

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Associations between neighbourhood social cohesion and subjective well-being in two different informal settlement types in Delhi, India: a quantitative cross-sectional study.
<b>AUTHORS</b>	Humble, Steve; Sharma, Aditya; Rangaraju, Baladevan; Dixon, Pauline; Pennington, Mark

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Kristjansson, Alfgeir West Virginia University, Department of Social and Behavioral Sciences
<b>REVIEW RETURNED</b>	07-Oct-2022

<b>GENERAL COMMENTS</b>	<p>BMJ Open review</p> <p>Thank you for the opportunity to review this manuscript. My comments follow the order of sections in the paper.</p> <p>Abstract</p> <p>Objectives: The order of IVs and DVs is reversed compared to the article title. Ideally these should align with one another.</p> <p>Results: Unsure if the correlations are bivariate or multivariate, including control variables.</p> <p>Introduction</p> <p>Overall a little unclear when the authors are referring to individual vs. collective levels, particularly in their discussion about neighbourhoods. Several times I found myself guessing if concepts were being discussed as holistic aspects of neighbourhoods or for individuals within them.</p> <p>P2L55-56: “..empower communities..” – perhaps this should be “empower individuals within communities...”</p> <p>P2L56-57: “Research has shown that neighbourhoods with higher levels of social cohesion can be beneficial to the well-being of their communities”. Perhaps this should end with “...beneficial to the well-being of their inhabitants” or similar.</p> <p>Methods</p> <p>Sample: A description of statistical power is provided. Neither this section nor the “Procedures” section however mention how many participants were accessed so neither completion rates nor response rates are included. I appreciate that the authors do mention that the study is based on a convenience sample, but how participants ended up being included vs. excluded in the sample is important. Presently the only assumption that can be drawn based on these sections is that new participants were approached until the power calculation numbers were acquired. This leads to questions about sample representativeness for the two neighbourhoods? Clarifying this is central to an assessment of if those findings can be assumed generalizable to these two neighbourhoods.</p> <p>Measures</p>
-------------------------	--

	<p>P5L25-26: "The Neighbourhood Cohesion Index (NCI) is used in this research to measure social cohesion with a focus on neighbourhood networks as well as causal interaction with neighbours". Unclear what "causal interaction" refers to.</p> <p>Procedures Please define "head of household". Please clarify process of data collection. The section mentions 18 interviewers but later that the survey was conducted online via Qualtrics. Please clarify how these are related. Unclear how confidentiality of participants was secured.</p> <p>Results P5 includes a section titled "Covariates", but it is not clear if the Results include any section where these were actually applied as covariates as opposed to merely testing the bivariate difference/similarities between the two neighbourhoods.</p>
--	--

<b>REVIEWER</b>	Oyebode, Oyinlola University of Warwick Warwick Medical School, Population, evidence and technology. Division of Health Sciences
<b>REVIEW RETURNED</b>	07-Nov-2022

<b>GENERAL COMMENTS</b>	<p>A very interesting topic area and rare data are presented in this manuscript.</p> <p>Unfortunately, the manuscript needs quite a bit of work before publication.</p> <p>The introduction is not very clear. For example: At one point the city is described as having been categorised into "planned" "special" or "unplanned" but then these terms are never explained or referred to again. It is hard to understand whether the following text somehow relates to these categories.</p> <p>"resettlement colonies are made up of families evicted from their squatter settlement by the SRA to randomly allocate housing"- it seems quite important that those reading this paper understand what is meant by *randomly* allocated housing. Does this mean each household could have ended up in any number of areas or houses? I wasn't sure whether the word random was really meant here.</p> <p>The paper mentions a convenience sample in several places, but then also says "households were selected by multi-stage random sampling, stratified on the population and geographic area". This doesn't make sense to me- is it convenience sampling or multi-stage random sampling? (Later in the results section there is no reporting of response rate, so I assume it really is convenience sampling- the full extent of the limitations of convenience sampling are not discussed, so if it is convenience sampling more text should be written about the biases involved here).</p> <p>Ethical approval in Newcastle UK is reported, but no local ethical approval is mentioned. Usually this is necessary, so would be good to know whether this has been left out, or wasn't actually sought or obtained (which would be worrying- Newcastle University should have made ethical approval conditional on obtaining local approval in my understanding).</p> <p>I didn't follow the sample size calculation but that doesn't mean it is wrong (hence ticking the need for statistical review from someone</p>
-------------------------	--

