PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	EDUCATIONAL DISPARITIES IN HYPERTENSION, DIABETES, OBESITY AND SMOKING IN BRAZIL: A TREND ANALYSIS OF 578,977 ADULTS FROM A NATIONAL SURVEY, 2007 TO 2018.
AUTHORS	Toteff Dulgheroff, Pedro; da Silva, Luciana; Madalena Rinaldi, Ana Elisa; Rezende, Leandro; Souza Marques, Emanuele; Azeredo, Catarina

VERSION 1 – REVIEW

REVIEWER	Strand, Bjørn Heine Norwegian Institute of Public Health
REVIEW RETURNED	08-Dec-2020

GENERAL COMMENTS	This interesting paper investigates educational inequalities in hypertension, diabetes, obesity and smoking in annually large cross-sectional Brazilian dataset including almost 580 000 adults during 2007-2018. There is a lack of studies of health inequalities from Brazil, so the study is very welcomed. The paper covers many dimensions (four outcomes, educational inequality, sex, time trends, skin color), and even if it seems well conducted, it is a bit hard to follow, and results could have been presented a bit easier. Add in two complex inequality metrics and it is even harder to follow. My main input to the authors is therefore to make results easier to follow, and analyses more transparent. Why was age in 2018 used as adjustment factor? This means that a person aged 60 in 2018 was in fact 49 at the time of measurement in 2007. I suspect that rather age was adjusted for it was age-standardized using the age distribution in 2018 as the standard? Why was the Prais-Winsten method applied? Was serial correlation an issue in the analyses? Could this be justified and described so the reader knows why this method (which probably is unfamiliar to most readers) was used. Why is both absolute and relative inequality reported? I would narrow it down to only one measure, preferably the absolute measure, which is the measure important for public health. This would make the paper easier to read. For example, it is stated that absolute educational inequalities during 2007 to 2018 is stable, while relative inequalities narrow. What message is this? How can readers interpret this? What is the more important metric? From figure on page 22 it is evident that the relative inequality decreases because obesity increases over time for all groups, both low and high education groups. This is key information, which should be

	provided along with the confusing results of stable SII and decreasing CIX. Furthermore, the choice of inequality indicator could be better justified. Why was the SII selected? This metric “adjusts” for change in the educational distribution over time, but is this what the authors want? Furthermore, it is not a comparison of the two groups 0-3 years of education with the group 12+ years of education, as claimed in methods; it is in fact a comparison of the one person at the top of the educational hierarchy with the one at the bottom (100 percentile vs 0 th percentile), so it is a hypothetical measure, an extrapolation of the actual 4 educational groups. SII is nice as you get one number for the inequality, and you do not have to report the difference between all 4 groups. Educational level often differ quite much by sex, as men often have higher average educational level. Therefore I am puzzled by the sex-collapsed analyses in table 1 and table 2, as well as in several of the figs. Also, in tables 1 and 2, results are age-standardized. Would it not be as important to adjust for sex? Sex-specific results are presented in appendix, but I suggest this to be incorporated in tables in the main paper. The straight regression lines in figs on page 23-24 are less transparent. They are also a bit messy with too much information in one graph and with the CI bands it gets quite difficult to interpret. Why present both skin color and sex in the same graph? Instead of CI bands, could you rather present a p-value for each line, which is a test of slope different from 0? Alternatively, you could add this information to table 2, for example after the 2018 column you could add a column with yearly-change and a corresponding p-value (or 95%CI). I would also expand table 2 to include sex specific results, along with the total results (age- and sex-adjusted). In table 1, mean age is presented, but should be accompanied by SD as well to get an idea of age distribution in the sample. Furthermore, it is stated that max age is 80 years but minimum age is not mentioned. The figure on page 22 is a really nice illustration of the results, but I would suggest to change X and Y axis so we see the time trend along the X-axis. You could also consider to make one set of graphs for women and one set for men. In results, there are several claims of different increases/decreases in educational inequalities over time by sex and by skin color. Were these claims tested statistically by the inclusion of interaction terms? The exceptional large prevalence of self-reported hypertension is striking, especially in low educated women (64% in 2018) but also in low educated men (53%). For high education the prevalence is much lower (24% in both men and women). It would be very informative to know how many of these are on blood pressure medication, because of treated the high BP might be less dangerous than if not treated. What are the reasons for this very high prevalence of hypertension? Diet? Physical inactivity? In tables in appendix the CIs for female, male, white and non-white have negative values, and they are very wide. It seems only for total is correctly calculated. The minimum value of a prevalence is zero and thereby the CIs lower bound is zero. I suspect the authors have used linear regression to estimate the prevalence and the CIs, but still I wonder why intervals are so wide? There are several methods in Stata for standardization (direct standardization, poststratification), which I suggest the authors to use.

