ABSTRACT: Objective: The study aimed to validate the estimates of adult smokers determined by Vigitel for small areas, defined by the Health Vulnerability Index (IVS). Methods: The database of the Health Survey of the Metropolitan Region of Belo Horizonte (RMBH) carried out in 2010 and the data from Vigitel in the period from 2010 to 2013 were used to obtain estimates of adult smokers by IVS. With Vigitel, the estimate of smokers by IVS was obtained by the indirect estimation method in small areas. The prevalence of adult smokers was compared, considering RMBH as the gold standard. The t test was used to evaluate the difference between the means and the Pearson correlation, with a significance level of 5%. Results: When stratifying by IVS in the household survey, the prevalence of adult smokers ranged from 13.39% (95%CI 11.88 – 14.91) for residents in a low-risk area to 22.9% (95%CI 12.33 – 33.48) among residents in a very high-risk area. With Vigitel, according to IVS, the prevalence of adult smokers ranged from 11.98% (95%CI 10.75 – 13.21) for residents in the low-risk area to 22.31% (95%CI 18.25 – 26.1) in very high-risk areas. The prevalence was similar between the two surveys, showing good Pearson correlation (r = 0.93). Conclusion: The study points out that the estimates of smokers were similar in both surveys, showing the external validity of Vigitel. There was a gradient in prevalence, with progressive increase, identifying a higher proportion of smokers in high-risk areas.

Keywords: Small-area analysis. Validation study. Health surveys. Logistic models. Selection bias.
INTRODUCTION

In Brazil, health information systems are decentralized, generating multiple information per municipality, and they still have problems in coverage because they are limited to the populations served, reliability of information, quality of data filled in and low standardization\(^1\).

On the other hand, national health surveys have great advantages, due to their scope, their breadth of topics, as they cover the entire population, and for assessing health conditions in their numerous themes, including determinants and conditions\(^2\).

National surveys, such as the National Health Survey\(^3\), Risk and Protection Factors Surveillance System for Chronic Diseases by telephone survey (Vigitel)\(^4\) and Longitudinal Study of the Health of Elderly Brazilians (Elsi)\(^5\), are important examples of the use of information for health planning and management. Because of they have operational difficulties and are expensive and time-consuming, these surveys do not allow estimations according to municipality, or for less populated areas, such as census tracts or aggregates of them.

However, there is a growing demand for health information made available by small areas, as it allows to verify intra-urban inequalities and support the formulation of local public policy programs\(^6,7\).

Accordingly, the Municipal Health Secretariat of Belo Horizonte developed in 1998 the Health Vulnerability Index (IVS) with the objective of guiding the planning of health measures. The IVS is a measure that associates socioeconomic and environmental variables in the same index and allows the analysis of the characteristics of population groups residing...
in the census sectors. In 2012, this index was updated with data from the 2010 Census, being divided into four clusters of health risk called low, medium, high and very high. It is worth mentioning that the area of statistics has contributed to methods for obtaining reliable estimates for smaller areas not initially covered in research sampling plans, just as demographics has used estimation methods in small areas to project the population by neighborhood, district or regional health through the use of information available for a larger area, such as the state.

The indirect method for estimates in small areas has gained popularity, where it uses available auxiliary variables to predict the variable of interest in smaller areas. This method uses the variables available in the demographic census to construct the predictive model. However, the quality of fit of the model depends on the existence of one or more auxiliary variables correlated with the small areas to obtain a good predictive model.

In epidemiological studies, there is a concern with producing valid estimates for the study population. Validity refers to the property of an instrument to measure exactly what is proposed. In this sense, the use of appropriate statistical techniques contributes to external validation, with regard to the generalization of these results to an external population. In several epidemiological studies, authors have compared the results obtained with a gold standard as a validation measure, called criterion validity. In this type of study, Pearson’s correlation coefficient is used to measure the correlation between the study’s estimates and the gold standard.

In view of the above, the present study aimed to validate the estimates of the prevalence of adult smokers, obtained by Vigitel Belo Horizonte, by IVS, using as reference the Health Survey of the Metropolitan Region of Belo Horizonte (RMBH), home-based survey with face-to-face interview. In this way, we expected to contribute to the methodological advance, aiming to determine health estimates for small areas.