	<p>else to check that).</p> <p>There is a lack of information about the covariables (strange that for household occupation/employment the options suggest there was no-one who was unemployed. Can that be right?).</p> <p>Some findings are presented that are irrelevant to the aim of the paper (e.g.: these wealth indicators show positive associations with monthly income).</p> <p>Psychometric properties are suddenly introduced in the results section- these should have been mentioned in the methods section.</p> <p>There are many findings and some of them seem potentially contradictory so it would be good to think about how to present these more carefully.</p> <p>Many typographical errors (e.g.: NCI11 (NEI) I borrow things and exchange... is listed in the table twice; lots of grammar needs attention).</p> <p>I expect there are more errors and omissions in this paper than I have written about in this review, as it seemed necessary that the paper is re-written with improved structure and clarity of reporting which would help me to perform as second review more carefully if requested to look at the revision.</p>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Dr. Alfgeir Kristjansson, West Virginia University, Icelandic Center for Social Research and Analysis

Comments to the Author: BMJ Open review

Thank you for the opportunity to review this manuscript. My comments follow the order of sections in the paper.

Where in the document	Reviewer's comment	Authors' response and changes made
Abstract	Objectives: The order of IVs and DVs is reversed compared to the article title. Ideally these should align with one another.	The objectives are now in line with the title "To evaluate the relationship between neighbourhood cohesion and subjective well-being"
Abstract	Results: Unsure if the correlations are bivariate or multivariate, including control variables.	In the abstract under the results heading we have added 'bivariate'. Thank you for pointing this out.
Introduction	Overall a little unclear when the authors are referring to individual vs. collective levels, particularly in their discussion about neighbourhoods. Several times I found myself guessing if	We have made changes in the first paragraph to clarify that we are talking about individuals within communities rather than the community as a whole.

	<p>concepts where being discussed as holistic aspects of neighbourhoods or for individuals within them.</p> <p>P2L55-56: “..empower communities..” – perhaps this should be “empower individuals within communities...”</p> <p>P2L56-57: “Research has shown that neighbourhoods with higher levels of social cohesion can be beneficial to the well-being of their communities”. Perhaps this should end with “...beneficial to the well-being of their inhabitants” or similar.</p>	
Methods	<p>Sample: A description of statistical power is provided. Neither this section nor the “Procedures” section however mention how many participants were accessed so neither completion rates nor response rates are included. I appreciate that the authors do mention that the study is based on a convenience sample, but how participants ended up being included vs. excluded in the sample is important. Presently the only assumption that can be drawn based on these sections is that new participants were approached until the power calculation numbers were acquired. This leads to questions about sample representativeness for the two neighbourhoods? Clarifying this is central to an assessment of if those findings can be assumed generalizable to these two neighbourhoods.</p>	<p>We have added details on the response rate as follows in the section ‘sample size calculation and sampling techniques’.</p> <p>“In order to achieve the power calculation, 660 households were approached. In total 21 households did not agree to participate, with an overall response rate of 97% -94% and 99% in Sanjay Colony and Bhalswa respectively”</p>
Measures	<p>P5L25-26: “The Neighbourhood Cohesion Index (NCI) is used in this research to measure social cohesion with a focus on neighbourhood networks as well as causal interaction with neighbours”. Unclear what “causal interaction” refers to.</p>	<p>Under the heading Measures, we have changed this to the following:</p> <p>“The Neighbourhood Cohesion Index (NCI) is used in this research to measure social cohesion with a focus on neighbourhood networks and the degree of neighbourliness; that is the emotional social support within the neighbourhood which includes visiting neighbours and friendships”.</p>
Procedures	<p>Please define “head of household”.</p>	<p>Thank you. We have now made this consistent with Table 3 as the interviewers always interviewed “the main household wage earner”.</p>
Procedures	<p>Please clarify process of data collection. The section mentions 18 interviewers but later that the survey was conducted online via Qualtrics. Please clarify how</p>	<p>Ah, yes you are right. This looks a little confusing, actually the data administrators used Qualtrics to gather the data in real time. We have clarified this now in data processing and analysis section:</p>

