CSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH



# Women who have undergone abortion in the city of Rio de Janeiro, Brazil: application of a Bayesian hierarchical model

## Mulheres que fizeram aborto no Município do Rio de Janeiro, Brasil: aplicação de um modelo hierárquico bayesiano

## Mujeres que abortaron en el Municipio de Río de Janeiro, Brasil: aplicación de un modelo jerárquico bayesiano

*Natália Santana Paiva* [1,2]
*Daniel Antunes Maciel Villela* [3]
*Leonardo Soares Bastos* [3]
*Francisco I. Bastos* [4]

doi: 10.1590/0102-311X00190718

**Abstract**

*Estimates of number of women who have undergone induced abortion in jurisdictions with restrictive abortion laws are still scarce in the scientific literature, and the disparate estimates from currently used methods call for the application of innovative estimation techniques such as new indirect methods. This need is especially acute in more densely populated areas, such as Brazil's state capitals, given the magnitude of unsafe abortions and the resulting risks and harms. The article aims to estimate the number of women who had induced abortions in the city of Rio de Janeiro in 2011, based on a Bayesian hierarchical model. The model was applied to data from a household survey that supported the use of the network scale-up method in the city of Rio de Janeiro, a Bayesian hierarchical model using indirect information based on the contact networks of randomly selected participants from the general population. Among the 1,758,145 women 15-49 years of age living in the city of Rio de Janeiro, 13,025 women (95%CrI: 10,635; 15,748) had induced abortions in 2011, resulting in a mean cumulative incidence of 7.41 (95%CrI: 6.05; 8.96) for every 1,000 women 15-49 years of age. The model's self-validation process identified patterns of underestimation in stigmatized subpopulations with low social visibility, such as women who have undergone induced abortion. Induced abortion is a common practice among women in the city of Rio de Janeiro. New indirect estimation methods can contribute to more precise measurement of this event, considering the context of illegality, and thereby contribute to appropriate health policies.*

*Induced Abortion; Population Estimates; Statistical Data Analyses; Multilevel Analysis; Estimation Thechniques*

**Correspondence**
*N. S. Paiva*
*Instituto de Estudos em Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil*
*Av. Horácio Macedo s/n, Cidade Universitária, Rio de Janeiro, RJ 21941-598, Brasil.*
*natalia_uff@hotmail.com*

*[1] Instituto de Estudos em Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.*
*[2] Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.*
*[3] Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.*
*[4] Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.*

## Introduction

The high incidence of induced and often unsafe abortions is still a public health problem and one of the principal indices in reproductive health. The global rate of unintended pregnancy was 62 (90% credible interval or 90%CrI: 59; 72) per 1,000 women 15 to 44 years of age from 2010 to 2014, and 56% (90%CrI: 53; 60) of all unintended pregnancies ended in abortion during the same period [1].

According to Bearak et al. [1], during the same period, developing (low and middle-income) countries reported a higher unintended pregnancy rate (mean of 65 per 1,000 women 15-44 years of age; 90%CrI: 62; 76) than developed (high-income) countries (mean of 42 per 1,000 women 15-44 years of age; 90%CrI: 38; 56). However, there was no significant difference at 90% credibility level in the percentage of unintended pregnancies that ended in abortion in developing (55%; 90%CrI: 52; 60) versus developed countries (59%; 90%CrI: 54; 65).

The World Health Organization (WHO) defines induced abortions as those initiated by deliberate action with the intention of terminating the pregnancy [2]. Induced abortions are often performed by persons without the minimum necessary skills or in locations that fail to comply with the most basic medical and hygienic standards, or both, thus characterizing unsafe abortions and posing avoidable risks to the woman's health and life [3].

Induced abortion is a serious problem in the areas of reproductive rights and public health and is one of the principal causes of maternal mortality. Developing countries show higher numbers of unsafe abortions and higher maternal mortality than developed countries [4].

Brazil, within a broad set of more than 65 countries, has quite restrictive laws on induced abortion. As in several other countries, induced abortion is subject to criminal prosecution (i.e., under the penal code) [5]. Such restrictive laws tend to be more present in low and middle-income countries and are uncommon in high-income countries (for example, the great majority of Western European countries and various states of the USA), for overlapping but distinct reasons. For example, in Brazil, abortion is a matter addressed by both the legislative and judiciary branches and is now under review in the Supreme Court (STF).

However, studies have shown that unsafe abortion is a common event in the reproductive life of Brazilian women, even in the context of illegality [6,7]. Meanwhile, the lack of official records on induced abortions hinders the measurement of the number of women who have undergone induced abortions in Brazil.

In the context of stigmatization and/or criminalization, nationwide studies [7,8,9] for decades have used indirect techniques to estimate unsafe abortions based on secondary data from Health Information Systems with the application of correction factors developed in the 1990s by the Alan Gutmacher Institute [10]. In 2010 and 2017, the most comprehensive Brazilian national surveys on abortion used the ballot-box technique to estimate the number of literate women 18-39 years of age in urban areas who had undergone at least one abortion any time in life [6,10].

An alternative methodology for estimating the size of hard-to-reach populations (often stigmatized and/or criminalized) has been employed in various public health settings both in international [11,12,13,14,15] and Brazilian studies [16,17,18]. The methodology is still under development and is called the network scale-up method (NSUM).

NSUM draws on indirect information on contact networks of participants from the general population. By transcending the exclusively individual dimension and incorporating the dimension of the interviewees' social networks, NSUM differs from other previously used direct and indirect methods.

The study aimed to estimate the number of women who had undergone induced abortion in the city of Rio de Janeiro in the year 2011. A Bayesian hierarchical model based on indirect information from the contact networks of randomly selected participants from the general population is applied This article takes an innovative approach in Brazil by applying a Bayesian hierarchical model based on indirect information from the contact networks of randomly selected participants from the general population to estimate the of number of women who have undergone abortion.

