CSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH

ConVid – Behavior Survey by the Internet during the COVID-19 pandemic in Brazil: conception and application methodology

ConVid – Pesquisa de Comportamentos pela Internet durante a pandemia de COVID-19 no Brasil: concepção e metodologia de aplicação

ConVid – Encuesta de Comportamientos por Internet durante la pandemia de COVID-19 en Brasil: concepción y metodología de aplicación

Celia Landmann Szwarcwald [1]
Paulo Roberto Borges de Souza Júnior [1]
Giseli Nogueira Damacena [1]
Deborah Carvalho Malta [2]
Marilisa Berti de Azevedo Barros [3]
Dalia Elena Romero [1]
Wanessa da Silva de Almeida [1]
Luiz Otávio Azevedo [1]
Ísis Eloah Machado [4]
Margareth Guimarães Lima [3]
André Oliveira Werneck [5]
Danilo Rodrigues Pereira da Silva [6]
Crizian Saar Gomes [7]
Arthur Pate de Souza Ferreira [1]
Renata Gracie [1]
Maria de Fátima de Pina [1]

Abstract

The ConVid – Behavior Survey was conducted in Brazil from April 24 to May 24, 2020, aiming to investigate changes in lifestyles and health conditions during the COVID-19 pandemic. In this article, we present the conception and methodology of the research. We used a cross-sectional study using an Internet questionnaire, with questions validated in previous health surveys. The sampling method "virtual snowball" was used, as well as post-stratification procedures. The results related to chronic non-communicable diseases and pre-pandemic lifestyles were compared with estimates from the 2013 Brazilian National Health Survey and 2019 Surveillance of Risk and Protective Factors for Chronic Diseases by Telephone Survey. The total sample was 45,161 people. After data weighing, the sample distributions of demographic variables were similar to population variables. Only people with a low schooling level were underrepresented. The comparison with the previous results showed similarity in most estimates: recommended consumption of fruits and vegetables (22.1%), recommended physical activity (35.2%), tobacco smoking habit (12.3%), frequent and abusive alcohol consumption (6.7%), obesity (21.2%), self-reported prevalence of hypertension (18.6%), diabetes (7.1%), and heart disease (4.4%). The online survey made it possible to know the population's health conditions during the pandemic. The similarity of the indicators with those obtained in traditional research allowed the validation of the mean estimates. Studies are needed to investigate how the endogenous effects of virtual social networks can be considered when estimating variance.

COVID-19; Internet; Health Surveys

Correspondence

C. L. Szwarcwald
Departamento de Informações para a Saúde, Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz.
Av. Brasil 4365, Biblioteca de Manguinhos, sala 225, Rio de Janeiro, RJ 21040-360, Brasil.
celia.szwarcwald@icict.fiocruz.br

[1] Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.
[2] Escola de Enfermagem, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
[3] Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brasil.
[4] Escola de Medicina, Universidade Federal de Ouro Preto, Ouro Preto, Brasil.
[5] Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, Brasil.
[6] Centro de Ciências Biológicas e da Saúde, Universidade Federal de Sergipe, São Cristóvão, Brasil.
[7] Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

Introduction

The great expansion of the Internet since the 1990s has made distance communication much more comprehensive and accessible, bringing people closer to different groups and cultures. In Brazil, the Internet emerged in the late 1980s as a means of disseminating information in the area of education and research, but quickly reached the general population. In 2018, it was estimated that 79.1% of permanent private households had Internet access [1].

Over the past 20 years, information and communication technologies have grown intensely, promoting profound changes in the way people communicate, capture and share information, and issue opinions [2,3,4]. There is a great proliferation of websites, chats, and social networks, making the world a huge network of connected people, regardless of their geographical locations [5].

Considering that traditional approaches to collect information from population-based surveys are not always feasible due to budgetary constraints, time limits for obtaining information, or absence of a list or reference system, research in a virtual environment is a current trend for data collection [6], and it has encouraged researchers to use online questionnaires as an alternative method for obtaining answers in scientific research, including health surveys [7,8]. The information can be collected by a single platform for data collection [9] (such as Facebook or an institutional site) or using a chain sampling procedure, such as the methodology known as "snowball" adapted to virtual social networks [10].

Based on the theory of the six degree of separation [11], chain sampling techniques have often been used to facilitate access to populations considered difficult to access because they adopt socially stigmatized behaviors or engage in illegal activities [12]. The "snowball" method is a non-probabilistic sampling procedure, functioning based on the indication of an initial group of people who are part of the target population (called seeds), which indicate pairs of the same population group, and so on, similar to the formation of a snowball [13].

Similarly, the method of data collection by "virtual snowball" is initiated by sending/presenting the access link to an electronic questionnaire, by e-mail or some virtual social network. In the message body, in addition to the presentation of the survey, there is a request for sharing the invitation to participate with the person's network of contacts. This non-probabilistic technique of data collection allows the definition of a sample with references made by people, which indicate others that meet the inclusion criteria in the research, and so on [10].

With the onset of the COVID-19 pandemic, the adoption of measures to restrict physical contact in several countries of the world limited the performance of health research with face-to-face interviews [14]. On the other hand, the need to acquire greater knowledge about the disease and other health problems directly or indirectly related to the pandemic, stimulated the use of the Internet as a means to obtain health information in an agile manner [15,16]. In addition to addressing adherence to measures to protect and to restrict physical contacts [17], the studies addressed other issues related to the pandemic period, such as symptoms [18], psychological disorders [19], difficulties in accessing health services [20], and the adoption of unhealthy behaviors [21,22].

In Brazil, from March 2020, strict measures to restrict contact between people were imposed by states and municipalities. The distancing of family and friends, uncertainty about the disease, substantial changes in the socioeconomic context, and lack of control over one's own life have caused damage to physical and mental health [23,24].

