International Journal of Drug Policy 74 (2019) 285–291

Contents lists available at ScienceDirect

International Journal of Drug Policy

journal homepage: www.elsevier.com/locate/drugpo

Research Paper

Reproducibility on science: Challenges and advances in Brazilian alcohol surveys

Raquel B. De Boni[a,*], Mauricio T.L. de Vasconcellos[b], Pedro N. Silva[b], Carolina Coutinho[c], Jurema Mota[c], Julia N.B. Peixoto[c], Neilane Bertoni[c,d], Francisco I. Bastos[c]

[a] *Instituto Nacional de Infectologia Evandro Chagas – FIOCRUZ, Rio de Janeiro, Brazil*
[b] *Escola Nacional de Ciências Estatísticas – IBGE, Rio de Janeiro, Brazil*
[c] *Instituto de Comunicação e Informação Cientifica em Saúde-FIOCRUZ, Rio de Janeiro, Brazil*
[d] *Instituto Nacional do Câncer, Rio de Janeiro, Brazil*

ARTICLE INFO

Keywords:
Alcohol use
Alcohol dependence
Health surveys
Reproducibility of findings
Brazil

ABSTRACT

Background: Reproducibility in Science is challenging and may be hard to achieve in alcohol research. Previous general population surveys in Brazil have estimated the prevalence of alcohol dependence to be around 10%. We aim to estimate alcohol use and dependence using different methods and definitions.

Methods: The 3rd Brazilian Household Survey on Substance Use (BHSU-3) was a nationwide, probability sample survey that interviewed 16,273 individuals. DSM-IV-TR criteria were used to determine alcohol dependence. In the BHSU-2 (covering only Brazil's 108 largest municipalities), alcohol dependence was defined as fulfilling 2/6 DSM-III criteria. Using the BHSU-3 data, alcohol use was estimated at: [1] the national level, [2] BHSU-2 municipalities, taking into consideration the sample design, and [3] BHSU-2 municipalities, ignoring the sample design. Alcohol dependence was calculated using: BHSU-3 and BHSU-2 definitions, two denominators ([A] population and [B] 12-month drinkers), and [1], [2], [3].

Results: Lifetime alcohol use ranged from 66.4% (95%CI:64.8–68.0 [1]) to 70.1% ([95%CI:69.1–71.0], [3]). The estimated population presenting with alcohol dependence ranged from N = 2.3 million (BHSU-3 definition, [1]) to N = 4.3 million (BHSU-2 definition, [1]). In the first case, the prevalence among the general population [A] and drinkers [B] was 1.5% (95%CI:1.2–1.8) and 3.5% (95%CI:2.8–4.2), respectively. In second case, prevalence was 2.8% (95%CI:2.4–3.3, [A]) and 6.6% (95%CI:5.6–7.6, [B]).

Conclusions: Prevalence of alcohol dependence may vary as much as 4.3 times, analyzing the same dataset in different ways. Brazilian research on alcohol is funded by governmental research and policy-making agencies, providing subsidies for alcohol policy in the country. It is crucial that sufficient methodological information is provided in order to guarantee reproducibility and consistency over time.

Introduction

Harmful alcohol consumption is a major public health problem worldwide (Rehm et al., 2009; WHO, 2018). In 2016, to "strengthen prevention and treatment of […] harmful alcohol use" was defined as one of the Sustainable Developmental Goals (United Nations, 2016). That was an important step in fostering better policies to minimize associated harms and risks. Reliable and reproducible data is fundamental to evaluate current policies success and to identify areas of improvement. However, reproducibility in Science may be challenging, as increasingly highlighted over the last decade (Oakden-rayner, Beam, & Palmer, 2018), specially in areas presenting multiple definitions and outcomes (Ioannidis, 2005).

Alcohol research is a broad, multidisciplinary area, where it may be difficult to obtain optimal reproducibility. Comparability of alcohol data can be problematic because of different classifications of what is considered harmful drinking (such as the International Classification of Diseases (ICD- OMS, 2007) and the Diagnostic Statistical Manual for Mental Disorders (DSM-APA, 2013)), as well as a myriad of ways to measure and classify use and patterns of use (Babor, Higgins-Biddle, Saunders, & Monteiro, 2001; NIAAA, 2004; World Health Organization, 2016). In addition, clinical data may be biased because of the stigma associated with problem drinking and a lack of personnel trained to adequately diagnose alcohol use disorders (Mitchell, Meader, Bird, &

* Corresponding author: INI Evandro Chagas –FIOCRUZ, Av. Brasil 4365. STD/AIDS Laboratory, suite 32, Rio de Janeiro, Zip Code 21040-360, Brazil.
E-mail address: raquel.boni@ini.fiocruz.br (R.B. De Boni).

https://doi.org/10.1016/j.drugpo.2019.07.029

Rizzo, 2012). Some countries have succeeded in overcoming most of the barriers that can compromise the reproducibility of data at the national level by using standardized and well documented methods over the years (CDC, 2016; SAMHSA, 2014).

In Brazil, the first national estimates of alcohol dependence are as recent as the 2000's. The first Brazilian Household Survey on Substance Use (BHSU-1) was conducted in 2001. It included individuals aged 12 to 65 years from 107 Brazilian municipalities with populations of more than 200,000, reporting a prevalence of 11.2% (95%CI:9.1–13.3) for "concern about alcohol consumption" (defined as the presence of 2 out of 6 DSM-III-R) (Galduróz & Carlini, 2007). The BHSU-2 (2005) used the same criteria but estimated a prevalence of "alcohol dependence" at 12.3% (95%CI:9.1–15.6) (Fonseca, Galduroz, Noto, & Carlini, 2010). These surveys were conducted only in the largest cities and thus remained open to criticism regarding under-coverage and bias. Bias could be introduced if the prevalence of alcohol use varies between large cities and elsewhere, and may be key in a country as large and heterogeneous as Brazil, with over 5500 municipalities of different sizes, located in both urban and rural areas.