## Material and methods

### Study design

This study is a subproject of a broad national survey concluded in 2012, entitled *Profile of Crack Cocaine Users in the 26 State Capitals, Federal District, 9 Metropolitan Areas, and Brazil* (the latter term corresponding to all other municipalities and the rural areas), developed by the Oswaldo Cruz Foundation (Fiocruz), whose main objective was to estimate crack cocaine consumption and which also covered a series of other public health issues, such as induced abortion, illegal drug use in general (except marijuana – *cannabis*), and female sex work, among others.

The data collected in the national household survey on the *Profile of Crack Cocaine Users in the 26 State Capitals, Federal District, 9 Metropolitan Areas, and Brazil* allow estimating the size of hard-to-reach populations, for example women who have undergone induced abortion in the city of Rio de Janeiro in the year prior to the interview, using the network scale-up method (NSUM) [17].

In the above-mentioned survey, the city of Rio de Janeiro (whose total population in 2010 was 6,320,446 according to the 2010 *Brazilian Demography Census* [20]) had a sample of 1,547 participants, randomly selected according to the complex sampling plan adopted in the household survey [17]. The sample from the city of Rio de Janeiro consisted of 991 women (63.53%), 563 men (36.09%), and 6 individuals for whom information as to their sex was missing (0.38%).

The household survey whose design backed the application of the network scale-up method included questions about 20 subpopulations whose sizes were known in advance, or so-called "countable populations", for which it is possible to draw on databases that serve as the benchmark for calibration of the estimates, as well as on subpopulations with unknown sizes, to generate estimates of their sizes, which is the primary objective of the model's application. Participants from the general population are asked, "How many people do you know from subpopulation *k*?" The current study's question was, "How many women do you know who have undergone induced abortion in the last 12 months?".

In the household survey, based on the NSUM, formulated and described in detail by Bastos & Bertoni [17], it was determined that the resident 18 years or older and selected to answer the questionnaire would be the individual with his or her birthday closest to the date of the first visit. It was also determined that "knowing someone" from the target subpopulation would be operationalized as follows: persons who met the definition were: (i) persons who lived in the same city as the interviewee; (ii) whom the interviewee knew by sight and by name; (iii) who also knew the interviewee by sight and by name; and (iv) whom the interviewee had contacted either personally or by telephone, correspondence, or e-mail in the previous 12 months.

Since the NSUM is based on the social network of each participant from the general population, it uses indirect information from the participant's contact network concerning the subpopulations with known sizes to estimate the degree of the individual's contact network, also known as the "size of the individual's contact network", besides serving as the benchmark for the initial estimates, which can and should be compared to the available data in the corresponding databases (see Salganik et al. [16], especially the topic on reference populations analyzed in detail in the appendix). The social network's degree is the individual's total number of contacts, according to the definition of "knowing someone" adopted in the study.

All stages in this study were approved by the Ethics Research Committee of the Sergio Arouca National School of Public Health (ENSP), Fiocruz, protocol number CAAE 84961418.4.0000.5240.

### Modeling

The hierarchical model, formulated from a Bayesian perspective, was applied in this article with the purpose of estimating the number of women who had undergone abortions in the 12 months prior to the interview, based on indirect information on the contact networks of randomly selected individuals in the household survey, conducted from April to November 2012 in the city of Rio de Janeiro. However, the model can be applied to estimation of the size of any hard-to-reach populations whose research methodology employs NSUM.

Although applications of NSUM have varied since its initial use to estimate the number of victims in an earthquake in Mexico [19], this methodology has been widely employed to address public health issues such as the estimation of HIV prevalence in the homeless population [11], prevalence of heroin use in 14 communities [12], and the size of the illicit drug-using subpopulation (other than marijuana) in Curitiba, Paraná, Brazil [16], among others.

The Bayesian hierarchical model applied in this article incorporated the network scale-up methodology, allowing the inclusion of variability in the estimation of the contact network's degree. However, it did not consider the household survey's complex sampling design in the modeling, an issue still unresolved in the international literature, since the attribution of weights and calibration were not incorporated in the prevailing models due to the complexity of such issues as potential overlapping of networks and the possibility that clustering modulates the cross-section of the networks reported by the interviewees.

## Hierarchical model

Let $Y_i$ be the number of women who had undergone an abortion in the previous 12 months living in the city of Rio de Janeiro known by the participant *i*. We assume that $Y_i$ follows a binomial distribution with parameters $\delta_i$ and $\theta$, which represent the degree of the contact network of individual i and the proportion of the target subpopulation in the general population in the city of Rio de Janeiro, estimated with indirect information on the contacts of the individuals selected in the survey, respectively. That is,

$$Y_i \mid \delta_i, \theta \sim \text{Binomial}(\delta_i, \theta) \qquad (1)$$

for $I = 1, 2, \ldots, n =$ 1,547 selected individuals.

Let the variable $X_{ik}$ be the number of contacts known by the participant i, residing in the city of Rio de Janeiro, from subpopulation k. We also assume that $X_{ik}$ follows a binomial distribution with parameters $\delta_i$ and $\varpi_k$, for $I = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, K = 20$. That is,

$$X_{ik} \mid \delta_i, \varpi_k \sim \text{Binomial}(\delta_i, \varpi_k) \qquad (2)$$

where $\varpi_k$ is the prevalence of subpopulation k in the city of Rio de Janeiro and is previously known.

The study considered *K* subpopulations of previously known sizes, such as bus drivers, bank employees, foreigners, etc. The information on subpopulations with known sizes was used to estimate the degree of the survey participants' networks, since the degree of the contact network was considered unknown. Considering that the size of a person's contact network is generally "large" and often hard to estimate, the respondent tends not to actually count each of his or her contacts (according to the definition of "knowing"), but to guess or make a rough estimate (guesstimate) of this number, which could generate a bias in the NSUM estimate.

