



FACULDADE DE MEDICINA  
UNIVERSIDADE DE  
**COIMBRA**

MESTRADO INTEGRADO EM MEDICINA – TRABALHO FINAL

ÁLVARO JOSÉ VICENTE BRANDÃO

***Estudo da fiabilidade e concordância intra-rater da versão  
portuguesa dos critérios STOPP/START***

ARTIGO CIENTÍFICO

ÁREA CIENTÍFICA DE MEDICINA GERAL E FAMILIAR

Trabalho realizado sob a orientação de:

DOUTOR LUÍS MIGUEL ANDRÉ MONTEIRO

DOUTORA INÊS ROSENDO CARVALHO E SILVA

FEVEREIRO/2022



# Índice

|                                 |    |
|---------------------------------|----|
| Resumo.....                     | 5  |
| Abstract.....                   | 5  |
| Índice de Abreviaturas.....     | 7  |
| Índice de Tabelas.....          | 8  |
| Introdução.....                 | 9  |
| Materiais e Métodos.....        | 13 |
| Resultados.....                 | 16 |
| Discussão e Conclusão.....      | 18 |
| Financiamento.....              | 19 |
| Conflitos de Interesse.....     | 19 |
| Referências Bibliográficas..... | 20 |

## **ESTUDO DA FIABILIDADE E CONCORDÂNCIA *INTRA-RATER* DA VERSÃO PORTUGUESA DOS CRITÉRIOS STOPP/START**

Álvaro José Vicente Brandão[1], Doutor Luís Miguel André Monteiro[2,3] e Doutora Inês Rosendo Carvalho e Silva[1,2,4].

[1] Faculdade de Medicina, Universidade de Coimbra, Portugal.

[2] CINTESIS – *Centre for Health Technology and Services Research*; Faculdade de Medicina, Universidade do Porto, Porto Portugal.

[3] USF Esgueira +.

[4] USF Coimbra Centro.

R. Dr. Plácido da Costa, 4200-450 Porto. [luismonteiro.net@gmail.com](mailto:luismonteiro.net@gmail.com).

## Resumo

**Introdução:** A multimorbilidade e a polifarmácia representam um desafio importante à prescrição racional de medicação na população idosa. Os critérios *Screening Tool of Older People's Prescriptions* (STOPP) – *Screening Tool to Alert to Right Treatment* (START) são associados a melhor qualidade de prescrição e melhores resultados clínicos. Este estudo consistiu na primeira tradução dos critérios STOPP/START para Português de Portugal e na avaliação da sua validade para os médicos portugueses de Medicina Geral e Familiar.

**Objetivos:** A tradução, adaptação e validação dos critérios STOPP/START para a língua Português de Portugal.

**Materiais e Métodos:** O estudo foi constituído por 4 fases. A primeira fase (fase I) foi a tradução dos critérios STOPP/START para Português de Portugal. A fase II foi um estudo analítico transversal e incluiu a colheita de dados clínicos de doentes. A fase III foi o estudo da fiabilidade e concordância *intra-rater*. A fase IV foi o estudo da fiabilidade e concordância *inter-rater* através da comparação dos resultados de 3 avaliadores.

**Resultados:** Para os critérios STOPP, o teste de fiabilidade *intra-rater* obteve um coeficiente Kappa de Cohen ( $\kappa$ ) de 0,70 (0,65-0,74), sendo considerado um valor substancial. As proporções positiva e negativa obtidas foram, respetivamente, 94,2% e 75,2%. Para os critérios START, o teste de fiabilidade *intra-rater* obteve um coeficiente Kappa de 0,60 (0,52-0,68), sendo considerado um valor moderado. As proporções positiva e negativa obtidas foram, respetivamente, 88,2% e 71,1%, com  $\kappa$  próximo de 1 em todas as avaliações.

**Discussão:** A ferramenta de rastreio STOPP/START é uma forma simples de gerir o tratamento de adultos mais velhos, sendo mais fácil de utilizar diariamente quando adaptada à linguagem do prescritor. Este estudo é inovador porque é o primeiro desenvolvimento de uma versão Portuguesa dos critérios STOPP/START com o objetivo de facilitar a prescrição apropriada de medicação em idosos e prevenir erros.

**Conclusão:** A ferramenta de rastreio STOPP/START é reprodutível e fiável.

**Palavras-chave:** critérios STOPP/START, adaptação, validação e Português de Portugal.

## Abstract

**Introduction:** Multimorbidity and polypharmacy represent a major challenge to rational medication prescribing in the elderly population. The Screening Tool of Older People's

Prescriptions (STOPP) – Screening Tool to Alert to Right Treatment (START) criteria are associated with better prescribing quality and clinical outcomes. This study consisted of the first translation of the STOPP/START criteria into Portuguese and the assessment of their validity for Portuguese General Practitioners.

