



Article

Inappropriate Prescriptions in Older People—Translation and Adaptation to Portuguese of the STOPP/START Screening Tool

Luís Monteiro ^{1,2,*} , Matilde Monteiro-Soares ^{1,3,4}, Cristiano Matos ^{5,6} , Inês Ribeiro-Vaz ^{1,3,7}, Andreia Teixeira ^{1,3,8} and Carlos Martins ^{1,3}

- ¹ CINTESIS—Centre for Health Technology and Services Research, Faculty of Medicine, University of Porto, 4099-002 Porto, Portugal; matsoares@med.up.pt (M.M.-S.); inesvaz@med.up.pt (I.R.-V.); andreasofiat@med.up.pt (A.T.); carlosmartins20@gmail.com (C.M.)
 - ² USF Esgueira +, 3800-322 Aveiro, Portugal
 - ³ MEDCIDS—Department of Community Medicine, Information and Decision in Health, Faculty of Medicine, University of Porto, 4099-002 Porto, Portugal
 - ⁴ Escola Superior de Saúde da Cruz Vermelha Portuguesa, 1300-125 Lisbon, Portugal
 - ⁵ Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal; cristiano.r.matos@gmail.com
 - ⁶ Escola Superior de Tecnologia da Saúde de Coimbra, Instituto Politécnico de Coimbra, 3045-093 Coimbra, Portugal
 - ⁷ Porto Pharmacovigilance Centre, Faculty of Medicine, University of Porto, 4099-002 Porto, Portugal
 - ⁸ IPVC—Instituto Politécnico de Viana do Castelo, 4900-347 Viana do Castelo, Portugal
- * Correspondence: luismonteiro.net@gmail.com

Abstract: Inappropriate prescribing, which encompasses the prescription of potentially inappropriate medications (PIMs) and potential prescribing omissions (PPOs), is a common problem for older people. The STOPP/START tool enables general practitioners, who are the main prescribers, to identify and reduce the incidence of PIMs and PPOs and appraise an older patient's prescribed drugs during the diagnosis process to improve the clinical care quality. This study aimed to translate and validate the STOPP/START screening tool to enable its use by Portuguese physicians. A translation-back translation method including the validation of the obtained Portuguese version was used. Intra- and inter-rater reliability and agreement analyses were used in the validation process. A dataset containing the information of 334 patients was analyzed by one GP twice within a 2-week interval, while a dataset containing the information of 205 patients was independently analyzed by three GPs. Intra-rater reliability assessment led to a Kappa coefficient (κ) of 0.70 (0.65–0.74) for the STOPP criteria and 0.60 (0.52–0.68) for the START criteria, considered to be substantial and moderate values, respectively. The results of the inter-rater reliability rating were almost perfect for all combinations of raters ($\kappa > 0.93$). The version of the STOPP/START criteria translated into Portuguese represents an improvement in managing the medications prescribed to the elderly. It provides clinicians with a screening tool for detecting potentially inappropriate prescribing in patients older than 65 years old that is reliable and easy to use.

Keywords: geriatric medicine; quality in health care; general medicine; STOPP/START



Citation: Monteiro, L.; Monteiro-Soares, M.; Matos, C.; Ribeiro-Vaz, I.; Teixeira, A.; Martins, C. Inappropriate Prescriptions in Older People—Translation and Adaptation to Portuguese of the STOPP/START Screening Tool. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6896. <https://doi.org/10.3390/ijerph19116896>

Academic Editor: Paul B. Tchounwou

Received: 17 March 2022

Accepted: 2 June 2022

Published: 4 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Today, it is globally accepted that adverse drug reactions (ADRs) are a public health problem and have a significant clinical impact related to morbidity and mortality, which results in the increased use of health services in developed countries [1,2]. ADRs are responsible for about 7% of all hospital admissions, many of which are considered preventable [2,3]. Additionally, about 2–3% of patients admitted with an ADR die as a result [2,4].

ADRs may occur in 6–20% of patients admitted to hospitals, increasing their hospitalization period; highly increasing the costs associated with healthcare [5]; indirectly impacting patients' and their families' economic, social context, and psychological well-being [6,7]; and leading to the discussion of patient participation and involvement in pharmacovigilance [8,9].

The number of older adults is increasing [10], as is their life expectancy [11,12], and these patients are more likely to have more than one chronic disease [13,14] and be prescribed multiple drugs, increasing their susceptibility to inappropriate medication use [15–21].

Inappropriate prescribing that encompasses potentially inappropriate medications (PIMs) and potential prescribing omissions (PPOs) is a common problem for older people and is closely related to adverse events and ADRs [22]. Older adults are more prone to drug-related problems, as most take several medicines for multiple comorbidities, described as polypharmacy [23,24].

It is necessary to reduce PIMs and PPOs and improve clinical care quality [25]. The STOPP/START (Screening Tool of Older Persons' Prescriptions—STOPP; Screening Tool to Alert to Right Treatment—START) criteria for the use of potentially inappropriate medication in older people recognize the dual nature of inappropriate prescribing by including a list of PIMs (STOPP criteria) and PPOs (START criteria).

STOPP/START is a valid, reliable, and comprehensive screening tool that enables the prescribing physician to appraise an older patient's prescribed drugs in the context of their diagnosis [26]. Since the first publication of the STOPP/START criteria in 2008 [26], the tool has been widely disseminated and validated in many countries at different levels of healthcare (primary care, hospitals, nursing homes). The latest version (version 2) was published in 2014 and consists of 114 criteria, including 80 STOPP criteria and 34 START criteria [20,27]. These criteria are based on an up-to-date literature review and consensus validation among a European panel of experts [20]. The STOPP/START criteria were translated and adapted from English into several languages such as Czech, French [28–30], Italian, Spanish [31,32], and Dutch [33] to facilitate the local application of the criteria worldwide and have had a positive impact on patient evaluation [26].