	these are related.	“Data were collected by the administrators who inputted, in real time, the responses into Qualtrics during the household survey, which were then exported into Stata 17 for analysis”.
Results	Unclear how confidentiality of participants was secured.	We have changed the Informed consent section to clarify how we ensured confidentiality of the participants:  “Verbal informed consent was provided by participants who were willing to take part. All participants were informed before the start of the household survey that participation was voluntary and anonymous with no personal identifiable data captured and the results would be kept strictly confidential and for research purposes only. Data were transferred and stored securely at Newcastle University. No incentives were provided for participation”.
Results	P5 includes a section titled “Covariates”, but it is not clear if the Results include any section where these were actually applied as covariates as opposed to merely testing the bivariate difference/similarities between the two neighbourhoods	We have changed the ‘covariate’ title to ‘Socio-demographic characteristics’

Reviewer: 2 Dr. Oyinlola Oyeboode, University of Warwick Warwick Medical School

Comments to the Author: A very interesting topic area and rare data are presented in this manuscript.

Where in the document	Reviewer’s comment	Authors’ response and changes made
Introduction	The introduction is not very clear. For example: At one point the city is described as having been categorised into "planned" "special" or "unplanned" but then these terms are never explained or referred to again. It is hard to understand whether the following text somehow relates to these categories.	Thank you. We have added clarity in the text that Sanjay Colony would be classified as unplanned in the Delhi Master Plan and that Bhalswa would be classified as planned.  In the introduction we now say: “In this paper we investigate similarities and differences in neighbourhood social cohesion and well-being for households living in two different settlement types in Delhi - Sanjay Colony, Okhla Phase II a squatter settlement (unplanned) and Bhalswa a resettlement colony (planned)”.  In the discussion we now say: “This research considered two different informal settlement types in Delhi, India, where both communities were built on unauthorised land, with one spontaneously developed by individual families (Sanjay) and the

		<p>other 'planned' by the government to reallocate slum dwellers away from the city (Bhalswa)".</p> <p>AND</p> <p>"Part of the reason for this, which we cannot substantiate, may relate to the more cramped living conditions in Sanjay in comparison to those in the 'planned' resettlement community of Bhalswa."</p>
Introduction	<p>"resettlement colonies are made up of families evicted from their squatter settlement by the SRA to randomly allocate housing"- it seems quite important that those reading this paper understand what is meant by *randomly* allocated housing. Does this mean each household could have ended up in any number of areas or houses? I wasn't sure whether the word random was really meant here.</p>	<p>We have re-worded this whole sentence and taken out the word 'random' to clarify our meaning to:</p> <p>"Squatter settlements are unauthorised occupations of vacant land, mostly public, with minimum access to civic services and amenities. Resettlement colonies are made up of families 'evicted' from their original squatter settlement to plots allotted by the Slum Rehabilitation Authority (SRA). Resettlement colonies, reflect the systematic process of relocating poor residents to the periphery to facilitate the gentrification of urban spaces".</p>
	<p>The paper mentions a convenience sample in several places, but then also says "households were selected by multi-stage random sampling, stratified on the population and geographic area". This doesn't make sense to me- is it convenience sampling or multi-stage random sampling? (Later in the results section there is no reporting of response rate, so I assume it really is convenience sampling- the full extent of the limitations of convenience sampling are not discussed, so if it is convenience sampling more text should be written about the biases involved here).</p>	<p>Are, we are sorry that this wasn't clear.</p> <p>We used convenience sampling to select our <b>two</b> communities. That is Sanjay Colony, Okhla Phase II and Bhalswa owing to "our long-term relationships with those communities."</p> <p>In the 'Sample size calculation and sampling techniques' section we state:  "Sanjay Colony and Bhalswa were selected through convenience sampling owing to our long-term relationships with the communities in these areas".</p> <p>Then within each of these two locations we used multi-stage random sampling to select our household. And we state that in the same paragraph:  "Households were selected by multi-stage random sampling, stratified on the population and geographic area".</p> <p>We have now added the response</p>