The first nationwide probability sample survey (1 st Brazilian nationwide survey on Alcohol Consumption Patterns - BNADS-2005-6) found a 9% prevalence of alcohol dependence among adults (n = 2346) (Laranjeira, Pinsky, Sanches, Zaleski, & Caetano, 2010) – using 3 positive criteria (out of 7) from the Composite International Diagnostic Interview (CIDI- Cottler, Robins, & Helzer, 1989). Finally, in the II BNADS (2012), the prevalence of DSM-V alcohol use disorders (AUD) were estimated at 11% (95%CI:9.6–12.8) (Abdalla et al., 2018) using the CIDI and an additional question on craving (2 out 11 criteria indicate an AUD).

These studies were funded with public resources and used to inform public policies. However, after 10 or more years the survey datasets are not yet publicly available (government official sites may collate reports, but none includes microdata, as has been offered by the National Institute of Geography and Statistics [see, for instance: https://ww2.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2015/microdados.shtm, as well as by the authors themselves in their collaborative work with Princeton University [https://opr.princeton.edu/archive/NSUM/]). In addition, methodological details on sample design, weighting and calibration, as well as derived variables' dictionary and definitions, data management (including how missing data were handled, for example) and scripts for data analysis have never made available – nor audited or monitored by competent governmental agencies, as far as known by the authors.

The aforementioned issues have two major consequences. First, the results have placed Brazilian estimates of alcohol dependence and AUD at among the highest in the world; the World Health Organization (WHO) global estimates are 2.3% and 4.1%, respectively. Such figures have a huge impact on the design and evaluation of public policies for prevention and treatment. Second, it is almost impossible for an independent research group to reproduce the methods in order to either confirm or refute the estimates – and, even the same group of investigators would not be able to provide reliable trends.

Thus, beyond the public health implications, it is timely and essential to discuss using standard definitions to estimate trends and to decrease bias in Brazilian alcohol research. In this paper, using data collected in the 3rd Brazilian Household Survey on Substance Use (BHSU-3), we estimated alcohol use and dependence by applying the methods and definitions used in BHSU-2 and in BHSU-3. This allowed us to assess the sensitivity of such estimates to the different criteria and methods used.

Methods

The BHSU-3 was a nationwide, multi-stage, probability sample survey conducted between May and December 2015. The survey was commissioned by the National Secretariat on Drug Policies after an open public bid. One of the key specifications was that data should be compared with the BHSU-2, as well as with international surveys, to provide UNODC with a national (in the strictist sense of the word) survey, with full statistical power to discern both major and less popular substance consumption patterns (e.g. heroin injection is a major concern worldwide but a rare event in Brazil). Our proposal was selected, and then approved by the Escola Politécnica de Saúde Joaquim Venâncio – FIOCRUZ Ethics Review Board (CAAE # 35283814.4.0000.5241). Participating individuals who were 18 years old or older signed an informed consent and for those under 18 years, the informed consent was signed by a parent or guardian while the individual signed an assent form.

Survey population

The survey population comprised individuals aged 12–65 years, living in private households in both urban and rural areas, across Brazil. Native individuals living in indigenous villages, foreigners living in Brazil, non-Portuguese speakers, those who were imprisoned or institutionalized, and those presenting physical or mental disabilities that precluded answering the interviews were not eligible..

Sample design

BHSU-3 used a stratified four-stage clustered probability sample. Municipalities were sampled in the first stage. Census enumeration areas (CEA) were sampled in the second stage. Households were sampled within each sampled CEA. Finally, one eligible (12–65-year-old) resident was sampled at random within each selected and participating household.

To facilitate comparison with BHSU-2, each state capital and large city (≥ 200,000 inhabitants in 2010) was included in the sample with certainty, becoming a selection stratum (therefore, in these municipalities the design was three-stage cluster sampling). All other municipalities were stratified into the five Brazilian macro-regions (North, Northeast, Southeast, South and Center-west). Within each macro-region, municipalities were stratified in three strata: (1) border municipalities (with part of area within 200 km from terrestrial border of the country); (2) municipalities within metropolitan regions; and (3) other municipalities. In these three strata, municipalities were further stratified by population size - small (≤ 11,000 inhabitants), medium (11,000 to 200,000 inhabitants) and large (≥ 200,000 inhabitants). Thus, the sample had 138 strata.

The municipalities not included with certainty in the sample were selected in each stratum with probability proportional to size (PPS), based on their population in 2010. In every selected municipality, CEAs were first sorted by their average household income, and then sampled with systematic PPS (size was the number of private households). The households were sampled using equal probability inverse sampling (Haldane, 1945; Vasconcellos, Silva, & Szwarcwald, 2005). This method samples households sequentially for contact and adopted as a stopping rule within each CEA either having reached 10 complete interviews or having reached 50 contacted households, irrespective of the number of complete interviews. In the last stage, one eligible resident was selected with equal probability among the eligible residents in each participating household.

Sample size

The total sample size was calculated to estimate a minimum proportion (prevalence) of 2% with a relative error of 30%, confidence level of 95% and design effect (Kish, 1965) of 1.5. Power allocation (with power = 3/4) was used to distribute the total sample size among the strata, using population as the size measure. After the allocation, the sample size reached 16,400 residents (or households) spread in 1640 CEAs and 351 municipalities.

Sample implementation, weighting and calibration

When data collection was concluded, a total of 16,273 eligible residents were interviewed, corresponding to an effective sample size that reached 99.2% of the required sample size. Applying the sequential inverse sampling procedure within the selected CEAs, a total of 27,906 households were selected and screened to identify eligible residents. Of these, 4036 could not be contacted after exhausting the attempts of the contact protocol (out of which, 24 addresses were not found). A total of 3180 were empty or used only as a non-permanent residence (e.g. rental flats, beach houses, etc.), another 1052 did not have eligible residents, and five had residents who had contagious diseases during the interview period. Of the remaining 19,633 contacted eligible households, 3057 refused to participate, 271 selected individuals refused to participate, and 32 interviews were not completed, giving a non-response rate of 17.1% [100*(3057 + 271 + 32) / 19,633)].