**METHODS**

This is an external validation study that compared two cross-sectional surveys, carried out between 2010 and 2013, in the city of Belo Horizonte, Minas Gerais.

**DEFINITION OF SMALL AREAS**

In this study, the IVS, updated in 2012 with data from the 2010 Census, was used, which delimits the areas of health vulnerability in the municipality, highlighting the inequalities of different social groups. In the composition of the IVS, variables were used, such as sanitary sewage and inadequate water supply, head of household, number of residents and family composition.

According to data from the 2010 Census, Belo Horizonte had 3,830 census sectors, which were grouped according to the IVS into four health risk clusters: low (1,330 sectors), medium (1,460 sectors), high (737 sectors) and very high (303 sectors).
VIGITEL DATABASE

Vigitel held in Belo Horizonte by the Ministry of Health, from 2010 to 2013\textsuperscript{12-15}, interviewed annually about 2,000 adults over 18 years of age.

This study selected adults aged 20 years and over, in the period from 2010 to 2013, to obtain estimates of the prevalence of smokers by IVS. Of the 3,830 census sectors, 2,790 (73\%) had at least one Vigitel interview. Bernal et al.\textsuperscript{7} reported the details of the linkage of Vigitel and the National Register of Addresses for Statistical Purposes.

HOUSEHOLD SURVEY DATABASE

The RMBH health survey, home-based, interviewed 7,778 adults aged 20 years and over, living in Belo Horizonte. Data were collected using a supplementary questionnaire of the Employment and Unemployment Survey of the Metropolitan Region of Belo Horizonte (PED/RMBH), conducted in 2010 by the João Pinheiro Foundation, an organ of the Government of the State of Minas Gerais\textsuperscript{16}.

Participants in the RMBH survey were selected according to a probabilistic sample, stratified in two stages. The census sectors of the Brazilian Institute of Geography and Statistics (IBGE) were the primary selection unit, and the sample unit was the home in the urban area of the 26 municipalities that comprised the RMBH. All residents of the selected households, aged 20 or over, were invited to participate in the interview. The final sample included 5,798 households in the RMBH, with 3,440 located in the city of Belo Horizonte.

The basis of the RMBH Health Survey had 7,479 (96\%) interviews with information from the census sector and the 2003 IVS. From the census sector code, it was possible to add the 2012 IVS in 7,438 (99\%) interviews, out of a total of 7,479 sectors.

The household survey was used to carry out the external validation of Vigitel, comparing the estimates of the prevalence of smokers by IVS 2012.

INDIRECT ESTIMATION IN SMALL AREAS

In this study, the methods proposed by Bernal et al.\textsuperscript{7} were used to estimate the prevalence of adult smokers according to IVS using data from Vigitel. This method consists in using statistical models for imputing missing data. The logistic regression model was used to impute the variable dichotomous response, smoker (1) or non-smoker (0), in the set of census sectors without any Vigitel interview, which corresponded to 1,040 (27\%) in the period from 2010 to 2013.

The adjusted model was applied to the set of sectors without a Vigitel interview to estimate the probability that an adult in the census sector without a Vigitel interview would be classified as a smoker or non-smoker and thus complete the variable smoker for the set of sectors without any Vigitel interview.
EXTERNAL VALIDATION

This study sought to measure external validity or the extent to which Vigitel results can be generalized or applied to other contexts. Accordingly, external parameters are required, in this case, the household survey carried out in 2010, which was taken as a reference standard. Therefore, a comparison was made between the two surveys to validate Vigitel’s ability to estimate the prevalence of smokers in similar population groups, being taken as having achieved external validity.

In Vigitel, smokers are those who answer “yes” to the question: Do you smoke? (“Yes, daily”; “yes, but not daily”; and “no”). In the RMBH survey, the analysis is performed according to the answers to the following question: Which of the following phrases best defines your habits regarding smoking cigarettes?