		<p>rate in the same section:          “In order to achieve the power calculation, 660 households were approached. In total 21 households did not agree to participate, with an overall response rate of 97% - 94% and 99% in Sanjay Colony and Bhalswa respectively”.</p>
Ethical approval	<p>Ethical approval in Newcastle UK is reported, but no local ethical approval is mentioned. Usually this is necessary, so would be good to know whether this has been left out, or wasn't actually sought or obtained (which would be worrying- Newcastle University should have made ethical approval conditional on obtaining local approval in my understanding).</p>	<p>We have added information on local ethical approval in both the body of the paper and ethical approval section. Thank you for pointing this out:  <b>Ethical approval:</b> The study was approved by the ethics committee of Newcastle University (NCL: 12353/2020) and local community leaders' approval through Indus Information Initiatives (III), a registered social research data collection organisation, Delhi, India (IRB Certification protocol number of the head of data collection: 35478464).          AND  <b>“Informed consent</b>          Verbal informed consent was provided by participants who were willing to take part. All participants were informed before the start of the household survey that participation was voluntary and anonymous with no personal identifiable data captured and the results would be kept strictly confidential and for research purposes only. Data were transferred and stored securely at Newcastle University”.</p>
Sample size	<p>I didn't follow the sample size calculation but that doesn't mean it is wrong (hence ticking the need for statistical review from someone else to check that).</p>	<p>Please see Point 4 from reviewer 1, who says – ‘A description of statistical power is provided’.</p> <p>Our calculations for this are referenced in [65] - Fuller WA. <i>Sampling Statistics</i>. New Jersey, US: John Wiley and Sons 2009</p>
Results	<p>There is a lack of information about the covariables (strange that for household occupation/employment the options suggest there was no-one who was unemployed. Can that be right?).</p>	<p>We say in Table 2 that data were given on the main household wage earner. The caption for Table 2 is “Sociodemographic characteristics of main household wage earner in two settlements”</p> <p>Hence the table relates to the main household wage earner.</p> <p>We have clarified this in the</p>

		<p>procedures section where we have used the term 'the main household wage earner' rather than 'head of household'.</p>
Results	<p>Some findings are presented that are irrelevant to the aim of the paper (e.g.: these wealth indicators show positive associations with monthly income).</p>	<p>We are sorry that you find some findings 'irrelevant'. However, we have kept the wealth indicators and the income in the paper to illustrate the robustness of our data and the convergent validity of 'income' as a measure of wealth. We see here that costly wealth items (refrigerator, washing machine, scooter/motorcycle) correlate with income. We feel it is important when asking about income in a survey validation is needed for robustness (see for example Deaton, A (2018) <i>The Analysis of Household Surveys</i>, Washington DC: World Bank)</p> <p>"The average monthly income in Sanjay Colony was statistically significantly less at Rs. 16,681.70/- (£172.82 (£1=Rs.96.52/- conversion rate)) compared with Bhalswa at Rs. 18,935.98/- (£196.18). Monthly income was positively correlated with the household owning a refrigerator with a freezer (<math>r=0.280</math>, <math>p&lt;0.01</math>), washing machine (<math>r=0.331</math>, <math>p&lt;0.01</math>) and scooter/motorcycle (<math>r=0.367</math>, <math>p&lt;0.01</math>) in both communities. These wealth indicators show positive associations with monthly income."</p>
Results	<p>Psychometric properties are suddenly introduced in the results section- these should have been mentioned in the methods section.</p>	<p>We are giving results of both a pilot and the main study, and so we have reported the psychometric properties together in the results section.</p> <p>We are following the structure of other BMJ Open published article including Yu et al., (2018) (doi:10.1136/bmjopen-2018-023332) and Millar et al., (2021) PLOS ONE (doi:10.1371/journal.pone.0250336) where the psychometric properties are given in the results section only.</p>
Discussion	<p>There are many findings and some of them seem potentially contradictory so it would be good to think about how to present these more carefully.</p>	<p>We have completely rewritten the 'discussion', 'limitations' and 'conclusion' sections of the paper in order to give greater clarity as suggested. Many thanks for this.</p> <p>Please see the track changes in the document.</p>



