Prevalence of corrected arterial hypertension based on the self-reported prevalence estimated by the Brazilian National Health Survey

Prevalência de hipertensão arterial corrigida com base na prevalência autorreferida estimada pela Pesquisa Nacional de Saúde

Prevalencia de la hipertensión arterial corregida, basada en la prevalencia autoinformada, estimada por la Encuesta Nacional de Salud Brasileña

Abstract

The objective was to correct the self-reported prevalence of systemic arterial hypertension (SAH) obtained from the Brazilian National Health Survey (PNS 2013). SAH prevalence estimates were corrected by means of sensitivity/specificity of information. Sensitivity and specificity values from a similar study (same self-report question, age range and gold standard) were used to this end. A sensitivity analysis was also performed, by using the upper and lower limits of confidence intervals as sensitivity and specificity parameters. The corrected prevalence of SAH for Brazil as a whole was 14.5% (self-reported: 22.1%). Women presented a higher rate of self-reported SAH but, after correction, men were found to have a higher prevalence. Among younger women (18-39 age range), the self-reported prevalence was 6.2%, a value that, after correction, dropped to 0.28%. There was not much difference between self-reported and corrected SAH among the elderly (51.1% vs. 49.2%). For certain groups the corrected results were greatly different from the self-reported prevalence, what may severely impact public health policy strategies.

Prevalence; Cross-Sectional Studies; Hypertension; Diagnosis; Self Report

Correspondence

J. P. L. Moreira
Rua Joaquim Rego 29, Rio de Janeiro, RJ 21021-590, Brasil.
jessica@iesc.ufrj.br

1 Instituto de Estudos em Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
2 Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.
3 Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

doi: 10.1590/0102-311X00033619
Introduction

Systemic arterial hypertension (SAH) is a very important public health problem, given its characteristics both as a disease and as a risk factor for other diseases. It has been estimated that its global prevalence among people above 18 years old is around 22%3, and, in the Americas, 14% to 40% among those above 35 years old. Information on this prevalence is useful in the definition of public policies towards its control, with a significant impact over a population health profile.

In Brazil, comprehensive coverage surveys commonly use self-reported measurements in order to classify individuals into disease categories, such as "hypertensive". For instance, the Risk and Protection Factors Surveillance System for Chronic Non-communicable Diseases Through Telephone Interview (VIGITEL) provided a self-reported prevalence of SAH among adults in Brazilian state capitals of 24.1% in 2013 and 24.8% in 2014. Also the health-related supplement of the Brazilian National Household Sample Survey (PNAD) used the "self report" methodology to estimate the prevalence of SAH in 1998, 2003 and 2008, and, more recently (2013), the Brazilian National Health Survey (PNS) made estimates available for the self-reported prevalence of chronic diseases, including SAH. This is currently the most recent and large-coverage survey in this country, and, according to these data, the prevalence of self-reported SAH in Brazil is 21.4%.

However, the real magnitude of arterial hypertension in Brazil is unknown, given that there are no population-based studies with national coverage that used measurement devices to make actual clinical diagnoses in the population. Nonetheless, PNS incorporated some blood pressure measurements from 2013 on, together with data from the self-reporting question.

Although self-reported measurements may be useful, they may present significant bias with no predictable direction or magnitude, and differential or non-differential classification errors. Thus, some individuals may be erroneously classified with the condition; while others, with the disease, may be classified as not having the condition, with a resulting need to correct these estimates, so that they can be brought closer to the real prevalence rates. These corrections could enable public policies that are more effective for the population, especially in what concerns heterogeneities of prevalence relating to age and sex.

A strategy for correcting self-reported measurements is through the use of the sensitivity and specificity values for the question that gave rise to their estimates. Validation studies conducted on self-reported SAH have reported these sensitivities/specificities, but studies may present differences in their gold standards, in the calibration of measurement devices, in the type of measurement performed ("last measurement" or "mean of the last three measurements", for instance), or in the question used to obtain the self-report. In Brazil, six studies on validation of self-reported SAH have been conducted: Lima-Costa et al., Campos, Chrestani et al., Selem et al., Louzada et al., and Menezes et al., in Minas Gerais State; Chrestani et al., in Rio Grande do Sul State; Selem et al. and Louzada et al., in São Paulo State; and Menezes et al. in Paraiba State. Around the world, a variety of studies to validate SAH self-reporting questions, and Vargas et al. and Martin et al. are among the most cited of these. In a systematic review of self-reported hypertension validation studies, Gonçalves et al. included 22 studies and observed a great deal of heterogeneity among countries and age groups.