Four core strategies were used to deal with non-response. First, the survey used inverse sampling to screen for eligible households, implying that any contact attempts would be recorded and that the final sample size in each selected CEA would match the required sample size per CEA (10 complete interviews). Second, the contact protocol ensured that households would only be declared ‘lost for interview’ after a substantial contact effort had been made (i.e., at least four visits on different days of the week and times of day). Third, interviewer selection, training and supervision were designed to ensure that interviewer contacts with selected households were effective and would lead to small refusal rates. Finally, calibration was applied to the basic sampling design weights to compensate for observed differential non-response.

In order to represent the survey population and recover the estimated population size (N), weights reflecting both the probabilistic sampling design and calibration to known population counts were calculated (Kish, 1965; Lumley, 2010). Basic sampling weights were calculated as reciprocals of each person’s sample inclusion probability. Calibration considered population totals by sex, age groups, macro-region and household size, obtained from the Brazilian Continuous National Household Sample Survey for the third quarter of 2015 (IBGE, 2015). In this sense, BHSU-3 is fully comparable, from the point of view of its sampling plan, to other official datasets generated by IBGE, which have been regularly downloaded and analyzed by both national and international experts (Cowell, Ferreira, & Litchfield, 1998; Madalozzo, 2012).

Variables

Demographic variables were collected using questions from the Brazilian Demographic Census 2010 (IBGE, 2010), and included: Sex at birth (male vs. female), Age (categorized as 12–17 years, 18–24, 25–34, 35–44, 45–54, 55–65), Color/Race (white, black, mixed, other), Schooling (Up to incomplete fundamental, Fundamental/incomplete high school, High school/Incomplete college education, Graduation or more) and Partnership (Steady = Yes vs. No). Brazil’s geographic macro-regions (North, Northeast, Southeast, South, Center-west) and Urban vs. Rural region were ascertained at the sampling definition.

Alcohol Use was self-reported as follows: lifetime use, prior 12-months, prior 30 days, alongside additional assessment of binge drinking (i.e. 4 (women) or 5 (men) alcoholic beverages in a couple of hours (NIAAA, 2004)). Alcohol dependence was estimated using the BHSU-3 definition (DSM-IV-TR criteria, i.e., individuals should answer positively to 3/7 criteria to be considered as having alcohol dependence) and using the same definition used in the BHSU-2 (fulfilling 2/6 DSM-III criteria).

Statistical analysis

A dataset in the form of an R survey package object, for the present and future analysis, was created including all the relevant information about the sampling weights and sample design structure (such as strata and primary sampling unit anonymized identifiers).

The prevalence of alcohol use (lifetime, 12-months, 30-days) was estimated at: [1] the national level and [2] the 108 municipalities previously included in BHSU-2, accounting for the complex sample design; and [3] the 108 BHSU-2 municipalities, considering a self-weighted sample – as it was performed in the BHSU-2.

The prevalence of alcohol dependence was estimated using BHSU-3 definition and BHSU-2 definition, considering two denominators ([A] the entire survey population and [B] 12-month drinkers), and [1], [2], [3].

The above prevalences were estimated with 95% confidence intervals (CI) and extrapolations were based on the total survey population.

All the analyses were performed in R v 3.5.1 (R Core Team, 2013), using the packages survey (Lumley, 2018), srvyr (Ellis & Lumley, 2018) and tidyverse (Wickham & RStudio, 2017). R scripts are available in the Supplementary material.

Results

The municipalities included in BHSU-3 (n = 351 [1]), as well as the 108 municipalities from BHSU-2 sample frame ([2] and [3]), are depicted in Fig. 1.

After sample weighting, the estimated population size and relative frequencies for the main demographic characteristics are presented in Table 1 - BHSU-3 [1]. Similar estimates were generated for domain containing the 108 municipalities from the BHSU-2 (Table 1, [2] and [3]). Table 1 shows that selecting only the 108 largest municipalities in the country, as was done in BHSU-2 ([2] and [3], determines an increase in the proportions of individuals from the Southeast, urban areas, and with higher education levels, irrespectively of sampling weights.

The prevalence of lifetime, 12-month and 30-day alcohol use, as well as binge drinking (Table 2), were also higher when considering only the 108 largest municipalities ([2] and [3]). Comparing estimates for the entire country [1] and the 108 largest municipalities (accounting for the complex sample design [2]), the difference found in lifetime alcohol use was not statistically significant–as can be inferred by the overlapping 95% confidence intervals (95% CI). However, the prevalence of 12-month alcohol use was 43.1% (95% CI 41.8–44.4) for the entire country [1], whereas it was 47.4% (95% CI 45.8–49.0) for the largest municipalities [2]. Although this prevalence was lower, the total estimated population using alcohol in previous 12-month was 66 million in the entire country [1], which was more than twice that estimated for the 108 largest municipalities [2]. Similar effects appear when comparing 30-day alcohol use and binge drinking (Table 2).

When comparing the entire country [1] with the 108 largest municipalities, ignoring the complex sample design in the latter [3], the only statically significant difference was found in lifetime alcohol use (66.4%; IC 95%:64.8–68.0 in the entire country and 70.1%; 95%CI:69.1–71.0 in the largest municipalities) – Table 2.

The prevalence of alcohol dependence was similar across the country [1] and in the 108 largest municipalities ([2] and [3]). However, the prevalence was 1.5% (95%CI:1.2-1.8) with BHSU-3 definition and 2.8% (95%CI:2.4-3.3) with BHSU-2 definition (using the entire population as denominator [A]). Hence, the number of alcohol dependent individuals was estimated at 2.3 million and 4.3 million, respectively (Table 2).