Thus, the likelihood, $L(\theta, \delta_1, \ldots, \delta_n; y_i, x_{ik})$, can be written as

$$L(\theta, \delta_1, \ldots, \delta_n; y_i, x_{ik}) = p(y_i \mid \delta_1, \theta) \times p(x_{ik} \mid \delta_i, \varpi_k) \qquad (3)$$

where $p(y_i \mid \delta_1, \theta)$ is the probability density given in equation (1) and is the $p(x_{ik} \mid \delta_i, \varpi_k)$ probability density given in equation (2).

One can thus express the number of women who had undergone an abortion in the city of Rio de Janeiro in the 12 months prior to the interview ($N\Upsilon$) as

$$N_y = N \times \theta \qquad (4)$$

where *N* is the size of the population in the city of Rio de Janeiro, available in the 2010 *Brazilian Demography Census* [20], and $\theta$ is the proportion of the target subpopulation in the general population of the city of Rio de Janeiro in the year 2011, estimated with indirect information on the contacts of the individuals selected in the survey.

Based on $N_Y$, given in equation (4), one can calculate the cumulative incidence of women in 15 to 49 years if age who had undergone an abortion in the city of Rio de Janeiro in the 12 months prior to the interview.

$$\theta_Y = \frac{N_Y}{POP_{M15\text{-}49}} \times 1.000 \qquad (5)$$

where $POP_{M15\text{-}49}$ is the size of the population of women 15 to 49 years of age living in the city of Rio de Janeiro, available in the 2010 *Brazilian Demography Census* [20], that is, $POP_{M15\text{-}49}$ = 1,758,145.

The following independent prior distribution where assumed:

$$\theta \sim \text{Beta}\,(a_\theta, b_\theta) \qquad (6)$$

$$\delta_i \sim Gamma\,(\alpha_\delta, b_\delta), \text{ para } i = 1, 2, \ldots, n \qquad (7)$$

where the quantities $a_\theta$ and $b_\theta$ were chosen to be equals to 1 representing a uniform prior for the proportion of induced abortions, and each individual's degree, although possessing a discrete nature, a gamma distribution was assumed with parameters $a_\delta$ and $b_\delta$ chosen according to the Hill & Dunbar's [21] number, since it is believed that there is a limited number of contacts/friends that one can establish in a relatively stable social relationship, which in turn, according to the authors, is based on human capacity for retention and nomination. Therefore, based on the Hill & Dunbar [21] number, the quantities $a_\delta$ and $b_\delta$ were defined such that the most likely value, that is the mode, would be 100 contacts, besides defining a relatively low probability (5%) that an individual would have more than 300 contacts.

The inference process for the Bayesian hierarchical model is based on the joint posterior distribution of all the parameters, which can be obtained via Bayes theorem, that is,

$$p(\theta,\delta_i,\ldots,\delta_n|y,x) \propto p(\theta)\prod_{i=1}^{n} p(\delta_i)\prod_{i=1}^{n}\prod_{k=1}^{k} L(\theta,\delta_1,\ldots,\delta_n;y_i,x_{ik}) \qquad (8)$$

where the prior distributions for θ e $\delta_i$ are respectively given in equations (6) and (7), and the likelihood function is presented in equation (3) which under conditional independence was given by the product of the joint likelihood function.

Samples from the posterior distribution of the Bayesian hierarchical model's parameters were obtained via MCMC (Markov chain Monte Carlo) [22] sampler. The routines were implemented in the R software, version 3.3.1 (http://www.r-project.org), and convergence was visually verified by analysis of the trace plot of the resulting sampling chains.

## The model's self-validation

The network scale-up method allows the model's self-validation in the "leave-one-out format", where a known population is removed from the model and treated as unknown, and a uniform prior is used for the removed subpopulation's prevalence, and we use the NSUM to estimate the size of this removed subpopulation. This process is repeated until all the subpopulations are estimated as if they were unknown. In the self-validation process, the population of women who had undergone abortion in the 12 months prior to the interview was not used.

The objectives of self-validation of the hierarchical model applied to the survey data were: (i) to verify whether the model is able to reasonably predict the size of any unknown population; (ii) to assess whether the proposed model is able to appropriately represent uncertainty; and (iii) to verify whether there is a pattern of underestimation or overestimation of the target estimates for the hard-to-reach populations based on the Bayesian hierarchical modeling in NSUM.

More detail on the Bayesian hierarchical model that uses indirect information based on the social networks of selected participants from the general population, the inference process, and the model's self-validation applied in this article can be seen in Paiva [23].

## Results

### The model's self-validation

Table 1 shows the comparison of the true prevalence values, the estimates of the mean posterior prevalence, and their respective 95%CrI for the 20 subpopulations of known size in the city of Rio de Janeiro, Brazil.

Based on Table 1, the estimated prevalence rates of the 20 subpopulations were relatively close to the true prevalence rates. The mean relative error was 0.05, and the median was -0.10. The mean exceeded the median because of the subpopulations that presented discrepant relative differences, such as "widows/widowers" (relative difference = 2.24) and "persons 15 years or older and who cannot read or write", that is, "illiterates" (relative difference = 1.73).

However, the 95%CrI are narrow, that is, exhibiting a narrow range around the estimated prevalence. The only subpopulations in which 95%CrI contained the true prevalence rates were "men who married in civil ceremonies in the last 12 months", "women who married in civil ceremonies in the last 12 months", "women over 70 years of age", "private primary and secondary schoolteachers", and "public primary and secondary schoolteachers (Table 1).