The *ConVid – Behavior Survey* was performed in a virtual environment in order to investigate the changes in lifestyles and health conditions of the Brazilian population during the COVID-19 pandemic. In this work, we present the conception, objectives, and methodology, and we analyze the representation of the population in the sample reached.

Methodology

Conception and design of the study

ConVid – Behavior Survey (https://convid.fiocruz.br/) was conducted nationwide by the Oswaldo Cruz Foundation (Fiocruz), in partnership with the Minas Gerais State Federal University and the Campinas State University. This is a cross-sectional study using a virtual questionnaire, self-completed by mobile phone or computer with Internet access. Data collection occurred from April 24 to May 24, 2020.

The research had the following objectives: to describe the intensity of adherence of the Brazilian population to social restriction measures; to investigate changes in the work and income situation; to evaluate the difficulties in performing routine activities; to analyze health conditions, and to describe the behavioral changes during the COVID-19 pandemic. Questions validated in previously applied health surveys in Brazil were used. The questions about the state of mind were adapted from the *World Health Survey* [25], and those questions regarding the diagnosis of chronic non-communicable diseases and lifestyles (smoking habits, food, alcohol consumption, physical activity, and sedentary lifestyle) were based on the questions of the 2013 *Brazilian National Health Survey* (PNS 2013, in Portuguese) [26] and monitored by 2019 *Surveillance of Risk and Protective Factors for Chronic Diseases by Telephone Survey* (Vigitel 2019, in Portuguese) [27]. Regarding sleep disorders, a single question of sleep self-assessment was used, adapted from the set of questions on sleep of the *Population-Based Health Survey in Campinas, 2014-2015* (ISA-CAMP, in Portuguese) [28]. To evaluate the difficulties of older adults (people aged 60 years or older) in performing activities of daily living, only one question was used, synthesis of a set of questions of the PNS 2013 [29]. Regarding the work situation and loss of income, the questions were based on those used in the research *Quality of Life and Adherence to HAART in HIV-infected Patients* conducted in 2008 [30].

For most topics addressed, questions were asked about the situation before and after the pandemic in Brazil. Changes in the work situation, difficulties in performing routine activities, health conditions, and lifestyles were evaluated. The full questionnaire is available on the ConVid website (https://convid.fiocruz.br/).

For the preparation of the questionnaire, the RedCap (Research Electronic Data Capture. https://www.project-redcap.org/) application was used. The information was collected directly with the Internet and it was stored on the server of the Institute of Scientific and Technological Communication and Information (ICICT/Fiocruz). The inclusion criteria for participation in the research were: aged 18 years or older and living in the Brazilian territory during the COVID-19 pandemic. All responses were anonymous and without any type of identification of the participants. Further details of *ConVid – Behavior Survey* are on the search site (https://convid.fiocruz.br/).

The project was approved by the Brazilian National Research Ethics Committee (CONEP) on April 19, 2020 (Opinion n. 3,980,277).

Sampling

The "virtual snowball" was the sampling method, initiated from sending invitations with the access link to the electronic questionnaire through the virtual social network WhatsApp or by email. The survey was called "ConVid" [a pun with the word *"convide"*, "to invite" in Brazilian Portuguese] to suggest that the participant should invite other people from their social networks: *"If you found it important, ConVide more people to participate!"*. At the end of the questionnaire, there was also a request for sharing the invitation to participate in the survey with the person's network of contacts: *"Join the ConVid Network and share this survey with three or more guests from your social network"*.

In the first stage of the participants' recruitment, 15 researchers who composed the study team chose approximately 200 other researchers from different states of Brazil in order to obtain national coverage. Furthermore, each researcher on the study team selected 20 people from their own social network, totaling about 500 people who were called seeds because they started the network of participants. After answering the questionnaire, the seeds composed the first wave of the recruitment chain. In order to obtain diversity of sociodemographic characteristics in the sample [31], the seeds sent

the research link to at least three people from their social networks in each of the 12 strata formed by sex, age group (18-39; 40-59; 60+ years), and schooling level (incomplete high school or less; complete high school or more). The people invited by the seeds composed the second wave of the recruitment chain. Each person in the second wave was asked to invite at least three other people from their social networks and so on. At the end of the period of information collection (from April 24 to May 24, 2020), the total sample size reached was 45,161 people.

Post-stratification procedure

Since the selection probabilities are unknown, network sampling is not probabilistic, and it is not possible to calculate the natural weights of the sampling design.

To obtain a representative sample of the population according to geographic location and sociodemographic characteristics, weightings calculated by post-stratification procedures [32] were performed by: Federation Unit (UF, in Portuguese), capital/other parts of the UF, gender, age group (18-29; 30-39; 40-49; 50-59; 60+ years), schooling level (incomplete higher education; complete higher education) and race/skin color based on population estimates from the 2019 *Brazilian National Household Sample Survey* (PNAD 2019, in Portuguese) [33].

Mathematically, whether *Nh* is the population estimate of the PNAD 2019 [33] in each *h* stratum, composed of UF, capital/other parts of the UF, sex, age group, schooling level, and race; and *N* the total number of observations. Considering *nh* the number of observations per *h* stratum in the ConVid sample and *n* the total sample, the weighting factor (*Wh*) in each stratum was calculated by:

$$Wh = (Nh/nh).(n/N)$$

Before the beginning of the weighting procedures, the data were analyzed for the presence of duplicates (all answers exactly equal) and missing data in the variables used for post-stratification.