Discussion

Our results show that the prevalence of alcohol use (lifetime, 12-month and 30-days) was lower for the entire country [1] when compared to the 108 largest municipalities ([2] and [3]), likely due to a lower prevalence of use for individuals from rural areas and from

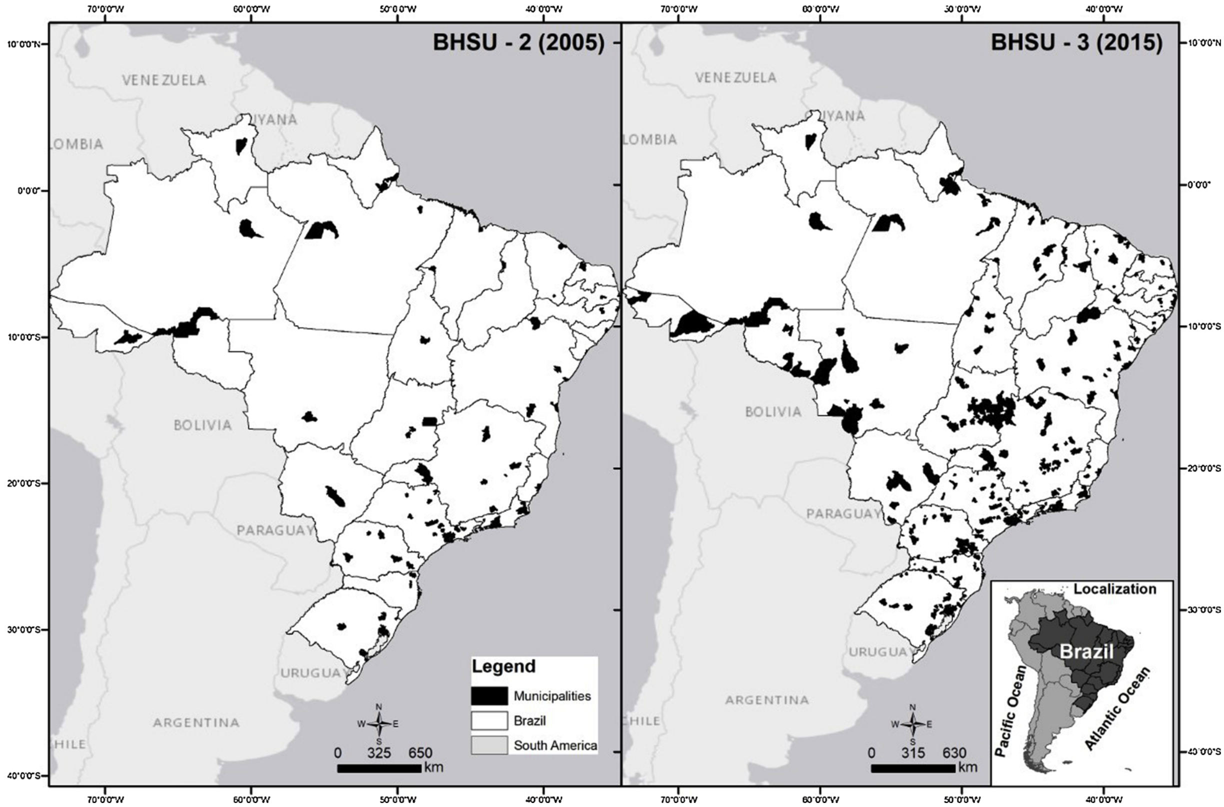

Fig. 1. Distribuition of the Brazilian municipalities included in the 2nd Brazilian Household Survey on Substance Use (BHSU-2, n = 108, 2005) and the 3rd Brazilian Household Survey on Substance Use-3 (BHSU-3, n = 351, 2015).

middle-sized and smaller municipalities. The same has been observed in several other countries (Amin-Esmaeili et al., 2017; Grant et al., 2017; Islam et al., 2017; Katulanda et al., 2014; Taype-Rondan et al., 2017; Williams, Mcfarland, & Nelson, 2012; Zhou et al., 2006) and in the 2013 Brazilian National Health Survey, where people living in rural areas (compared to urban) were significantly less likely to have drunk alcohol in the previous 30 days (Macinko, Mullachery, Silver, Jimenez, & Libanio Morais Neto, 2015). One must note here that Brazil uses a definition of “Rural” which has been criticized by its own experts (IBGE, 2017), but has remained the same over the years due to political and legal norms. The official definition of “rural” does not allow for the cross-comparison of Brazil’s data on rural areas with data from member states of the OCDE (Organization for Economic Co-operation and Development; as discussed in detail in IBGE, 2017), as well as from other countries (e.g. Argentina).

We did not find such differences when evaluating the prevalence of alcohol dependence. We did find, however, an up to 4.3 times higher prevalence using the previous BHSU-2 definition and alcohol users [B] as denominators compared to BHSU-3 and the entire population [A] as denominator- which reinforces the need for clear definitions when presenting the data. In addition, the alcohol dependence estimate in the BHSU-3 was much lower (1.5%) than found in previous Brazilian surveys (Abdalla et al., 2014; Fonseca et al., 2010; Galduróz & Carlini, 2007), but closer to the WHO global average (WHO, 2018), indicating the need to be extremely careful before interpreting those results together, i.e., it is incorrect to evaluate trends using these different studies. The above difference has multiple, public health and policy, implications. As mentioned before, all the surveys (BHSU, BHSU-2 and BHSU-3) were funded by the Brazilian government and are expected to inform initiatives addressing the needs of the public health system (SUS), as well as other policies regarding alcohol, as foreseen in the National Policy on Alcohol (Presidencia da República, 2007). In addition, these data are intended to help monitor the efficacy of such initiatives and to be shared with international organizations, such as WHO and UNODC, to monitor alcohol and substance use related harm globally.