As shown in Figure 1, there was an overestimation pattern in the subpopulations with low prevalence in the general population in the city of Rio de Janeiro, such as "women 20 years or older who gave birth in the last 12 months" (1.24%; 95%CrI: 1.17; 1.32), "bank employees" (0.91%; 95%CrI: 0.84; 0.97), and

**Table 1**

Self-validation of the Bayesian hierarchical model using indirect information on social networks in the city of Rio de Janeiro, Brazil.

| Subpopulation with known size | True prevalence (%) | Estimated prevalence * (%) | ICr95% | Erro relativo ** |
|---|---|---|---|---|
| Bank employees | 0.40 | 0.91 | 0.84; 0.97 | -0.56 |
| Women under 20 years of age who gave birth in the last 12 months | 0.99 | 1.73 | 1.64; 1.82 | -0.43 |
| Women 20 years and older who gave birth in the last 12 months | 0.16 | 1.24 | 1.17; 1.32 | -0.87 |
| Mothers who received the Brazilian Income Transfer Program | 2.75 | 1.96 | 1.87; 2.06 | 0.40 |
| Bus drivers | 0.52 | 1.17 | 1.10; 1.25 | -0.56 |
| Persons 15 years or older who cannot read or write ("illiterates") | 2.29 | 0.84 | 0.78; 0.90 | 1.73 |
| Women with four or more children (only living biological children) | 3.48 | 2.07 | 1.97; 2.17 | 0.68 |
| Foreigners | 1.10 | 1.20 | 1.13; 1.28 | -0.08 |
| Women who married in civil ceremonies in the last 12 months | 0.44 | 0.50 | 0.44; 0.55 | -0.12 |
| Men who married in civil ceremonies in the last 12 months | 0.44 | 0.45 | 0.41; 0.50 | -0.02 |
| Women over 70 years of age | 4.25 | 4.55 | 4.41; 4.70 | -0.07 |
| Men over 70 years of age | 2.31 | 3.47 | 3.35; 3.61 | -0.33 |
| 5th to 8th grade students in private schools | 3.06 | 2.51 | 2.40; 2.62 | 0.22 |
| Private secondary school students | 1.16 | 1.85 | 1.76; 1.95 | -0.37 |
| Public secondary school students | 3.45 | 4.38 | 4.24; 4.53 | -0.21 |
| Secondary and primary private schoolteachers | 1.58 | 1.53 | 1.44; 1.61 | 0.03 |
| Secondary and primary public schoolteachers | 2.35 | 2.34 | 2.23; 2.45 | 0.00 |
| Girls under 5 years of age | 2.85 | 4.33 | 4.19; 4.48 | -0.34 |
| Boys under 5 years of age | 2.92 | 4.08 | 3.95; 4.23 | -0.28 |
| Widows/widowers (men or women whose last spouse died and who did not remarry) | 5.93 | 1.83 | 1.74; 1.92 | 2.24 |

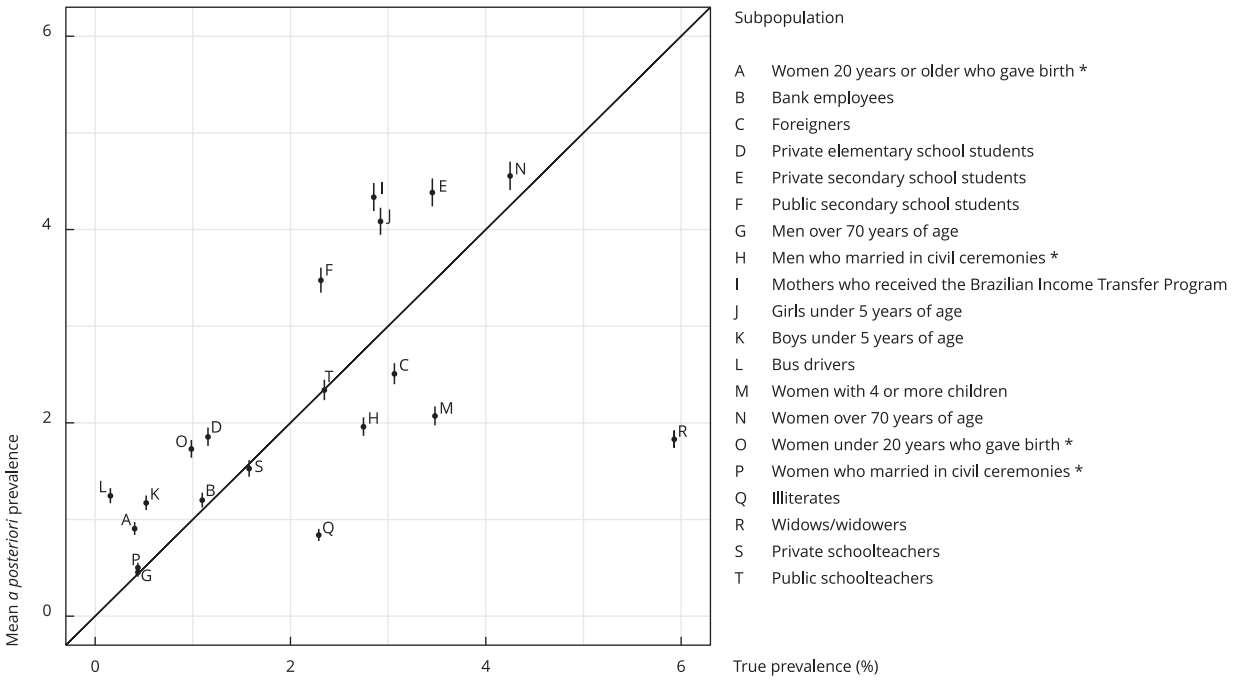
95%CrI: 95% credible interval a posteriori.

* Mean a posteriori prevalence;

** Relative error = (true prevalence – estimated prevalence)/estimated prevalence.

**Figure 1**

Self-validation of the Bayesian hierarchical model: true prevalence versus mean a posteriori prevalence (bar = 95%CrI). City of Rio de Janeiro, Brazil.




95%CrI: 95% credible interval.

"bus drivers" (1.17%; 95%CrI: 1.10; 1.25), whose prevalence rates in the population are 0.16%; 0.40%; and 0.52%, respectively.