This tool identifies potentially inappropriate prescriptions (PIPs) [34,35], identifying potentially inappropriate medicines through the STOPP criteria and identifying potential prescription omissions through the START criteria. The prevalence of patients with at least one instance of PIP identified by the STOPP criteria ranges from 21% [36] to 79% [37]. However, this range should be interpreted cautiously due to the heterogeneity of the sample population and study design between the different studies where this tool was assessed. The START criteria have identified at least one instance of PPO in 23% [36] to 74% [37] of patients.

A recent comparison of tools used to identify PIMs showed that the STOPP version 2 criteria identified substantially more PIMs than the EU (7)-PIM list [38], PRISCUS—Potentially Inappropriate Medications in the Elderly list [39–41], FORTA [39,40], and Beers criteria [25,42–46]. The STOPP/START criteria were found to be significantly associated with detecting adverse events in acutely ill older people, unlike the Beers criteria [20,42–45]. Compared to the Beers criteria or the prescribing indicators provided in the Elderly Australia criteria, the number and scope of drug-related problems identified were found to be best represented by the STOPP/START criteria [20,47,48]. Another advantage of this tool is that it considers PIMs and the indications to start an appropriate medication (START) [18].

A previous study from Gallagher et al., concluded that the STOPP/START criteria are generalizable across different European countries and languages [49]. Despite this, in other countries, such as in those with resource-limited healthcare settings, the original STOPP/START criteria might not be directly applicable; thus, modified versions of the STOPP/START criteria have been developed and validated recently [24]. In Portugal, this tool has already been used by Portuguese authors, but the translation and adaptation of the criteria have never been carried out, and the original tool is still used [25,50–53]. The current study aimed to translate and validate the STOPP/START screening tool to enable its use by Portuguese general practitioners/family physicians.

2. Materials and Methods

This study was conducted in four phases: The first phase (phase I) was the translation and adaptation of the STOPP/START screening tool to the Portuguese language, followed

by patient data collection (phase II). Phase III consisted of an intra-rater reliability and agreement study, and phase IV consisted of an inter-rater reliability and agreement study. Pre-registration on the ‘Open Registries Network’ was conducted ([DOI10.17605/OSF.IO/SK2RJ](https://doi.org/10.17605/OSF.IO/SK2RJ)) (accessed on 31 March 2021), and the translation and adaptation of the STOPP/START screening tool to Portuguese has been described elsewhere [18].

2.1. Phase I: Translation and Adaptation of the STOPP/START Screening Tool to the Portuguese Language

The translation and adaptation of the STOPP/START screening tool followed the Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes Measures [30]. The adaptation and translation were carried out based on the 2014 O’Mahony et al., version of STOPP/START [20]. Permission from the STOPP/START’s authors to translate, adapt, and validate this tool for use in Portuguese was obtained by email. The final version was distributed to 15 general practitioners to verify if there were any interpretation issues and improve clarity. The research team analyzed the results obtained from applying the STOPP/START tool and prepared the final version. As a translation was needed, the chance for possible disagreements between raters was reduced by validating these translations before studying the intra-rater and inter-rater agreements. The detailed procedure was published previously in the protocol [18].

2.2. Phase II: Data Collection of Patient Data

Patients were randomly selected from a list of patients aged > 65 years old from a primary care center in the Centre Region of Portugal, following which a total of 8165 patients were followed, with 1625 aged over 65 years old. The sample size was calculated in the published protocol, and 334 subjects were randomly selected to participate in the study [26,54]. Exclusion criteria included incapacity or unwillingness to provide written informed consent, diagnosis of psychotic disorder, institutionalization, and the presence of terminal illness. Patients were interviewed during previously scheduled medical appointments. Every participant signed a written consent form (Supplementary Materials File S1). The identity of all participants was protected throughout the study.

Sociodemographic data such as age, sex, and educational level were collected and are shown in Table 1. Clinical data were collected by consulting health record registries and conducting interviews of clinical patients, including the identification of the total number of medications used for chronic diseases, any prescribed drugs, dosage, pharmaceutical dosage, pharmaceutical form and route of administration, reason for taking medication, allergies, drug-related conditions, history of adverse drug reactions, and current or past conditions/diseases. Other clinical information was also collected and described in the protocol but not used in the adaptation of the STOPP/START tool [18]. The information collected was input into a database, each patient was numbered from 1 to 334 by the main investigator, and the record of the coding was stored offline in an Excel 2016[®] spreadsheet. All data recorded during this study will be stored for 5 years after the closure of the investigation, following the Portuguese Clinical Research Law. After this period, data containing participant codes will be destroyed.

2.3. Phase III: Intra-Rater Reliability and Agreement Study

As previously proposed by Kottner et al. [55], reliability may be defined as the ability of a measurement to differentiate among subjects or objects, comprising the ratio of variability between subjects or objects to the total variability of all measurements in the sample [56,57]. Intra-rater agreement assesses the extent to which the two responses from the same rater are concordant [58]. By definition, intra-rater reliability refers to the consistency of data recorded by the same rater, using the same scale, classification, instrument, or procedure, to assess the same subjects or objects at different times one rater over several trials. It is best determined when multiple trials are administered over a short period [55]. An independent researcher physician (named investigator/rater ‘A’) applied the Portuguese version of the STOPP/START criteria to all patient data collected in phase II. The investigator/rater ‘A’

was a family doctor with more than 10 years of experience in primary care, which included caring for and making daily prescriptions for older adults. To ensure intra-rater reliability and agreement, two weeks later, investigator A re-administered the tool. Both assessments of rater 'A' were used to study the intra-rater reliability.

Table 1. Characteristics of patients ($n = 334$).