In addition, especially when dealing with large samples, one has to deal with operational and computational problems related to these corrections, and specific strategies are required in such situations. Therefore, the present study had the objective of correcting the self-reported prevalence of SAH in Brazil, using data available from PNS 2013, presenting the SAH results according to age and sex.

Materials and methods

Brazilian National Health Survey

The PNS is a population-based survey covering the entire territory of Brazil. It uses a complex sampling scheme consisting of cluster sampling and stratification according to census tracts. So, for the present study, we used a subsample selected in three stages: (1) stratification of census tracts; (2)
selection of homes in each tract; and (3) selection of one person aged 18 years or over in each home, by means of simple random sampling. A total of 60,202 people were thus interviewed. Further information on the PNS sampling scheme can be obtained from Souza-Júnior et al. 21.

**Correction of prevalence**

Correction of SAH prevalence can be done algebraically using 10:

\[
p_r = \frac{p_a + Sp}{Se + Sp - 1} \tag{1}
\]

where: \( p_r \) = real prevalence (corrected); \( p_a \) = self-reported prevalence; \( Sp \) = specificity; and \( Se \) = sensitivity.

However, this solution is not unrestrictedly applicable to every sensitivity and specificity value, being limited to the interval \( 1 - Sp \leq p_a \leq Se \) (the complement of specificity and the sensitivity). If this condition is not respected, the solution will present prevalence results that are negative or greater than 1.

With the aim of dealing with this problem, Lew & Levy 11 proposed an adjustment to the formula above, such that the correction would only result in possible and interpretable results. Therefore, this estimator ensures that for any self-reported prevalence values, a correction will be possible. Essentially, this strategy consists in replacing the self-reported prevalence in the previous expression with \( d \):

\[
d = \frac{Se}{1 - Sp} p_a^{x+1} (1 - p_a)^{n-x} dp_a \tag{2}
\]

where: \( n \) = size of the sample; and \( x \) = number of self-reported subjects with the condition.

Therefore, the corrected prevalence is given by:

\[
p_r' = \frac{d + Sp - 1}{Se + Sp - 1} \tag{3}
\]

Notwithstanding the analytical solution, the calculation for \( d \) is not immediate and depends on specific software, capable of complex numerical integration. Moreover, this integration presents a computational limitation relating to the sample size, making it impossible the use of computers when the sample is large (usually greater than 1,000 cases). Therefore, a simplification has been suggested, taking into account a Bayesian method for large samples. This method consists in proportionally reducing the sample size and the number of people with the condition, until reaching the maximum number that can be calculated using the available hardware/software 22.

Thus, in order to calculate the 95% confidence interval (95%CI) for corrected prevalences, the approximation suggested by Lew & Levy 11 was used:

\[
p_r' \pm 1.96SE(p_r') \tag{4}
\]

and \( SE = \) standard error.

However, data for the present study was originated from a complex sample, and therefore the design effect (\( deff \)) needs to be taken into consideration in order to incorporate the estimate loss of precision. The design effect is the ratio between the estimate of the variance from the sampling level actually used and the estimate of the variance if it had been obtained through a simple random sample of the same size. Thus, the variance of the estimate for the corrected prevalence is multiplied by the design effect, which is obtained from the survey data with the sampling strategies mentioned before. Therefore, the standard error that takes into account the design effect follows equation (5):

\[
SE(p_r') = \frac{SE(p_r' (1 - p_r'))}{\sqrt{n(Se + Sp - 1)^2}} \cdot deff \tag{5}
\]

where: \( SE = \) standard error; and \( deff = \) design effect.
Sensitivity and specificity of the question

Correcting the self-reported prevalence of an event demands information on both its sensitivity and specificity, measurements that (e.g.) can be obtained from similar studies in the literature (similar populations, methods, measuring equipment and survey types). Among the four available Brazilian articles on the validation of self-reported SAH, Lima-Costa et al.\textsuperscript{12} used the same question as the PNS, had subjects above 18 years and used a table sphygmomanometer as gold standard. Therefore, that study provided the basic measurements needed for corrections (general and age-related sensitivities and specificities – see Table 1). Combined values for sex and age were obtained from the article raw data (available from the authors).

Variables used

The question used for diagnosing self-reported prevalence was: “Has a doctor or other healthcare professional ever told you that you have high blood pressure or hypertension?” (variable Q004). There were three response categories: “Yes”; “Yes, only during pregnancy” (only for women) and “No”. Women who reported SAH only during pregnancy were included in the category “No”.