Likewise, we found a pattern of underestimation in the subpopulations with high prevalence in the general population, such as "widows/widowers" (1.83%; 95%CrI: 1.74; 1.92), "persons 15 years or older who cannot read or write" (0.84%; 95%CrI: 0,78; 0.90), and "mothers who receive the Brazilian Income Transfer Program" (1.96%; 95%CrI: 1.87; 2.06), whose prevalence rates were 5.93%; 2.29%, and 2.75%, respectively.

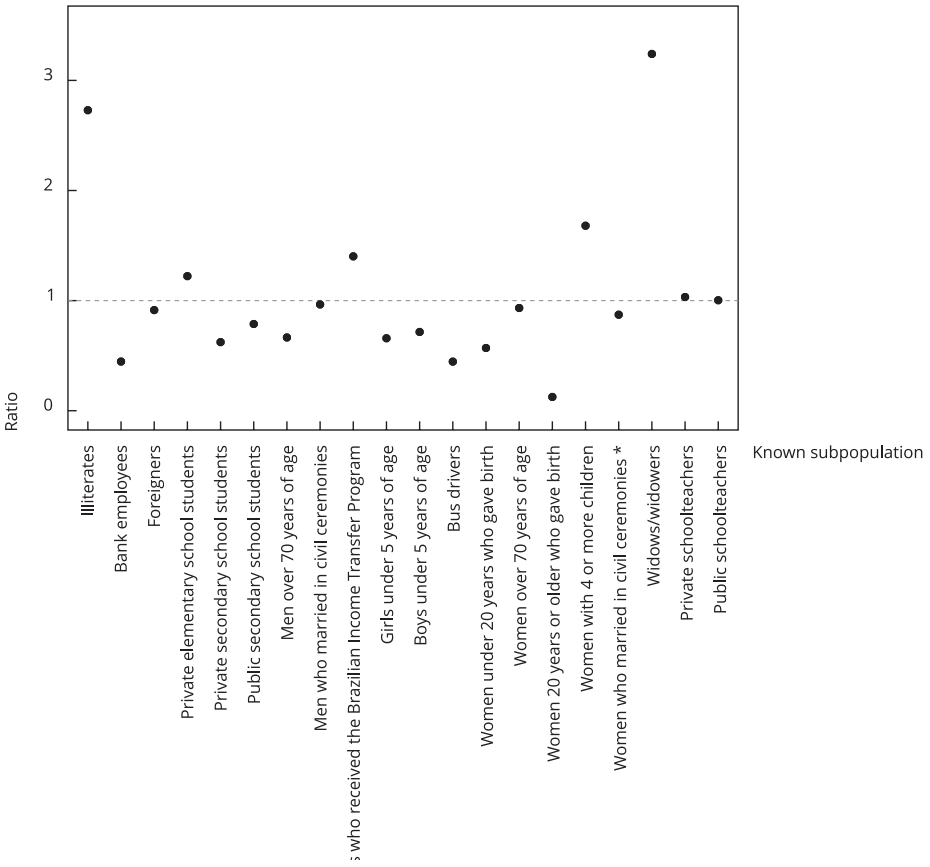
As shown in Figure 2, based on the ratio between the subpopulation's true prevalence (%) and estimated prevalence (%), via the posterior mean, we concluded that the posterior estimates for the subpopulations "illiterates" (odds = 2.73) and "widows/widowers" (odds = 3.24) were those that most overestimated the true prevalence.

As shown in Figure 2, in general the results of the hierarchical model based on the NSUM were not highly discrepant when comparing the true and predicted values, since the minimum ratio was 0.13 (overestimating the subpopulation of "women 20 years or older that gave birth in the last 12 months") and the maximum value was 3.24 (underestimating the subpopulation of "widows/widowers"). The ideal result, in Figure 2, is a ratio equal to 1, as seen in the subpopulation of "primary and secondary public schoolteachers".

Despite the intrinsic imprecision in these estimation processes via the hierarchical model based on the NSUM, considering the variability associated with different subpopulations, it was expected that the correlation between the estimates and the true values would not be perfect. Thus, the posterior estimates from the Bayesian hierarchical model proved reasonable when compared to the true prevalence rates of these subpopulations.

**Figure 2**

Self-validation of the Bayesian hierarchical model: ratio between the true prevalence and the mean *a posteriori* prevalence for the 20 known subpopulations. City of Rio de Janeiro, Brazil.




We now present the current study's target estimate, the mean number of women who had undergone induced abortion in the city of Rio de Janeiro, as well as the cumulative incidence, knowing that the 95%CrI may be underestimating the uncertainty.

## Women who underwent abortions in the city of Rio de Janeiro in 2011

We estimated a cumulative incidence of 7.41 abortions (95%CrI: 6.05; 8.96), for every 1,000 women 15 to 49 years of age in the city of Rio de Janeiro in the year 2011. This translates as a total of 13,025 women (95%CrI: 10,635; 15,748) who underwent induced abortions that year. That is, of the 1,758,145 [20] women 15 to 49 years of age living in the city of Rio de Janeiro, an estimated 13,025 women underwent induced abortions that year.

## Discussion

The act of inducing an abortion, according to Brazilian legislation, is a crime as specified in Decree-*Law n. 2.848* of the Penal Code [24] in force since 1940. However, it is known that induced abortion, often in unsafe conditions, still persists as a crosscutting phenomenon in society, differing according to race or color, religion, marital status, and socioeconomic status [6,25], not only in Brazil but throughout the world [4].

The current study's innovative approach represents a significant step in relation to previously published Brazilian and international studies on the estimation of abortion, for several reasons.

First, as far as we know, this is a pioneering study in the Americas with the use of the NSUM for estimating the number of women who have undergone abortion.

