

## Análise da preferência digital de idade no sistema de farmacovigilância do Brasil, 2008-2013

Analysis of the digital preference for age in the pharmacovigilance system in Brazil, 2008-2013

Daniel Marques Mota<sup>1</sup>  
Álvaro Vigo<sup>1</sup>  
Ricardo de Souza Kuchenbecker<sup>1</sup>

**Abstract** *The scope of this article is to analyze the digital preference for the "age" variable in the database of cases of adverse drug events reported between 2008 and 2013 in the pharmacovigilance system in Brazil. The database was analyzed in three stages: 1) Initial exploration; 2) standardization of variables; and 3) duplicate records management. The digital preference for the "age" variable according to sex and grouped regions of the country, was determined using the Whipple and Myers methods and also measured by the adhesion test based on statistical  $\chi^2$  at 5% significance level. The female population was predominant (60.3%), with a female/male ratio of 1.5:1 and a female majority in 8 of the 12 age groups. Throughout the period under analysis, no preference for a terminal digit was observed in the "age" variable. The ages that were registered in the database appear not to reveal any imprecision and, therefore, this information cannot be interpreted as unreliable. These results point to a good performance in collecting data and information related to the "age" variable of the Brazil's pharmacovigilance system.*

**Key words** *Pharmacovigilance, Adverse Drug Reaction Reporting Systems, Bias, Health vigilance*

**Resumo** *Analisar a preferência digital na variável "idade" da base de dados de casos de eventos adversos a medicamentos notificados entre 2008 e 2013 no sistema de farmacovigilância do Brasil. A base de dados foi analisada em três etapas: 1) exploração inicial; 2) padronização de variáveis; e 3) manejo dos registros repetidos. A preferência digital da variável "idade", segundo sexo e regiões agrupadas do país, foi determinada usando os métodos de Whipple e Myers e medida, também, pelo teste de aderência baseado na estatística  $\chi^2$  ao nível de 5% de significância. Predominou a população feminina (60,3%), perfazendo uma razão feminino/masculino de 1,5:1 e com maioria feminina em oito das 12 faixas etárias estudadas. Ao longo do período analisado não foi observado qualquer preferência de dígito terminal na variável "idade". As idades registradas nas notificações parecem não demonstrar imprecisão e, por conseguinte, esta informação pode ser interpretada como de qualidade não duvidosa. Tais resultados apontam para um bom desempenho na coleta de dados e informações relacionados à variável "idade" do sistema de farmacovigilância do Brasil.*

**Palavras-chave** *Farmacovigilância, Sistemas de notificação de reações adversas a medicamentos, Viés, Vigilância sanitária*

<sup>1</sup> Programa de Pós-Graduação em Epidemiologia, Faculdade de Medicina da Universidade Federal do Rio Grande do Sul. R. Ramiro Barcelos 2.400/2º, Bom Fim. 90035-003 Porto Alegre RS Brasil. dmarques2003@yahoo.com.br

## Introdução

As informações da idade são utilizadas para caracterização demográfica de populações, realocação de indivíduos a grupos de interesse para estudo e como fator de risco para agravos à saúde e fator de confusão em estudos epidemiológicos<sup>1,2</sup>. Em farmacovigilância, a idade é apreciada como fator predisponente para à ocorrência de eventos adversos a medicamentos (EAMs), de modo especial, às reações adversas a medicamentos (RAMs)<sup>3</sup>. Formulários para notificação de EAMs têm a idade como uma variável de preenchimento na caracterização dos pacientes<sup>4</sup> e na análise de causalidade das RAMs<sup>4,5</sup>.

Há pelos menos três tipos de erros relacionados à qualidade dos dados da variável “idade” em populações estudadas<sup>6</sup>. São eles: i) omissão diferencial por idade – um erro de cobertura de certas idades que é difícil a detecção (sub-registros); ii) idade ignorada – corresponde os registros de pessoas que não se tem informação sobre os anos de vida, ou seja, um problema de não resposta (incompletude); e iii) má declaração – refere-se à informação de uma idade equivocada, voluntária ou involuntariamente, por parte do entrevistado<sup>6</sup>. Em várias ocasiões, a idade é coletada de modo errôneo<sup>7,8</sup>. Um erro comum ao relatar a variável idade é a ocorrência de preferência por algum dígito, seja por desconhecimento da informação, por problemas de comunicação entre entrevistador e entrevistado, hábitos sociais e culturais e erros durante o registro e processamento dos dados<sup>9,10</sup>. Em tais situações, pode ocorrer a tendência de arredondamento para um valor mais próximo que termina em “0” ou “5”, o que é conhecido como “preferência digital”, que produz uma concentração da idade em determinados valores, podendo ocasionar vieses e imprecisões, afetando negativamente a qualidade dos dados<sup>11</sup>.

Este estudo analisou a ocorrência de preferência digital da variável “idade no momento do evento” (denominada de “idade”) nas notificações de suspeitas de EAMs feitas ao Sistema Nacional de Notificações para a Vigilância Sanitária, formulário medicamento que ocasionou danos à saúde do paciente (Notivisa-medicamento), registradas entre 2008 e 2013.