	I would skip decimals in the text of the results section as this will increase readability. I would remove subjective words such as “important” in the results chapter (page 8, line 50). Such words can be saved for the discussion chapter. To my knowledge, Body Mass Index is mostly abbreviated as BMI.

REVIEWER	Matranga, Domenica University of Palermo, Italy, Department of Health promotion Sciences and Mother- Child Care
REVIEW RETURNED	05-Jan-2021

GENERAL COMMENTS	It is a well-written paper about the observed trend of educational inequality in major NCDs (hypertension and diabetes) and their risk factors (smoking and obesity) in Brazil between 2007 and 2018. This paper gives a relevant contribution to scarce literature about this topic in Brazil. My main concern is about the representativeness of the sample used, although its magnitude is impressive (from 54,271 people interviewed in 2007 to 48,463 interviewed in 2018). Besides, it is also acknowledged by the authors due to the survey technique used (telephone interview) that introduces severe selection bias in the sample. Therefore, I suggest not saying that the sample is representative, as every correction proposed (individual weighting for instance) is questionable and in my opinion, cannot give representativeness to this sample. Specifically, Page 4, Lines 11-12: Write “We used large National samples from Brazilian adults…”(replace representative by national) Page 6, lines 37-39: Delete the sentence “This weighting ensured the representativeness for the general adult population of each city in all years” Page 9, lines 32-36: unclear, please reformulate Pages 22-23, Figures 2-3: use a caption to explain the categories with an asterisk Finally, I recommend a stylistic review by a Native speaker

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Dr. Bjørn Heine Strand, Norwegian Institute of Public Health

1) This interesting paper investigates educational inequalities in hypertension, diabetes, obesity and smoking in annually large cross-sectional Brazilian dataset including almost 580 000 adults during 2007-2018. There is a lack of studies of health inequalities from Brazil, so the study is very welcomed. The paper covers many dimensions (four outcomes, educational inequality, sex, time trends, skin color), and even if it seems well conducted, it is a bit hard to follow, and results could have been presented a bit easier. Add in two complex inequality metrics and it is even harder to follow. My main input to the authors is therefore to make results easier to follow, and analyses more transparent.

We appreciate the positive comments on our paper. According to your following comments, we have revised the text and figures to make the results easier to follow.

2) Why was age in 2018 used as adjustment factor? This means that a person aged 60 in 2018 was in fact 49 at the time of measurement in 2007. I suspect that rather age was adjusted for it was age-standardized using the age distribution in 2018 as the standard?

We apologize for the lack of clarity. We have used age distribution in 2018 as a standard to age-standardize the outcomes of interest. Age is a variable associated with all the outcomes of interest, and the age distribution of the sample was different throughout the period assessed (2007 to 2018), which could induce confounding if not adjusted. We have included a sentence in the methods section to clarify this issue (page 6, paragraph 2):

“Prevalence of hypertension, diabetes, smoking and obesity (2007-2018) was age-standardized using the age distribution of the year 2018.”

3) Why was the Prais-Winsten method applied? Was serial correlation an issue in the analyses? Could this be justified and described so the reader knows why this method (which probably is unfamiliar to most readers) was used.