**Objectives:** The translation, adaptation and validation of the STOPP/START criteria for the Portuguese language from Portugal.

**Materials and Methods:** The study consisted of 4 phases. The first phase (phase I) was the translation of the STOPP/START criteria into Portuguese. Phase II was a cross-sectional analytical study and included the collection of clinical data from patients. Phase III was the study of intra-rater reliability and agreement. Phase IV was the study of inter-rater reliability and agreement by comparing the results of 3 raters.

**Results:** For the STOPP criteria, the intra-rater reliability test obtained a Cohen's Kappa coefficient ( $\kappa$ ) of 0.70 (0.65-0.74), which is considered a substantial value. The positive and negative proportions obtained were 94.2% and 75.2%, respectively. For the START criteria, the intra-rater reliability test obtained a Kappa coefficient of 0.60 (0.52-0.68), which is considered a moderate value. The positive and negative proportions obtained were 88.2% and 71.1%, respectively, with  $\kappa$  close to 1 in all evaluations.

**Discussion:** The STOPP/START screening tool is a simple way to manage the treatment of older adults, being easier to use on a daily basis when adapted to the prescriber's language. This study is innovative because it is the first development of a Portuguese version of the STOPP/START criteria with the aim of facilitating appropriate medication prescribing in older adults and preventing errors.

**Conclusion:** The STOPP/START screening tool is reproducible and reliable.

**Keywords:** STOPP/START criteria, adaptation, validation and Portuguese from Portugal.

## Índice de Abreviaturas

EAM – eventos adversos evitáveis com medicamentos.

IC – intervalo de confiança.

$\kappa$  – coeficiente Kappa de Cohen.

MGF – Medicina Geral e Familiar.

MPI – medicamentos potencialmente inapropriados.

POP – potenciais omissões de prescrição.

PPI – prescrição potencialmente inapropriada.

START – *Screening Tool to Alert to Right Treatment*.

STOPP – *Screening Tool of Older People's Prescriptions*.

## Índice de Tabelas

1. Características dos doentes (n=334) ..... 16
2. Fiabilidade e concordância *intra-rater* a partir da análise dos critérios STOPP ..... 17
3. Fiabilidade e concordância *intra-rater* a partir da análise dos critérios START ..... 17

## Introdução

A população idosa é definida como pessoas com 65 anos ou mais e está a aumentar.[1-6] Com o avançar da idade, a acumulação gradual de danos moleculares e celulares causa a diminuição das reservas fisiológicas, o aumento do risco de doenças, o declínio geral da capacidade física e mental e, por último, a morte.[2,3,6-8] As doenças crónicas são a carga dominante dos cuidados de saúde e estão fortemente relacionadas com o envelhecimento.[9,10] Em Portugal, a prevalência de multimorbilidade é cerca de 72,7% na população adulta e pode ser maior do que 90% na população idosa.[1,7,10]

Os idosos são os maiores consumidores de farmacoterapia a nível mundial.[11] A polifarmácia pode ser definida como a toma diária e crónica de 5 ou mais medicamentos.[12,13] A polimedicação está associada a piores resultados clínicos, aumenta o risco de eventos adversos evitáveis com medicamentos (EAM), internamento hospitalar e custos relacionados e a sua prevalência está a aumentar.[6,8,12-17] Um EAM pode ser definido como qualquer efeito nocivo, involuntário e indesejável de um medicamento, excluindo falhas terapêuticas, envenenamento intencional ou acidental e abuso de drogas.[18] A prevalência de EAM pode ir até 35% na população idosa, sendo responsáveis por até 30% das admissões hospitalares de adultos mais velhos.[6,11]

A multimorbilidade e a polifarmácia representam um desafio importante à prescrição racional de medicação na população idosa.[1,9,13,19] Os idosos são mais suscetíveis ao uso inapropriado de medicação, uma vez que podem apresentar: mudanças relacionadas com a idade na farmacocinética e na farmacodinâmica; multimorbilidade e polimedicação; baixa adesão terapêutica a regimes de tratamento complexos e défice cognitivo e/ou físico.[1,6,8,11,13-15,17-20] Por outro lado, os doentes idosos são sub-representados ou excluídos da maioria dos ensaios clínicos sobre segurança e eficácia, resultando em evidência científica pobre.[17,19,20] Prescrição potencialmente inapropriada (PPI) é a prescrição de medicação em quantidade excessiva, insuficiente ou de forma incorreta e inclui medicamentos potencialmente inapropriados (MPI) e potenciais omissões de prescrição (POP).[1,4,6,11,17,20-22] Na Europa, 1 em cada 5 doentes mais velhos (22,6%) está exposto a PPI, um dos principais fatores de risco de EAM.[4,6,11]