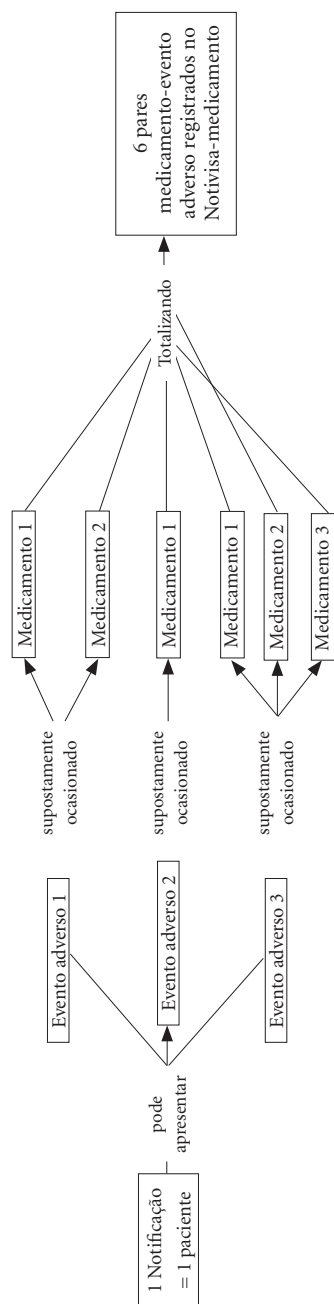
## Métodos

### Desenho do estudo e caracterização do Notivisa-medicamento

Realizou-se estudo de avaliação em saúde com base de dados nominais extraída do Notivisa-medicamento. O Notivisa-medicamento é um sistema de informação informatizado *online* da Agência Nacional de Vigilância Sanitária (Anvisa) que recebe notificações espontâneas de casos suspeitos de EAMs. Criado em 2008, abrange o território brasileiro e é considerado o maior e mais importante repositório de dados de EAMs do Sistema Nacional de Farmacovigilância (SINAFA)<sup>12</sup> do país. Cada notificação registrada no formulário padronizado do Notivisa-medicamento contém informações sobre eventos adversos (sinais e sintomas), medicamentos suspeitos, paciente, entre outros dados. O formulário apresenta três variáveis que caracterizam o paciente quanto ao seu tempo de vida: data de nascimento, idade e grupo de idade (criança, adolescente, adulto, etc.).

### Estrutura organizacional da base de dados

A base de dados do Notivisa-medicamento apresenta uma peculiaridade em sua estrutura de registros dos casos de EAMs em comparação com àquelas oriundas de sistemas de vigilância epidemiológica, pois não obedece ao seguinte arranjo: uma notificação = um registro = um paciente. Uma notificação advinda de um único paciente pode resultar em mais de um registro na base de dados, pois a sua organização no Notivisa-medicamento baseia-se em pares de fenômenos, como medicamento-evento adverso ou evento adverso-classificação da sua gravidade, como, também, na multiplicação de registros, conforme o número de sinais e sintomas, entre outros arranjos. Cada arranjo consta como um registro na base de dados. Exemplificando, a Figura 1 apresenta uma notificação com quatro pares representados por duas reações adversas e dois diferentes medicamentos suspeitos. Neste caso, a base de dados do Notivisa-medicamento contém quatro registros oriundos de um único paciente que se diferencia apenas nas informações peculiares para cada par de fenômeno. As informações, como idade e sexo, entre outras, se repetem em cada um dos registros.



**Figura 1.** Modelo exemplificativo do arranjo par medicamento-evento adverso no Notivisa-medicamento (1 par = 1 registro).

## Preparação e manejo da base de dados

Para fins do estudo, organizou-se a base de dados nominais prevendo três etapas: 1) exploração inicial; 2) padronização da base de dados; e 3) manejo dos registros repetidos. A intenção do estudo foi analisar a base de dados tendo como referência a variável idade na estrutura de organização das notificações, impedindo que esta fosse alterada em função da necessidade de repetição dos registros.

### Exploração inicial

Identificou um grande número de registros com o campo “nome do paciente” não preenchido. Trata-se da variável mais importante no processo de identificação e exclusão de registros repetidos. No Notivisa-medicamento, a fonte notificadora tem quatro possibilidades de preenchimento de variáveis relacionadas à identificação do paciente: a) inserir o nome do paciente; b) registrar apenas as iniciais do nome; c) preencher os itens “a” e “b”; e d) deixar os campos vazios. Considerando-se essas possibilidades, o estudo dividiu a base de dados em três bases menores a serem utilizadas na etapa de detecção e exclusão de registros repetidos. Este procedimento objetivou minimizar a perda de registros a serem analisados. As bases de dados criadas foram: 1) BD1 – com o nome dos pacientes preenchidos; 2) BD2 – somente com as iniciais do nome informadas; e 3) BD3 – com os campos nome dos pacientes e suas iniciais vazios. Após a exclusão dos registros repetidos, as três bases de dados foram unidas, formando a base de dados de interesse deste estudo, denominada de “BD final”.

### Padronização da base de dados

Antes de proceder a divisão da base de dados, a mesma foi submetida, a padronização, geral e específica, que envolveu um conjunto de variáveis, denominada de variáveis identificadoras (VI), escolhidas para detectar os registros repetidos e únicos de um mesmo paciente, não considerando as diferenças nos tipos de arranjos supracitados.

De maneira a permitir o relacionamento entre o maior número possível de pares verdadeiros com o menor número de pares errados, a escolha das VI foi baseada na literatura científica<sup>13</sup>, avaliando seu potencial de apresentar as seguintes características: i) menor propensão a alterações ao longo do tempo em comparação com outras variáveis, como endereço residencial<sup>14</sup>; e ii) maior poder discriminatório para representar um pa-

ciente, ou seja, variáveis que têm mais opções de respostas possíveis<sup>15</sup>. A padronização geral das VI incluiu: a) eliminação de acentos ortográficos, cedilha e pontos; b) mudança das letras para maiúsculas; e c) exclusão de excessos de espaços entre duas letras ou palavras.