We used the Prais-Winsten method modified by Durbin and Watson as a linear regression model because it reduces the autoregressive problem found in social serial data trends analysis in comparison with traditional regression analysis. We included more explanation in our methods (page 7, paragraph 2):

“The time trend of the indicators was analyzed using the Prais-Winsten method modified by Durbin and Watson instead of traditional linear regression to avoid the autoregressive problem common in this social serial trend analysis”

4) Why is both absolute and relative inequality reported? I would narrow it down to only one measure, preferably the absolute measure, which is the measure important for public health. This would make the paper easier to read. For example, it is stated that absolute educational inequalities during 2007 to 2018 is stable, while relative inequalities narrow. What message is this? How can readers interpret this? What is the more important metric? From figure on page 22 it is evident that the relative inequality decreases because obesity increases over time for all groups, both low and high education groups. This is key information, which should be provided along with the confusing results of stable SII and decreasing CIX.

The use of both absolute and relative inequality measures is usual in inequality studies because they provide complementary information and thus should be presented together. They show different dimensions of inequality and one measure is not more important than the other (WHO, 2013). Absolute measure (SII) results from the regression line and represent the absolute difference, in predicted values, on disease prevalence between the highest and the lowest education level, considering the entire distribution of the stratification variable. Relative measure (CIX – concentration index) shows how concentrated are the diseases towards the poor groups (Barros et al., 2013; Silva et al., 2018).
CIX values can highlight how unequal the estimates among groups are, even being very sensitive to the occurrence of the outcome studied. Indicators with lower frequencies may present high relative inequalities, overestimating these differences in a public health overview. Similarly, high prevalence

outcomes may present important inequalities but are not grasped by relative measures (Silva et al., 2018).
For instance, the absolute educational inequalities (SII) from 2007 to 2018 was stable for obesity, while relative inequality (CIX) has narrowed. We agree with the reviewer that the message is not clear. The increase in the prevalence of obesity was higher in people with more years of study, especially in those with 9 to 11 years that was almost twice the change in those with 0 to 3 years. Because of that, relative inequality has reduced. However, it does not reflect a beneficial change in inequality.
We have included more information regarding these two measures in methods and the interpretation of these conflicting results in the discussion section (page 6, paragraph 2).

References:

WHO. Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. Geneva: WHO press, 2013:105.

Barros AJ, Victora CG. Measuring coverage in MNCH: determining and interpreting inequalities in coverage of maternal, newborn, and child health interventions. PLoS Med 2013;10(5):e1001390. doi: 10.1371/journal.pmed.1001390 [published Online First: 2013/05/15]

Silva I, Restrepo-Mendez MC, Costa JC, et al. Measurement of social inequalities in health: concepts and methodological approaches in the Brazilian context. Epidemiol Serv Saude 2018;27(1):e000100017. doi: 10.5123/S1679-49742018000100017 [published Online First: 2018/03/08]

5) Furthermore, the choice of inequality indicator could be better justified. Why was the SII selected? This metric “adjusts” for change in the educational distribution over time, but is this what the authors want? Furthermore, it is not a comparison of the two groups 0-3 years of education with the group 12+ years of education, as claimed in methods; it is in fact a comparison of the one person at the top of the educational hierarchy with the one at the bottom (100 percentile vs 0 th percentile), so it is a hypothetical measure, an extrapolation of the actual 4 educational groups. SII is nice as you get one number for the inequality, and you do not have to report the difference between all 4 groups.

We have chosen to present inequality using SII and CIX because they represent, respectively, absolute and relative inequality measures. SII and CIX are complex measures that assess disparities between all the categories of education, not being influenced by education level's extremes.
We agree with the revisor, however, that SII does not consider the change in the educational distribution over time, but this was not our aim. Our aim was to identify if the health gap between socioeconomic groups has increased or decreased over time, not to assess whether the number of low educated people have reduced or increased over time.
We have made it clear in text (page 6, paragraph 2; page 11, paragraph 2; page 14, paragraph 2):