The use of standardized methods and reliable statistics are key for domestic policies, international organizations and for the estimates of burden of disease (GBD group, 2018). A former accusation about incomplete/inchoate data on Brazil’s health system has ignited a harsh debate (Almeida et al., 2001), which fortunatelly has been nowadays rather a warningon the need for methodological informationto be publicly available and open to criticism and continuous improvement. The discrepancies found, however, are not restricted to the Brazilian context. In 2017/2018, two papers evaluating US trends on AUD (i.e., the combination of DSM-IV alcohol abuse and dependence) presented contradictory results. Whereas Grant et al. (2017) found an increase in AUD – comparing the 2002 and 2014 National Epidemiologic Surveys on Alcohol and Related Conditions (NESARC), Cheng, Kaakarli, Breslau, and Anthony (2018)) found the prevalence to be stable (in the 13 editions of the National Surveys on Drug Use and Health - NSDUH). The later authors highlighted that even minor methodological differences in psychiatric surveys may bring different estimates influencing trends and called for the collaboration of epidemiologists and survey methodologists to solve such issues.

Major differences found in the US and Brazilian contexts are related to transparency, documentation and dissemination of survey’s results, which are present in both NESARC and NSDUH. To properly evaluate and interpret the results, it is crucial to provide comprehensive methodological information, and clearly point/test the necessary changes made in definitions and concepts. Of course, an important reason for the lack of information found in Brazil is the lack of financial resources – *a sine qua non* condition to produce the highly specialized and time demanding information. However, as in other countries, both Brazilian scientists and contracting governmental agents and funders must do better to achieve reproducible Science (Ioannidis, 2018).

Table 1
Estimates on demographic characteristics at the 3rd Brazilian Household Survey on Substance Use (BHSU-3) using different methods, Brazil, 2015.

	[1]* BHSU-3			[2]* BHSU-2 municipalities			[3]* BHSU-2 municipalities		
	N(x1000)	%	(95%CI)	N(x1000)	%	(95%CI)	N(x1000)	%	(95%CI)
Region									
North	12,612	8.2	na	4,862	7.8	(7.0-8.5)	5,027	8.0	(7.4-8.6)
Northeast	41,736	27.3	na	12,464	19.9	(17.8-22.1)	13,343	21.3	(20.5-22.2)
Southeast	64,968	42.4	na	32,729	53.0	(50.8-55.2)	30,845	50.0	(49.0-51.1)
South	22,160	14.5	na	7,217	11.5	(10.6-12.5)	7,944	12.7	(12.0-13.4)
Center-West	11,619	7.6	na	4,789	7.6	(6.6-8.7)	4,903	7.9	(7.3-8.4)
Urban-Rural									
Urban	126,692	82.8	(80.6-84.9)	60,517	97.5	(96.5-98.6)	60,262	97.1	(96.7-97.4)
Rural	26,404	17.2	(15.1-19.4)	1,545	2.5	(1.4-3.5)	1,800	2.9	(2.6-3.3)
Sex at birth									
Male	74,179	48.5	na	30,730	49.4	(47.7-51.2)	23,646	38.0	(37.0-39.0)
Female	78,916	51.5	na	31,333	50.6	(48.8-52.3)	38,416	62.0	(61.0-63.0)
Age bracket									
12-17 years	20,276	13.2	na	6,346	10.2	(8.4-12.1)	2,048	3.3	(2.9-3.6)
18-24 years	22,327	14.6	na	9,114	14.7	(13.6-15.8)	7,882	12.7	(12.0-13.4)
25-34 years	31,646	20.7	na	13,720	22.1	(21.0-23.2)	13,343	21.4	(20.6-22.3)
35-44 years	30,400	19.9	na	12,516	20.2	(19.1-21.2)	13,095	21.1	(20.3-22.0)
45-54 years	26,465	17.3	na	11,062	17.8	(16.8-18.8)	12,226	19.7	(18.9-20.6)
55-65 years	21,980	14.4	na	9,303	14.9	(13.9-16.0)	13,467	21.7	(20.8-22.6)
Color/race									
White	67,778	44.3	(42.6-46.0)	28,539	45.8	(44.2-47.5)	28,797	46.3	(45.3-47.4)
Black	15,497	10.1	(9.3-11.0)	6,937	11.2	(10.2-12.1)	6,703	10.8	(10.1-11.5)
Mixed	68,083	44.5	(42.7-46.2)	25,834	41.7	(40.0-43.3)	25,818	41.7	(40.6-42.7)
Other	1,737	1.1	(0.9-1.4)	752	1.2	(0.9-1.5)	745	1.1	(0.9-1.4)
Schooling									
Up to incomplete fundamental	55,390	36.2	(34.6-37.5)	15,311	24.6	(23.0-26.3)	16,260	26.1	(25.2-27.0)
Fundamental/incomplete high school	34,744	22.7	(21.7-23.7)	13,928	22.5	(21.2-23.9)	12,847	20.8	(20.0-21.6)
High school/Incomplete college education	47,581	31.1	(30.0-32.1)	23,255	37.5	(35.9-39.0)	23,584	38.0	(37.0-39.0)
College or higher	15,380	10.0	(9.2-10.9)	9,568	15.3	(13.8-16.9)	9,371	15.0	(14.3-15.8)
Steady partner									
Yes	93,660	61.2	(60.0-62.4)	37,162	59.9	(58.3-61.5)	38,416	61.9	(60.9-62.9)
No	59,436	38.8	(37.6-40.0)	24,900	40.1	(38.4-41.7)	23,646	38.1	(37.1-39.1)

* BHSU-3 data estimates at: [1] the national level and [2] for BHSU-2 municipalities only, considering the complex sample design; and [3] for BHSU-2 municipalities, considering a self-weighted sample. Na = those variables were used for calibration.

Discussion on how to increase reproducibility and transparency in health Sciences, in ways that protect the privacy of research subjects and intellectual propriety, is not near to have an end.. Different initiatives are being implemented across the globe, including the European projects to foster open Science (https://www.fosteropenscience.eu/) and the increasing debate and availability of pre-print platforms (Kleinert & Horton, 2018; Maslove, 2018; Peiperl, 2018). In Brazil, many official statistics have microdata and methodological reports publicly disseminated, as can be seen in the IBGE website (http://doi.org/: http://www.sidra.ibge.gov.br). Academic research data, however, are not subject to the same laws and practices. Therefore, many details about their production may be lost because of the publication pressure in relevant, higly cited Scientific journals (which usually have limited space). It is noteworthy that many journals already ask for datasets, international funders may ask for data management plans, and review boards and reviewers may ask for registered protocols before project results can be evaluated (the later is already mandatory for international clinical trials, for instance, and there are national (www.rebec.org) and international (www.clinicaltrials.org) online repositories that may be consulted at any time).