O Quadro 1 sintetiza as VI utilizadas em cada base de dados, as regras adotadas e a padronização específica. Persistindo os registros com campos vazios ou preenchidos como ignorados das VI “sexo”, “cidade de ocorrência do evento” e “nome do notificador”, após essa última padronização, procedeu-se a exclusão dos mesmos da base de dados. Apesar da variável “idade” não ter sido relacionada como VI, a mesma sofreu a seguinte padronização: i) definiu-se a unidade “anos” para todas as idades, variando de < 1 ano até 109 anos; ii) valor zero foi atribuído aos registros com idades < 1 ano de idade; e iii) os registros com idades não válidas foram considerados

sem preenchimento, ou seja, o campo “idade” permaneceu vazio.

### Manejo dos registros repetidos

Utilizou-se a técnica de relacionamento determinístico para a detecção dos registros repetidos que obedeceu à combinação exata das VI para cada base de dados. A detecção e exclusão automática dos registros repetidos se deu em um único passo mediante o uso do *software* SPSS, versão 18ª, optando pela permanência do primeiro registro na base de dados. Os registros foram identificados como iguais, formando pares, trios, etc, ou como diferentes, ou seja, havia apenas um único registro. Os principais motivos da escolha da técnica determinística foram: i) simplificação do processo de relacionamento dos registros; ii) praticidade e rapidez na execução; e iii) arranjo organizacional da base de dados no Notivisa-medicamento.

**Quadro 1.** Variáveis identificadoras utilizadas na detecção de registros repetidos e únicos, segundo a divisão da base de dados do Notivisa-medicamento.

Variável identificadora	Regra e padronização específica	Base de dados (BD)		
		BD1	BD2	BD3
Nome do paciente	Foram considerados os registros com pelo menos dois nomes próprios. Correções de erros tipográficos e de inconsistências. Termos como Irmã e Senhor(a) foram excluídos. Os registros iniciados com o “RN de...” ou “Lactante de...” ou “Filha(o) de...” permaneceram na base de dados sem alterações. Remoção das preposições dos nomes (de, da, do, das e dos).	x		
Iniciais do nome do paciente	Foram considerados os registros com pelos menos duas letras do alfabeto. Inseriu-se as iniciais que constavam no campo “nome do paciente” que não se apresentavam adequadamente registradas no campo “iniciais do nome”. Registros da BD1 que não atenderam às regras foram inseridos na BD2.		x	
Nome da fonte notificadora	Exclusão de termos específicos, como Dr e Dra. Os registros das bases BD1 e BD2 que não atenderam às regras definidas foram inseridos na BD3.			x
Sexo	Definição do sexo a partir do nome do paciente julgado ser masculino ou feminino para os registros com preenchimento ignorado. Correção de alguns registros quanto ao sexo. Os registros com nomes duvidosos permaneceram sem alteração, sendo excluídos com aqueles que permaneceram ignorados.	x	x	x
Data de nascimento	Não foi padronizada. Utilizou-se os registros sem preenchimento da data de nascimento (N= 7.698) que corresponderam a 5,7%, 15,5% e 35,0% das BD1, BD2 e BD3, respectivamente.	x	x	x
Cidade de ocorrência do evento	Utilizou-se cidade da instituição notificadora para preenchimento dos campos vazios. Os campos com nomes de regiões administrativas do Distrito Federal foram modificados para Brasília.	x	x	x
Data de início do evento	A data de recebimento pelo notificador ou a de notificação no Notivisa-medicamento complementaram os registros com campos vazios e com datas inconsistentes.		x	x

### Preferência digital

A preferência digital da variável “idade”, segundo sexo e regiões agrupadas do país, foi determinada usando os métodos de Whipple e Myers<sup>8,10,16-18</sup>. A preferência digital, também, foi medida pelo teste de aderência baseado na estatística qui-quadrado de Pearson, utilizando a distribuição  $\chi^2$  com nove graus de liberdade, como referência<sup>11</sup>. A hipótese nula diz respeito a que, no conjunto dos dados do Notivisa-medicamento, a distribuição de todos os dígitos terminais das idades entre 10 e 99 anos (0, 1, 2, 3, 4, 5, 6, 7, 8 e 9) é uniforme.

O método de Whipple é aplicado para detectar preferências por dígitos terminais “0” e “5” no grupo populacional de 23 a 62 anos de idade. Há duas razões da escolha deste intervalo de idades. Primeiramente, o método assume uma distribuição uniforme da população estudada em intervalos de cinco anos, ou seja, não há preferência por nenhuma idade em particular. A segunda razão, as populações mais jovens e mais idosas, com frequência, são mais afetadas por outros tipos de erros e problemas do que por preferência digital. Os resultados é medido pelo índice de Whipple (IW) que podem variar entre 0 (zero), representando nenhum registro dos dígitos “0” e “5” e 500, indicando que apenas as idades terminadas em tais dígitos foram registradas no formulário do Notivisa-medicamento<sup>10</sup>. A fórmula abaixo foi utilizada para o cálculo do IW, sendo que P é o tamanho da população para cada idade<sup>10,18</sup>.

$$W = \frac{\Sigma (P25 + P30 + P35 + \dots P60)}{1/5 \times \Sigma (P23 + P24 + P25 + \dots P62)} \times 100$$

Utilizou-se a escala qualitativa proposta pelas Nações Unidas, como parâmetro para medir a qualidade dos dados da variável “idade”, distribuída da seguinte forma: i) < 105 – dados muito precisos; ii) 105 a 109,9 – dados relativamente precisos; iii) 110 a 124,9 – dados aproximados; iv) 125 a 174,9 – dados caracterizados como “ruins”; v)  $\geq 175$  – dados definidos como “muito ruins”<sup>10,18,19</sup>.