“These measures of inequality are complementary and were calculated” … “The SII results from a linear regression of the cumulative population proportional distribution in each one of the four educational groups in this study and represents the absolute difference, in predicted values, on disease prevalence between the least and the most favored person, with no education and the highest possible education, taking into consideration the entire distribution of the stratification variable.” … “CIX values should be read with caution because it can overestimate inequalities when the outcome of interest has a low frequency and may not be able to identify important inequalities when the outcome prevalence is high”

“The gap in obesity prevalence between least and the most educated groups reduced over time, but it was not sufficient to impact SII indicator. However, due to an increase in obesity prevalence in all education groups, specially in those with 9 to 11 study years (53,1% while prevalence raised 33,3% in people with less than 4 years of study), relative inequality reduced. This reduction in relative inequality is an artificial change that should not be read as an achievement because does not reflect a beneficial change in inequality, but rather a worsening scenario for all strata of education."

“The reduction in inequality for obesity should be read with caution because it reflects increases in obesity prevalence in all groups.”

6) Educational level often differ quite much by sex, as men often have higher average educational level. Therefore I am puzzled by the sex-collapsed analyses in table 1 and table 2, as well as in several of the figs.

Our main objective was to investigate whether health inequalities have increased or not over time for the four outcomes. We also explored some sex and race/color differences in stratified analyses, but this was a subgroup analysis. We believe that to understand how the overall inequality trend is in Brazil is of major importance, but we have not neglected specific vulnerabilities (according to sex and skin color).
We reformulated/created tables and figures to make them easier to understand.

7) Also, in tables 1 and 2, results are age-standardized. Would it not be as important to adjust for sex? Sex-specific results are presented in appendix, but I suggest this to be incorporated in tables in the main paper.

We decided to standardize analyses for age because our population has aged during our study, which could affect the prevalence of the diseases assessed, however the sex distribution has not changed over time. We decided to keep sex-specific information only in supplementary tables because despite being informative, the prevalence of risk factors are not part of our primary outcomes and theses tables are too detailed to be included as main tables.

8) The straight regression lines in figs on page 23-24 are less transparent. They are also a bit messy with too much information in one graph and with the CI bands it gets quite difficult to interpret. Why present both skin color and sex in the same graph? Instead of CI bands, could you rather present a p-value for each line, which is a test of slope different from 0? Alternatively, you could add this information to table 2, for example after the 2018 column you could add a column with yearly-change and a corresponding p-value (or 95%CI). I would also expand table 2 to include sex specific results, along with the total results (age- and sex-adjusted).

We have created/reformulated figures, excluding CI bands and separating sex and skin color information. We included annual change and/or p-values as suggested in our figures and tables. More detailed data about sex and skin color is now present in table 2.

9) In table 1, mean age is presented, but should be accompanied by SD as well to get an idea of age distribution in the sample. Furthermore, it is stated that max age is 80 years but minimum age is not mentioned.

Suggestion accepted. We have included Standard Error in all table 1 to facilitate the estimation of CIs and included the minimum age in text (page 5, paragraph 2).

“This weighting aimed to achieve representativeness for population aged 18 years and over of each state capital in Brazil, including DF in all years”

10) The figure on page 22 is a really nice illustration of the results, but I would suggest to change X and Y axis so we see the time trend along the X-axis. You could also consider to make one set of graphs for women and one set for men.

Suggestion accepted. Two sets of graphs were made: one for sex and another for skin color, presented as supplementary figures 1 and 2. We appreciate your suggestion to change X and Y axis. However, these figures were made using the Equiplot for Stata, and this function does not allow axis inversion (please see: https://www.equidade.org/ineq-measures). Therefore, our objective with these figures is to verify the “distance” between the different groups. That is easier two see if prevalence is present in X axis.

11) In results, there are several claims of different increases/decreases in educational inequalities over time by sex and by skin color. Were these claims tested statistically by the inclusion of interaction terms?