Some of the measures adopted to guarantee reproducibility of BHSU-3 include the extensive documentation of the survey methods, including questionnaires, SOPs and coding. In addition, microdata was prepared for dissemination and further analysis by creating a dataset where all the relevant information about the sampling weights and sample design structure (such as strata and primary sampling unit

Table 2
Estimates on alcohol use in the 3rd Brazilian Household survey on Substance Use (BHSU-3) using different methods, Brazil, 2015.

	[1]* BHSU-3		[2]* BHSU-2 municipalities		[3]* BHSU-2 municipalities	
	N (x1000)	%(95%CI)	N (x1000)	%(95%CI)	N (x1000)	%(95%CI)
Lifetime	101,615	66.4(64.8-68.0)	43,463	69.6(67.9-71.2)	43,506	70.1(69.1-71.0)
12 months	65,943	43.1(41.8-44.4)	29,585	47.3(45.7-49.0)	28,238	45.5(44.4-46.5)
30 days	46,036	30.1(28.9-31.3)	21,041	33.7(32.1-35.2)	19,798	31.9(31.0-32.9)
Binge drinking	25,310	16.5(15.5-17.5)	11,942	19.1(17.9-20.3)	10,985	17.8(17.0-18.6)
Dependence DSM-IV criteria- General population[A]	2,328	1.5(1.2-1.8)	865	1.4(1.1-1.7)	807	1.3(1.1-1.6)
Dependence DSM-IV criteria -Alcohol users [B]	2,328	3.5(2.8-4.2)	865	2.9(2.3-3.6)	807	2.9(2.4-3.4)
Dependence BHSU-2 criteria- General population [A]	4,342	2.8(2.4-3.3)	1,780	2.8(2.4-3.3)	1,613	2.6(2.3-3.0)
Dependence BHSU-2 criteria- Alcohol users[B]	4,342	6.6(5.6-7.6)	1,780	6.0(5.0-7.0)	1,613	5.8(5.1-6.5)

* BHSU-3 data, alcohol use (lifetime, 12-months, 30-days) was estimated at: [1] the national level and [2] for BHSU-2 municipalities only, considering the complex sample design; and [3] for BHSU-2 municipalities, considering a self-weighted sample. Alcohol dependence was calculated using: BHSU-3 and BHSU-2 definitions, two denominators ([A] population and [B] 12-month drinkers).

anonymized identifiers) are included with the survey responses. This dataset was prepared in the form of an R survey package object that can be easily analyzed, and for which the analyst needs only to specify the required analysis, not having to worry about all the details of the complex calculations involved, provided he/she uses tools available in the R *survey* package. Both R and its *survey* package are freely and openly available worldwide, without the need for software licensing fees.

This study is not free of limitations, and some of these may have led to underestimation of alcohol dependence. The first is the face-to-face interview, which may increase social desirability bias. However, regarding alcohol, our previous studies have found it is a highly tolerated and socially accepted substance in Brazil (De Boni, Pechansky, Silva, de Vasconcellos, & Bastos, 2013; De Boni, Bertoni, Bastos, & Bastos, 2014; Pechansky et al., 2009). In addition, our interviewers were experienced in conducting face-to-face household surveys and were trained and supervised to deal with sensitive questions. The second limitation refers to the decreased likelihood of finding severe substance users at their homes, as they may have become hard-to-reach populations, spending most of their time on the streets, where they make their living and interact with their core networks (Linton et al., 2017; Sudman, Sirken, & Cowan, 1988; Sypsa et al., 2015). This constitutes a limitation of all household-based surveys. To reach such individuals, additional methods are required, but have not been used in the present study.

The above limitations apply equally to the three survey methods we compared. Nevertheless, our study has made clear the following: a) national estimates for alcohol usage and dependence cannot be derived safely from samples which cover only the largest municipalities; b) unbiased and 'representative' survey estimates for alcohol usage and dependence must take into account the underlying survey design and estimation methods; c) reproducible surveys require substantial effort to document and disseminate the methods used, and the results achieved, as well as access to both the survey microdata and to the core survey data processing scripts, written in an open source and freely available software, as we have done for the BHSU-3.

Contributors

RBDB and FBI conceived the present manuscript. RBDB wrote the first draft; MTLV and PNS conceived the sample and supervised the present data analysis; RBDB, FBI, MVTL, CC, JM and NB worked in the original study conception and coordination; JNBP geocoded data and produced the maps. All authors read, provided important intellectual content and approved the final manuscript.

Funding

Brazilian Household Survey on Substance Use–III was funded by the Brazilian National Secretariat for Drug Policies (SENAD). Its contents are solely the responsibility of the authors and do not necessarily represent the official position of the funding agency.

Francisco I. Bastos acknowledges funding by CNPq (Universal 2012 No 056626560009452) and FAPERJ (2015/2014, E-26/010.001755)

Declaration of Competing Interest

Authors' have no conflict to declare.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.drugpo.2019.07.029.

References

Abdalla, R. R., Madruga, C. S., Ribeiro, M., Pinsky, I., Caetano, R., & Laranjeira, R. (2014). Prevalence of cocaine use in Brazil: Data from the II brazilian national alcohol and drugs survey (BNADS). *Addictive Behaviors, 39*(1), 297–301. https://doi.org/10.1016/j.addbeh.2013.10.019.

Abdalla, R. R., Massaro, L., de Queiroz Constantino Miguel, A., Laranjeira, R., Caetano, R., & Madruga, C. S. (2018). Association between drug use and urban violence: Data from the II Brazilian National Alcohol and Drugs Survey (BNADS). *Addictive Behaviors Reports, 7*, 8–13. https://doi.org/10.1016/j.abrep.2017.11.003.