- Has not smoked more than a hundred cigarettes during whole life;
- Has smoked more than hundred cigarettes during whole life, but stopped smoking;
- Smokes some days, but not everyday;
- Smokes everyday but less than a pack of cigarettes;
- Smokes everyday between one and two packs of cigarettes;
- Smokes everyday at least two packs of cigarettes;
- Has never smoked.

The study compared the prevalence of adult smokers according to the 2012 IVS, in the two population surveys, using the RMBH health survey, performed on a household basis, as the gold standard. For this purpose, it was necessary to calculate the estimates of the household survey for the 2012 IVS, using post-stratification weights to match the sample distribution to that of the 2010 population by IVS. These weights were determined using the rake method using data from the 2010 Census as the reference population. The Vigitel survey also used post-stratification weights, as described in other publications. These weights were determined with the rake method using the SURVWGT package available in the STATA program.

For the purpose of comparing the surveys, the data from the health survey and Vigitel were adjusted for the population of the 2010 Census by IVS using the post-stratification weights calculated in STATA. For external validity, the hypothesis test was used for the difference between the means of the two surveys (Vigitel and Health Survey), as well as Pearson’s correlation, with a significance level of 5%.

This project was approved by the Human Research Ethics Committee of the Federal University of Minas Gerais (UFMG), in the project “Inequalities in small geographical areas of indicators of chronic non-communicable diseases and violence and their risk factors”.

RESULTS

Figure 1 compares the distribution of the RMBH health survey estimates with the 2010 Census, using the 2012 IVS. It is noted that the low and medium IVS are...
overrepresented, and the high and very high IVS are underrepresented in the sample when compared to the 2010 Census (Figure 1A). After adjusting the post-stratification weights by the 2012 IVS, the distribution of the sample is equal to the distribution of the 2010 Census (Figure 1B).

Table 1 compares the classification of the census sectors by IVS (low, medium, high and very high) in the Belo Horizonte health survey, according to the application of the IVS in 2003 and in 2012. In the comparison between the 2003 and 2012 IVS, most of the interviews categorized as low- (94.8%) or medium- (82.3%) risk remained in the same classification, while in the high- and very high-risk groups, 46.1 and 56.9%, respectively. However, 48.5% of the interviews in the high-risk group migrated to the medium-risk group, and 5.4%, to the very high-risk group. Similarly, 32.8% of the interviews in the very high-risk group migrated to the high-risk group, and 10.3%, to the medium-risk group (Table 1).

![Image: Figure 1A and B]

IVS: Health Vulnerability Index; RMBH: Metropolitan Region of Belo Horizonte.

**Figure 1.** (A) Sample of adults (≥20 years old) by Health Vulnerability Index 2012 and (B) adult population (≥20 years old) adjusted by Health Vulnerability Index 2012, both stratified by data source. Belo Horizonte.

**Table 1.** Comparison of the distribution of census sectors by Health Vulnerability Index (low, medium, high and very high) in the Belo Horizonte health survey, according to the Health Vulnerability Index 2003 and 2012.

<table>
<thead>
<tr>
<th>IVS 2003</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Very high</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Low</td>
<td>2,808</td>
<td>94.8</td>
<td>144</td>
<td>4.9</td>
<td>–</td>
</tr>
<tr>
<td>Medium</td>
<td>364</td>
<td>11.5</td>
<td>2,605</td>
<td>82.3</td>
<td>198</td>
</tr>
<tr>
<td>High</td>
<td>–</td>
<td>–</td>
<td>578</td>
<td>48.5</td>
<td>550</td>
</tr>
<tr>
<td>Very high</td>
<td>–</td>
<td>–</td>
<td>12</td>
<td>10.3</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>3,172</td>
<td>42.6</td>
<td>3,339</td>
<td>44.9</td>
<td>786</td>
</tr>
</tbody>
</table>

IVS: Health Vulnerability Index.
The results of Vigitel, according to the IVS, in the period of 2010 to 2013, estimated a mean prevalence of 15.06% (95% confidence interval — 95%CI 14.24 – 15.89), and showed a gradient in prevalence of adult smokers, with 11.98% (95%CI 10.75 – 13.21) among residents in the low-risk area, reaching 22.21% (95%CI 18.25 – 26.10) in the high-risk IVS (Table 2).