MPI são medicamentos cujo uso tem, potencialmente, mais riscos do que benefícios, apesar de alternativas não-farmacológicas e farmacológicas mais seguras estarem disponíveis.[1,4,6,8,11,17,19-21] A utilização de MPI apresenta taxas de prevalência elevadas, desde 18 até mais de 40%, numa variedade de contextos de cuidados de saúde, sendo predominante em doentes idosos.[6,13,18-20] O uso de MPI pode causar EAM, tais como quedas, fraturas e delírium, e está associado a utilização ineficiente de recursos,

internamento hospitalar, incapacidade, institucionalização e mortalidade.[4,8,12,13,16,19,20,22] POP são medicamentos potencialmente benéficos, clinicamente indicados, mas que não foram prescritos, sem uma razão válida.[4,11,20,21] Na população idosa, a prevalência de POP é cerca de 22,7% nos cuidados de saúde primários e 57,8% nos cuidados hospitalares.[18] A omissão de medicamentos potencialmente benéficos pode ter vários motivos: falta de conhecimento de terapias preventivas secundárias; níveis baixos de expectativa terapêutica em doentes frágeis; desejo de evitar a polifarmácia; maior foco na palição dos sintomas do que na prevenção secundária de doenças e atitudes idadistas e sexistas negativas que conduzem ao niilismo terapêutico.[21]

Uma revisão sistemática sobre EAM em idosos em cuidados ambulatoriais referiu uma taxa de prevenção de até 53%, o que indica que uma gestão adequada da medicação nesta população pode ajudar a prevenir EAM graves.[6,19] A identificação de MPI em idosos pode ser facilitada pelo uso de ferramentas de rastreio como: Critérios de *Beers*; *Potentially Inappropriate Medications in the Elderly list* e a combinação de *Screening Tool of Older People's Prescriptions* (STOPP) – *Screening Tool to Alert to Right Treatment* (START).[1,4,13,23] Os critérios STOPP/START são uma ferramenta de rastreio baseada na evidência, desenvolvida por um painel de especialistas em Medicina e Farmacoterapia Geriátrica e validada usando a metodologia de consenso *Delphi*. [4,11,13,15,17,18,22]

Os critérios STOPP compreendem 81 casos comuns e evitáveis de PPI em idosos que predis põem a EAM, incluindo interações medicamentosas e entre medicamento-doença, medicamentos que aumentam o risco de queda e de declínio cognitivo e medicamentos de classes duplicadas.[4,11,13,15,17,18] Estão organizados por sistema fisiológico para facilitar a sua utilização, incluem explicações sobre a razão pela qual a medicação é inapropriada e estão significativamente associados com EAM em idosos que causam ou contribuem para a hospitalização urgente.[4,11,15,17,18,22] Os critérios START incluem 34 cenários de POP comuns no tratamento de idosos, divididos por sistema fisiológico, onde medicamentos específicos são recomendados.[4,11,15,17,18,21] A justaposição de MPI com POP dá uma avaliação holística e abrangente da higiene de prescrição.[11] Os critérios STOPP/START permitem: a identificação de MPI; a deteção de POP com conseqüente início da medicação apropriada; melhor adequação da medicação; a diminuição de EAM, do número de visitas aos cuidados de saúde primários e ao serviço de urgência, da duração de internamento hospitalar e de custos em medicação.[1,17,20-22]

A ferramenta de rastreio STOPP/START é considerada uma das mais exaustivas em termos do tipo de prescrição de MPI e de POP, do número de especialidades médicas abordadas e de temas investigados.[17] Em comparação com outras ferramentas, os critérios

STOPP/START permitem: identificar uma proporção significativamente maior de doentes idosos que necessitam de internamento hospitalar em consequência de eventos adversos relacionados com MPI; reduzir o maior número de medicamentos e identificar maior número de potenciais problemas clínicos importantes.[1,15,18] Ao contrário dos Critérios de *Beers*, esta ferramenta de rastreio está significativamente associada a EAM em idosos gravemente doentes.[22] Os critérios STOPP/START são associados a melhor qualidade de prescrição e melhores resultados clínicos.[1]

São necessários estudos controlados randomizados prospetivos para testar se a aplicação por rotina dos critérios STOPP/START pode melhorar significativamente a adequação da prescrição e reduzir a morbilidade relacionada com medicamentos.[4] Contudo, antes de demonstrar efeitos nos resultados dos doentes, uma ferramenta de rastreio deve ser generalizável e fiável.[4] Os estudos de fiabilidade e concordância fornecem informações sobre a quantidade de erro inerente a qualquer diagnóstico, medição ou pontuação, onde a quantidade de erro determina a validade dos resultados do estudo.[24] Fiabilidade é o rácio de variabilidade entre resultados dos mesmos participantes (por avaliadores diferentes ou em momentos diferentes) para a variabilidade total de todos os resultados na amostra.[1,24] A fiabilidade refere-se à consistência com que algo ocorre ou, inversamente, quanto impacte o erro ou o viés de medição tem nos dados colhidos.[25,26] A fiabilidade *intra-rater* (o mesmo avaliador) permite determinar o grau em que os resultados obtidos por um procedimento de medição podem ser replicados, sendo integral na conceção de estudos robustos.[25-27] Se o estudo for mal definido e houver ambiguidade nas suas categorias, então a qualidade dos dados será provavelmente pobre devido a um elevado erro de medição.[25] Concordância é o grau em que os resultados são idênticos: “*As observações são semelhantes ou diferem em que grau?*”. [1,24]