Statistical analysis

The prevalence of self-reported SAH was estimated for the population as a whole, according to sex and to age group. Cases in which no information on self-reported SAH was available were excluded from the analysis. In addition, prevalences were also corrected by taking into account the upper and lower values of the 95%CIs for sensitivity and specificity. The adjusted expression for the Bayesian estimator and its adaptation for large samples were used in cases in which the condition $1 - Sp \leq p_a \leq Se$ was not met.

The Maple v.5 software (https://www.maplesoft.com/) was used for integration and other algebraic manipulation.

Results

Table 2 presents the prevalence of self-reported SAH in Brazil according to sex and age, from the PNS 2013, together with the corrected prevalence values developed in the present study. The adjusted expression for the Bayesian estimator and its adaptation for large samples were used among women.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>18-39 years</th>
<th>40-59 years</th>
<th>&gt; 60 years</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (95%CI)</td>
<td>% (95%CI)</td>
<td>% (95%CI)</td>
<td>% (95%CI)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>37.5 (16.8-62.4)</td>
<td>60.0 (44.4-74.2)</td>
<td>76.0 (56.6-89.7)</td>
<td>60.5 (55.8-65.2)</td>
</tr>
<tr>
<td>Female</td>
<td>50.0 (26.6-73.4)</td>
<td>81.5 (70.7-89.6)</td>
<td>82.8 (72.1-90.6)</td>
<td>78.6 (75.2-82.1)</td>
</tr>
<tr>
<td>Total</td>
<td>43.8 (39.3-48.2)</td>
<td>73.3 (68.6-78.1)</td>
<td>80.9 (74.6-87.2)</td>
<td>72.1 (69.3-75.0)</td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>94.6 (90.9-97.2)</td>
<td>84.8 (76.8-90.9)</td>
<td>86.5 (72.6-94.9)</td>
<td>90.9 (88.2-93.7)</td>
</tr>
<tr>
<td>Female</td>
<td>88.8 (84.4-92.3)</td>
<td>74.2 (66.1-81.2)</td>
<td>65.4 (45.9-81.6)</td>
<td>82.6 (79.5-85.8)</td>
</tr>
<tr>
<td>Total</td>
<td>91.4 (88.9-93.9)</td>
<td>78.9 (74.5-83.3)</td>
<td>77.8 (71.2-84.4)</td>
<td>86.4 (84.3-88.6)</td>
</tr>
</tbody>
</table>

95%CI: 95% confidence interval.
aged 18-39 years, among subjects 18-39 years as a whole and for the 95% CIs for sensitivity and specificity among men and women 18-39 years and women 40-59 years. In some categories, large differences between the self-reported and corrected values for SAH could be seen. Across Brazil, regardless of sex and age, the prevalence of corrected SAH was 14.5%, 7.6 percentage points lower than the self-reported value (22.1%). The corrected prevalence for men did not change much, but became higher than that of women (19.5% among men vs. 11.8% among women). This sex difference was especially visible among non-elderly people.

Among males, the corrected prevalence in the 18-39 age group was 0.9%, and, in the 40-59 group, 20.6%. Among elderly men, corrected prevalence increased from 45.8% to 51.7%. Among women, self-reported prevalence was more than three times higher in the age group 40 to 59 years (31% vs. 9.3%). Regardless of sex, self-reported SAH was overestimated in all age groups, but overestimation error decreased with increasing age (Table 2).

As expected, lower prevalences were found when combining the upper and lower limits for sensitivity and specificity (and higher prevalences in the opposite case). Table 3 presents these prevalences. The new prevalence interval varies from 10.8% to 18.5%, and in the age group 18 to 39, a variation between 0.1% and 20.7% could be seen among men. Among women in the same age group, this range was much smaller, going from 0.1% to 1.1%. In the age group 40 to 59, a larger variation for the corrected prevalence also was seen among men, from 2.4% to 43.3%, and, among the elderly, this range was larger among women, whose corrected prevalence varied from 3% to 68.5%.

Discussion

Knowledge on the real magnitude of a disease in a specific population, for instance estimated by correcting self-reported prevalence, is extremely relevant for public health purposes, and Brazilian validation studies on SAH have shown that self-reported prevalence of SAH is usually overestimated by 10% to 15% (without stratification), with larger variations according to sex and age group. The present study indicates that the prevalence of SAH is really overestimated, such that in some categories the self-reported magnitude may even be twice the real prevalence.

It is interesting that in the age group 40 to 59 (a group frequently targeted for health campaigns), self-reported prevalence was more than twofold overestimated. On the other hand, among elderly people (over 60 years of age), overestimation was only 4%. The only category in which self-reported prevalence was underestimated was males, but the degree of underestimation was small.