Second, the application of Bayesian hierarchical modeling in the context of NSUM allowed incorporating uncertainty and prior information in the process of estimating the number of women who have undergone abortion in the city of Rio de Janeiro.

Third, the use of indirect information on the participants' contact networks allowing reducing possible embarrassments related to stigmatizing issues such as induced abortion, since participants selected randomly from the general population were not reporting their own social behavior, but the social behavior of others.

Finally, any household survey can incorporate the network scale-up methodology using questions such as, "How many persons do you know who belong to subpopulation *k*?" This allows simultaneously estimating the size of various hard-to-reach populations. At least theoretically, it is possible to nest NSUM in any survey, respecting the limitation that the method was designed for populations whose frequency (or fraction of the general population) is relatively low.

However, with the application of an innovative method, comparisons of the current article with previous studies entail a relevant degree of uncertainty due to the differences in the studies' design, the estimation methods employed, and the definition of the objects of study, as for example induced or unsafe abortion in women who have undergone abortion at least once in life or in the last 12 months.

Previous studies that have estimated the number of women who have undergone abortion or the number of induced or unsafe abortions are approximately comparable to the current article, since studies that refer to Brazil or its major geographic regions, states, or municipalities generally have estimated the occurrence of abortions in a given period based on secondary data on hospitalizations for abortion complications [7,8,9] or the number of literate women 18-39 years of age living in "urban" Brazil who have had an abortion any time in life, through the ballot-box technique in the *Brazilian National Abortion Survey* (PNA) [6,10].

The range of induced abortions per 1,000 women 15-49 years in Southeast Brazil in 2011 was 12-15 induced abortions per 1,000 women [8], according to the estimation method proposed by the Guttmacher Institute [10], based on data on hospitalizations for abortion complications, applying correction factors. In the city of Rio de Janeiro, the current study found a cumulative incidence of women who had undergone abortion of 7.41 per 1,000 (95%CrI: 6.05; 8.96). Note however that the population of women 15-49 years of age living in the city of Rio de Janeiro, according to the 2010 *Brazilian Demography Census* [20], was 7.6% of the women in the same age bracket living in the Southeast region of Brazil, which totals 23,061,700 women 15-49 years of age.

The estimate of the mean annual number of unsafe abortions, based on records of hospital admissions and correction factors [26], reported by Martins-Melo et al. [9] for the state of Rio de Janeiro was 8,700. The current study estimated that 13,025 women (95%CrI: 10.635; 15.748) had undergone induced abortions in the city of Rio de Janeiro in 2011. The number of induced abortions may have been even greater than estimated, since a woman may have undergone more than one abortion in 12 months.

According to the PNA in 2010 [25], 15% (95% confidence interval – 95%CI: 13%; 17%) of literate women 18-39 years of age living in "urban" Brazil reported having had an abortion at least once in their lives. According to the PNA [6], 12% (95%CI: 10%; 14%) of literate women 18-39 years of age living in Southeast Brazil reported having had an abortion at least once in life, as of 2010.

Self-validation of the model described in this study identified a pattern of overestimation of the size of subpopulations with low prevalence rates in the general population and underestimation of the size of subpopulations with high prevalence rates in the general population, as in previous studies [13,16,27]. According to Killworth et al. [11], NSUM becomes imprecise for events with prevalence higher than 4%,

since it violates the basic assumption of non-juxtaposition of the interviewees' networks, as seen in the current study through the self-validation process when estimating the size of the subpopulation of "widow/widowers", whose prevalence in the general population is approximately 6%.

Although self-validation aimed at verifying the estimates' consistency in subpopulations of unknown sizes is somewhat encouraging, the process does not allow assessing the estimates' precision.

We believe that the Bayesian hierarchical model based on indirect information from contact networks, as applied in this study, underestimates the prevalence of stigmatized subpopulations (not necessarily criminalized, such as women who have undergone abortion) as well as of subpopulations of "persons 15 years or older who cannot read or write" and "mothers that received the Brazilian Income Transfer Program", since persons often do not disclose such behaviors (or characteristics), even to the closest members of their contact network.

According to Salganik et al. [28], if the target subpopulation has low social visibility, as is probably the case of women who have undergone abortion, the estimates based on indirect information from contact networks will be very low.

We thus believe that the estimate from the Bayesian hierarchical model based on NSUM underestimated the number of women who had undergone abortion in the city of Rio de Janeiro.

An alternative to deal with this limitation to the NSUM is the application of correction methods, for example an adaptation of the "game of contacts" used by Salganik et al. [28] to estimate the social visibility of groups of drug users. However, in the current study this was not feasible, since the game was not performed during the household survey that provided the data. Another alternative is to use independent surveys, which allow correcting the "transmission error", lack of information in the context of social networks involving a heavily stigmatized characteristic across different strata and networks in society, as in the study by Ezoe et al. [29] in men who have sex with men in Japan.

The self-validation process revealed that the 95%CrI presented deficient coverage. Although this result suggests that the proposed model underestimates the uncertainty, the result corroborates those reported by Salganik et al. [16]. According to Maltiel et al. [30], by using external information [30], the network scale-up method can improve the estimates and their associated uncertainty. We thus suggest that in the future, the sources of uncertainty should be captured in order to minimize the methodology's intrinsic biases.

One can thus state that induced abortion is a common practice among large Brazilian cities like Rio de Janeiro, even in an illegal and extremely stigmatizing context where women suffer moral and religious prejudices.

According to Ganatra et al. [4], approximately 45% (90%CrI: 40.6; 50.1) of abortions in the world in 2010-2014 were performed in unsafe conditions. The authors also identified a disparity in the safety of abortion between developed and developing regions during the same period. In South America, 75% (90%CrI: 46.3; 95.4) of abortions were performed unsafely, generating avoidable risks to the woman's health and life.

There is a need to know the real number of women who have undergone induced abortions in Brazil, because due to its illegality, there are no official data on the procedure, often performed clandestinely in inadequate locations and in subhuman conditions.