We apologize for the lack of clarity. We have not included interaction terms using statistical tests because in Brazil there is a priori knowledge that health and access to health care are unequally distributed according to sex and skin color. Our intention was to identify whether the inequality according to these subgroups was reducing or increasing over time and if vulnerable subgroups were catching more privileged subgroups. These changes over time were tested statistically. We have included a sentence in the section of methods to explain the rational.

12) The exceptional large prevalence of self-reported hypertension is striking, especially in low educated women (64% in 2018) but also in low educated men (53%). For high education the prevalence is much lower (24% in both men and women). It would be very informative to know how many of these are on blood pressure medication, because of treated the high BP might be less dangerous than if not treated. What are the reasons for this very high prevalence of hypertension? Diet? Physical inactivity?

We appreciate the opportunity to discuss the high prevalence of hypertension. We used self-reported diagnostic of hypertension. Questionnaires containing self-reported measurements have been widely used in other countries and in Brazil. They are also low cost and easy to execute. Validation studies of different diagnostic criteria for hypertension (self-reported, measured by instrument and/or use of antihypertensive drugs) suggested that self-reported hypertension can be used as a valid population estimate (Lima-Costa et al, 2004; Chrestani et al, 2009; Malta et al, 2018). Only 3% of the Brazilian population declared never having measured blood pressure, which increased confidence in adopting self-referred measures as a proxy of the population prevalence rates (Malta et al, 2018). In addition, Brazilian studies have shown that the frequency of individuals with hypertension who reported drug treatment for the disease is greater than 80%, higher in women than in men, increasing with age and less education (Malta et al, 2015; Brasil, 2019). This result can be justified by the wide access to drugs for hypertension and diabetes in the Brazilian Health System (Sistema Único de Saúde) and the free drug distribution programs such as “Aqui tem Farmácia Popular”. Regarding the factors that contribute to the increase in the prevalence of hypertension, a study carried out with Brazilian adults

found that ageing, black skin color, low education, obesity, being a former smoker, self-reported diabetes and high cholesterol and high salt intake were associated with a higher prevalence of hypertension. In addition to individual factors, a study conducted with the North American population indicated that states with greater socioeconomic vulnerability, such as low family income and high percentages of the population below the poverty line were significantly associated with a high prevalence of self-reported hypertension, which corroborates with the inequality findings in our study. We have included a sentence in the discussion section to explain the reasons for the very high prevalence of hypertension (pages 10; paragraph 2).

“Hypertension had higher prevalence (33.3% in 2018) and the highest absolute educational inequality (-37.8 in 2018). The prevalence in the least educated group was 60.7%. A study carried out with Brazilian adults found that aging, black skin color, low education, obesity, being a former smoker, self-reported diabetes, high cholesterol and high salt intake were associated with a higher prevalence of hypertension19. In addition to individual factors, a study conducted with the North American population indicated that states with greater socioeconomic vulnerability, such as low family income and high percentages of the population below the poverty line were significantly associated with a high prevalence of self-reported hypertension20, which corroborates with the inequality findings in our study. However, although we found the highest educational inequality for hypertension, it remained constant in the period.”

References:

Lima-Costa MF, Peixoto SV, Firmo JOA. Validity of self-reported hypertension and its determinants (the Bambuí study). Rev. Saúde Pública. 2004; 38(5): 637-642. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102004000500004&lng=en. https://doi.org/10.1590/S0034-89102004000500004.

Chrestani MAD, Santos IS, Matijasevich AM. Hipertensão arterial sistêmica auto-referida: validação diagnóstica em estudo de base populacional. Cad Saúde Pública. 2009; 25(11): 2395-406. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2009001100010&lng=en. https://doi.org/10.1590/S0102-311X2009001100010.

Malta DC, Gonçalves RPF, Machado IE, Freitas MIF, Azeredo C, Szwarcwald CL. Prevalence of arterial hypertension according to different diagnostic criteria, National Health Survey. Rev. bras. epidemiol. 2018; 21(Suppl 1): e180021. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2018000200419&lng=en. Epub Nov 29, 2018. https://doi.org/10.1590/1980-549720180021.supl.1.