Representativeness

To verify the representation of the Brazilian population in the sample obtained at ConVid, the sample distributions of sociodemographic characteristics used and not used in the post-stratification (number of residents in the household and condition in the workforce) were compared with those of the PNAD 2019 [33].

Considering ConVid questions about the situation before the pandemic in Brazil, the indicators related to healthy behaviors were compared to those of the PNS 2013 [29,34], considered the gold standard of national health surveys. Moreover, Vigitel 2019 [27] indicators were also used to compare with more current data.

The following lifestyle indicators were considered: proportion of people aged 18 years or older with the recommended consumption of fruits and vegetables (FLV); proportion of people aged 18 years or older with the consumption of beans for 5 days or more per week (beans 5 days or more); proportion of people aged 18 years or older with abusive and frequent consumption of alcoholic beverages, defined as the consumption of eight or more doses of alcoholic beverage per week for women and 15 or more doses for men (heavy drinking); proportion of people aged 18 years or older who are smokers (smokers); proportion of people aged 18 years or older who practice the recommended level of physical activity during leisure time (physical activity ≥ 150 minutes); proportion of people aged 18 years or older who watch TV or use tablet in their free time for 3 hours or more (TV/tablet 3 hours or more).

Regarding the practice of physical activity in the recommended time (150 minutes of moderate physical activity or 75 minutes of intense physical activity per week), since the questions related to the number of days (1-2; 3-4; 5 or more days), and the time of physical activity (< 30; 30-45; 46-59; 60 or more minutes) are categorical in ConVid and there were no questions about the type of physical activity practiced, the estimate corresponded to the average between the minimum proportion (individuals who practice 3-4 days a week for 46 minutes or more and 5 or more days per week and 30 or

more minutes) and the maximum (individuals who practice 1-2 days a week and 60 or more minutes or 3-4 days a week for 30 minutes or more and 5 or more days per week).

Among older adults (60+ years), it was considered the indicator of functional limitation that consists of needing help to perform daily activities (ADL limitation), such as eating, bathing, using the bathroom, dressing, getting around the house, or lying down.

Furthermore, the indicators of reported morbidity were also compared to those obtained in the PNS 2013 [29,34] and Vigitel 2019 [27]. The following indicators were considered: proportions of people aged 18 years or older who report a diagnosis of hypertension, diabetes, heart disease and depression; proportion of people aged 18 years or older who report a diagnosis of noncommunicable chronic diseases at risk of worsening COVID-19 (NCD COVID-19 risk); proportion of people aged 18 years or older with excess weight, defined by the body mass index (BMI) greater than or equal to 25kg/m2 (overweight); proportion of people aged 18 years or older with obesity, defined by BMI greater than or equal to 30kg/m2 (obesity).

The presence of NCD COVID-19 risk was calculated based on the reported diagnosis of at least one of the following diseases: diabetes, hypertension, asthma/emphysema/chronic respiratory disease or other lung disease, heart disease or cancer.

For each of the indicators of lifestyles and prevalence of self-reported NCDs, three different estimates were considered: total sample without weighting; total sample with weighting and, to reduce the selection bias [35] of the snowball recruitment procedure, the sample was considered without the first two waves, that is, excluding the first 18,500 participants (500 seeds that invited three people in each of the 12 strata).

Since no information was collected about the contacts network, it was not possible to consider the dependence of observations on the estimation of variance. In this study, the proposal of Goel & Salganik [35] was adopted, and the estimates of lifestyle indicators and self-reported non-communicable diseases prevalence was estimated based on the ConVid sample, without considering the first two recruitment chains. The indicators of each research were described using estimates of proportions and 95% confidence intervals (95%CI).

A map was elaborated with all the municipalities in Brazil addressed by the research in order to verify the diversity of ConVid sample in geographical terms.

Results

In total, 47,184 people agreed to participate in the survey. After excluding the questionnaires with missing data in the variables used for post-stratification (4.3%), a total of 45,161 questionnaires were considered for data analysis.

Table 1 shows the distributions of the variables used for database weighting obtained with data from PNAD 2019 and ConVid. There are few differences between the distributions of the variables used for weighting. Regarding the variables not used for post-stratification, both for the number of residents in the household and in the workforce condition, the distributions were similar.

Three different estimates of lifestyle indicators and prevalence of diagnosis of NCDs are presented in Table 2. The first estimate performed with the total sample without weight compared to the weighted sample shows the influence of the higher schooling level among ConVid participants: for regular consumption of fruits and vegetables, the percentage was higher by seven points, and the percentage of smokers was lower by 1.3 points. As for NCD, the greatest difference occurred for depression. When comparing the estimates corresponding to the sample with the exclusion of the first two waves with those obtained in the total weighted sample, small differences are perceived and shown that the estimates are significant, even after the exclusion of participants from the first two recruitment chains.

In Table 3, the estimates of lifestyle indicators performed by the ConVid are compared with those obtained in the PNS 2013 [29,34] and Vigitel 2019 [27]. Regarding the indicator of recommended consumption of fruits and vegetables, the results of the three surveys are very close, ranging from 22.1% to 22.9%. On the other hand, the percentage of consumption of beans for five days or more of ConVid (42.4%) was much lower than that of the PNS 2013 (71.9%) and Vigitel 2019 (59.7%).

Table 1

Comparison of distributions of the variables used and not used for the post-stratification of *ConVid – Behavior Survey* data with the distributions of the variables in 2019 *Brazilian National Household Sample Survey* (PNAD 2019, in Portuguese) [33].