Variable	<i>n</i> (%)
Age, years mean (SD); min–max	74.2 (6.9); 65–99
65–69 years	105 (31.4%)
70–74 years	71 (21.3%)
75–79 years	78 (23.4%)
80–84 years	50 (15.0%)
85+ years	30 (9.0%)
Sex	
Women	159 (47.6%)
Men	175 (52.4%)
Education level	
Early childhood, primary and lower secondary education (level 0–2)	316 (94.6%)
Upper secondary and post-secondary non-tertiary education (levels 3–4)	17 (5.1%)
Short-cycle tertiary education, Bachelor's, Master's, or Doctorate (levels 5–8)	1 (0.3%)
Number of medicines used	
0–1	57 (17.1%)
2 to 4 (Minor polypharmacy)	66 (19.8%)
5 to 9 (Major polypharmacy)	210 (62.8%)
10+ (Severe polypharmacy)	1 (0.3%)

Discrepancies in totals are due to rounding.

2.4. Phase IV: Inter-Rater Reliability and Agreement Study

Inter-rater reliability refers to the consistency of data recorded by different raters, using the same scale, classification, instrument, or procedure, to assess the same subjects or objects. Inter-rater agreement assess the extent to which the responses of two or more independent raters are concordant [58]. In this specific study, intra-rater and inter-rater reliability assist in determining if the measurement tool produces results that can be used by a clinician to make decisions confidently [55]. Three independent researchers (named investigators/raters 'B', 'C', and 'D') independently applied the Portuguese version of STOPP/START using the data collected in phase II. The investigators/raters 'B', 'C', and 'D' were family doctors with more than 10 years of experience in primary care. For the total of 334 subjects who participated in the intra-rater study, 205 patients were randomly selected for the inter-rater assessment [59]. These three physicians were independent investigators and only had contact with the authors to access the collected data. These investigators independently assessed the STOPP/START criteria in each of the 205 patients and were invited to provide written comments if necessary. Inter-rater agreement was assessed by comparing the results of the three raters. Between raters 'B', 'C', and 'D', an inter-rater reliability test was performed. Inter-rater reliability assessment is useful because observers will not necessarily interpret answers (or tools) in the same way and may disagree on how the constructed tool is used [60,61].

2.5. Statistical Analysis

Data were stored with Microsoft Excel 2016[®] software (Microsoft Corporation, Redmond, WA, USA). Data analyses were conducted using SPSS[®] V.27.0 (SPSS Inc, Chicago, IL, USA) and R Studio[®] V 1.3.1093 (Integrated Development for R. RStudio, PBC, Boston, MA, USA). Categorical variables were described using absolute and relative frequencies, n (%). Quantitative variables were summarized by means and their respective standard deviations (SDs), along with minimum and maximum values (min–max). According to the 'Guidelines for reporting reliability and agreement studies', reliability analyses and agreement analyses (intra- and inter-rater) were performed using Kappa statistics and the proportions of a

specific agreement, respectively [62–64]. The Kappa statistics were interpreted as poor if the score was ≤ 0.2 , fair if it was 0.21–0.40, moderate if it was 0.51–0.6, substantial if it was 0.61–0.8, and good if it was 0.81–1.00. The proportion of specific agreement distinguishes agreement on positive (PPos) or negative (PNeg) proportions, which might have different implications in clinical practice [65]. The 95% confidence intervals (95% CI) were presented for Kappa statistics and agreement proportions [55].

Kappa statistics were used for the calculation of both inter- and intra-rater reliability [66]. The Kappa statistic is a coefficient of reliability for categorical data [67]. As the Kappa coefficient is known to be affected by rare observations, it may not always reflect the true agreement rates and will provide an underestimation of the actual agreement [68]. A simple solution for this problem is calculating the proportions of agreement and separating the agreement rates into positive and negative agreements, thus making it easier for readers to interpret the results [63,69].

3. Results

A total of 334 patients were enrolled in this study. The patients' characteristics (age, sex, educational level, and number of medicines used) are described in Table 1.

Educational level was grouped according to the International Standard Classification of Education (ISCED 2011) [70]. The number of medicines used was grouped according to the definition of polypharmacy, grouping the number of medicines [23,71–73].

Intra-rater reliability and agreement involved the analysis of Rater A's evaluation of 334 patients' records and re-evaluation after a 2-week interval. Results are reported in Table 2 (STOPP) and 3 (START). Inter-rater reliability and agreement analyses were performed by three different raters ('B', 'C', and 'D') who evaluated 205 randomized patients from the database. Each rater evaluated the same patients to allow for their comparison. The results obtained for the inter-rater reliability and agreement using the STOPP and START tools are shown in Tables 2 and 3, respectively.

Table 2. Intra- and inter-rater reliability and agreement based on the analysis of the STOPP criteria.

STOPP Criteria	Rater Combination	Agreement (%)		Reliability
		Ppos † (95% CI)	PNeg † (95% CI)	Kappa (95% CI)
Intra-rater	Rater A Rater A	94.2 (93.1–95.1)	75.2 (70.9–79.1)	0.70 (0.65–0.74)
Inter-rater	Rater B Rater C	99.8 (99.4–99.9)	98.9 (97.1–99.7)	0.99 (0.97–1.00)
	Rater B Rater D	99.6 (99.1–99.8)	97.8 (95.5–99.1)	0.97 (0.95–0.99)
	Rater C Rater D	99.5 (99.1–99.8)	97.5 (95.0–98.8)	0.97 (0.95–0.99)
	Rater B Rater C Rater D	99.6 (99.3–99.8)	98.1 (96.6–99.2)	0.98 (0.94–1.00)

† Ppos, agreement on positive proportions. † PNeg, agreement on positive negative proportions.

Table 3. Intra- and Inter-rater reliability and agreement from the analysis of the START criteria.