Malta DC, Stopa SR, Andrade SSCA, Szwarcwald CL, Silva-Júnior JB, Reis AAC. Health care in adults with self-reported hypertension in Brazil according to the National Health Survey, 2013. Rev. bras. epidemiol. 2015; 18(Suppl 2): 109-122. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-790X2015000600109&lng=en. https://doi.org/10.1590/1980-5497201500060010.

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise em Saúde e Vigilância de Doenças não Transmissíveis. Vigitel Brasil 2018: 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 2018 / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise em Saúde e Vigilância de Doenças não Transmissíveis. – Brasília: Ministério da Saúde, 2019.

13) In tables in appendix the CIs for female, male, white and non-white have negative values, and they are very wide. It seems only for total is correctly calculated. The minimum value of a prevalence is zero and thereby the CIs lower bound is zero. I suspect the authors have used linear regression to estimate the prevalence and the CIs, but still I wonder why intervals are so wide? There are several methods in Stata for standardization (direct standardization, poststratification), which I suggest the authors to use.

Thanks for your observation, in fact, the information presented before in CIs was wrong. We have corrected them.

14) I would skip decimals in the text of the results section as this will increase readability.

We appreciate your suggestion. However, to skip decimals, we would have to round numbers, and they would, eventually, do not match numbers from tables which could confuse readers.

15) I would remove subjective words such as “important” in the results chapter (page 8, line 50). Such words can be saved for the discussion chapter.

Thank you, we have removed these words from the results.

16) To my knowledge, Body Mass Index is mostly abbreviated as BMI.

We have corrected this mistake.

Reviewer(s)' Comments to Author:

Reviewer: 2

Dr. Domenica Matranga, University of Palermo, Italy

1) It is a well-written paper about the observed trend of educational inequality in major NCDs (hypertension and diabetes) and their risk factors (smoking and obesity) in Brazil between 2007 and 2018. This paper gives a relevant contribution to scarce literature about this topic in Brazil. My main concern is about the representativeness of the sample used, although its magnitude is impressive (from 54,271 people interviewed in 2007 to 48,463 interviewed in 2018). Besides, it is also acknowledged by the authors due to the survey technique used (telephone interview) that introduces severe selection bias in the sample.

We appreciate your positive comments on our paper.
The estimates obtained from the samples in each Brazilian capital and the DF can be expanded to adults living in these cities without a landline telephone after using specific weighting factors (rake weighting method). This weighting factor makes it possible to match the sociodemographic composition of VIGITEL sample with the adult population residing in the Brazilian capitals and the Distrito Federal without landline telephone. However, the weighting performance is reduced in capitals from North and Northeast regions due to the lower landline telephone coverage, according to a previous study (Bernal et al., 2017). The addition of cell phone lines would reduce bias in estimates, mainly in North and Northeast, but this method has not yet been applied to VIGITEL. We have included information in methods and acknowledged this limitation in our discussion (page: 5,

paragraph 2; page 13, paragraph 3).

“the number of adults living in the household and the socio-demographic composition of the sample, based on the 2000 and 2010 demographic censuses. This weighting aimed to achieve representativeness for population aged 18 years and over of each state capital in Brazil, including DF in all years10, but it cannot be used as a representative sample of the whole country”

“VIGITEL survey collected data only from the population with landlines and included only the adults living in Brazilian capitals and the federal district. Despite using weighting measures for the general population, we would expect some small differences in the prevalence of our outcomes if we had assessed a sample that was not limited by landline access”

Reference:

Bernal Regina Tomie Ivata, Malta Deborah Carvalho, Claro Rafael Moreira, Monteiro Carlos Augusto. Effect of the inclusion of mobile phone interviews to Vigitel. Rev. Saúde Pública [Internet]. 2017 [cited 2021 Mar 07] ; 51(Suppl 1): 15s. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102017000200304&lng=en. Epub June 01, 2017. https://doi.org/10.1590/s1518-8787.2017051000171.

2) Therefore, I suggest not saying that the sample is representative, as every correction proposed (individual weighting for instance) is questionable and in my opinion, cannot give representativeness to this sample.