Characteristic/Category	Distribution	
	ConVid	PNAD 2019
Characteristics used in post-stratification		
Region		
North	7.7	7.9
Northeast	25.0	26.3
Southeast	45.3	43.4
South	15.1	14.7
Central	6.9	7.6
Capital/Other parts of the UF		
Capital	25.1	24.7
Other parts of the UF	74.9	75.3
Sex		
Male	47.0	47.3
Female	53.0	52.7
Age group (years)		
18-29	24.7	23.6
30-39	21.0	20.7
40-49	18.1	18.4
50-59	15.9	16.2
60+	20.3	21.0
Schooling level		
Complete high school	83.6	84.3
Complete higher education	16.4	15.7
Race/Skin color		
White	44.7	43.8
Black	9.6	9.7
Mixed race	44.9	45.4
Other	0.8	1.1
Characteristics not used in post-stratification		
Condition in relation to the workforce *		
Works	71.6	70.7
Does not work	28.4	29.3
Number of residents in the household		
1	9.5	7.2
2	26.4	24.1
3	27.3	27.8
4	22.6	23.0
5	9.5	10.4
6+	4.7	7.4

UF: Federation Unit (acronym in Portuguese).

* Among individuals aged 18-69 years.

Table 2

Estimates of the indicators using the sample without weighting, the weighted sample and the sample not considering the first two waves.

Indicator	Estimate 1 *	Estimate 2 **	Estimate 3 ***
	%	%	%
Life habits			
FLV	29.1	22.1	22.1
Beans 5 days or more	37.7	43.2	42.4
Physical activity ≥ 150 minutes	34.2	35.51	35.2
TV/tablet ≥ 3 hours	66.0	60.6	61.6
Heavy drinking	7.1	6.5	6.7
Smokers	10.7	12.0	12.3
ADL limitation	5.7	6.1	6.2
Prevalence of chronic non-communicable diseases (self-reported data)			
Hypertension	18.5	19.3	18.6
Diabetes	7.0	7.2	7.1
Heart disease	3.8	4.3	4.4
Depression	16.6	14.9	14.9
NCD COVID-19 risk	33.3	33.9	33.3
Overweight	54.0	56.3	55.7
Obesity	19.8	21.4	21.2

ADL limitation: functional limitation that consists of needing help to perform daily activities among older adults; Beans 5 days or more: consumption of beans for 5 days or more per week; FLV: recommended consumption of fruits and vegetables; Heavy drinking: abusive and frequent consumption of alcoholic beverages, defined as the consumption of eight or more doses of alcoholic beverage per week for women and 15 or more doses for men; NCD COVID-19 risk: report a diagnosis of noncommunicable chronic diseases at risk of worsening COVID-19; Physical activity ≥ 150 minutes: practice the recommended level of physical activity during leisure time; Smokers: who are smokers; TV/tablet ≥ 3 hours: watch TV or use tablet in their free time for 3 hours or more.
* No weighting;
** Weighted;
*** With weighting, not considering the first two waves.

When comparing leisure-time physical activity in the recommended time per week (≥ 150 minutes), ConVid percentage (35.2%) is at the same level as Vigitel 2019 (39%). The small difference of 3.8 percentage points can be justified by the method of estimating the indicator with the ConVid data, corresponding to the mean between the minimum proportion (30%) and the maximum (40.5%), very close to the Vigitel 2019 data. Regarding TV/tablet time for 3 hours or more, the proportion found in ConVid, 61.6%, was similar to Vigitel, 62.7% (Table 2).

Regarding drinking habits, the prevalence of abusive and frequent alcohol consumption (heavy drinking) calculated using the ConVid data of 6.7% was similar to that estimated with the PNS 2013 data of 6.1%.

Regarding the proportion of smokers, the ConVid estimate (12.3%) was among the proportion found in Vigitel 2019 (9.8%) and that verified in the PNS 2013 (14.7%).

The comparative analysis of the indicator of functional limitation among people aged 60 years or older who need help to perform activities of daily living (Table 2) presented very similar percentages in ConVid (6.2%) and in the PNS 2013 (6.1%).

Table 4 the estimates of indicators of reported morbidity are compared. The proportions of people who report hypertension diagnosis are similar in the three studies, with values of 18.6% (ConVid), 21.4% (PNS 2013), and 24.5% (Vigitel 2019). Regarding the proportions that refer to the diagnosis of diabetes, the percentages were 6.2% (PNS 2013), 7.1% (ConVid), and 7.4% (Vigitel 2019), consistent with the estimates obtained with the results of laboratory tests of glycated hemoglobin of the PNS (2014-2015), which ranged from 6.6% to 9.4% using different criteria for the presence of this disease [36]. The comparison of the prevalence of diagnosis of ConVid heart disease with the PNS

Table 3

Comparison of lifestyle indicators obtained with data from *ConVid – Behavior Survey*, 2013 *National Health Survey* (PNS 2013, in Portuguese) [26] and 2019 *Surveillance of Risk and Protective Factors for Chronic Diseases by Telephone Survey* (Vigitel 2019, in Portuguese) [27].

Prevalence	ConVid		PNS 2013		Vigitel 2019	
	%	95%CI	%	95%CI	%	95%CI
FLV	22.1	21.6-22.6	22.8	22.1-23.5	22.9	22.1-23.7
Beans 5 days or more	42.4	41.8-43.0	71.9	71.2-71.6	59.7	58.8-60.6
Physical activity ≥ 150 minutes	35.21	34.6-35.8	22.5	21.8-23.1	39.0	38.0-39.9
TV/tablet ≥ 3 hours	61.6	61.0-62.2	-	-	62.7	61.8-63.6
Heavy drinking	6.7	6.4-7.0	6.1	5.8-6.4	-	-
Smokers	12.3	11.9-12.7	14.5	14.5-15.0	9.8	9.2-10.5
ADL limitation	6.2	5.9-6.5	6.1	5.4-6.8	-	-

95%CI: 95% confidence interval; ADL limitation: unctional limitation that consists of needing help to perform daily activities among older adults; Beans 5 days or more: consumption of beans for 5 days or more per week; FLV: recommended consumption of fruits and vegetables; Heavy drinking: abusive and frequent consumption of alcoholic beverages, defined as the consumption of eight or more doses of alcoholic beverage per week for women and 15 or more doses for men; Physical activity ≥ 150 minutes: practice the recommended level of physical activity during leisure time; Smokers: who are smokers; TV/tablet ≥ 3 hours: watch TV or use tablet in their free time for 3 hours or more.
* In ConVid, the number of days and time practiced in physical activity were categorical variables and the estimate was made by mean among the minimum percentages (30.1%) and maximum (40.9%).