A validation study in Pelotas (Brazil) also found that SAH self-reported prevalence was underestimated among men. However, among women, the self-reported prevalence of SAH was overesti-

Table 2
Prevalences of self-reported and corrected systemic arterial hypertension (SAH) in Brazil, according to sex and age group, from the Brazilian National Health Survey (PNS), 2013.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Self-reported SAH [% (95%CI)]</th>
<th>Corrected SAH [% (95%CI)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-39 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5.7 (5.1-6.4)</td>
<td>0.90 (0.07-1.7)</td>
</tr>
<tr>
<td>Female</td>
<td>6.7 (6.1-7.4)</td>
<td>0.21 (0-0.32) *</td>
</tr>
<tr>
<td>Total</td>
<td>6.2 (5.8-6.7)</td>
<td>0.28 (0-0.56) *</td>
</tr>
<tr>
<td>40-59 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>24.4 (22.7-26.2)</td>
<td>20.5 (17.0-24.0)</td>
</tr>
<tr>
<td>Female</td>
<td>31.0 (29.5-32.5)</td>
<td>9.3 (7.6-11.0)</td>
</tr>
<tr>
<td>Total</td>
<td>27.9 (26.8-29.1)</td>
<td>13.0 (11.3-14.7)</td>
</tr>
<tr>
<td>≥ 60 years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>45.8 (43.2-48.5)</td>
<td>51.7 (47.5-55.9)</td>
</tr>
<tr>
<td>Female</td>
<td>55.2 (53.0-57.3)</td>
<td>42.7 (38.2-47.2)</td>
</tr>
<tr>
<td>Total</td>
<td>51.1 (49.4-52.9)</td>
<td>49.2 (46.2-52.3)</td>
</tr>
<tr>
<td>Total</td>
<td>19.1 (18.2-20.0)</td>
<td>19.5 (17.7-21.2)</td>
</tr>
</tbody>
</table>

95%CI: 95% confidence interval.
* Lower limit of the confidence interval rounded to 0.
mated more than twofold. This result is in line with other data from the literature, in which women presented higher self-reported prevalence and lower measured prevalence of SAH\(^\text{3,8,23,24}\). On the other hand, at least for men, the corrected values developed here (19.5% for men and 11.8% for women) were very close to the SAH prevalence estimates made by the World Health Organization (WHO) for the Americas region in 2014 (21% for men and 16% for women)\(^3\).

Despite the need for such corrections, only one other study could be found (osteoarthritis in France) in which self-reported prevalence was corrected by means of sensitivity/specificity information. In that case, the authors found that the prevalence was underestimated when self-reported measurements were used (7.9% for self-reporting; 9.1% for the corrected estimates)\(^25\).

The self-reported prevalence found in this study, using data from PNS, was 22.1%. With the same data, Andrade et al.\(^8\) found a value of 21.4%. This difference is due to the cases in which information was not available, which were excluded from the present study.

As mentioned, the present study used sensitivity/specificity values from another study\(^12\) in order to obtain corrected estimates for self-reported SAH in Brazil as a whole. Although the populations studied in Lima-Costa et al. and here are not specifically the same, it should be noted that both studies used the same question for ascertaining SAH, included subjects above 18 years old and used the same gold standard for validating SAH. These similarities (and the fact that questions were asked in the same language) guarantee a degree of methodological consistency for the use of those estimates. Nevertheless, a limitation of the present study is that further validation should be sought using sensitivity/specificity values from more recent/more comprehensive data, including different regions of the country. Also, in the present study, prevalences were corrected by simulating different combinations of sensitivity and specificity, taking into account their lower and upper confidence interval limits. Although this is an interesting strategy for the inclusion of uncertainties, it does not consider the plausibility of the results and should be considered as a "worst case" scenario, since it analyzes the combinations of the upper and lower limits of sensitivity/specificity as if these values had the same likelihood to occur. Therefore, an excessively pessimistic or conservative image of the results might be obtained\(^9\). Another means of obtaining representative and plausible sensitivity/specificity values would be through a meta-analysis, in which all the articles validating the self-report question would be included.

This study presented the corrected prevalence of SAH in Brazil, according to age and sex, taking as its basis the sensitivity and specificity values of a self-report question posed in 2013. The resulting estimates are therefore closer to the real prevalence, and it was observed that, in all categories except men, the prevalence of SAH was overestimated when the subjects were asked about the disease. In addition, the corrected values were closer to and in the same direction of worldwide estimates for the prevalence of SAH. This result is extremely important, since it would enable the formulation of public policies that take into account the proportion of individuals in the Brazilian population that actually have this condition.

### Table 3

Corrected prevalences of systemic arterial hypertension (SAH), taking into consideration the lower and upper limits of the 95% confidence intervals for sensitivity and specificity of the validation study of Lima-Costa et al.\(^12\), according to sex and age group.