This sample weight (Rake weight) is part of the original VIGITEL design. It is based on the Brazilian demographic census and the postal codes of the telephone lines. As previously explained, the use of these sample weights approximates the sample's distribution from the source population of these cities, aiming representativeness. We included this information in methods - study design and source of data and sample-to clarify (page: 5, paragraph: 2). In addition, we have acknowledged that even using sample weights. Representativeness may be compromised.

3) Specifically,Page 4, Lines 11-12: Write “We used large National samples from Brazilian adults…”(replace representative by national)

We rewrote this sentence. We excluded the term “representative”, but we did not include “national”, because samples from Vigitel did not have national representativeness (page: 5, section: ‘Study design and source of data and sample’).

4) Page 6, lines 37-39: Delete the sentence “This weighting ensured the representativeness for the general adult population of each city in all years”

To make the information easier to understand, we changed the sentence to: “This weighting ensured the representativeness for the population aged 18 and over of each state capital in Brazil, including DF in all years. However, it cannot be used as a representative sample of the whole country” (page: 5, paragraph: 2).

5) Page 9, lines 32-36: unclear, please reformulate

We have reformulated the sentence (page: 9, paragraph: 1).

“The exception was smoking, where SII and CIX were higher in men.”

6) Pages 22-23, Figures 2-3: use a caption to explain the categories with an asterisk

We have reformulated the figures and removed the asterisks.

7) Finally, I recommend a stylistic review by a Native speaker

Suggestion accepted. The English of the entire manuscript has been revised.

VERSION 2 – REVIEW

REVIEWER	Strand, Bjørn Heine Norwegian Institute of Public Health
REVIEW RETURNED	11-Mar-2021

GENERAL COMMENTS	Thank you for the nice review and answers to my concerns. Most of my concerns were addressed and checked out. However, a couple of things remain. I still think analyses should be adjusted by sex, in addition to age – even if the distribution of sex is stable over time (shown in table 1). Why not do this, what is the catch? For single years, we see that there are more females than males (in 2007 53.2% vs 46.8%), it is also of interest to know whether the other characteristics in the table differs by sex, especially education. I strongly encourage the authors to split table 1 by sex. Table 2 should have a column with p-values for change from 2007 to 2018. Figure 1 and and 2 should be adjusted by sex in addition to age. The results section do not have any p-values of CIs.As commented on previously, in results, there are several claims of different increases/decreases in educational inequalities over time by sex and by skin color. However, these claims are not accompanied by formal testing statistically by the inclusion of interaction terms or p-values. Authors claim: “We have not included interaction terms using statistical tests because in Brazil there is a priori knowledge that health and access to health care are unequally distributed according to sex and skin color”. You should not rely on apriori knowledge, but rather investigate this and perform the necessary statistical analyses to back up your claims in the results section. Why bother doing the research, if you already apriori know the results? Minor: Page 8, line 14: Rather than presenting the educational difference in diabetes as a difference of 281.3%, the authors might want to

	report this solely as percentage point difference of 18.0 percentage points (24.4-6.4=18.0). Good luck with your review!

REVIEWER	Matranga, Domenica University of Palermo, Italy, Department of Health promotion Sciences and Mother- Child Care
REVIEW RETURNED	16-Mar-2021

GENERAL COMMENTS	I am very pleased with the changes that the authors made to the manuscript. I recommend it for publication on BMJ Open

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Dr. Bjørn Heine Strand, Norwegian Institute of Public Health

1) I still think analyses should be adjusted by sex, in addition to age – even if the distribution of sex is stable over time (shown in table 1). Why not do this, what is the catch? For single years, we see that there are more females than males (in 2007 53.2% vs 46.8%).

We appreciate your suggestion, and we apologize for not have being able to fully explain why we did not adjust our analyses for sex in the first round of revision. More importantly than the stability of sex over the years is that our aim was not to treat sex as a confounder variable; our aim was to assess if the educational inequality in our outcomes of interest differed by sex, considering sex as a variable of importance that could explain, rather than confuse, our research (please, see figure 3 and supplementary figure 1). We hypothesized that educational inequality would affect the sexes differently, and the adjustment for sex would have masked this difference.