Table 4

Comparison of estimates of the prevalence of non-communicable chronic diseases obtained with self-reported data from *ConVid – Behavior Survey*, 2013 *National Health Survey* (PNS 2013, in Portuguese) [26] and 2019 *Surveillance of Risk and Protective Factors for Chronic Diseases by Telephone Survey* (Vigitel 2019, in Portuguese) [27].

Prevalence	ConVid		PNS 2013		Vigitel 2019	
	%	95%CI	%	95%CI	%	95%CI
Hypertension	18.6	18.1-19.1	21.4	20.8-22.0	24.5	23.8-25.3
Diabetes	7.1	6.8-7.4	6.2	5.9-6.6	7.4	7.0-7.9
Heart disease	4.4	4.2-4.6	4.2	3.9-4.5	-	-
Depression	14.9	14.5-15.3	7.6	7.2-8.1	-	-
NCD COVID-19 risk	33.3	32.7-33.9	34.0	33.2-34.7	-	-
Overweight	55.7	55.1-56.3	56.4	55.6-57.2	55.4	54.4-56.3
Obesity	21.2	20.7-21.7	20.8	20.2-21.4	20.3	19.5-21.0

95%CI: 95% confidence interval; NCD COVID-19 risk: report a diagnosis of noncommunicable chronic diseases at risk of worsening COVID-19.

2013 shows similar values of 4.4% and 4.2%, respectively. In relation to the percentage of people with chronic diseases at risk of worsening the COVID-19 (33.3%), the estimate was also very similar to that calculated with the PNS 2013 data (34%) [37]. Regarding overweight and obesity, the prevalence of the three studies is similar. The only considerable difference in relation to the PNS was in relation to the report of depression with a higher proportion in ConVid (14.9% vs. 7.6%).

The map shown in Figure 1 shows the wide geographical scope of ConVid. In one month of research, all ufs and 1,696 municipalities with at least one participant were reached.

Discussion

The accomplishment of the *ConVid – Behavior Survey* made it possible to describe the changes in the behaviors of Brazilians, in health conditions and access to health services, to evaluate psychological disorders, and socioeconomic problems related to rigid measures of social restriction imposed by state and municipal governments after the intense dissemination of COVID-19 in Brazil.

The study used a virtual questionnaire with participants recruited by chain sampling – virtual snowball type – which enabled a large sample of more than 45,000 people in a month, without the need for external resources. The results were also published by the research website (https://convid.fiocruz.br/) and by the media, in an agile way, which made it possible to guide the population to maintain healthy behaviors and to seek telemedicine care to mitigate psychological disorders.

The comparison of indicators of healthy behaviors and morbidity reported with the results of traditional studies – such as PNS 2013 and Vigitel 2019 – showed great similarity in most estimates, making it possible to affirm that there is representation of the Brazilian population in the ConVid sample. In the set of lifestyle indicators, ConVid estimates regarding smoking and regular physical activity were closer to those of Vigitel 2019 and indicate the advances achieved in these behaviors over the years [27,38]. Only the percentage of regular consumption of beans showed dissonant value, probably associated with the higher schooling level of ConVid participants, being compatible with the percentage found among individuals with higher schooling level in Vigitel 2019 [27]. In the set of indicators of reported morbidity, the only discrepant indicator was the self-reported prevalence of depression when compared to the PNS 2013 estimate. In view of the large percentage of people who sought care for mental health issues during the pandemic, the proportion of people diagnosed with depression may, in fact, have increased.

Figure 1

Municipalities with at least one convid respondent *ConVid – Behavior Survey*. Brazil, from April 24 to May 24, 2020.

Although Internet searches have a number of advantages, problems of different orders can occur in network sampling. Regarding ConVid, people who do not have Internet access have zero probability of being selected and the snowball procedure was used by virtual networks, which is a non-probabilistic sampling technique in which participants are volunteers, and both selection probabilities and non-response rate cannot be estimated. Moreover, as the recruitment logic is not managed by the researcher, the connections between participants are not known, and it is not possible to consider the dependence of observations on the estimation of variance.

To obtain a representative sample of the population, weightings calculated by post-stratification procedures based on the PNAD 2019 were performed. The problem of post-stratification is that there may be factors related to the indicators of interest, but not considered in the set of variables used in stratification due to limitations in the sample size for stratification by various variables or the lack of reach of the sample of certain strata. In recent articles [39,40], alternatives were discussed for estimating the probabilities of selection in non-probabilistic samples based on a probabilistic reference sample or demographic census. In the case of Brazil, where PNAD is performed annually, the application of these methodologies is feasible and of great interest for providing statistical inference in research with non-probabilistic samples performed over the Internet.