START Criteria	Rater Combination	Agreement (%)		Reliability
		PPos † (95% CI)	PNeg † (95% CI)	Kappa (95% CI)
Intra-rater	Rater A Rater A	88.2 (85.4–90.6)	71.1 (64.5–76.8)	0.60 (0.52–0.68)
Inter-rater	Rater B Rater C	98.7 (97.3–99.4)	94.4 (88.6–97.7)	0.93 (0.87–0.99)
	Rater B Rater D	98.7 (97.3–99.4)	94.4 (88.6–97.7)	0.93 (0.87–0.99)
	Rater C Rater D	99.1 (97.9–99.7)	96.1 (90.1–98.7)	0.95 (0.91–1.00)
	Rater B Rater C Rater D	98.8 (97.9–99.6)	94.9 (90.8–98.2)	0.94 (0.87–1.00)

† PPos, agreement on positive proportions. † PNeg, agreement on positive negative proportions.

For the STOPP criteria, the intra-rater reliability showed a Kappa coefficient of 0.70 [95% CI 0.65–0.74], considered substantial; the positive and negative proportions of agreement obtained were 94.2% [95% CI 93.1–95.1] and 75.2% [95% CI 70.9–79.1], respectively. The results obtained for the inter-rater reliability were almost perfect, with κ near to

one in all possible combinations of raters. Inter-rater agreement determines the agreement between pairs of raters and all raters' judgments regarding the STOPP criteria.

For the START criteria, the intra-rater reliability showed a Kappa coefficient of 0.60 [0.52–0.68], considered a moderate value; the positive and negative proportions of agreement obtained were, respectively, 88.2% [85.4–90.6] and 71.1% [64.5–76.8]. The inter-rater reliability results were almost perfect, with κ near to one in all possible combinations of raters. Inter-rater agreement determines the agreement between pairs of raters and all raters' judgments regarding the START criteria.

The final version of the Portuguese adaptation of STOPP/START is presented in Supplementary Material File S2.

4. Discussion

This is to the best of our knowledge, the first study to translate and adapt the STOPP/START screening tool to Portuguese. The intra-rater reliability and inter-rater reliability scores obtained were not inferior to those obtained in previous studies conducted in other languages [28–33].

When testing reliability, several approaches are taken to determine consistency [74,75]. However, according to Innes et al., test–retest reliability, intra-rater reliability, and inter-rater reliability are the most common measures used among work-related assessments [74,75].

The first source of intra-rater inconsistency could be explained by various factors related to the assessment process. Rater A presented a high disagreement between two STOPP and START criteria evaluations. A major explanation was based on the analysis of the discrepancies. From 129 discrepancies seen between the first and second evaluation, 94 were related to proton pump inhibitors (F2 or A1 criteria without further investigation). In the second evaluation, with a better knowledge of the tool, the drug was properly assessed. Out of 119 discrepancies found in the evaluation on the START criteria, 51 were related to the introduction of vaccines in the second evaluation (I1 or I2 criteria were used). According to previous studies, a high level of familiarity is required to efficiently apply the STOPP/START criteria in clinical practice [49]. Additionally, raters could differ concerning their experience, specialties, and professional skills and have different perceptions regarding the knowledge required to use a particular item of the assessment tool. It is therefore important to highlight that the professionals that perform medication reviews with the STOPP/START tool should receive adequate training in order to use the tool appropriately [76,77].

One strength of this study is its innovation, with it representing the first development of a Portuguese version of the STOPP/START criteria. Our research was not merely a translation, but also an adaptation carried out by independent general practitioners that will hopefully increase the use of this version in the primary care setting. To ensure intra-rater reliability and agreement, the same doctor re-evaluated patients' records by applying the same criteria 2 weeks later, avoiding recall bias. Additionally, this study provides evidence for a near-perfect inter-rater reliability, meaning that raters almost always agree on whether to exclude/include medicines, although the reasons for these decisions were not necessarily similar. Finally, this version translated into Portuguese can be used by general practitioners or any other medical practitioner and could be used in countries where Portuguese is the main language. However, the differences in healthcare systems between countries; the different ranges of medicines available; and differences in population characteristics, such as genetic or racial differences, should be considered.

One potential limitation was related to the fact that the adapted version of STOPP/START exclusively focuses on primary healthcare centers. The authors deliberately did not include patients with specific pathologies. It is important to clarify that this tool may not be appropriate for use in all population groups or in different healthcare settings, and the assessment tool should be evaluated in future studies, including in other populations with specific pathologies and in different contexts.

Furthermore, some of the randomized patients ($n = 26, 8\%$) did not have any drugs prescribed, which would have reduced discrepancies between the raters when evaluating the STOPP criteria. Another potential limitation is the fact that the current tool was originally published in 2014, which means that there may be new medication and/or additional therapeutic indications that do not fit the current tool. Finally, the raters' decision to stop or start a drug based on this tool was a dichotomous decision and was not validated as either right or wrong from a clinical point of view. No assessment of clarity was performed, so a quality appraisal study should be conducted in the future to improve the clarity of clinical practice guidelines on a language level and enhance its clinical applicability [78].

In addition, using this tool, raters can point out different reasons for withdrawing or adding drugs without this changing the final decision. Since the criteria were applied to data from files in the absence of a clinical evaluation of patients by raters and prescriptions are subject to a certain variation in interpretation concerning the clinical heterogeneity observed in the elderly population, clinical evaluation was not performed by PPOs and the reasons for stopping and starting drugs were not compared [79].

5. Conclusions

The major research result of the current study was the adaptation of the STOPP/START (2014) criteria into Portuguese.

The objective of our research was not to test the tool in a Portuguese population. The use of this tool in this context may not lead to clinical differences for patients, or, at least, this was not the main objective for its use in this study.

The STOPP/START criteria have been proven to be a good tool for detecting potentially inappropriate prescriptions and improving prescription quality in older people in all healthcare settings, therefore leading to improved quality of life in patients, reducing the incidence of PIMs and PPOs, and improving clinical care quality. This research provides clinicians with a screening tool with which to detect potentially inappropriate prescribing in patients older than 65 years old that is easier to use for Portuguese native speakers. The tool is also useful for improving the training of medical students in managing polypharmacy [76] and can have a positive economic impact by reducing medicine expenditure in older patients [80,81]. This version in Portuguese represents a step forward in improving the management of medications in the elderly. The adaptation of this tool will be useful not only for Portugal but also for other Portuguese-language countries.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijerph19116896/s1>: Supplementary Materials File S1: Informed Consent; Supplementary Materials File S2: STOPP/START criteria—Portuguese version.