	<p>Many typographical errors (e.g.: NCI11 (NEI) I borrow things and exchange... is listed in the table twice; lots of grammar needs attention) I expect there are more errors and omissions in this paper than I have written about in this review, as it seemed necessary that the paper is re-written with improved structure and clarity of reporting which would help me to perform as second review more carefully if requested to look at the revision.</p>	<p>Thank you. We have deleted the duplication of NCI 11. Our grammar has been checked and corrected as can be seen in track changes.</p>
--	---	--

**VERSION 2 – REVIEW**

<b>REVIEWER</b>	Kristjansson, Alfgeir West Virginia University, Department of Social and Behavioral Sciences
<b>REVIEW RETURNED</b>	23-Feb-2023

<b>GENERAL COMMENTS</b>	The authors have addressed all my comments to my satisfaction
-------------------------	---

<b>REVIEWER</b>	Oyebode, Oyinlola University of Warwick Warwick Medical School, Population, evidence and technology. Division of Health Sciences
<b>REVIEW RETURNED</b>	23-Feb-2023

<b>GENERAL COMMENTS</b>	<p>Thank you for thoughtfully addressing many of my concerns. This is much clearer than the first version and enjoyable to read. I have a few remaining concerns.</p> <p>Methods: Please can you give more detail on the sampling procedure. It is really helpful that you have more clearly described the convenience sampling of the two neighbourhoods and then multi-stage random sampling for the households but it would be good to understand what your sampling frame was. It is very surprising that you would have such high response rate when you were specifically surveying the main household wage-earner (who are often absent from the household for much of the day), this is worth mentioning in the discussion along with any factors that you think were relevant in achieving this high response rate (I assume fieldworkers went out at various times of day and made multiple visits to the randomly selected households). This is a strength of the study.</p> <p>Limitations: Given that you surveyed the main household wage earner, was your sample mainly men? You don't give the gender split in the participant characteristics. If it is mainly men, it would be worth mentioning this in the discussion section as associations between social cohesion and subjective well-being might be different between men and women.</p> <p>Some minor things Pilot: I appreciate that other papers have been published with the structure you have used- but personally if I was reporting the psychometric properties in the results section, I would include an aim relating to this in the introduction, and the methods used in the methods section. The editor can decide if that's necessary or not.</p> <p>Language relating to associations: You do state clearly that this is</p>
-------------------------	---

	<p>cross-sectional and you cannot make claims about causality but some of the statements do veer towards causality. For example "The longer a resident lived in the community, there was a greater sense of cohesion". Also you provide some discussion suggesting that "The longer the resident had lived in Bhalswa the greater the negative effect on their subjective wellbeing". Both of these associations could be due to causality running in the opposite direction e.g.: residents who feel there is a greater sense of cohesion (who fit in to the neighbourhood better) stick around for longer (feel less need to move than people who don't feel they fit in?); residents with poor wellbeing do not do so well at work which means they do not earn enough to allow them to leave for alternative accomodation and are forced to remain in Bhalswa (others may progress to better areas of the city more quickly?).</p> <p>Much clearer writing, especially the discussion- just some tiny minor typos that I thought I should pass on: "Well-being was also statically significantly correlated with employment..." "The average monthly income is Sanjay Colony was statistically significantly less..." " having chosen were to live..."</p>
--	--

### VERSION 2 – AUTHOR RESPONSE

Reviewer: 1 Dr. Alfgeir Kristjansson, West Virginia University, Icelandic Centre for Social Research and Analysis

Where	Reviewer's comment	Authors' response
	"The authors have addressed all my comments to my satisfaction"	Thank you!