Os critérios STOPP/START foram traduzidos e adaptados para várias línguas europeias.[1] Gallagher et al. (2009) demonstrou que a fiabilidade *inter-rater* (avaliadores diferentes) dos critérios STOPP/START é boa quando testada entre múltiplos médicos que pratiquem independentemente em centros médicos europeus diferentes e que esta ferramenta de rastreio pode ser generalizada para países e línguas europeias diferentes.[1,4,18] Os critérios STOPP/START foram aplicados com sucesso, tanto para fins de investigação como para fins clínicos práticos, em vários países da Europa, Ásia, Austrália, América do Norte e América do Sul.[22] Como tal, também podem ter relevância global.[22]

Este estudo consistiu na primeira tradução dos critérios STOPP/START para Português de Portugal e na avaliação da sua validade para os médicos portugueses de Medicina Geral e Familiar (MGF).[1] Teve como objetivo a validação da tradução que incluiu: fiabilidade e

concordância *intra-rater* (o mesmo avaliador utilizou a tradução portuguesa dos critérios STOPP/START e avaliou os mesmos doentes em 2 momentos diferentes) e fiabilidade e concordância *inter-rater* (avaliadores diferentes utilizaram a tradução portuguesa dos critérios e avaliaram os mesmos doentes).[1,24]

## Materiais e Métodos

O estudo foi constituído por 4 fases.[1]

A primeira fase (fase I) foi a tradução dos critérios STOPP/START para Português de Portugal, seguindo os *Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes Measures*. [1,28] Os autores originais desta ferramenta de rastreio permitiram a sua tradução, adaptação e validação para Português de Portugal. [1] Um consultor-chave Português-nativo, fluente em Inglês, com formação em investigação na área de saúde e experiência em traduzir documentos em Inglês, realizou e ajudou com a tradução. [1,28] Foram obtidas 2 traduções independentes dos critérios STOPP/START: uma do consultor-chave e outra de um contra-tradutor, também Português-nativo e fluente em Inglês. [1,28] As 2 traduções foram reconciliadas pela equipa de investigação para obter a tradução consensual final. [1,28] A tradução consensual foi contra-traduzida, ou seja, traduzida de Português de Portugal para Inglês, por um tradutor profissional Inglês-nativo e fluente em Português que não teve qualquer conhecimento prévio da versão Inglesa original. [1,28] A contra-tradução foi comparada com a versão original para identificar quaisquer diferenças relevantes. [1,28] Por último, a versão Portuguesa reconciliada dos critérios STOPP/START foi distribuída por um grupo de 15 médicos de MGF para identificar quaisquer problemas com a sua interpretação. [1] A equipa de investigação analisou os resultados da aplicação da ferramenta de rastreio STOPP/START para preparar a versão final. [1]

A fase II foi um estudo analítico transversal e incluiu a colheita de dados clínicos de doentes. [1] Foi realizada numa unidade de saúde de cuidados primários localizada em Aveiro, na região Centro de Portugal, onde 5 médicos de MGF seguem um total de 8165 doentes, dos quais 1625 têm 65 anos ou mais. [1] O tamanho da amostra foi calculado com a função *CIBinary* do pacote *kappaSize* do *software R*. [1,29] O tamanho da amostra do estudo *intra-rater* foi 334 participantes, considerando os parâmetros: valor  $\kappa$  (coeficiente Kappa de Cohen) estimado 0,68; margem de erro 0,1; prevalência de cada item dos critérios START 0,25; número de momentos 2 e nível de significância 5%. [1] O tamanho da amostra do estudo *inter-rater* foi 205 participantes, considerando os parâmetros: valor  $\kappa$  estimado 0,68; margem de erro 0,1; prevalência de cada item dos critérios START 0,25; número de avaliadores 3 e nível de significância 5%. [1] Os 205 participantes do estudo *inter-rater* foram selecionados aleatoriamente dos 334 participantes do estudo *intra-rater*. [1]

Foi realizada uma amostragem aleatória independente, utilizando dígitos aleatórios gerados por computador, para seleccionar os participantes a partir de uma lista de doentes idosos com 65 anos ou mais de uma unidade de saúde de cuidados primários. [1] Os doentes seleccionados foram convidados por chamada telefónica para participar no estudo. [1] O investigador ou um

colega de investigação previamente treinado entrevistaram os doentes no consultório.[1] Cada participante assinou um formulário de consentimento escrito.[1] O recrutamento continuou até à inscrição de 334 doentes.[1] Os critérios de exclusão incluíram: incapacidade ou recusa em fornecer consentimento informado por escrito; diagnóstico de perturbação psicótica; institucionalização e a presença de doença terminal.[1]