In line with this explanation, a recent study published in BMJ Global Health (Shapiro, Klein & Morgan, 2021) explained the issues related to the adjustment by sex and gender in scientific papers. According to the authors, “controlling for sex or gender means treating these variables as confounding factors, rather than variables of importance to the research question. Technically, this usually means that a term was included in a regression model to account for the fact that sex, gender, or both, might influence the predictor and the outcome, and possibly confuse the relationship under investigation. While this allows for sex or gender differences in the outcome at baseline, it also forces this difference to be the same at all levels of the predictor”. The authors also reinforce that “instead of controlling for sex and gender—be it statistically or in the application of an intervention— we urge those working in global health to consider sex and gender as variables of importance that can explain, rather than confuse, their research. A first, and necessary, step is to disaggregate data to interrogate how sex and gender intersect with each other or with the predictors and outcomes under investigation. Disaggregation of data is a trigger for sex-responsive and gender-responsive research that allows for understanding how the true relationship between a predictor and outcome differs between males and females or among men, women and gender minorities. This avoids the pitfalls and unintended consequences of ignoring sex as a biological variable and gender as a social variable, and adds

richness and depth to the field of global health, which undoubtedly benefits the populations we serve". Therefore, treating sex and gender as confounding variables to be controlled, instead of significant sources of variation in the population, it is detrimental to the equitable improvement of global health.

In addition, due to the descriptive nature of inequality analyses, adjustments are not usual, nor appropriate (WHO, 2013).

References:
Shapiro JR, Klein SL, Morgan R. Stop 'controlling' for sex and gender in global health research. BMJ Global Health 2021;6:e005714.
WHO. Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. Geneva: WHO press, 2013:105.

2) It is also of interest to know whether the other characteristics in the table differs by sex, especially education. I strongly encourage the authors to split table 1 by sex.

Suggestion accepted. We have inserted this information at the new Supplementary Table 1 with mean age, education years, and prevalence of hypertension, diabetes, smoking and obesity by sex. We also presented this information according to skin color in this new table.

3) Table 2 should have a column with p-values for change from 2007 to 2018.

Suggestion accepted. We have included the p-values in Table 2 (Page: 21).

4) Figure 1 and and 2 should be adjusted by sex in addition to age.

Thanks for your suggestion, but as we have explained in Q1 we believe that the adjustment for sex would not be adequate to our analyses. Age, on the other hand, is a variable that could confound our analyses and we did not have a hypothesis of effect modification; therefore, we have standardized our analyses by age.

5) The results section do not have any p-values of CIs. As commented on previously, in results, there are several claims of different increases/decreases in educational inequalities over time by sex and by skin color. However, these claims are not accompanied by formal testing statistically by the inclusion of interaction terms or p-values. Authors claim: "We have not included interaction terms using statistical tests because in Brazil there is a priori knowledge that health and access to health care are unequally distributed according to sex and skin color". You should not rely on apriori knowledge, but rather investigate this and perform the necessary statistical analyses to back up your claims in the results section. Why bother doing the research, if you already apriori know the results?

Suggestion accepted. P-values inserted in the text. (Page: 7. Paragraph: 5; Page: 8. Paragraphs: 1 and 2; Page: 9. Paragraph: 1)

6) Page 8, line 14: Rather than presenting the educational difference in diabetes as a difference of 281.3%, the authors might want to report this solely as percentage point difference of 18.0 percentage points (24.4-6.4=18.0).

We appreciate your observation and we have added this information in the text (Page: 8. Paragraph: 2).

VERSION 3 – REVIEW

REVIEWER	Strand, Bjørn Heine Norwegian Institute of Public Health
REVIEW RETURNED	12-May-2021

GENERAL COMMENTS	I thank the authors for careful interpretation and response of all my comments. Congratulations with a very nice paper! I have no further comments.