Author Contributions: Conceptualization, L.M.; methodology, L.M., C.M. (Carlos Martins), A.T. and M.M.-S.; writing—original draft preparation, L.M., C.M. (Cristiano Matos), A.T. and M.M.-S.; writing—review and editing, L.M., C.M. (Cristiano Matos), I.R.-V., A.T., M.M.-S. and C.M. (Carlos Martins). All authors have read and agreed to the published version of the manuscript.

Funding: This article was supported by National Funds through FCT—Fundação para a Ciência e a Tecnologia, I.P. within CINTESIS, R&D Unit (reference UIDB/4255/2020).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved on 30 July 2020 by the Ethics Committee of the Central Health Region of Portugal with the reference number 034-2020.

Informed Consent Statement: Informed consent was obtained from all the subjects involved in the study.

Data Availability Statement: The datasets presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Lazarou, J.; Pomeranz, B.H.; Corey, P.N. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA* **1998**, *279*, 1200–1205. [CrossRef] [PubMed]
2. Pirmohamed, M.; James, S.; Meakin, S.; Green, C.; Scott, A.K.; Walley, T.J.; Farrar, K.; Park, B.K.; Breckenridge, A.M. Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18,820 patients. *BMJ* **2004**, *329*, 15. [CrossRef] [PubMed]
3. Herdeiro, M.T.; Figueiras, A.; Polónia, J.; Gestal-Otero, J.J. Influence of pharmacists' attitudes on adverse drug reaction reporting. *Drug Saf.* **2006**, *29*, 331–340. [CrossRef] [PubMed]
4. Wester, K.; Jönsson, A.; Spigset, O.; Hägg, S. Spontaneously reported fatal suspected adverse drug reactions: A 10-year survey from Sweden. *Pharmacoepidemiol. Drug Saf.* **2007**, *16*, 173–180. [CrossRef]
5. Dormann, H.; Muth-Selbach, U.; Krebs, S.; Criegee-Rieck, M.; Tegeder, I.; Schneider, H.T.; Hahn, E.G.; Levy, M.; Brune, K.; Geisslinger, G. Incidence and costs of adverse drug reactions during hospitalisation. *Drug Saf.* **2000**, *22*, 161–168. [CrossRef]
6. Van Hunsel, F.; Härmark, L.; Pal, S.; Olsson, S.; van Grootheest, K. Experiences with Adverse Drug Reaction Reporting by Patients. *Drug Saf.* **2012**, *35*, 45–60. [CrossRef]
7. Matos, C.; van Hunsel, F.; Joaquim, J. Are consumers ready to take part in the Pharmacovigilance System?—A Portuguese preliminary study concerning ADR reporting. *Eur. J. Clin. Pharmacol.* **2015**, *71*, 883–890. [CrossRef]
8. Chinchilla, K.; Matos, C.; Hall, V.; van Hunsel, F. Patient organizations' barriers in pharmacovigilance and strategies to stimulate their participation. *Drug Saf.* **2021**, *44*, 181–191. [CrossRef]
9. Matos, C.; Weits, G.; van Hunsel, F. The Role of European Patient Organizations in Pharmacovigilance. *Drug Saf.* **2019**, *42*, 547–557. [CrossRef]
10. OECD. Elderly Population (Indicator). 2019. Available online: <https://data.oecd.org/pop/elderly-population.htm#indicator-chart> (accessed on 12 May 2019).
11. World Health Organization. *World Report on Ageing and Health*; World Health Organization: Geneva, Switzerland, 2015.
12. World Health Organization. 2019. Available online: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (accessed on 21 November 2021).
13. Formiga, F.; Ferrer, A.; Sanz, H.; Marengoni, A.; Albuquerque, J.; Pujol, R.; on behalf of the Octabaix Study Members. Patterns of comorbidity and multimorbidity in the oldest old: The Octabaix study. *Eur. J. Intern. Med.* **2013**, *24*, 40–44. [CrossRef]
14. Prazeres, F.; Santiago, L. Prevalence of multimorbidity in the adult population attending primary care in Portugal: A cross-sectional study. *BMJ Open* **2015**, *5*, e009287. [CrossRef] [PubMed]
15. Marengoni, A.; Angleman, S.; Melis, R.; Mangialasche, F.; Karp, A.; Garmen, A.; Meinow, B.; Fratiglioni, L. Aging with multimorbidity: A systematic review of the literature. *Ageing Res. Rev.* **2011**, *10*, 430–439. [CrossRef] [PubMed]
16. Weng, M.-C.; Tsai, C.-F.; Sheu, K.-L.; Lee, Y.-T.; Lee, H.-C.; Tzeng, S.-L.; Ueng, K.-C.; Chen, C.-C.; Chen, S.-C. The impact of number of drugs prescribed on the risk of potentially inappropriate medication among outpatient older adults with chronic diseases. *QJM Int. J. Med.* **2013**, *106*, 1009–1015. [CrossRef] [PubMed]
17. Motter, F.R.; Fritzen, J.S.; Hilmer, S.N.; Paniz, É.V.; Paniz, V.M.V. Potentially inappropriate medication in the elderly: A systematic review of validated explicit criteria. *Eur. J. Clin. Pharmacol.* **2018**, *74*, 679–700. [CrossRef]
18. Monteiro, L.; Monteiro-Soares, M.; Mendonça, L.V.; Ribeiro-Vaz, I.; Martins, C.; Teixeira, A. Translation and adaptation of the STOPP-START screening tool to Portuguese for detecting inappropriate prescriptions in older people: A protocol. *BMJ Open* **2021**, *11*, e043746. [CrossRef]
19. Atkin, P.A.; Veitch, P.C.; Veitch, E.M.; Ogle, S.J. The epidemiology of serious adverse drug reactions among the elderly. *Drugs Aging* **1999**, *14*, 141–152. [CrossRef]
20. O'Mahony, D.; O'Sullivan, D.; Byrne, S.; O'Connor, M.N.; Ryan, C.; Gallagher, P. STOPP/START criteria for potentially inappropriate prescribing in older people: Version 2. *Age Ageing* **2014**, *44*, 213–218. [CrossRef]
21. Onder, G.; Van Der Cammen, T.J.M.; Petrovic, M.; Somers, A.; Rajkumar, C. Strategies to reduce the risk of iatrogenic illness in complex older adults. *Age Ageing* **2013**, *42*, 284–291. [CrossRef]
22. O'Connor, M.N.; Gallagher, P.; O'Mahony, D. Inappropriate prescribing. *Drugs Aging* **2012**, *29*, 437–452. [CrossRef]
23. Masnoon, N.; Shakib, S.; Kalisch-Ellett, L.; Caughey, G.E. What is polypharmacy? A systematic review of definitions. *BMC Geriatr.* **2017**, *17*, 230. [CrossRef]
24. Samaranayake, N.R.; Balasuriya, A.; Fernando, G.H.; Samaraweera, D.; Shanika, L.G.T.; Wanigasuriya, J.K.P.; Wijekoon, C.N.; Wanigatunge, C.A. 'Modified STOPP-START criteria for Sri Lanka'; translating to a resource limited healthcare setting by Delphi consensus. *BMC Geriatr.* **2019**, *19*, 282. [CrossRef] [PubMed]
25. Monteiro, C.; Canário, C.; Ribeiro, M.Â.; Duarte, A.P.; Alves, G. Medication Evaluation in Portuguese Elderly Patients According to Beers, STOPP/START Criteria and EU (7)-PIM List—An Exploratory Study. *Patient Prefer. Adherence* **2020**, *14*, 795. [CrossRef] [PubMed]
26. Gallagher, P.; Ryan, C.; Byrne, S.; Kennedy, J.; O'Mahony, D. STOPP (screening tool of older person's prescriptions) and START (screening tool to alert doctors to right treatment). Consensus validation. *Int. J. Clin. Pharmacol. Ther.* **2008**, *46*, 72. [CrossRef] [PubMed]
27. Barenholtz Levy, H.; Marcus, E.-L. Potentially inappropriate medications in older adults: Why the revised criteria matter. *Ann. Pharmacother.* **2016**, *50*, 599–603. [CrossRef]