In ConVid, for the demographic variables (gender, age group, and race) the necessary diversity was obtained to weigh data, in order to obtain the representativeness of the Brazilian population. Regarding the scope of geographic distribution, the work reached all UFs and about 1,700 municipalities, allowing to represent all regions and the location of the municipality of residence (capital/other parts of the state). Due to the fact that the participants had the tendency to invite peers to participate in the research with characteristics similar to theirs – called the effect of homophily [41], the stratification of seeds, as performed at ConVid, was a significant factor to achieve diversity in the sample [31].

However, as the percentage of people with incomplete high school in the sample (11.1%) was much lower than that of the PNAD 2019 (48.7%), only two categories of schooling level were used in the post-stratification (higher level or not), in order to not attribute very high weights to a small group of individuals. According to data from PNAD Contínua 2018 [1], among people who did not access the Internet in 2018, 41.6% claimed that they did not know how to use. The difficulties in participating in Internet surveys also affect the speed of development of networks, which develop faster among individuals who have greater ease in filling out a questionnaire and end up being predominant in the sample.

In data collected by chain sampling, the effect of homophily also contributes to the dependence of observations, and it is necessary to use procedures that consider the connections between participants for the estimation of variance and drawing effect [42]. When the recruiter/recruiting pairs are registered, the bootstrap procedure has been recommended for estimating the variance of the indicators of interest by several simulations of samples generated with the same process that originated the total sample [43,44], but it was not possible to use it in ConVid because information about the network of contacts was not collected. Another proposal to estimate variance in samples by networks is to disaggregate the endogenous and contextual effects by comparing the covariances of the pairs' results with the covariances of the results of the pairs of pairs [45].

Studies point to a greater diversity of people in social networks, regarding geographic location and socioeconomic characteristics, than in the classic clusters used in traditional surveys by home sampling [3,10]. Under the assumption that the effect of homophily decreases as the participant's "distance" to seed increases [45], the expansion of the network in several waves of recruitment would converge to the composition of a large and comprehensive sample of population characteristics. In the case of ConVid, the estimates made with the total sample without weighting show the influence of the schooling level on the estimates of behavior indicators. Probable explanations are the social exclusion of Internet access and the faster development of the network among individuals with higher education.

Although the use of online surveys is increasingly frequent to estimate the prevalence of interest events to the public health, inferences about the structural properties of social networks are rarely made [46]. Goel & Salganick [35] proposed to exclude participants from the first waves of the network in order to correct the selection bias of the chain recruitment processes. The estimates of lifestyle and non-communicable diseases indicators excluding the first two waves, were significant and very simi-

lar to those calculated with the total weighted sample, so that this may be a more appropriate option for use in virtual research and it was adopted in this study.

Online research is very promising for the health-related fields; especially regarding the acquisition of knowledge in an agile and immediate manner, reflecting the situation of health conditions in real time. To assure more quality to research conducted on the Internet, a list of recommendations was proposed ranging from the identification of people who fill out the electronic questionnaire more than once, the preparation of the questionnaire, to the calculation of the participation rate [47]. However, there is still concern as to whether the group of people reached in the sample by networks represents the general population. The comparison of distribution by schooling level in the ConVid sample with that obtained in the PNAD 2019 showed the Brazilian social division in Internet access, considered as one of the most significant markers of socioeconomic inequalities [48]. Despite the gap of people with low schooling, the estimates of several health indicators were, mostly, compatible with the values found in the PNS 2013 and Vigitel 2019. Regarding the estimation of variance, studies are necessary to investigate how the endogenous effects of virtual social networks can considered, so that statistical inference can be used appropriately and with more reliability.

Contributors

Additional informations

ORCID: Celia Landmann Szwarcwald (0000-0002-7798-2095); Paulo Roberto Borges de Souza Júnior (0000-0002-8142-4790); Giseli Nogueira Damacena (0000-0002-7059-3353); Deborah Carvalho Malta (0000-0002-8214-5734); Marilisa Berti de Azevedo Barros (0000-0003-3974-195X); Dalia Elena Romero (0000-0002-2643-9797); Wanessa da Silva de Almeida (0000-0002-5164-8603); Luiz Otávio Azevedo (0000-0002-4876-5948); Ísis Eloah Machado (0000-0002-4678-2074); Margareth Guimarães Lima (0000-0001-6996-0745); André Oliveira Werneck (0000-0002-9166-4376); Danilo Rodrigues Pereira da Silva (0000-0003-3995-4795); Crizian Saar Gomes (0000-0001-6586-4561); Arthur Pate de Souza Ferreira (0000-0002-6686-0105); Renata Gracie (0000-0003-0225-3696); Maria de Fátima de Pina (0000-0002-1521-7865).

Acknowledgments

References

1. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios Contínua: acesso à Internet e à televisão e posso de telefone móvel celular para uso pessoal 2018. PNAD Contínua. https://agenciadenoticias.ibge.gov.br/media/com_mediaibge/arquivos/10d5c0576ff8d726467f1d4571dd8e62.pdf (accessed on 01/Sep/2020).

2. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016; (Suppl 1):S48-61.

3. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012; 14:e20.

4. De Cocker K, Chastin SFM, De Bourdeaudhuij I, Imbo I, Stragier J, Cardon G. Citizen science to communicate about public health messages: the reach of a playful online survey on sitting time and physical activity. Health Commun 2019; 34:720-5.

5. Milgram S. The small-world problem. Psychol Today 1967; 2:60-7.

6. Faleiros F, Käppler C, Pontes FAR, Silva SSC, Goes FSN, Cucik CD. Uso de questionário online e divulgação virtual como estratégia de coleta de dados em estudos científicos. Texto Contexto Enferm 2016; 25:e3880014.

7. Ekman A, Litton JE. New times, new needs; e-epidemiology. Eur J Epidemiol 2007; 22:285-92.

8. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010; 172:1292-8.