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ULSe and LLSp</td>
<td>LLSe and ULSp</td>
<td>ULSe and LLSp</td>
</tr>
<tr>
<td>18-39 years</td>
<td>0.10</td>
<td>20.7</td>
<td>0.10</td>
</tr>
<tr>
<td>40-59 years</td>
<td>2.4</td>
<td>43.3</td>
<td>1.0</td>
</tr>
<tr>
<td>≥ 60 years</td>
<td>29.5</td>
<td>79.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Total</td>
<td>13.7</td>
<td>25.9</td>
<td>6.7</td>
</tr>
</tbody>
</table>

LLSe: lower limit of the 95% confidence interval for sensitivity; LLSp: lower limit of the 95% confidence interval for specificity; ULSe: upper limit of the 95% confidence interval for sensitivity; ULSp: upper limit of the 95% confidence interval for specificity.
Contributors
The authors contributed equally to this work.

Additional informations
ORCID: Jessica Pronestino de Lima Moreira (0000-0003-1987-3584); Renan Moritz Varnier Rodrigues de Almeida (0000-0003-0158-7432); Nei Carlos dos Santos Rocha (0000-0002-7493-6069); Ronir Raggio Luiz (0000-0002-7784-9905).

Acknowledgments
We are grateful for the original data from the study conducted in Bambuí, which were made available to us by Professor Maria Fernanda Lima-Costa.

References
15. Selem SSAC, Castro MA, César CLG, Marchioni DML, Fisberg RM. Validity of self-reported hypertension is inversely associated with the level of education in Brazilian individuals. Arq Bras Cardiol 2013; 100:52-9.
Resumo

O estudo teve como objetivo corrigir a prevalência autorreferida de hipertensão arterial sistêmica (HAS) obtida pela Pesquisa Nacional de Saúde (PNS 2013). As estimativas de prevalência de HAS foram corrigidas pelos dados de sensibilidade/especificidade. Foram utilizados os valores de sensibilidade e especificidade de um estudo semelhante (mesma pergunta autorreferida, faixa etária e padrão de ouro). Foi utilizada também a análise de sensibilidade, com os limites superiores e inferiores dos intervalos de confiança enquanto parâmetros de sensibilidade e especificidade. A prevalência corrigida de HAS para o Brasil como um todo foi de 14,5% (autorreferida: 22,1%). As mulheres apresentaram uma prevalência mais alta de HAS autorreferida, mas depois da correção, os homens mostraram uma prevalência mais alta. Entre as mulheres mais jovens (18-39 anos), a prevalência autorreferida foi de 6,2%, caindo para 0,28% depois da correção. Nos idosos, não houve muita diferença entre a HAS autorreferida e a corrigida (51,1% vs. 49,2%). Para determinados grupos, os resultados corrigidos foram muito diferentes da prevalência autorreferida, o que pode ter um impacto relevante nas estratégias de saúde pública.

Prevalência; Estudos Transversais; Hipertensão; Diagnóstico; Autorrelato

Resumen

El objetivo fue corregir la prevalencia autoinformada de hipertensión arterial sistémica (HAS), obtenida de la Encuesta Nacional de Salud Brasileña (PNS 2013). Las estimaciones de prevalencia HAS se corrieron mediante información de sensibilidad/especificidad. Los valores de sensibilidad y especificidad de un estudio similar (la misma pregunta autoinformada, rango de edad y estándar de excelencia) se usaron hasta el final. También se realizó un análisis de sensibilidad, usando los límites superiores e inferiores de los intervalos de confianza como parámetros de sensibilidad y especificidad. La prevalencia corregida de HAS para Brasil como un todo fue 14,5% (autoinformada: 22,1%). Las mujeres presentaron una tasa más alta de HAS autoinformada pero, tras la corrección, los hombres fueron quienes tuvieron una prevalencia más alta. Entre mujeres más jóvenes (con un rango de edad entre 18-39), la prevalencia autoinformada fue 6,2%, un valor que, tras la corrección, cayó al 0,28%. No hubo mucha diferencia entre la HAS autoinformada y la corregida entre ancianos (51,1% vs. 49,2%). Para ciertos grupos los resultados corregidos fueron considerablemente diferentes respecto a la prevalencia autoinformada, lo que quizás tiene un impacto severo en las estrategias de las políticas públicas de salud.

Prevalencia; Estudios Transversales; Hipertensión; Diagnóstico; Autoinforme

Submitted on 21/Feb/2019
Final version resubmitted on 27/Jun/2019
Approved on 09/Jul/2019