28. Lang, P.-O.; Dramé, M.; Guignard, B.; Mahmoudi, R.; Payot, I.; Latour, J.; Schmitt, E.; Pepersack, T.; Vogt-Ferrier, N.; Hasso, Y.; et al. Les critères STOPP/START. v2: Adaptation en langue française. *NPG Neurol.* **2015**, *15*, 323–336.
29. Lang, P.-O.; Hasso, Y.; Belmin, J.; Payot, I.; Baeyens, J.-P.; Vogt-Ferrier, N.; Gallagher, P.; O'Mahony, D.; Michel, J.P. STOPP-START: Adaptation of a French language screening tool for detecting inappropriate prescriptions in older people. *Can. J. Public Health Rev. Can. Sante Publique* **2009**, *100*, 426–431. [[CrossRef](#)]
30. Wild, D.; Grove, A.; Martin, M.; Eremenco, S.; McElroy, S.; Verjee-Lorenz, A.; Erikson, P. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value Health* **2005**, *8*, 94–104. [[CrossRef](#)]
31. Silveira, E.D.; García, M.M.; Errasquin, B.M.; Castellano, C.S.; Gallagher, P.F.; Cruz-Jentoft, A.J. Prescripción inapropiada de medicamentos en los pacientes mayores: Los criterios STOPP/START. *Rev. Esp. Geriatr. Gerontol.* **2009**, *44*, 273–279. [[CrossRef](#)]
32. Silveira, E.D.; Errasquin, B.M.; García, M.M.; Vélez-Díaz-Pallarés, M.; Montoya, I.L.; Sánchez-Castellano, C.; Cruz-Jentoft, A.J. Mejorando la prescripción de medicamentos en las personas mayores: Una nueva edición de los criterios STOPP-START. *Rev. Esp. Geriatr. Gerontol.* **2015**, *50*, 89–96. [[CrossRef](#)]
33. Knol, W.; Verduijn, M.M.; der Zande, A.C.; van Marum, R.J.; Brouwers, J.R.; der Cammen, T.J.; Petrovic, M.; Jansen, P.A. Detecting inappropriate medication in older people: The revised STOPP/START criteria. *Ned. Tijdschr. Voor Geneesk.* **2015**, *159*, A8904.
34. Renom-Guiteras, A.; Thümann, P.A.; Miralles, R.; Klaaßen-Mielke, R.; Thiem, U.; Stephan, A.; Bleijlevens, M.H.C.; Jolley, D.; Leino-Kilpi, H.; Hallberg, I.R.; et al. Potentially inappropriate medication among people with dementia in eight European countries. *Age Ageing* **2018**, *47*, 68–74. [[CrossRef](#)] [[PubMed](#)]
35. Sennesael, A.-L.; Dalleur, O.; Henrard, S.; Artoisenet, C.; Schoevaerdt, D.; Spinewine, A. Implementing a screening tool to improve prescribing in hospitalized older patients: A pilot study. *Int. J. Clin. Pharm.* **2018**, *40*, 15–19. [[CrossRef](#)] [[PubMed](#)]
36. Ryan, C.; O'Mahony, D.; Kennedy, J.; Weedle, P.; Byrne, S. Potentially inappropriate prescribing in an Irish elderly population in primary care. *Br. J. Clin. Pharmacol.* **2009**, *68*, 936–947. [[CrossRef](#)] [[PubMed](#)]
37. Garcia-Gollarte, F.; Baleriola-Júlvez, J.; Ferrero-López, I.; Cruz-Jentoft, A.J. Inappropriate drug prescription at nursing home admission. *J. Am. Med. Dir. Assoc.* **2012**, *13*, 83.e9–83.e15. [[CrossRef](#)]
38. Renom-Guiteras, A.; Meyer, G.; Thümann, P.A. The EU (7)-PIM list: A list of potentially inappropriate medications for older people consented by experts from seven European countries. *Eur. J. Clin. Pharmacol.* **2015**, *71*, 861–875. [[CrossRef](#)]
39. Lavrador, A.M.A.A.P. *Avaliação da Medicação Inapropriada em Idosos Institucionalizados: Comparação entre Diversos Critérios (Beers, Start/Stop, Forta e Priscus)*; University of Coimbra: Coimbra, Portugal, 2015.