O método de Myers permite determinar a preferência para cada dígito terminal (0 a 9) nas idades que variam de 10 a 99 anos. Baseia-se no princípio de que a população de cada dígito terminal da variável “idade” apresenta uma distribuição uniforme que representa 10% da população total. A violação deste princípio é medida pelo índice de Myers (IM), que sintetiza a pre-

ferência digital para todos os dígitos da variável “idade”<sup>19</sup>.

O IM é calculado pela soma do número de pessoas das idades terminadas em um dígito particular na população entre 10 e 99 anos, e, outro, na população com idade entre 20 e 99 anos. Aplica-se pesos para cada dígito, de acordo com a divisão populacional (pesos 1 e 9 para 0 dígito, pesos 2 e 8 para 1, pesos 3 e 7 para 2, etc). Os resultados são somados para obter uma população fictícia em cada dígito e convertidos os valores em percentual em relação ao total da população. Tais percentuais são subtraídos do valor teórico (10%), obtendo os desvios como resultados para cada dígito<sup>10,17</sup>. A soma dos desvios encontrados, desconsiderando o sinal negativo dos valores, produz o IM. Teoricamente, o IM abrange um intervalo entre 0 (zero) que representa nenhuma preferência das idades e 180, indicando uma preferência de todas as idades em um único dígito, por exemplo cinco<sup>18,20</sup>. A seguinte escala qualitativa de valores foi utilizada para a classificação da preferência geral pelo IM: 0,0 a 5,0 – Baixo; 5,1 a 15,0 – Mediano; 15,1 a 30,0 – Alto; 30,0 a 180 – Muito alto<sup>18-20</sup>.

### Análise dos dados e considerações éticas

Os *softwares* Office Excel® 2013 e RecLink III, versão 3.1.6â foram utilizados na etapa de padronização das VI. O excesso ou déficit da população masculina como percentual do total da população foi calculado como sendo a diferença entre os sexos dividido pelo total da população e, em seguida, multiplicado por 100. O IW e IM foram calculados em planilha eletrônica do Office Excel®. Para ambos os índices foram calculadas as taxas de variação anual média (TVAM). A TVAM proporciona informação sobre as variações anuais acumuladas, em termos médio, entre as distintas variações de cada ano do período estudado. Matematicamente, a TVAM é definida como  $TVAM (\%) = (n - 1) \sqrt{VF/VI} - 1) \times 100$  em que “n” representa o número de anos estudados e VI e VF representam o valor inicial e o valor final do período, respectivamente. O teste U de Mann-Whitney foi utilizado para comparar as distribuições da idade, segundo o sexo. O teste baseado na estatística qui-quadrado de Pearson foi utilizado para comparação de proporções por sexo. O nível de 5% de significância estatística foi utilizado nas análises. O estudo foi aprovado pelo comitê de ética em pesquisa do Hospital de Clínicas de Porto Alegre da Universidade Federal do Rio Grande do Sul.

## Resultados

Entre 2008 e 2013 foram identificados 63.512 registros de EAMs na base de dados. Após a padronização das VI, em que a variável “sexo” foi a única que permaneceu com campos ignorados, foram excluídos 537 registros. Um registro realizado como simulação-teste foi identificado e igualmente excluído, resultando em 62.974 registros. Excluídos os registros repetidos, a população do estudo foi formada por 28.875 pacientes, representando 45,8% do total de registros presentes na base de dados pós-padronização.

Na BD1 foi encontrado um intervalo de registros repetidos para um único paciente, que

variou de 2 a 44. Os intervalos nas BD2 e BD3 foram de 2 a 64 e 2 a 125 registros repetidos, respectivamente. Em todas as bases de dados, a quantidade mais frequente de registros foi igual a 2 (moda). Segundo a combinação de VI utilizada, 16.956 (56,4%) dos registros foram considerados como repetidos na BD1. Na BD2 e BD3 os percentuais corresponderam a 52,5% (n = 14.960) e 49,3% (n = 2.183), respectivamente.

A Tabela 1 apresenta as características demográficas da população de pacientes do estudo. Houve diferença entre a média de idade e o total de pacientes entre os sexos ( $p < 0,001$ ). Predominou a população do sexo feminino (60,3%), perfazendo razão feminino/masculino de 1,5:1 e

**Tabela 1.** Características demográficas da população de pacientes da base de dados do Notivisa-medicamento, 2008 a 2013 (N = 28.875).