Na inclusão, o investigador principal colheu dados sociodemográficos dos doentes, tais como: idade; sexo; nível de educação; situação laboral e estado civil.[1] O nível de educação foi agrupado de acordo com a *International Standard Classification of Education* (ISCED 2011). A colheita de dados clínicos incluiu: número total de medicamentos para doenças crónicas; quaisquer medicamentos prescritos e respetiva dosagem, forma farmacêutica, via de administração e justificação para a toma; alergias, condições relacionadas com medicamentos e história de reações adversas a medicamentos; doenças/condições atuais e passadas.[1] O número de medicamentos utilizados foi agrupado de acordo com a definição de polifarmácia. Acessoriamente, também foram colhidas as seguintes informações: presença ou ausência de edema do tornozelo; T-scores de densidade mineral óssea; história de vacinação contra a gripe e pneumocócica; frequência cardíaca e pressão arterial sistólica e diastólica.[1] Os dados foram colhidos utilizando registos eletrónicos de consultas de saúde e entrevistas clínicas a doentes.[1]

A informação colhida não incluiu informação que pudesse identificar os participantes.[1] Cada doente foi numerado de 1 a 334 e os documentos utilizados para colher os dados contêm apenas esse código de identificação, para proteger a sua identidade.[1] Adicionalmente, foi criada uma comissão de monitorização da proteção de dados composta por 2 investigadores externos com experiência nesta área clínica e em atividades académicas e científicas.[1] De acordo com a Lei Portuguesa de Investigação Clínica, todos os dados registados durante o estudo serão armazenados durante 5 anos num local seguro e adequado no centro de saúde do investigador principal, após o encerramento da investigação.[1] Todos os dados contendo códigos de identificação de participantes serão destruídos após este período.[1]

A fase III foi o estudo da fiabilidade e concordância *intra-rater*. [1] Um investigador/médico de MGF independente (nomeado investigador “A”) aplicou a tradução Portuguesa dos critérios STOPP/START a todos os doentes, utilizando a informação colhida na fase II.[1] Duas semanas depois, o mesmo investigador reavaliou os registos dos doentes, aplicando os mesmos critérios, para evitar o viés de recordação.[1]

A fase IV foi o estudo da fiabilidade e concordância *inter-rater* através da comparação dos resultados de 3 avaliadores.[1] Três investigadores/médicos de MGF (nomeados investigadores “B”, “C” e “D”) aplicaram independentemente a tradução Portuguesa dos

critérios STOPP/START a 205 doentes selecionados aleatoriamente, utilizando a informação colhida na fase II.[1] Os médicos pertencem a unidades de saúde diferentes e só tiveram contacto com o autor correspondente que lhes forneceu os dados compostos.[1]

Os dados foram armazenados com o *software Microsoft Excel* e a sua análise foi feita com *SPSS Statistics V.27.0* e o *software R*. [1] As variáveis categóricas foram descritas por frequências absolutas e relativas.[1] As variáveis contínuas foram descritas por média e desvio-padrão, se normalmente distribuídas, ou por mediana e dispersão interquartil, se não-normalmente distribuídas.[1] A normalidade foi avaliada pela observação de histogramas e implementação do teste Kolmogorov-Smirnov.[1] A fiabilidade *intra-rater/inter-rater* foi medida utilizando o coeficiente Kappa de Cohen ( $\kappa$ ) e o respetivo intervalo de confiança (IC) de 95%. [1] O coeficiente Kappa de Cohen é a percentagem de acordo, ajustada ao acaso, e foi interpretado como: pobre ( $\kappa \leq 0,2$ ); justo ( $0,21 \leq \kappa \leq 0,4$ ); moderado ( $0,41 \leq \kappa \leq 0,6$ ); substancial ( $0,61 \leq \kappa \leq 0,8$ ) e bom ( $0,81 \leq \kappa \leq 1,00$ ). [1,25] A concordância *intra-rater/inter-rater* foi avaliada utilizando proporções de acordo e proporções de acordo específicas (positivas e negativas) e o respetivo IC de 95%. [1] Um valor p inferior ou igual a 0,05 foi considerado estatisticamente significativo.[1]

O protocolo foi aprovado em 30 de julho de 2020 pela Comissão de Ética da Administração Regional de Saúde do Centro de Portugal com o número de referência 034-2020.[1]

## Resultados

Foram inscritos 334 participantes no estudo. As características dos doentes, tais como idade, sexo, nível de educação e número de medicamentos utilizados, estão descritas na Tabela 1.