40. Wickop, B.; Härterich, S.; Sommer, C.; Daubmann, A.; Baehr, M.; Langebrake, C. Potentially inappropriate medication use in multimorbid elderly inpatients: Differences between the FORTA, PRISCUS and STOPP ratings. *Drugs Real World Outcomes* **2016**, *3*, 317–325. [[CrossRef](#)]
41. Holt, S.; Schmiedl, S.; Thümann, P.A. Potentially inappropriate medications in the elderly: The PRISCUS list. *Dtsch Arztebl. Int.* **2010**, *107*, 543.
42. Hamilton, H.; Gallagher, P.; Ryan, C.; Byrne, S.; O'Mahony, D. Potentially inappropriate medications defined by STOPP criteria and the risk of adverse drug events in older hospitalized patients. *Arch. Intern. Med.* **2011**, *171*, 1013–1019. [[CrossRef](#)]
43. Lam, M.P.S.; Cheung, B.M.Y. The use of STOPP/START criteria as a screening tool for assessing the appropriateness of medications in the elderly population. *Expert Rev. Clin. Pharmacol.* **2012**, *5*, 187–197. [[CrossRef](#)]
44. Hill-Taylor, B.; Sketris, I.; Hayden, J.; Byrne, S.; O'Sullivan, D.; Christie, R. Application of the STOPP/START criteria: A systematic review of the prevalence of potentially inappropriate prescribing in older adults, and evidence of clinical, humanistic and economic impact. *J. Clin. Pharm. Ther.* **2013**, *38*, 360–372. [[CrossRef](#)]
45. Tosato, M.; Landi, F.; Martone, A.M.; Cherubini, A.; Corsonello, A.; Volpato, S.; Bernabei, R.; Oner, G. Potentially inappropriate drug use among hospitalised older adults: Results from the CRIME study. *Age Ageing* **2014**, *43*, 767–773. [[CrossRef](#)] [[PubMed](#)]
46. American Geriatrics Society. Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. *J. Urol.* **2016**, *195*, 667–669.
47. Curtain, C.M.; Bindoff, I.K.; Westbury, J.L.; Peterson, G.M. A comparison of prescribing criteria when applied to older community-based patients. *Drugs Aging* **2013**, *30*, 935–943. [[CrossRef](#)] [[PubMed](#)]
48. Boland, B.; Guignard, B.; Dalleur, O.; Lang, P.-O. Application of STOPP/START and Beers criteria: Compared analysis on identification and relevance of potentially inappropriate prescriptions. *Eur. Geriatr. Med.* **2016**, *7*, 416–423. [[CrossRef](#)]
49. Gallagher, P.; Baeyens, J.-P.; Topinkova, E.; Madlova, P.; Cherubini, A.; Gasperini, B.; Cruz-Jentoft, A.; Montero, B.; Lang, P.O.; Michel, J.-P.; 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*, 603–606. [[CrossRef](#)]
50. Silva, C.; Ramalho, C.; Luz, I.; Monteiro, J.; Fresco, P. Drug-related problems in institutionalized, polymedicated elderly patients: Opportunities for pharmacist intervention. *Int. J. Clin. Pharm.* **2015**, *37*, 327–334. [[CrossRef](#)]
51. Da Costa, F.A.; Silvestre, L.; Periquito, C.; Carneiro, C.; Oliveira, P.; Fernandes, A.I.; Cavaco-Silva, P. Drug-related problems identified in a sample of Portuguese institutionalised elderly patients and pharmacists' interventions to improve safety and effectiveness of medicines. *Drugs Real World Outcomes* **2016**, *3*, 89–97. [[CrossRef](#)]
52. Da Costa, F.A.; Periquito, C.; Carneiro, M.C.; Oliveira, P.; Fernandes, A.I.; Cavaco-Silva, P. Potentially inappropriate medications in a sample of Portuguese nursing home residents: Does the choice of screening tools matter? *Int. J. Clin. Pharm.* **2016**, *38*, 1103–1111. [[CrossRef](#)]