Almeida, C., Braveman, P., Gold, M. R., Szwarcwald, C. L., Ribeiro, J. M., Miglionico, A., ... Viacava, F. (2001). Methodological concerns and recommendations on policy consequences of the World Health Report 2000. *Lancet, 357*(9269), 1692–1697.

Amin-Esmaeili, M., Rahimi-Movaghar, A., Sharifi, V., Hajebi, A., Mojtabai, R., Radgoodarzi, R., ... Motevalian, A. (2017). Alcohol use disorders in Iran: Prevalence, symptoms, correlates, and comorbidity. *Drug and Alcohol Dependence, 176*, 48–54. https://doi.org/10.1016/j.drugalcdep.2017.02.018.

APA (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*.

Babor, T., Higgins-Biddle, J., Saunders, J., & Monteiro, M. (2001). *AUDIT-the alcohol use disorders identification test. Guidelines for use in primary care.* Geneva: WHO.

CDC (2016). *Behavioral risk factors survaillance survey.* Retrieved fromhttps://www.cdc.gov/brfss/data_documentation/statistic_brief.htm.

Cheng, H. G., Kaakarli, H., Breslau, J., & Anthony, J. C. (2018). Assessing changes in alcohol use and alcohol use disorder prevalence in the United States: Evidence from national surveys from 2002 through 2014. *JAMA Psychiatry, 75*(2), 211–213. https://doi.org/10.1001/jamapsychiatry.2017.4008.

Cottler, L. B., Robins, L. N., & Helzer, J. E. (1989). The reliability of the CIDI-SAM: A comprehensive substance abuse interview. *British Journal of Addiction, 84*(7), 801–814.

Cowell, F. A., Ferreira, F. H. G., & Litchfield, J. A. (1998). Income Distribution in Brazil 1981-1990. *Journal of Income Distribuition, 8*, 63–76.

De Boni, R., Pechansky, F., Silva, P. L. D. N., de Vasconcellos, M. T. L., & Bastos, F. I. (2013). Is the prevalence of driving after drinking higher in entertainment areas? *Alcohol and Alcoholism, 48*(3), 86–91. https://doi.org/10.1093/alcalc/ags126.

De Boni, R. B., Bertoni, N., Bastos, L. S., & Bastos, F. I. (2014). Unrecorded alcohol in Rio De Janeiro: Assessing its misusers through respondent driven sampling. *Drug and Alcohol Dependence, 139*, 169–173. https://doi.org/10.1016/j.drugalcdep.2014.03.015.

Ellis, G. F., & Lumley, T. (2018). *Srvyr-Like Syntax for summary statistics of survey data*.

Fonseca, A. M., Galduroz, J. C., Noto, A. R., & Carlini, E. L. (2010). Comparison between two household surveys on psychotropic drug use in Brazil: 2001 and 2004. *Cienc. Saúde coletiva, 15*(3), 663–670. https://doi.org/10.1590/S1413-81232010000300008.

Galduróz, J. C. F., & Carlini, E.a. (2007). Use of alcohol among the inhabitants of the 107 largest cities in Brazil–2001. *Brazilian Journal of Medical and Biological Research, 40*(3), 367–375.

GBD (2018). 2016 Alcohol and Drug Use Collaborators. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016 (2018). *The Lancet Psychiatry, 5*(December (12)), 987–1012.

Grant, B. F., Chou, S. P., Saha, T. D., Pickering, R. P., Kerridge, B. T., Ruan, W. J., ... Hasin, D. S. (2017). Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001-2002 to 2012-2013: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *JAMA Psychiatry, 74*(9), 911–923. https://doi.org/10.1001/jamapsychiatry.2017.2161.

Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika, 33*, 222–225.

IBGE (2010). *Censo Demográfico 2010.* http://doi.org/:http://www.sidra.ibge.gov.br/bda/orcfam/default.asp.

IBGE (2015). *Indicadores IBGE: Pesquisa Nacional por Amostra de Domicilios Continua. 3 trimestre.* Rio de Janeiro. Retrieved fromhttps://biblioteca.ibge.gov.br/visualizacao/periodicos/2421/pnact_2015_3tri.pdf.

IBGE (2017). *Classificação e caracterização dos espaços rurais e urbanos do Brasil : uma primeira aproximação.* Rio de Janeiro. Retrieved fromhttps://biblioteca.ibge.gov.br/visualizacao/livros/liv100643.pdf.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), 0696–0701. https://doi.org/10.1371/journal.pmed.0020124.

Ioannidis, J. P. A. (2018). All science should inform policy and regulation. *PLoS Medicine, 15*(5), e1002576. https://doi.org/10.1371/journal.pmed.1002576.

Islam, J. Y., Zaman, M. M., Bhuiyan, M. R., Hasan, M. M., Ahsan, H. N., Rahman, M. M., ... Jalil Chowdhury, M. A. (2017). Alcohol consumption among adults in Bangladesh: Results from STEPS 2010. *WHO South-East Asia Journal of Public Health, 6*(1), 67–74. https://doi.org/10.4103/2224-3151.206168.

Katulanda, P., Ranasinghe, C., Rathnapala, A., Karunaratne, N., Sheriff, R., & Matthews, D. (2014). Prevalence, patterns and correlates of alcohol consumption and its' association with tobacco smoking among Sri Lankan adults: A cross-sectional study. *BMC Public Health, 14*(1), 612. https://doi.org/10.1186/1471-2458-14-612.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Kleinert, S., & Horton, R. (2018). Preprints with The Lancet : joining online research discussion platforms. *Lancet, 391*(10139), 2482–2483. https://doi.org/10.1016/S0140-6736(18)31125-5.

Laranjeira, R., Pinsky, I., Sanches, M., Zaleski, M., & Caetano, R. (2010). Alcohol use patterns among Brazilian adults. *Revista Brasileira de Psiquiatria, 32*(3), 214–215.