Características dos pacientes		Sexo (%)		p-valor
		Masculino	Feminino	
Idade (anos) (N = 24.393)				
Média (IC95%)	44,4 (44,1;44,6)	42,8 (42,4;43,3)	45,4 (45,1;45,8)	-
Mediana [mínimo-máximo]	46 [0 – 109]	45 [0 – 109]	47 [0 – 104]	< 0,001 <sup>†</sup>
Faixa etária (anos)				
	<b>Total (%)</b>			<b>p-valor<sup>§</sup></b>
< 1 ano	526 (1,8)	272 (51,7)	254 (48,3)	0,43
1 a 4	931 (3,2)	513 (55,1)	418 (44,9)	0,002
5 a 9	724 (2,5)	378 (52,2)	346 (47,8)	0,23
10 a 14	833 (2,9)	474 (56,9)	359 (43,1)	< 0,001
15 a 19	965 (3,3)	458 (47,5)	507 (52,5)	0,11
20 a 29	2.643 (9,2)	1.010 (38,2)	1.633 (61,8)	< 0,001
30 a 39	3.256 (11,3)	1.183 (36,3)	2.073 (63,7)	< 0,001
40 a 49	3.567 (12,4)	1.218 (34,1)	2.349 (65,9)	< 0,001
50 a 59	4.118 (14,3)	1.586 (38,5)	2.532 (61,5)	< 0,001
60 a 69	3.384 (11,7)	1.416 (41,8)	1.968 (58,2)	< 0,001
70 a 79	2.200 (7,6)	920 (41,8)	1.280 (58,2)	< 0,001
80 e mais	1.246 (4,3)	485 (38,9)	761 (61,1)	< 0,001
Ignorados/vazios	4.482 (15,5)	1.554 (34,7)	2.928 (65,3)	< 0,001
Região				
Sudeste	17.342 (60,1)	6.401 (36,9)	10.941 (63,1)	< 0,001
Sul	4.866 (16,9)	2.093 (43,0)	2.773 (57,0)	< 0,001
Nordeste	4.571 (15,8)	2.031 (44,4)	2.540 (55,6)	< 0,001
Norte	1.701 (5,9)	752 (44,2)	949 (55,8)	< 0,001
Centro-Oeste	395 (1,4)	190 (48,1)	205 (51,9)	0,45
Ano de notificação				
2008	2.524 (8,7)	1.106 (43,8)	1.418 (56,2)	< 0,001
2009	3.449 (11,9)	1.412 (40,9)	2.037 (59,1)	< 0,001
2010	4.496 (15,5)	1.788 (39,8)	2.702 (60,2)	< 0,001
2011	5.517 (19,1)	2.233 (40,5)	3.284 (59,5)	< 0,001
2012	5.789 (20,0)	2.200 (38,0)	3.589 (62,0)	< 0,001
2013	7.106 (24,6)	2.728 (38,4)	4.378 (61,6)	< 0,001
Total	28.875 (100)	11.467 (39,7)	17.408 (60,3)	< 0,001

com maioria feminina em oito das 12 faixas etárias estudadas. As faixas etárias de < 1 ano, 5 a 9 anos e 15 a 19 anos não apresentaram diferenças estatisticamente significativas entre os sexos ( $p > 0,05$ ). Uma maior proporção do sexo feminino, também prevaleceu em todas as regiões geográficas e ano de notificação, sendo estatisticamente significativas em ambas as situações ( $p < 0,001$ ), à exceção da região Centro-Oeste. O *déficit* de homens como percentual da população total estudada foi de -20,6%.

A informação sobre a idade não foi registrada no formulário em 4.482 das notificações (15,5%). Deste total, a parcela das mulheres correspondeu a 65,3% ( $p < 0,001$ ). As notificações provenientes das regiões Sudeste, Sul e Nordeste corresponderam a, respectivamente, 60,1%, 16,9% e 15,8%. Os dois últimos anos estudados, 2012 e 2013, contribuíram com um maior percentual de notificações registradas, nesta ordem, 20,0% e 24,6% (Tabela 1). Os índices de preferência digital foram obtidos, segundo sexo e regiões agrupadas do país, para cada ano e estão apresentados na Tabela 2.

A avaliação quantitativa da variável “idade”, mediante os IW e IM, demonstrou que os valores das diferentes idades foram, em geral, bem coletados. Analisando a TVAM ao longo do período estudado em relação ao IW, verificou-se uma

melhoria na qualidade dos dados advindos das regiões Nordeste/Norte (-0,6%). Quanto ao IM, verificou-se melhorias na TVAM para todas as variáveis analisadas, à exceção do conjunto regiões Nordeste/Norte que registrou um valor de 1,2% (Tabela 2). Não foi observado, em nenhum dos anos, qualquer padrão de preferência ou repulsão para um ou outro dígito terminal na variável “idade” dentro de cada grupo de comparação (sexo e regiões), segundo o IM. Como forma de ilustrar essa heterogeneidade digital, a Figura 2 retrata o IM para cada dígito terminal nos anos de 2008 e 2013, segundo sexo e regiões do país.

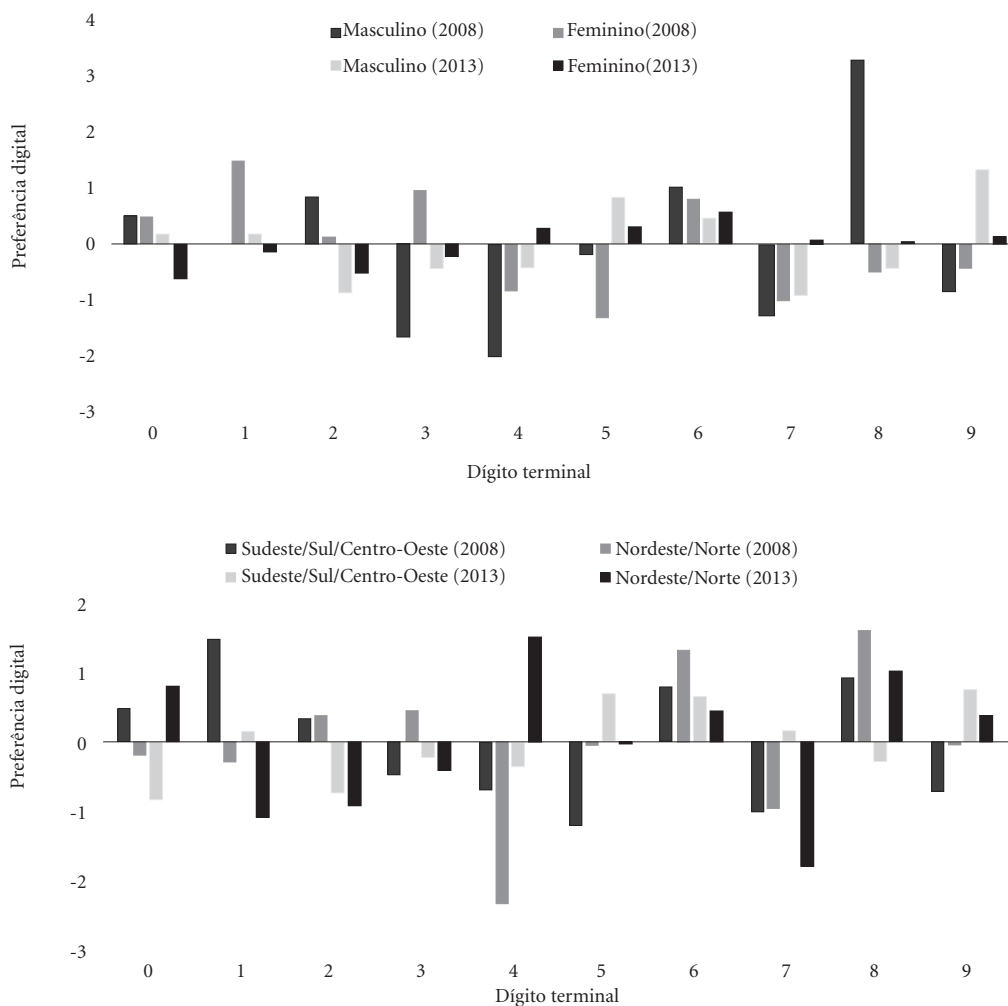
Em relação à variável sexo, houve um destaque nos dígitos 8 (masculino, 2008) e 1 (feminino, 2008). Destaques negativos (repulsões) foram para os dígitos 4 (masculino, 2008) e 3 (masculino, 2008). Os dígitos 4 (Nordeste/Norte, 2008) e 7 (Nordeste/Norte, 2013) foram identificados como os mais evitados enquanto que os dígitos 8 (Nordeste/Norte, 2008) e 1 (Sudeste/Sul/Centro-Oeste, 2008) e 4 (Nordeste/Norte, 2013), ambos empatados, como os menos evitados, segundo as regiões agrupadas do país.