9. Nolte MT, Shauver MJ, Chung KC. Analysis of four recruitment methods for obtaining normative data through a web-based questionnaire: a pilot study. Hand (N Y) 2015; 10:529-34.

10. Costa BRL. Bola de neve virtual: o uso das redes sociais virtuais no processo de coleta de dados de uma pesquisa científica. Revista Interdisciplinar de Gestão Social 2018; 7:15-37.

11. Aglipay M, Wylie JL, Jolly AM. Health research among hard-to-reach people: six degrees of sampling. CMAJ 2015; 187:1145-9.

12. Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS 2005; 19 Suppl 2:S67-72.

13. Faugier J, Sargeant M. Sampling hard to reach populations. J Adv Nurs 1997; 26:790-7.

14. Ali SH, Foreman J, Capasso A, Jones AM, Tozan Y, Di Clemente RJ. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Med Res Methodol 2020; 20:116.

15. Gharpure R, Hunter CM, Schnall AH, Barrett CE, Kirby AE, Kunz J, et al. Knowledge and practices regarding safe household cleaning and disinfection for COVID-19 prevention – United States, May 2020. Morb Mortal Wkly Rep 2020; 69:705-9.

16. Reuben RC, Danladi MMA, Saleh DA, Ejembi PE. Knowledge, attitudes and practices towards COVID-19: an epidemiological survey in north-central Nigeria. J Community Health 2020; [Epub ahead of print].

17. Machida M, Nakamura I, Saito R, Nakaya T, Hanibuchi T, Takamiya T, et al. Adoption of personal protective measures by ordinary citizens during the COVID-19 outbreak in Japan. Int J Infect Dis 2020; 94:139-44.

18. Denis F, Galmiche S, Dinh A, Fontanet A, Scherpereel A, Benezit F, et al. Epidemiological observations on the association between anosmia and COVID-19 infection: analysis of data from a self-assessment web application. J Med Internet Res 2020; 22:e19855.

19. Guo Y, Cheng C, Zeng Y, Li Y, Zhu M, Yang W, et al. Mental health disorders and associated risk factors in quarantine adults during the COVID-19 outbreak in China: a cross-sectional study. J Med Internet Res 2020; 22:e20328.

20. Vordenberg SE, Zikmund-Fisher BJ. Older adults' strategies for obtaining medication refills in hypothetical scenarios in the face of COVID-19 risk. J Am Pharm Assoc (2003) 2020; 60:915-22.

21. Steele EM, Rauber F, Costa CS, Leite MA, Gabe KT, Louzada MLC, et al. Dietary changes in the NutriNet Brasil cohort during the COVID-19 pandemic. Rev Saúde Pública 2020; 54:91.

22. Bommele J, Hopman P, Walters BH, Geboers C, Croes E, Fong GT, et al. The double-edged relationship between COVID-19 stress and smoking: implications for smoking cessation. Tob Induc Dis 2020; 18:63.

23. Mukhtar S. Psychological health during the coronavirus disease 2019 pandemic outbreak. Int J Soc Psychiatry 2020; 66:512-6.

24. Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. J Popul Econ 2020; [Online ahead of print].

25. Szwarcwald CL, Viacava F. *Pesquisa Mundial de Saúde* no Brasil, 2003. Cad Saúde Pública 2005; 21 Suppl 1:S4-5.

26. Szwarcwald CL, Malta DC, Pereira CA, Vieira ML, Conde WL, Souza Júnior PR, et al. *Pesquisa Nacional de Saúde* no Brasil: concepção e metodologia de aplicação. Ciênc Saúde Colet 2014; 19:333-42.

27. Departamento de Análise em Saúde e Vigilância de Doenças Não Transmissíveis., Secretaria de Vigilância em Saúde, Ministério da Saúde. Vigitel Brasil 2019. Vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2019. Brasília: Ministério da Saúde; 2020.

28. Barros MBA, Lima MG, Ceolin MF, Zancanella E, Cardoso TAMO. Quality of sleep, health status and subjective well-being in a population-based study. Rev Saúde Pública 2019; 53:82.

29. Pereira CA, Malta DC, Szwarcwald CL. Pesquisa Nacional de Saúde 2013: ciclos de vida. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2015.

30. Souza Júnior PR, Szwarcwald CL, Castilho EA. Self-rated health by HIV infected individuals undergoing antiretroviral therapy in Brazil. Cad Saúde Pública 2011; 27 Suppl 1:S56-66.

31. McKnight C, Des Jarlais D, Bramson H, Tower L, Abdul-Quader AS, Nemeth C, et al. Respondent-driven sampling in a study of drug users in New York City: notes from the field. J Urban Health 2006; 83 Suppl 6:i54-9.

32. Szwarcwald CL, Damacena GN. Amostras complexas em inquéritos: planejamento e implicações na análise estatística de dados. Rev Bras Epidemiol 2008; 11 Suppl 1:38-45.

33. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios (PNAD) 2019. Notas técnicas. Versão 1.5. 4ª Ed. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2019.

34. Pereira CA, Malta DC, Szwarcwald CL. Pesquisa Nacional de Saúde 2013: percepção do estado de saúde, estilos de vida e doenças crônicas. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2014.

35. Goel S, Salganik MJ. Respondent-driven sampling as Markov chain Monte Carlo. Stat Med 2009; 28:2202-29.

36. Malta DC, Duncan BB, Schmidt MI, Machado IE, Silva AGD, Bernal RTI, et al. Prevalence of diabetes mellitus as determined by glycated hemoglobin in the Brazilian adult population, National Health Survey. Rev Bras Epidemiol 2019; 22 Suppl 2:E190006.

