate prescribing and prescribing omissions in older Irish adults: findings from the Irish longitudinal study on ageing study (TILDA). *Eur J Clin Pharmacol* 2014;70:599–606.
14 Di Giorgio C, Provenzani A, Polidori P. Potentially inappropriate drug prescribing in elderly hospitalized patients: an analysis and comparison of explicit criteria. *Int J Clin Pharm* 2016;38:462–8.
15 Ryan C, O'Mahony D, Byrne S. Application of STOPP and start criteria: interrater reliability among pharmacists. *Ann Pharmacother* 2009;43:1239–44.
16 Gallagher P, Baeyens J-P, Topinkova E, et al. Inter-Rater reliability of STOPP (screening tool of older persons' prescriptions) and start (screening tool to alert doctors to right treatment) criteria amongst physicians in six European countries. *Age Ageing* 2009;38:603–6.
17 O'Mahony D, O'Sullivan D, Byrne S, et al. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing* 2014;44:213–8.
18 Mahony Denis O', Sullivan David O', Byrne S, et al. Corrigendum: STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing* 2018;47:489.
19 AGREE II Instrument. Appraisal of guidelines for research and evaluation Consortium (online), 2017. Available: https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf [Accessed 26 Mar 2019].
20 Brouwers MC, Makarski J, Kastner M, et al. The Guideline Implementability decision excellence model (GUIDE-M): a mixed methods approach to create an international resource to advance the practice guideline field. *Implement Sci* 2015;10:1–11.
21 Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. *Translat Bioinforma* 2009:116–20.
22 Eudralex. Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the community code relating to medicinal products for human use. *Official Journal of the European Union* 2001;L311:67.
23 Curtin D, Gallagher PF, O'Mahony D. Explicit criteria as clinical tools to minimize inappropriate medication use and its consequences. *Ther Adv Drug Saf* 2019;10:204209861982943.
24 Nederlof M, Stoker LJ, Egberts TCG, et al. Instructions for clinical and biomarker monitoring in the summary of product characteristics (SMPC) for psychotropic drugs: overview and applicability in clinical practice. *J Psychopharmacol* 2015;29:1248–54.
25 De Koning FHP, Egberts TCG, De Smet P, et al. Instructions on laboratory monitoring in 200 drug labels. *Clin Chem Lab Med* 2012;50:1351–8.
26 Farrell B, Pottie K, Rojas-Fernandez CH, et al. Methodology for developing deprescribing guidelines: using evidence and grade to guide recommendations for deprescribing. *PLoS One* 2016;11:e0161248.
27 Reeve E, Thompson W, Farrell B. Deprescribing: a narrative review of the evidence and practical recommendations for recognizing opportunities and taking action. *Eur J Intern Med* 2017;38:3–11.
28 Bruyère Research Institute. Benzodiazepine & Z-Drug (BZRA) Deprescribing Algorithm, 2019. Available: https://deprescribing.org/resources/deprescribing-guidelines-algorithms/ [Accessed 13 Nov 2019].
29 Siering U, Eikermann M, Hausner E, et al. Appraisal tools for clinical practice guidelines: a systematic review. *PLoS One* 2013;8:e82915.
30 Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. *BMJ* 2008;336:1049–51.
31 Anderson K, Stowasser D, Freeman C, et al. Prescriber barriers and enablers to minimising potentially inappropriate medications in adults: a systematic review and thematic synthesis. *BMJ Open* 2014;4:e006544.
32 Huibers CJA, Sallevelt BTGM, 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* 2019;125:110–7.
33 Anrys P, Boland B, Degryse J-M, et al. STOPP/START version 2—development of software applications: easier said than done? *Age Ageing* 2016;45:590–3.