Os resultados do teste de aderência, para cada ano, evidenciaram que não houve preferência por qualquer dígito terminal na variável “idade”: valor  $p$  – variou de  $p = 0,08$  (2010) a  $p = 0,81$

**Tabela 2.** Índices de preferência digital, segundo sexo e regiões agrupadas do Brasil, 2008 a 2013.

Ano	Total	Sexo		Regiões agrupadas	
		Masculino	Feminino	Sudeste/Sul/ Centro-Oeste	Nordeste/ Norte
<i>1. Índice de Wipple (Qualidade dos dados)</i>					
2008	98 (Muito precisos)	104	94	95	104
2009	92 (Muito precisos)	89	94	87	105
2010	109 (Relativamente precisos)	104	112	108	113
2011	96 (Muito precisos)	99	94	98	90
2012	97 (Muito precisos)	89	102	100	87
2013	99 (Muito precisos)	105	96	99	101
TVAM (%)	0,2	0,2	0,4	0,8	-0,6
<i>2. Índice de Myers (Nível de preferência)</i>					
2008	7,8 (Mediano)	11,7	7,9	8,4	8,0
2009	5,9 (Mediano)	10,6	9,0	7,9	7,3
2010	5,3 (Mediano)	9,2	4,7	6,0	11,4
2011	4,0 (Baixo)	5,9	5,2	4,2	8,6
2012	3,3 (Baixo)	7,6	4,3	2,4	8,5
2013	3,4 (Baixo)	6,1	3,0	4,5	8,5
TVAM (%)	-15,3	-12,2	-17,6	-11,7	1,2

TVAM – Taxa de variação anual média.



**Figura 2.** Índice de Myers para cada dígito terminal da variável “idade” dos anos de 2008 e 2013, segundo sexo e regiões do país.

(2012), corroborando com os achados obtidos pelo método de Myers.

## Discussão

O estudo fez uso de dois diferentes métodos utilizados para avaliar a qualidade dos dados da variável “idade” registrados no formulário de notificação do Notivisa-medicamento – principal repositório de dados sobre EAMs do SINAF. Combinada a esta avaliação, foi realizada uma caracterização demográfica dos pacientes notificados no Notivisa-medicamento entre 2008 e 2013.

A composição do sexo e idade dos pacientes são variáveis comumente utilizadas para caracterizar os EAMs registrados nas notificações de sistemas de farmacovigilância. Em geral, no conjunto destas notificações há um predomínio de pacientes do sexo feminino<sup>21-24</sup>, condição também observada no Notivisa-medicamento, em proporções que nas duas diferentes circunstâncias foram estatisticamente significativas quando comparadas àquelas obtidas para os homens. Um estudo francês mostrou que a diferença entre homens e mulheres não se mostrou estatisticamente significativa sob diversas condições, como diferentes faixas etárias e número total de RAMs



graves<sup>22</sup>. As variações observadas entre homens e mulheres podem ser explicadas por fatores biológicos e comportamentais, como peso e composição corporal, metabolismo hepático e renal, consumo de medicamentos, procura por serviços de saúde e taxas de adesão ao tratamento medicamentoso<sup>22,23,25</sup>.

Em relação à idade, um estudo envolvendo notificações do sistema de farmacovigilância de Portugal, entre 2009 e 2011, obteve uma média de 48,5±22,6 anos. A mediana foi de 51 anos (intervalo: 1 dia a 99 anos)<sup>24</sup>. Os dados do Notivisa-medicamento não se encontraram distintos dos valores registrados para o estudo português. A incidência de RAM notificadas em um centro regional de farmacovigilância da França não foi diferente em homens e mulheres dentro de faixas etárias<sup>22</sup>. Contrariando o resultado francês, os dados do Notivisa-medicamento mostraram que a partir da faixa etária entre 15 e 19 anos, os casos de EAMs passaram a ocorrer com maior frequência relativa no sexo feminino, sendo que as diferenças estatisticamente significativas entre homens e mulheres foram observadas a partir do grupo de 20 a 29 anos.