53. Candeias, C.; Gama, J.; Rodrigues, M.; Falcão, A.; Alves, G. Potentially Inappropriate Medications and Potential Prescribing Omissions in Elderly Patients Receiving Post-Acute and Long-Term Care: Application of Screening Tool of Older People's Prescriptions/Screening Tool to Alert to Right Treatment Criteria. *Front. Pharmacol.* **2021**, *12*, 2852. [[CrossRef](#)]
54. Rotondi, M.A. *kappaSize: Sample Size Estimation Functions for Studies of Interobserver Agreement, R Package version 1*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
55. Kottner, J.; Audigé, L.; Brorson, S.; Donner, A.; Gajewski, B.J.; Hróbjartsson, A.; Roberts, C.; Shoukri, M.; Streiner, D.L. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int. J. Nurs. Stud.* **2011**, *48*, 661–671. [[CrossRef](#)]
56. Streiner, D.L.; Norman, G.R.; Cairney, J. *Health Measurement Scales: A Practical Guide to Their Development and Use*; Oxford University Press: Oxford, UK, 2015.
57. Dunn, G. *Statistical Evaluation of Measurement Errors*; Hodder Education Publishers: London, UK, 2004.
58. Gisev, N.; Bell, J.S.; Chen, T.F. Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Res. Soc. Adm. Pharm.* **2013**, *9*, 330–338. [[CrossRef](#)] [[PubMed](#)]
59. Shoukri, M.M. Reliability for continuous scale measurements. In *Measures of Interobserver Agreement and Reliability*; CRC Press: Boca Raton, FL, USA, 2004.
60. Mohajan, H.K. Two criteria for good measurements in research: Validity and reliability. *Ann. Spiru Haret Univ. Econ. Ser.* **2017**, *17*, 59–82. [[CrossRef](#)]
61. McDonald, N.; Schoenebeck, S.; Forte, A. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proc. ACM Hum.-Comput. Interact.* **2019**, *3*, 72. [[CrossRef](#)]
62. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159–174. [[CrossRef](#)] [[PubMed](#)]
63. Cicchetti, D.V.; Feinstein, A.R. High agreement but low kappa: II. Resolving the paradoxes. *J. Clin. Epidemiol.* **1990**, *43*, 551–558. [[CrossRef](#)]
64. De Vet, H.C.W.; Mokkink, L.B.; Terwee, C.B.; Hoekstra, O.S.; Knol, D.L. Clinicians are right not to like Cohen's κ . *BMJ* **2013**, *346*, f2125. [[CrossRef](#)]
65. De Vet, H.C.W.; Dikmans, R.E.; Eekhout, I. Specific agreement on dichotomous outcomes can be calculated for more than two raters. *J. Clin. Epidemiol.* **2017**, *83*, 85–89. [[CrossRef](#)]
66. Cohen, J. *Statistical Power Analysis for the Behavior Science: Lawrence Erlbaum Association*; Academic Press: Cambridge, MA, USA, 1988.
67. Kierkegaard, M.; Tollbäck, A. Inter- and intra-rater reliability of the B. Lindmark Motor Assessment. *Adv. Physiother.* **2005**, *7*, 2–6. [[CrossRef](#)]
68. Viera, A.J.; Garrett, J.M. Understanding interobserver agreement: The kappa statistic. *Fam. Med.* **2005**, *37*, 360–363.
69. Feinstein, A.R.; Cicchetti, D.V. High agreement but low kappa: I. The problems of two paradoxes. *J. Clin. Epidemiol.* **1990**, *43*, 543–549. [[CrossRef](#)]
70. UNESCO Institute for Statistics. International standard classification of education: ISCED 2011. *Comp. Soc. Res.* **2012**, *30*, 85.
71. Housley, B.C.; Stawicki, S.P.A.; Evans, D.C.; Jones, C. Comorbidity-polypharmacy score predicts readmission in older trauma patients. *J. Surg. Res.* **2015**, *199*, 237–243. [[CrossRef](#)] [[PubMed](#)]
72. Nagaraju, B.; Padmavathi, G.V.; Dattathreya, G. Prevalence and assessment of polypharmacy in Sri Devraj URS medical college & hospital. *Kolar Int. J. Pharm. Pharm. Sci.* **2012**, *4*, 488–493.
73. Jorgensen, T.L.; Herrstedt, J.; Friis, S.; Hallas, J. Polypharmacy and drug use in elderly Danish cancer patients during 1996 to 2006. *J. Geriatr. Oncol.* **2012**, *3*, 33–40. [[CrossRef](#)]
74. Innes, E.; Straker, L. Reliability of work-related assessments. *Work* **1999**, *13*, 107–124. [[PubMed](#)]
75. Scheel, C.; Mecham, J.; Zuccarello, V.; Mattes, R. An evaluation of the inter-rater and intra-rater reliability of OccuPro's functional capacity evaluation. *Work* **2018**, *60*, 465–473. [[CrossRef](#)] [[PubMed](#)]
76. Keijsers, C.J.P.W.; van Doorn, A.B.D.; van Kalles, A.; de Wildt, D.J.; Brouwers, J.R.B.J.; de Kamp, H.J.; Jansen, P.A.F. Structured pharmaceutical analysis of the systematic tool to reduce inappropriate prescribing is an effective method for final-year medical students to improve polypharmacy skills: A randomized controlled trial. *J. Am. Geriatr. Soc.* **2014**, *62*, 1353–1359. [[CrossRef](#)]
77. Verdoorn, S.; Kwint, H.-F.; Faber, A.; Gussekloo, J.; Bouvy, M.L. Majority of drug-related problems identified during medication review are not associated with STOPP/START criteria. *Eur. J. Clin. Pharmacol.* **2015**, *71*, 1255–1262. [[CrossRef](#)]
78. Sallevelt, B.T.G.M.; Huibers, C.J.A.; Knol, W.; van Puijenbroek, E.; Egberts, T.; Wilting, I. Evaluation of clarity of the STOPP/START criteria for clinical applicability in prescribing for older people: A quality appraisal study. *BMJ Open* **2020**, *10*, e033721. [[CrossRef](#)]
79. Nelson, E.A.; Dannefer, D. Aged heterogeneity: Fact or fiction? The fate of diversity in gerontological research. *Gerontologist* **1992**, *32*, 17–23. [[CrossRef](#)]
80. Winit-Watjana, W.; Sakulrat, P.; Kespichayawattana, J. Criteria for high-risk medication use in Thai older patients. *Arch. Gerontol. Geriatr.* **2008**, *47*, 35–51. [[CrossRef](#)] [[PubMed](#)]
81. Bradley, M.C.; Fahey, T.; Cahir, C.; Bennett, K.; O'Reilly, D.; Parsons, C.; Hughes, C.M. Potentially inappropriate prescribing and cost outcomes for older people: A cross-sectional study using the Northern Ireland Enhanced Prescribing Database. *Eur. J. Clin. Pharmacol.* **2012**, *68*, 1425–1433. [[CrossRef](#)] [[PubMed](#)]