A idade média ( $\pm$  desvio-padrão) foi  $74,2 \pm 6,9$  anos. Cerca de 52,4% dos doentes eram do sexo masculino, enquanto 47,6% eram do sexo feminino. A grande maioria, cerca de 94,6%, apresentou o nível de educação “*early childhood, primary and lower secondary education (level 0-2)*”. A maior parte dos doentes, cerca de 63,2%, apresentou a toma diária e crónica de 5 ou mais medicamentos.

Tabela 1. Características dos doentes (n=334).

| Variável   | n (%)             |
|--|-------------------|
| <b>Idade – média (DP); mín-máx</b>   | 74,2 (6,9); 65-99 |
| 65-69 anos   | 105 (31,4%)       |
| 70-74 anos   | 71 (21,3%)        |
| 75-79 anos   | 78 (23,4%)        |
| 80-84 anos   | 50 (15,0%)        |
| 85+ anos   | 30 (9,0%)         |
| <b>Sexo</b>  |                   |
| Feminino   | 159 (47,6%)       |
| Masculino  | 175 (52,4%)       |
| <b>Nível de educação</b>   |                   |
| <i>Early childhood, primary and lower secondary education (level 0-2)</i>            | 316 (94,6%)       |
| <i>Upper secondary and post-secondary non-tertiary education (level 3-4)</i>         | 17 (5,1%)         |
| <i>Short-cycle tertiary education, Bachelor's, Master's or Doctorate (level 5-8)</i> | 1 (0,3%)          |
| <b>Número de medicamentos utilizados</b>   |                   |
| 0-1  | 57 (17,1%)        |
| 2-4 (Polifarmácia <i>minor</i> )   | 66 (19,8%)        |
| 5-9 (Polifarmácia <i>major</i> )   | 210 (62,9%)       |
| 10+ (Polifarmácia grave)   | 1 (0,3%)          |

Legenda: DP – desvio-padrão; máx – idade máxima; mín – idade mínima; n – número. As discrepâncias nos totais são devidas a arredondamentos.

No estudo da fiabilidade e concordância *intra-rater*, a avaliação dos registos dos 334 doentes pelo investigador A, bem como a sua reavaliação após um intervalo de 2 semanas, foram analisadas. Os resultados estão descritos na Tabela 2 (STOPP) e na Tabela 3 (START).

Para os critérios STOPP, o teste de fiabilidade *intra-rater* do investigador A obteve um coeficiente Kappa de 0,70 (0,65-0,74), sendo considerado um valor substancial. As proporções positiva e negativa obtidas foram, respetivamente, 94,2% e 75,2%.

Tabela 2. Fiabilidade e concordância *intra-rater* a partir da análise dos critérios STOPP.

| Critérios STOPP    | Combinação do avaliador            | Concordância     |                  | Fiabilidade      |
|--------------------|------------------------------------|------------------|------------------|------------------|
|                    |                                    | PPos (%)         | PNeg (%)         | Kappa (95% IC)   |
| <i>Intra-rater</i> | Investigador A *<br>Investigador A | 94,2 (93,1-95,1) | 75,2 (70,9-79,1) | 0,70 (0,65-0,74) |

Legenda: IC – intervalo de confiança; PNeg – proporção negativa; PPos – proporção positiva; STOPP – *Screening Tool of Older People's Prescriptions*.

Para os critérios START, o teste de fiabilidade *intra-rater* do investigador A obteve um coeficiente Kappa de 0,60 (0,52-0,68), sendo considerado um valor moderado. As proporções positiva e negativa obtidas foram, respetivamente, 88,2% e 71,1%, com  $\kappa$  próximo de 1 em todas as avaliações.

Tabela 3. Fiabilidade e concordância *intra-rater* a partir da análise dos critérios START.

| Critérios START    | Combinação do avaliador            | Concordância     |                  | Fiabilidade      |
|--------------------|------------------------------------|------------------|------------------|------------------|
|                    |                                    | PPos (%)         | PNeg (%)         | Kappa (95% IC)   |
| <i>Intra-rater</i> | Investigador A *<br>Investigador A | 88,2 (85,4-90,6) | 71,1 (64,5-76,8) | 0,60 (0,52-0,68) |

Legenda: IC – intervalo de confiança; PNeg – proporção negativa; PPos – proporção positiva; START – *Screening Tool to Alert to Right Treatment*.