Os pacientes idosos e pediátricos são mais vulneráveis à ocorrência de RAMs, pois a absorção e metabolismo dos fármacos são mais variáveis e menos previsíveis em ambos os grupos<sup>23,25</sup>. No nosso estudo, as idades extremas responderam por uma parcela importante das notificações de suspeitas de EAMs, particularmente, a população de idosos (60 a 69 anos) ocupou a 3ª posição entre as demais faixas etárias analisadas.

No formulário do Notivisa-medicamento há três possibilidades de preenchimento de variáveis relacionadas ao tempo de vida do paciente que, possivelmente, tenha contribuído com um aumento importante dos campos vazios na variável “idade”. Segundo os parâmetros de avaliação da incompletude de uma variável adotados por Romero e Cunha<sup>26</sup>, o campo “idade” na base de dados do Notivisa-medicamento seria classificado como regular (10 a 20%). Este fenômeno, denotado de incompletude, também, foi observado no estudo português que registrou um percentual ainda maior de 22,6%<sup>24</sup>. Várias razões explicam a origem do problema que pode envolver a fonte notificadora, o entrevistado e o formulário de notificação. A título exemplificativo, o entrevistado pode se recusar a responder o item. Às vezes, o entrevistado pode não ter as informações para responder à pergunta, e isso é mais comum quando o entrevistado é uma *proxy* do paciente<sup>27</sup>. Outra possibilidade envolve a má concepção do

formulário que pode confundir a fonte notificadora devido as instruções de preenchimento não serem claras ou suficientes, levando a um preenchimento inadequado<sup>28</sup>.

Inadequações no registro da idade são frequentes, sendo mais comum nos países em desenvolvimento, impossibilitando o uso instantâneo do dado<sup>6,16</sup>. Um dos motivos que explica a má declaração é a preferência digital decorrente do arredondamento sistemático no dígito terminal, por exemplo para “0” ou “5”, que muitas pessoas fazem ao declararem sua idade, bem como preferência ou repulsão por alguns dígitos entre 0 e 9<sup>6</sup>. A preferência digital é uma indicação bastante aceita para mensuração da baixa qualidade dos dados de diferentes bases de dados<sup>8,10,11,29,30</sup>. Os valores obtidos para os índices IW e IM apontam não haver preferência por dígitos na base de dados do Notivisa-medicamento, segundo sexo e regiões agrupadas do país. Também não foi identificada preferência digital para o conjunto dos registros estudados, avaliado mediante o teste de aderência  $\chi^2$ <sup>11</sup>.

Uma das possíveis explicações para a boa qualidade da informação referente à idade no Notivisa-medicamento é que a coleta do dado não ocorre por meio de declaração verbal do paciente ou familiares/cuidadores. A obtenção da informação, possivelmente, é feita diretamente do prontuário do paciente ou calculada a partir da data de nascimento contida neste documento ou em outro que identifica o paciente, como a carteira de identidade.

Estudos que utilizaram os IW e IM demonstraram que a preferência digital ocorreu mais frequentemente em pessoas do sexo feminino<sup>7,10,31</sup> e em dígitos terminados em “0” e “5”<sup>31,32</sup>. Estudos brasileiros mencionam que, em geral, as informações das regiões Norte e Nordeste têm apresentado problemas de qualidade<sup>33-35</sup>. Neste estudo, os dados da variável “idade” oriundos destas regiões não apresentaram problemas importantes de preferência digital que macule sua qualidade.

O uso de registros sem preenchimento da variável “data de nascimento” no processo de detecção de duplicidades pode ser considerado uma limitação deste estudo, haja vista que esta variável aumenta a possibilidade de formação de pares verdadeiramente repetidos. Com vistas a minimizar este problema, nas bases de dados com um percentual maior de não preenchimento (BD2 e BD3) fez-se uso de mais uma variável identificadora, a data de início do evento adverso. Esta variável não foi utilizada nos 5,7% dos registros sem preenchimento da “data de nascimento” na

BD1 por entender que tais registros continham a principal variável para identificar registros repetidos, a variável “nome do paciente”.

Sugere-se que as idades registradas nas notificações do Notivisa-medicamento entre 2008 e 2013 parecem não demonstrar imprecisão no registro e, por conseguinte, esta informação, segundo sexo e regiões agrupadas, pode ser interpretada como de qualidade não duvidosa. Tais resultados apontam para um bom desempenho do SINAF na coleta de dados relacionados à variável “idade.

### **Colaboradores**

DM Mota, A Vigo e RS Kuchenbecker participaram na concepção e desenho do estudo, análise e interpretação dos dados do manuscrito. DM Mota elaborou a redação do artigo e A Vigo e RS Kuchenbecker participaram da sua revisão crítica e aprovação final.