Reviewer: 2 Dr. Oyinlola Oyebode, University of Warwick, Warwick Medical School

Where	Reviewer's comment	Authors' response & changes
Methods	Please can you give more detail on the sampling procedure. It is really helpful that you have more clearly described the convenience sampling of the two neighbourhoods and then multi-stage random sampling for the households but it would be good to understand what your sampling frame was. It is very surprising that you would have such high response rate when you were specifically surveying the main household wage-earner (who are often absent from the household for much of the day), this is worth mentioning in the discussion along with any factors that you think were relevant in achieving this high response rate (I assume fieldworkers went out at various times of day and made multiple visits to the randomly selected households). This is a strength of the study.	Thank you. We have made this clear that multiple visits were made to interview the main household wage earner (page 6) "The main household wage earner was interviewed by the survey administrators in a random sample of households. When the main household wage earner was not available a repeat visit was made at a time suitable to the resident. Where there was a nonresponse, the team moved onto the next 'available' household."
Limitations	Given that you surveyed the main household wage earner, was your sample mainly men? You don't give the gender split in the participant characteristics. If it is mainly men, it would be worth mentioning this in the discussion section as	Yes, you are correct. Thank you. We have put this in the limitations section "Finally, associations between social cohesion and subjective well-being may vary between men and women, one limitation of this study is that data were

	associations between social cohesion and subjective well-being might be different between men and women.	collected from the main household wage earner, who in the Indian context is typically male”.
Pilot (minor)	I appreciate that other papers have been published with the structure you have used- but personally if I was reporting the psychometric properties in the results section, I would include an aim relating to this in the introduction, and the methods used in the methods section. The editor can decide if that's necessary or not.	We have put in a sentence in the introduction as suggested “In the present study we evaluate the psychometric properties of the Neighbourhood Cohesion Index (NCI) and the SWB items initially through a pilot in Bangalore, India.” As well as in the methods section “Initially a pilot was carried out with 150 residents in Hawadigar Colony, Karnataka, India (Delhi being in COVID-19 lockdown in early 2022) to test the cross-cultural transferability of the survey. Hawadigar Colony is an unplanned squatter settlement made up of 308 households. Four researchers working in pairs interviewed the main household wage earner in a random sample of households. The psychometric properties of the NCI and SWB are reported in the results section”. Thank you.
Language	<p>Language relating to associations: You do state clearly that this is cross-sectional and you cannot make claims about causality but some of the statements do veer towards causality.</p> <p>For example "The longer a resident lived in the community, there was a greater sense of cohesion".</p> <p>Also you provide some discussion suggesting that</p> <p>"The longer the resident had lived in Bhalswa the greater the negative effect on their subjective wellbeing".</p> <p>Both of these associations could be due to causality running in the opposite direction e.g.: residents who feel there is a greater sense of cohesion (who fit in to the neighbourhood better) stick around for longer (feel less need to move than people who don't feel they fit in?); residents with poor wellbeing do not do so well at work which means they do not earn enough to allow them to leave for alternative accomodation and are forced to remain in Bhalswa (others may progress to better areas of the city more quickly?).</p> <p>Much clearer writing, especially the discussion- just some tiny minor typos that I thought I should pass on: “Well-being was also statically significantly correlated with</p>	<p>Thank you. We have added the alternative associations thus:</p> <p>“The longer a resident lived in the community there was a greater sense of cohesion. This could imply that residents who feel there is a greater sense of cohesion are more likely to remain in the neighbourhood”.</p> <p>And</p> <p>“the longer the resident had lived in Bhalswa, the greater the negative effect on their subjective well-being. Residents with poor subjective wellbeing may be those unable to leave owing to lower incomes and employment possibilities”.</p> <p>Thank you we have fixed these typos.</p>

	employment..." "The average monthly income in Sanjay Colony was statistically significantly less..." "having chosen were to live..."	
--	--	--