## Discussão e Conclusão

A ferramenta de rastreio STOPP/START é uma forma simples de gerir o tratamento de adultos mais velhos, sendo mais fácil de utilizar diariamente quando adaptada à linguagem do prescritor.[1,11] Este estudo é inovador porque é o primeiro desenvolvimento de uma versão Portuguesa dos critérios STOPP/START com o objetivo de facilitar a prescrição apropriada de medicação em idosos e prevenir erros.[1] Esta investigação não foi apenas uma tradução, mas também uma adaptação feita por médicos de MGF independentes que irá, potencialmente, aumentar a utilização desta versão no contexto dos cuidados de saúde primários.[1]

No teste de fiabilidade *intra-rater*, o coeficiente Kappa foi 0,70 (0,65-0,74) para os critérios STOPP e 0,60 (0,52-0,68) para os critérios START, sendo considerado um valor substancial e moderado, respetivamente. Isto sugere que a ferramenta de rastreio STOPP/START é reprodutível e fiável. No teste de concordância *intra-rater*, as proporções positiva e negativa obtidas foram, respetivamente, 94,2% e 75,2% para os critérios STOPP e 88,2% e 71,1% para os critérios START.

Contudo, os critérios STOPP/START também apresentam limitações: não substituem o conhecimento do médico especialista; não fornecem considerações especiais de uso ou terapias alternativas para evitar MPI; necessitam de atualização regular e de mais estudos controlados randomizados.[4,15,17-19,21,23] Esta versão adaptada dos critérios STOPP/START está exclusivamente focada nos centros de saúde e, logo, nos cuidados primários e foi feita apenas numa unidade de saúde.[1]

## Financiamento

Este artigo foi apoiado por Fundos Nacionais através da Fundação para a Ciência e a Tecnologia (FCT), dentro de Centro de Investigação em Tecnologias e Serviços de Saúde (CINTESIS), Unidade de Investigação & Desenvolvimento (referência UIDB/4255/2020).[1]

## Conflitos de Interesse

Os autores declaram não haver interesse financeiro ou outro tipo de conflito de interesse.

## Referências Bibliográficas

1. Monteiro, L., Monteiro-Soares, M., Mendonça, L. V., Ribeiro-Vaz, I., Martins, C., & Teixeira, A. (2021). Translation and adaptation of the STOPP-START screening tool to Portuguese for detecting inappropriate prescriptions in older people: a protocol. *BMJ open*, 11(6), e043746. <https://doi.org/10.1136/bmjopen-2020-043746>.
2. World Health Organization. (2021, 4 de outubro). *Ageing and health*. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>.
3. World Health Organization. (2019). *The Global strategy and action plan on ageing and health*. Ageing and life-course. <https://www.who.int/ageing/global-strategy/en/>.
4. Gallagher, P., Baeyens, J. P., Topinkova, E., Madlova, P., Cherubini, A., Gasperini, B., Cruz-Jentoft, A., Montero, B., Lang, P. O., Michel, J. P., & O'Mahony, D. (2009). 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 and ageing*, 38(5), 603–606. <https://doi.org/10.1093/ageing/afp058>.
5. Organisation for Economic Co-operation and Development. (n.d.). *Elderly population*. OECD iLibrary. <https://doi.org/10.1787/8d805ea1-en>.
6. Tommelein, E., Mehuys, E., Petrovic, M., Somers, A., Colin, P., & Boussey, K. (2015). Potentially inappropriate prescribing in community-dwelling older people across Europe: a systematic literature review. *European journal of clinical pharmacology*, 71(12), 1415–1427. <https://doi.org/10.1007/s00228-015-1954-4>.
7. Prazeres, F., & Santiago, L. (2015). Prevalence of multimorbidity in the adult population attending primary care in Portugal: a cross-sectional study. *BMJ open*, 5(9), e009287. <https://doi.org/10.1136/bmjopen-2015-009287>.
8. Holt, S., Schmiedl, S., & Thürmann, P. A. (2010). Potentially inappropriate medications in the elderly: the PRISCUS list. *Deutsches Arzteblatt international*, 107(31-32), 543–551. <https://doi.org/10.3238/arztebl.2010.0543>.
9. Formiga, F., Ferrer, A., Sanz, H., Marengoni, A., Alburquerque, J., Pujol, R., & Octabaix study members (2013). Patterns of comorbidity and multimorbidity in the oldest old: the Octabaix study. *European journal of internal medicine*, 24(1), 40–44. <https://doi.org/10.1016/j.ejim.2012.11.003>.
10. Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B., & Fratiglioni, L. (2011). Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews*, 10(4), 430–439. <https://doi.org/10.1016/j.arr.2011.03.003>.
11. Gallagher, P., Ryan, C., Byrne, S., Kennedy, J., & O'Mahony, D. (2008). STOPP (Screening Tool of Older Person's Prescriptions) and START (Screening Tool to Alert