37. Rezende LFM, Thome B, Schveitzer MC, Souza Júnior PRB, Szwarcwald CL. Adults at high-risk of severe coronavirus disease-2019 (COVID-19) in Brazil. Rev Saúde Pública 2020; 54:50.

38. Pinto M, Bardach A, Palacios A, Biz A, Alcaraz A, Rodriguez B, et al. Burden of smoking in Brazil and potential benefit of increasing taxes on cigarettes for the economy and for reducing morbidity and mortality. Cad Saúde Pública 2019; 35:e00129118.

39. Elliott MR, Valliant R. Inference for nonprobability samples. Stat Sci 2017; 32:249-64.

40. Valliant R. Comparing alternatives for estimation from nonprobability samples. J Surv Stat Methodol 2020; 8:231-63.

41. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. Sociol Methodol 2004; 34:193-240.

42. Szwarcwald CL, Souza Júnior PR, Damacena GN, Barbosa Junior A, Kendall C. Analysis of data collected by RDS among sex workers in 10 Brazilian cities, 2009: estimation of the prevalence of HIV, variance, and design effect. J Acquir Immune Defic Syndr 2011; 57 Suppl 3:S129-35.

43. Spiller MW, Gile KJ, Handcock MS, Mar CM, Wejnert C. Evaluating variance estimators for respondent-driven sampling. J Surv Stat Methodol 2017; 2017:smx018.

44. Baraff AJ, McCormick TH, Raftery AE. Estimating uncertainty in respondent-driven sampling using a tree bootstrap method. Proc Natl Acad Sci U S A 2016; 113:14668-73.

45. Rose CD. Identification of peer effects through social networks using variance restrictions. Econom J 2017; 20:S47-60.

46. Verdery AM, Fisher JC, Siripong N, Abdesselam K, Bauldry S. New survey questions and estimators for network clustering with respondent-driven sampling data. Sociol Methodol 2017; 47:274-306.

47. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys. J Med Internet Res 2004; 6:e34.

48. Feehan DM, Cobb C. Using an online sample to estimate the size of an offline population. Demography 2019; 56:2377-92.

Resumo

A ConVid – Pesquisa de Comportamentos *foi realizada no Brasil de 24 de abril a 24 de maio de 2020, com o objetivo de investigar as mudanças nos estilos de vida e nas condições de saúde durante a pandemia de COVID-19. Neste artigo, apresentamos a concepção e metodologia da pesquisa. Estudo de corte transversal com a utilização de um questionário pela Internet, com questões validadas em inquéritos de saúde anteriores. O método de amostragem foi o "bola de neve virtual" e foram usados os procedimentos de pós-estratificação. Os resultados relativos às doenças crônicas não transmissíveis e estilos de vida pré-pandemia foram comparados às estimativas da* Pesquisa Nacional de Saúde *de 2013 e da* Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico *de 2019. A amostra total foi de 45.161 pessoas. Após a ponderação dos dados, as distribuições amostrais das variáveis demográficas foram semelhantes às populacionais. Apenas as pessoas de baixo nível de instrução foram sub-representadas. A comparação com os resultados anteriores mostrou similaridade na maioria das estimativas: consumo recomendado de frutas e legumes (22,1%), atividade física recomendada (35,2%), fumo de cigarros (12,3%), consumo frequente e abusivo de álcool (6,7%), obesidade (21,2%), prevalências autorreferidas de hipertensão (18,6%), diabetes (7,1%) e doença do coração (4,4%). O inquérito online possibilitou conhecer as condições de saúde da população durante a pandemia. A similaridade dos indicadores com os obtidos em pesquisas tradicionais permitiu validar as estimativas médias. Estudos são necessários para investigar como os efeitos endógenos das redes sociais virtuais podem ser levados em consideração na estimação da variância.*

COVID-19; Internet; Inquéritos Epidemiológicos

Resumen

La ConVid – Encuesta de Comportamientos *se realizó en Brasil del 24 de abril al 24 de mayo de 2020, con el objetivo de investigar los cambios en los estilos de vida y en las condiciones de salud durante la pandemia de COVID-19. En este artículo, se presenta la concepción y metodología de la encuesta. Estudio de corte transversal, utilizando un cuestionario por Internet, con cuestiones validadas en encuestas de salud anteriores. El método de muestra fue el de "bola de nieve virtual" y se usaron procedimientos de posestratificación. Los resultados relativos a las enfermedades crónicas no transmisibles y estilos de vida prepandemia se compararon con las estimaciones de la* Encuesta Nacional de Salud *de 2013 y de la* Vigilancia de Factores de Riesgo y Protección para Enfermedades Crónicas por Encuesta Telefónica *de 2019. La muestra total fue de 45.161 personas. Tras la ponderación de los datos, las distribuciones de la muestra de las variables demográficas fueron semejantes a las poblacionales. Solamente las personas con un bajo nivel de formación estuvieron sub-representadas. La comparación con los resultados anteriores mostró similitud en la mayoría de las estimaciones: consumo recomendado de frutas y legumbres (22,1%), actividad física recomendada (35,2%), tabaco de liar para cigarrillos (12,3%), consumo frecuente y abusivo de alcohol (6,7%), obesidad (21,2%), prevalencias autoinformadas de hipertensión (18,6%), diabetes (7,1%) y enfermedad del corazón (4,4%). La encuesta online posibilitó conocer las condiciones de salud de la población durante la pandemia. La similitud de los indicadores con los obtenidos en investigaciones tradicionales permitió validar las estimaciones medias. Se necesitan estudios para investigar cómo los efectos endógenos de las redes sociales virtuales pueden ser tenidos en consideración en la estimación de la variancia.*

COVID-19; Internet; Encuestas Epidemiológicas