The existing methodologies for estimating unsafe (or induced) abortions in Brazil through secondary data from the Health Information Systems, with the use of correction factors, proposed in 1994 by the Guttmacher Institute [10], should be preserved and used systematically, mainly to build a historical series and define basic parameters for comparability.

However, it is necessary to update these correction factors, since nearly 25 years have gone by since they were formulated to correct the estimate in such a heterogeneous and continental-sized society as Brazil.

In short, induced abortion is still common among Brazilian women, despite its illegality, and it is probably performed unsafely in most cases. New estimation methods can help measure the event more precisely in this context of illegality and thus contribute to the formulation of appropriate health policies.

## Contributors

N. S. Paiva participated in the method's elaboration, data analysis and interpretation, writing of the article, and approval of the final version. D. A. M. Villela and L. S. Bastos participated in the method's elaboration, data analysis, and writing of the article. F. I. Bastos contributed substantially to the study's conception and design, contributed to the data acquisition, conducted the content's critical revision, and participated in elaborating the manuscript.

## Additional Informations

ORCID: Natália Santana Paiva (0000-0003-0541-4686); Daniel Antunes Maciel Villela (0000-0001-8371-2959); Leonardo Soares Bastos (0000-0002-1406-0122); Francisco I. Bastos (0000-0001-5970-8896).

## Acknowledgements

The authors wish to acknowledge the financial support received from the National Secretariat for Drug Policies for the national survey *Profile of Crack Users in 26 State Capitals, the Federal District, 9 Metropolitan Areas, and Brazil*, which produced the current study's data. We also wish to thank the Graduate Studies Coordinating Board (CAPES) for the PhD scholarship and the Graduate Studies Program in Epidemiology and Public Health of the Sergio Arouca National School of Public Health, Oswaldo Cruz Foundation.

## References

1. Bearak J, Popinchalk A, Alkema L, Sedgh G. Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a Bayesian hierarchical model. Lancet Glob Health 2018; 6:e380-e9.
2. World Health Organization. Aborto espontáneo y provocado: informe de un Grupo Científico de la OMS. Geneva: World Health Organization; 1970.
3. Ganatra B, Tunçalp Ö, Johnston HB, Johnson BR, Gülmezoglu AM, Temmerman M. From concept to measurement: operationalizing WHO's definition of unsafe abortion. Bull World Health Organ 2014;92:155.
4. Ganatra B, Gerdts C, Rossier C, Johnson BR, Tunçalp Ö, Assifi A, et al. Global, regional, and subregional classification of abortions by safety, 2010-2014: estimates from a Bayesian hierarchical model. Lancet 2017; 390:2372-81.
5. Center For Reproductive Rights. The World's abortion laws. http://worldabortionlaws.com/map/ (acessado em 24/Mai/2016).
6. Diniz D, Medeiros M, Madeiro A. Pesquisa Nacional de Aborto 2016. Ciênc Saúde Colet 2017; 22:653-60.
7. Monteiro MFG, Adesse L. Estimativas de aborto induzido no Brasil e Grandes Regiões (1992-2005). In: Anais do XV Encontro Nacional de Estudos Populacionais. Caxambu: Associação Brasileira de Estudos Populacionais; 2006. p. 1-10.
8. Monteiro MFG, Adesse L, Drezett J. Atualização das estimativas da magnitude do aborto induzido, taxas por mil mulheres e razões por 100 nascimentos vivos do aborto induzido por faixa etária e grandes regiões. Brasil, 1995 a 2013. Reprod Clim 2015; 30:11-8.
9. Martins-Melo FR, Lima MS, Alencar CH, Ramos AN, Carvalho FHC, Machado MMT, et al. Temporal trends and spatial distribution of unsafe abortion in Brazil, 1996-2012. Rev Saúde Publica 2014; 48:508-20.
10. Guttmacher Institute. Clandestine abortion: a Latin American reality. New York: Guttmacher Institute; 1994.
11. Killworth PD, McCarty C, Bernard HR, Shelley GA, Johnsen EC. Estimation of seroprevalence, rape, and homelessness in the United States using a social network approach. Eval Rev 1998; 22:289-308.
12. Kadushin C, Killworth PD, Bernard HR, Beveridge AA. Scale-up methods as applied to estimates of heroin use. J Drug Issues 2006; 36:417-40.
13. Zheng T, Salganik MJ, Gelman A. How many people do you know in prison? J Am Stat Assoc 2006; 101:409-23.
14. McCormick TH, Salganik MJ, Zheng T. How many people do you know? Efficiently estimating personal network size. J Am Stat Assoc 2010; 105:59-70.