In the external validation of Vigitel, data from the health survey carried out in 2010 were used. The household survey showed that the prevalence of adult smokers was 15.12% (95%CI 14.12 – 16.10). When stratifying by IVS, this prevalence ranged from 13.39% (95%CI 11.88 – 14.91) among residents in a low-risk area to 22.09% (95%CI 12.33 – 33.48) among residents in high-risk areas (Table 2).

When comparing the prevalences estimated by Vigitel and the RMBH survey, both showed a gradient, increasing the prevalence for the areas of high and very high IVS. The prevalences were similar between the two surveys and there was no difference in the estimates of smokers per IVS, showing a high Pearson correlation (r = 93%) and good external validity of Vigitel (Table 2, Figure 2).

**DISCUSSION**

This was a study of the external validation of Vigitel for small areas, using the household survey conducted in Belo Horizonte in 2010 as the gold standard. Estimates of prevalence of adult smokers according to IVS were analyzed, which was not foreseen in the sample plan. Estimates were similar between the two surveys, showing that the Vigitel results have external validity. It was possible to observe the occurrence of a positive gradient in prevalence of smoking as the risk increases, where the proportion of smokers was about twice as high in areas of very high risk.

The study validated the new methodology to generate estimates of smokers for small areas, which can be useful to estimate prevalences of other risk factors and diseases in small areas and thus to identify subgroups with higher social risk and to help in the development, monitoring and evaluation of programs in the areas of public health policies aimed at tackling smoking.

Table 2. Estimate of prevalence* of adult smokers by Health Vulnerability Index, according to type of survey.

<table>
<thead>
<tr>
<th>IVS 2012</th>
<th>RMBH – Household survey</th>
<th>Vigitel Indirect method SAE</th>
<th>Hypothesis test**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (A) 95%CI</td>
<td>% (B) 95%CI</td>
<td>B – A p-value</td>
</tr>
<tr>
<td>Low</td>
<td>13.39 11.88 14.91</td>
<td>11.98 10.75 13.21</td>
<td>-1.42 0.127</td>
</tr>
<tr>
<td>Medium</td>
<td>15.36 13.98 16.73</td>
<td>15.61 14.30 16.92</td>
<td>0.26 0.785</td>
</tr>
<tr>
<td>High</td>
<td>15.56 12.30 18.83</td>
<td>17.90 15.78 20.02</td>
<td>2.33 0.194</td>
</tr>
<tr>
<td>Very high</td>
<td>22.90 12.33 33.48</td>
<td>22.21 18.25 26.16</td>
<td>-0.69 0.883</td>
</tr>
<tr>
<td>Total</td>
<td>15.12 14.12 16.13</td>
<td>15.06 14.24 15.89</td>
<td>-0.06 0.927</td>
</tr>
</tbody>
</table>

IVS: Health Vulnerability Index; RMBH: Metropolitan Region of Belo Horizonte; SAE: Small Area Estimation; 95%CI: 95% confidence interval; *adjusted for the 2010 Census by Health Vulnerability Index; **difference between two means.
In 2014, Malta et al.19 presented the first results of the Vigitel Belo Horizonte 2010 indicators for smaller areas. The authors pointed out intra-urban differences, with the districts Norte, Venda Nova, Barreiro, Nordeste and Noroeste having the worst sociodemographic indicators and the highest rates of consumption of milk with fat, low regular consumption of fruits and vegetables and low level of exercise in free time. These results already pointed out the importance of using Vigitel to produce estimates in small areas, which was further elaborated in the present study on the basis of estimates by census sector, combined in the four IVS, which managed to address the dimension of inequality in the municipality of Belo Horizonte.

The literature points out that low income and education are associated with higher prevalence of smoking both in Brazil20 and in other countries21, increasing vulnerability21. This fact corroborates the results found in the present study, in which the prevalences in high-risk areas were higher, while the low-risk areas, in which populations with better socioeconomic conditions reside, had the lowest prevalences.