Linton, S. L., Cooper, H. L., Kelley, M. E., Karnes, C. C., Ross, Z., Wolfe, M. E., ... National HIV Behavioral Surveillance Study Group (2017). Cross-sectional association between ZIP code-level gentrification and homelessness among a large community-based sample of people who inject drugs in 19 US cities. *BMJ Open, 7*(6), e013823. https://doi.org/10.1136/bmjopen-2016-013823.

Lumley, T. (2010). *Complex surveys: A guide to analysis using R.* New York: John Wiley & sons.

Lumley, T. (2018). *Package 'survey' -Analysis of complex survey samples.* Retrieved fromhttp://r-survey.r-forge.r-project.org/survey/.

Macinko, J., Mullachery, P., Silver, D., Jimenez, G., & Libanio Morais Neto, O. (2015). Patterns of alcohol consumption and related behaviors in Brazil: Evidence from the 2013 national health survey (PNS 2013). *PloS One, 10*(7), e0134153. https://doi.org/10.1371/journal.pone.0134153.

Madalozzo, R. (2012). Transitions in fertility for brazilian women: An analysis of impact factors. *PloS One, 7*(7), https://doi.org/10.1371/journal.pone.0040756.

Maslove, D. (2018). Medical preprints- a debate woth having. *JAMA : The Journal of the American Medical Association, 319*(5), 443–444. https://doi.org/10.1126/science.

Mitchell, A. J., Meader, N., Bird, V., & Rizzo, M. (2012). Clinical recognition and recording of alcohol disorders by clinicians in primary and secondary care: Meta-analysis. *The British Journal of Psychiatry : The Journal of Mental Science, 201*, 93–100. https://doi.org/10.1192/bjp.bp.110.091199.

NIAAA (2004). *NIAAA council approves definition of binge drinking. NIAAA Newsletter.* Bethesda: NIH.

Oakden-rayner, L., Beam, A. L., & Palmer, L. J. (2018). Medical journals should embrace preprints to address the reproducibility crisis. *International Journal of Epidemiology,* 1–3. https://doi.org/10.1093/ije/dyy105.

OMS (2007). *CID-10: Classificaçãoo Estatística Internacional de Doenças e Problemas Relacionados a Saúde, Vol. I.* São Paulo: Editora da Universidade de São Paulo.

Pechansky, F., De Boni, R., Diemen, L., Von Bumaguin, D., Pinsky, I., Zaleski, M., ... Laranjeira, R. (2009). Highly reported prevalence of drinking and driving in Brazil : Data from the first representative household study. *Revista Brasileira de Psiquiatria, 31*(2), 125–130.

Peiperl, L. (2018). Preprints in medical research: Progress and principles. *PLoS Medicine, 15*(4), 3–5. https://doi.org/10.1371/journal.pmed.1002563.

Presidencia da República (2007). *Politica Nacional sobre o Álcool. Decreto número 6117.* May 22nd. Retrieved fromhttp://www.planalto.gov.br/ccivil_03/_ato2007-2010/2007/decreto/d6117.htm.

R Core Team (2013). *R: A language and environment for statistical computing.* Retrieved fromVienna: R Foundation for Statistical Computing. http://www.r-project.org.

Rehm, J., Mathers, C., Popova, S., Thavorncharoensap, M., Teerawattananon, Y., & Patra, J. (2009). Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet, 373*(9682), 2223–2233. https://doi.org/10.1016/S0140-6736(09)60746-7.

SAMHSA (2014). *Results from the 2013 national survey on drug use and health: Summary of national findings. NSDUH series H-48, HHS publication No. (SMA) 14-4863.* Retrieved fromRockville, MD: Substance Abuse and Mental Health Services Administration. http://oas.samhsa.gov/NSDUH/2k10NSDUH/2k10Results.pdf.

Sudman, S., Sirken, M. G., & Cowan, C. D. (1988). Rare and elusive populations. *Science, 240*, 991–996.

Sypsa, V., Paraskevis, D., Malliori, M., Nikolopoulos, G. K., Panopoulos, A., Kantzanou, M., ... Hatzakis, A. (2015). Homelessness and Other Risk Factors for HIV Infection in the Current Outbreak Among Injection Drug Users in Athens, Greece. *American Journal of Public Health, 105*(1), 196–204. https://doi.org/10.2105/AJPH.2013.301656.

Taype-Rondan, A., Bernabe-Ortiz, A., Alvarado, G. F., Gilman, R. H., Smeeth, L., & Miranda, J. J. (2017). Smoking and heavy drinking patterns in rural, urban and rural-to-urban migrants: The PERU MIGRANT Study. *BMC Public Health, 17*(1), 165. https://doi.org/10.1186/s12889-017-4080-7.

United Nations (2016). *Sustainable development goals-Goal 3: Ensure healthy lives and promote well-being for all at all ages.* Retrieved fromhttps://www.un.org/sustainabledevelopment/health/.

Vasconcellos, M. T., Silva, P. L., & Szwarcwald, C. L. (2005). Sampling design for the world health survey in Brazil. *Cad Saude Publica, 21*(Suppl), 89–99 http://doi.org/S0102-311X2005000700010 [pii] /S0102-311X2005000700010.

WHO (2018). *Global status report on alcohol and health*Geneva. Retrieved fromhttp://www.who.int/substance_abuse/publications/global_alcohol_report/en/.

Wickham, H., & RStudio (2017). *Tidyverse: Easily install and load the "Tidyverse.".* Retrieved fromhttps://cran.r-project.org/package=tidyverse.

Williams, E. C., Mcfarland, L. V., & Nelson, K. M. (2012). Alcohol consumption among urban, suburban, and rural veterans affairs outpatients. *The Journal of Rural Health, 28*(2), 202–210. https://doi.org/10.1111/j.1748-0361.2011.00389.x.

World Health Organization (2016). *WHO - ASSIST V3.0*.

Zhou, X., Su, Z., Deng, H., Xiang, X., Chen, H., & Hao, W. (2006). A comparative survey on alcohol and tobacco use in urban and rural populations in the Huaihua District of Hunan Province, China. *Alcohol, 39*(2), 87–96. https://doi.org/10.1016/j.alcohol.2006.07.003.