15. Wang J, Yang Y, Zhao W, Su H, Zhao Y, Chen Y, et al. Application of network scale Up method in the estimation of population size for men who have sex with men in Shanghai, China. PLoS One 2015; 10:e0143118.
16. Salganik MJ, Fazito D, Bertoni N, Abdo AH, Mello MB, Bastos FI. Assessing network scale-up Estimates for groups most at risk of HIV/AIDS: evidence from a multi-method study of heavy drug users in Curitiba, Brazil. Am J Epidemiol 2011; 174:1190-6.
17. Bastos FI, Bertoni N. Pesquisa nacional sobre o uso de crack: quem são os usuários de crack e/ou similares do Brasil? Quantos são nas capitais brasileiras? Rio de Janeiro: Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz; 2014.
18. Bertoni N. Quantos usuários de crack e/ou similares existem nas capitais brasileiras? Resultados de um inquérito nacional com a utilização da metodologia Network Scale-Up [Doctoral Dissertation] Rio de Janeiro: Fundação Oswaldo Cruz; 2014.
19. Bernard HR, Johnsen EC, Killworth PD, Robinson S. Estimating the size of an average personal network and of an event subpopulation: some empirical results. Soc Sci Res 1991; 20:109-21.
20. Instituto Brasileiro de Geografia e Estatística. Censo demográfico 2010. Características da população e dos domicílios: resultados do universo. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2011.
21. Hill RA, Dunbar RIM. Social network size in humans. Human Nature: An Interdisciplinary Biosocial Perspective 2003; 14:53-72.
22. Gamerman D, Lopes HF. Markov Chain Monte Carlo-Stochastic Simulation for Bayesian Inference. Boca Raton: Chapman and Hall/CRC Press; 2006. (Texts in Statistical Science Series, 68).
23. Paiva NS. Quantas mulheres fizeram aborto provocado nas capitais brasileiras? Resultados de um inquérito nacional com a aplicação de modelos hierárquicos utilizando informações indiretas baseadas na rede de contatos [Doctoral Dissertation]. Rio de Janeiro: Fundação Oswaldo Cruz; 2018.
24. Brasil. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Código Penal. Diário Oficial da União 1940; 31 dez.
25. Diniz D, Medeiros M. Aborto no Brasil: uma pesquisa domiciliar com técnica de urna. Ciênc Saúde Colet 2010; 15 Suppl 1:959-66.
26. Guttmacher Institute. An overview of clandestine abortion in Latin America. New York: Gttmacher Institute; 1996.
27. Killworth PD, McCarty C, Bernard HR, Johnsen EC, Domini J, Shelley GA. Two interpretations of reports of knowledge of subpopulation sizes. Soc Networks 2003; 25:141-60.
28. Salganik MJ, Mello MB, Abdo AH, Bertoni N, Fazito D, Bastos FI. The game of contacts: estimating the social visibility of groups. Soc Networks 2011; 33:70-8.
29. Ezoe S, Morooka T, Noda T, Sabin ML, Koike S. Population size estimation of men who have sex with men through the network scale-up method in Japan. PLoS One 2012; 7:e31184.
30. Maltiel R, Raftery AE, McCormick TH, Baraff AJ. Estimating population size using the network scale up method. Ann Appl Stat 2015; 9:1247-77.

## Resumo

*Estimativas de mulheres que fizeram aborto provocado em localidades cujas leis são restritivas ainda são escassas na literatura científica, e a não coincidência de estimativas oriundas dos métodos hoje em uso clama pela aplicação de métodos inovadores, como novos métodos indiretos. Tal necessidade é especialmente aguda nas áreas mais densamente povoadas, como as capitais brasileiras, dada a magnitude do fenômeno e os danos e riscos daí decorrentes. O artigo objetiva estimar o número de mulheres que fez aborto provocado no Município do Rio de Janeiro, Brasil, em 2011, por meio de um modelo hierárquico bayesiano. Ele foi aplicado aos dados de um inquérito domiciliar que subsidiou a utilização do método* network scale-up*, no Município do Rio de Janeiro, um modelo hierárquico bayesiano utilizando as informações indiretas baseadas na rede de contatos dos participantes selecionados de forma aleatória da população. Das 1.758.145 mulheres de 15-49 anos residentes no Município do Rio de Janeiro (13.025; ICr95%: 10.635; 15.748) mulheres fizeram aborto provocado em 2011, resultando numa incidência acumulada média de 7,41 (ICr95%: 6,05; 8,96) para cada 1.000 mulheres de 15-49 anos. O estudo de autovalidação do modelo permitiu identificar padrões de subestimação em subpopulações estigmatizadas com baixa visibilidade social, como mulheres fizeram aborto provocado. O abortamento provocado é uma prática recorrente entre as mulheres no Município do Rio de Janeiro. Novos métodos de estimação indireta podem contribuir para a apreensão mais precisa do evento, considerando o contexto de ilegalidade, e contribuir para formulação de políticas de saúde.*



*Aborto Induzido; Estimativas de População; Análise Estatística de Dados; Análine Multinível; Técnicas de Estimação*



## Resumen

*Las estimaciones de mujeres que tuvieron un aborto provocado en localidades cuyas leyes son restrictivas todavía son escasas en la literatura científica, y la no coincidencia de las estimaciones procedentes de los métodos hoy en uso reclama urgentemente la aplicación de métodos innovadores, como los nuevos métodos indirectos. Tal necesidad es especialmente acuciante en las áreas más densamente pobladas, como las capitales brasileñas, dada la magnitud del fenómeno y los daños y riesgos derivados de allí. El artículo tiene como objetivo estimar el número de mujeres que realizaron un aborto provocado en el Municipio de Río de Janeiro, Brasil, en 2011, a partir de un modelo jerárquico bayesiano. Este se aplicó a los datos de una encuesta domiciliaria que fomentó la utilización del método* network scale-up*, en el Municipio de Río de Janeiro, un modelo jerárquico bayesiano utilizando información indirecta, basada en la red de contactos de los participantes seleccionados de forma aleatoria en la población. De las 1.758.145 mujeres de 15-49 años, residentes en el Municipio de Río de Janeiro, 13.025 (ICr95%: 10.635; 15.748) mujeres tuvieron un aborto provocado en 2011, resultando en una incidencia acumulada media de 7,41 (ICr95%: 6,05; 8,96) para cada 1.000 mujeres de 15-49 años. El estudio de autovalidación del modelo permitió identificar patrones de subestimación en subpoblaciones estigmatizadas con baja visibilidad social, como las mujeres que tuvieron un aborto provocado. El aborto provocado es una práctica recurrente entre mujeres en el municipio de Río de Janeiro. Nuevos métodos de estimación indirecta pueden contribuir a la aprehensión más precisa de este evento, considerando el contexto de ilegalidad, y contribuir a la formulación de políticas de salud.*



*Aborto Inducido; Estimativas de Población; Análisis Estadísticas de Datos; Análisis Multinível; Técnicas de Estimación*

Submitted on 01/Oct/2018
Final version resubmitted on 12/Feb/2019
Approved on 06/May/2019