Bernal and Silva22 showed that fixed telephone coverage is not evenly distributed among the population and that users are concentrated in the most favored social classes. In addition, possession of a landline is positively associated with education and white skin color. These findings coincide with the results found in this study, given that most of the census sectors identified in Vigitel are concentrated in groups of low and medium health risk, showing the importance of using post-stratification weights to reduce the possible bias in the estimates23.

RMBH: Metropolitan Region of Belo Horizonte; IVS: Health Vulnerability Index.

Figure 2. Correlation between the prevalence of smokers by Health Vulnerability Index, according to survey.
The prevalence of tobacco indicators in Brazil has been on the decline, which has been pointed out in other studies\textsuperscript{24,25} and attributed to the regulatory measures implemented by Brazil. Among the measures, we highlight the prohibition of advertising tobacco products; the country’s adherence to the Framework Convention for Tobacco Control\textsuperscript{24}, in 2006; Law No. 12.546 of 2011, which instituted tobacco-free environments\textsuperscript{26}, and Decree No. 8.262/2014, which regulated these environments, determined an increase in spaces for warnings and anti-smoking images on packaging, prohibiting the sale of these products to minors (under 18 years old), the increase in taxes, among other measures\textsuperscript{27}.

Despite these studies pointing to a decline in the prevalence of tobacco indicators, Bernal et al.\textsuperscript{7} showed a downward trend in the prevalence of adult smokers in the low and high risk areas between the periods 2006 to 2009 and 2010 to 2013, while in the medium and high risk areas, the indicator remained stable in the same period. This result shows that the poorest socioeconomically population is more affected by risk factors for chronic diseases, such as smoking, gets sicker and has worse access to health services, which further increases the picture of inequalities that affects Brazil.

In addition, the spatial theme and its relationship with health have been addressed in studies in the country, such as the Adult Health Survey of the Metropolitan Region of Belo Horizonte (Minas Gerais), which assessed the perception of the neighborhood’s social environment and self-reported morbidity,\textsuperscript{28}; the Saúde em Beagá Study, which investigated the psychometric qualities of contextual characteristics measured from the perception of the social and physical environments of residents participating in the survey; as well as the association of healthy eating in areas of better socioeconomic status\textsuperscript{6}. These studies contribute to the mapping of health inequalities in small areas.

The present study showed differences in the profile of smokers due to IVS, which shows the importance of identifying these subgroups for the formulation, implementation, monitoring and impact assessment of programs for each health risk group, in the areas of public health policies aimed at coping with smoking.

Vigitel has been an important tool for monitoring smoking indicators in Brazil, standing out for its agility and its annual monitoring of all capitals. The present study points out that, being accessible in databases containing information by census sector, one can also make estimations for smaller areas and define more specific risk areas.

The present study had some limitations. First, 14% of Vigitel’s interviews were excluded due to the non-identification of census sectors by linkage. Second, there was the allocation of adult smokers or non-smokers to the group of sectors without any Vigitel interview. Third, post-stratification weights were used, according to the 2010 Census population by IVS, to minimize potential bias due to the absence of sectors without interviews and, also, for the joint analysis of the interviews by period. In addition, there was a greater presence of adults in the low- and medium-risk groups, and less presence of those belonging to the high-risk and very high-risk groups, in both surveys. Thus, the household survey may not produce adequate estimates in areas of high risk, as well as Vigitel.
The present study highlights the external validity of Vigitel when compared to the household survey, showing its ability to estimate differences in the prevalence of tobacco use among vulnerable population groups. Thus, analyses for small areas can identify vulnerability and the need to invest in specific policies for this group. Vigitel data proved to be valid and can be used to estimate inequalities in small areas.

REFERENCES


Recebido em: 23/09/2020
Revisado em: 11/12/2020
Aceito em: 13/12/2020
Preprint on: 12/15/2020

Authors’ contributions: RTIB and DCM conceptualized the study, defined the objectives, performed the statistical analyses, and wrote the purpose, introduction, methods and discussion. SVP and MFLC contributed to the preparation of the methods, results and discussion. All authors contributed to the revising of the manuscript and approved the final article.