## Referências

- Consonni D, Bertazzi PA, Zocchetti C. Why and how to control for age in occupational epidemiology. *Occup Environ Med* 1997; 54(11):772-776.
- Mcnamee R. Confounding and Confounders. *Occup Environ Med* 2003; 60(3):227-234.
- Alomar MJ. Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharm J* 2014; 22(2):83-94.
- Bandekar MS, Anwikar SR, Kshirsagar NA. Quality check of spontaneous adverse drug reaction reporting forms of different countries. *Pharmacoepidemiol Drug Saf* 2010; 19(11):1181-1185.
- Singh A, Bhatt P. Comparative evaluation of adverse drug reaction reporting forms for introduction of a spontaneous generic ADR form. *J Pharmacol Pharmacother* 2012; 3(3):228.
- Lastra RP, Bolaños MV. *La declaración de la edad. Un análisis comparativo de su calidad en los censos generales de población y vivienda*. Toluca: El Colegio Mexiquense; 1999. Documentos de investigación.
- Bailey M, Makannah TJ. Patterns of digit preference and avoidance in the age statistics of some recent African censuses: 1970-1986. *J Off Stat* 1993; 9(3):705-715.
- Paes NA, Albuquerque ME. Evaluation of population data quality and coverage of registration of deaths for the Brazilian regions. *Rev Saude Publica* 1999; 33(1):33-43.
- Borkotoky K, Unisa S. Indicators to examine quality of large scale survey data: An example through District Level Household and Facility Survey. *PLoS One* 2014; 9(3):1-11.
- Yazdanparast A, Pourhoseingholi MA, Abadi A. Digit preference in Iranian age data. *Ital J Public Health* 2012; 9(1):64-70.
- Brestoff JR, Van Den Broeck J. Reporting data quality. In: Van den Broek J, Brestoff JR, editors. *Epidemiology: Principles and Practical Guidelines*. Dordrecht: Springer; 2013. p. 557-570.
- Agência Nacional de Vigilância Sanitária (Anvisa). *Diretrizes para o Gerenciamento do Risco em Farmacovigilância*. Brasília: Anvisa; 2008.
- Tregunno PM, Fink DB, Fernandez-Fernandez C, Lázaro-Bengoia E, Norén GN. Performance of Probabilistic Method to Detect Duplicate Individual Case Safety Reports. *Drug Saf* 2014; 37(4):249-258.
- European Medicines Agency. *CHMP Guideline on detection and management of duplicate individual cases and Individual Case Safety Reports (ICSRs)*. London: European Medicines Agency; 2012.
- Quantin C, Binquet C, Bourquard K, Pattisina R, Gouyon-Cornet B, Ferdynus C, Gouyon JB, François-André A. Which are the best identifiers for record linkage? *Med Inform Internet Med* 2004; 29(3-4):221-227.
- Denic S, Khatib F, Saadi H. Quality of age data in patients from developing countries. *J Public Health (Oxf)* 2004; 26(2):168-171.
- Zohry A. *Population Size and Change* [Internet]. 2012 [cited 2016 Mar 10]. Available from: <http://www.zohry.com/cdc103/>
- Bay G. Análisis de la estructura por sexo y edad de los censos de población. In: II Reunión regional sobre evaluación y estimaciones demográficas con base en información censal. Santiago: CEPAL; 2012.
- World Bank. *Acerca de la calidad de los datos* [Internet]. [cited 2016 Feb 23]. p. 1-3. Available from: <http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1181743055198/3877319-1181919778981/calidadatos.pdf>
- Alves LC, Andrade PG, Maria PF, Pereira ACR, Marins RL, Brusse GPL, Camargo KCM. *Uma proposta de utilização do software R para a construção de algoritmos de avaliação da qualidade da declaração da idade*. Campinas: Unicamp; 2016.
- Tran C, Knowles SR, Liu B a, Shear NH. Gender differences in adverse drug reactions. *J Clin Pharmacol* 1998; 38(11):1003-1009.
- Montastruc JL, Lapeyre-Mestre M, Bagheri H, Fooladi A. Gender differences in adverse drug reactions: Analysis of spontaneous reports to a Regional Pharmacovigilance Centre in France. *Fundam Clin Pharmacol* 2002; 16(5):343-346.
- Alomar MJ. Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharm J* 2014; 22(2):83-94.
- Silva JC, Soares M, Martins S. *Reações Adversas a Medicamentos-Análise da base de dados do Sistema Nacional de Farmacovigilância (SVIG)*. Lisboa: SVIG; 2012. Relatório Final.
- Resende SO, Theodoro E. Fatores de risco associados às reações adversas a medicamentos antituberculose. *J Bras Pneumol* 2015; 41(1):77-89.
- Romero DE, Cunha CB. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informações sobre Nascidos Vivos, 2002. *Cad Saude Publica* 2007; 23(3):701-714.
- Statistical Policy Office of Management and Budget. *Measuring and Reporting Sources of Error in Surveys*. Washington: Statistical Policy Office of Management and Budget; 2001.
- Household Sample Surveys in Developing and Transition Countries*. Washington: United Nations Department of Economic and Social Affairs; 2005. Vol. F, Studies in Methods.
- Hayes SJ. Terminal digit preference occurs in pathology reporting irrespective of patient management implication. *J Clin Pathol* 2008; 61(9):1071-1072.
- Hessel P. Terminal digit preference in blood pressure measurements: effects on epidemiological associations. *Int J Epidemiol* 1986; 15(1):122-125.
- Barua RK. *Detection of Digit Preference and Age Misreporting by using Demographic Techniques*. Dhaka: East West University; 2015.
- Dahiru T, Dikko HG. Digit preference in Nigerian censuses data of 1991 and 2006. *Epidemiol Biostat Public Heal* 2013; 10(2):6-10.
- Brasil. Ministério da Saúde (MS). *Manual para Investigação do Óbito com Causa Mal Definida*. Brasília: MS; 2009.

34. Kanso S, Romero DE, Leite IDC, Moraes EN. Diferenciais geográficos, socioeconômicos e demográficos da qualidade da informação da causa básica de morte dos idosos no Brasil. *Cad Saude Publica* 2011; 27(7):1323-1339.
35. Santo AH. Causas mal definidas de morte e óbitos sem assistência. *Rev Assoc Med Bras* 2008; 54(1):23-28.

---

Artigo apresentado em 03/08/2016

Aprovado em 25/11/2016

Versão final apresentada em 27/11/2016