- doctors to Right Treatment). Consensus validation. *International journal of clinical pharmacology and therapeutics*, 46(2), 72–83. <https://doi.org/10.5414/cpp46072>.
12. Gnjidic, D., Hilmer, S. N., Blyth, F. M., Naganathan, V., Waite, L., Seibel, M. J., McLachlan, A. J., Cumming, R. G., Handelsman, D. J., & Le Couteur, D. G. (2012). Polypharmacy cutoff and outcomes: five or more medicines were used to identify community-dwelling older men at risk of different adverse outcomes. *Journal of clinical epidemiology*, 65(9), 989–995. <https://doi.org/10.1016/j.jclinepi.2012.02.018>.
  13. 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. (2013). The impact of number of drugs prescribed on the risk of potentially inappropriate medication among outpatient older adults with chronic diseases. *QJM : monthly journal of the Association of Physicians*, 106(11), 1009–1015. <https://doi.org/10.1093/qjmed/hct141>.
  14. Payne, R. A., Avery, A. J., Duerden, M., Saunders, C. L., Simpson, C. R., & Abel, G. A. (2014). Prevalence of polypharmacy in a Scottish primary care population. *European journal of clinical pharmacology*, 70(5), 575–581. <https://doi.org/10.1007/s00228-013-1639-9>.
  15. Gallagher, P., & O'Mahony, D. (2008). STOPP (Screening Tool of Older Persons' potentially inappropriate Prescriptions): application to acutely ill elderly patients and comparison with Beers' criteria. *Age and ageing*, 37(6), 673–679. <https://doi.org/10.1093/ageing/afn197>.
  16. Lu, W. H., Wen, Y. W., Chen, L. K., & Hsiao, F. Y. (2015). Effect of polypharmacy, potentially inappropriate medications and anticholinergic burden on clinical outcomes: a retrospective cohort study. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, 187(4), E130–E137. <https://doi.org/10.1503/cmaj.141219>.
  17. 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. *European Geriatric Medicine*, 7(5), 416–423. <https://doi.org/10.1016/j.eurger.2016.03.010>.
  18. Hamilton, H., Gallagher, P., Ryan, C., Byrne, S., & O'Mahony, D. (2011). Potentially inappropriate medications defined by STOPP criteria and the risk of adverse drug events in older hospitalized patients. *Archives of internal medicine*, 171(11), 1013–1019. <https://doi.org/10.1001/archinternmed.2011.215>.
  19. Motter, F. R., Fritzen, J. S., Hilmer, S. N., Paniz, É. V., & Paniz, V. (2018). Potentially inappropriate medication in the elderly: a systematic review of validated explicit criteria. *European journal of clinical pharmacology*, 74(6), 679–700. <https://doi.org/10.1007/s00228-018-2446-0>.

20. Hill-Taylor, B., Walsh, K. A., Stewart, S., Hayden, J., Byrne, S., & Sketris, I. S. (2016). Effectiveness of the STOPP/START (Screening Tool of Older Persons' potentially inappropriate Prescriptions/Screening Tool to Alert doctors to the Right Treatment) criteria: systematic review and meta-analysis of randomized controlled studies. *Journal of clinical pharmacy and therapeutics*, 41(2), 158–169. <https://doi.org/10.1111/jcpt.12372>.
21. Barry, P. J., Gallagher, P., Ryan, C., & O'mahony, D. (2007). START (screening tool to alert doctors to the right treatment)--an evidence-based screening tool to detect prescribing omissions in elderly patients. *Age and ageing*, 36(6), 632–638. <https://doi.org/10.1093/ageing/afm118>.
22. O'Mahony, D., O'Sullivan, D., Byrne, S., O'Connor, M. N., Ryan, C., & Gallagher, P. (2015). STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age and ageing*, 44(2), 213–218. <https://doi.org/10.1093/ageing/afu145>.
23. By the 2019 American Geriatrics Society Beers Criteria® Update Expert Panel (2019). American Geriatrics Society 2019 Updated AGS Beers Criteria® for Potentially Inappropriate Medication Use in Older Adults. *Journal of the American Geriatrics Society*, 67(4), 674–694. <https://doi.org/10.1111/jgs.15767>.
24. Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of clinical epidemiology*, 64(1), 96–106. <https://doi.org/10.1016/j.jclinepi.2010.03.002>.
25. Harvey, N. D. (2021, December 4). A Simple Guide to Inter-rater, Intra-rater and Test-retest Reliability for Animal Behaviour Studies. <https://doi.org/10.31219/osf.io/8stpy>.
26. Gwet, K. L. (2008). Intrarater Reliability. In *Wiley Encyclopedia of Clinical Trials* (eds R.B. D'Agostino, L. Sullivan and J. Massaro). <https://doi.org/10.1002/9780471462422.eoct631>.
27. Health Technology Assessment. (n.d.). *Intra-rater Reliability*. HTA glossary. <https://htaglossary.net/intra-rater-reliability>.
28. Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., Erikson, P., & ISPOR Task Force for Translation and Cultural Adaptation (2005). 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 in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 8(2), 94–104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>.
29. Rotondi, M. A. (2013). *kappaSize: Sample Size Estimation Functions for Studies of Interobserver Agreement* (Versão 1.1) [Software de computador]. <https://CRAN.R-project.org/package=kappaSize>.