Johannesburg's 'poor housing, good health' paradox: the role of health status assessment, statistical modelling, residential context and migrant status

George TH Ellison[a,b]*and Thea De Wet[a]

[a]CAT, University of Johannesburg, Gauteng, South Africa; [b]CDI, Faculty of Science & Technology, University of Central Lancashire, Preston, UK.

*Corresponding author: George TH Ellison PhD DSc, Centre for Data Innovation, Faculty of Science & Technology, University of Central Lancashire, Preston PR1 2HE, UK; gthellison@uclan.ac.uk; +44 (0) 1772 201 201

Johannesburg's 'poor housing, good health' paradox: the role of health status assessment, statistical modelling, residential context and migrant status

Abstract:

Objectives - The 'poor housing, good health' paradox observed by De Wet et al. (2011) across eight of Johannesburg's poorest Wards (neighbourhoods), was re-examined using: a more generic measure of self-reported health; better-specified adjustment for measured confounders; household data from a range of Wards and income strata across Johannesburg and Gauteng; and adjustment for migrant status.

Study Design – The present study examined the relationship between (formal vs. informal) housing and self-reported health(-limited work and/or social activities) across four subsamples of respondents to the 2013 Quality of Life survey undertaken by the Gauteng City Regional Observatory: n=1,494 from households in the eight Wards examined by De Wet et al. (2011); n=3,059 from households with the lowest income tertile in Johannesburg; n=8,263 from households throughout Johannesburg; and n=24,727 from households throughout Gauteng Province, irrespective of Ward or income.

Methods - The relationship between housing and self-reported health in each of these subsamples was examined before and after adjustment for measured confounders identified using a temporally determined causal path diagram in the form of a directed acyclic graph.

Results - Following adjustment for measured confounders, 'informal housing' was only associated with 'good' self-reported health in Johannesburg's poorest Wards (OR:1.39; 95%CI:1.07,1.79) and Johannesburg as a whole (OR:1.15; 95%CI:1.00,1.31). These associations were substantially attenuated following adjustment for migrant status (OR:1.26; 95%CI:0.97,1.65; OR:1.07; 95%CI:0.93,1.22, respectively).

Conclusions - While the present study found that Johannesburg's 'poor housing, good health' paradox was still evident when using a more generic/subjective (self-reported) measure of health, the apparent paradox appears to reflect, at least in part: the differential aggregation of migrant-headed households in Johannesburg neighbourhoods exhibiting a high concentration of informal housing; and the likely impact of the health-related selection on the health of migrant-headed households.

Keywords: Johannesburg; Migration; Housing; Urban; Informal settlement; Self-reported Health

Introduction

Housing quality constitutes an important determinant of health, both because it (indirectly) reflects the availability of household resources to sustain/enhance health, and as a result of the (direct) risks poor quality housing can pose to health.[1,2] As such, various characteristics of housing quality (including: construction materials; permanency; service provision/amenities; and tenure) have been used as proxies for socioeconomic position in analyses exploring socioeconomic disparities in health.[3] These characteristics have also been used to assess the direct impact of housing on health.[4] However, because such characteristics capture/incorporate both *indirect* and *direct* pathways between housing and health, it can be challenging to separate the health effects of poverty *per se* from the effects of poor quality housing.[5,6]. This is because housing quality is often determined by socioeconomic position, and because the relation between socioeconomic position and health is often mediated by housing quality. For these reasons one might expect that any relationship between housing quality and health might predominantly reflect the broader impact of socioeconomic position on health; and that it might be safe to assume that households inhabiting poor quality housing would have worse health *primarily* as a result of their poorer socioeconomic position.

Nonetheless, there are at least three scenarios in which the relationship between housing quality and health might operate very differently to that of socioeconomic position. The first of these would occur where building regulations ensure that all dwellings attain a quality commensurate with good health; the second where the absence of suitable materials or infrastructure mean that no dwellings achieve the standard required to sustain good health; and the third where the social or structural determinants of poor quality housing are associated with good quality health. De Wet et al.'s[7] recent analyses of housing quality and health in eight of the poorest Wards within the City of Johannesburg appear to have identified an example of the third scenario, where respondents living in the poorest quality ('informal') housing were found to have better health than those living in better quality ('formal') dwellings. Indeed, since the Wards examined by De Wet et al.[7] displayed substantial variation in the prevalence of informal dwellings, they lacked the level of homogeneity in housing quality that would characterise either of the first or second scenarios.

De Wet et al.[7] speculated that their findings might reflect differential selection for (poor) housing and (good) health amongst recent migrants to the City of Johannesburg. Such a situation would occur were migrants to be subject to: *negative* discrimination in the housing market – as evident in the segregated nature of South African communities[8,9] and the impact of xenophobia on the exclusion of migrants from all

but the most marginalised neighbourhoods;[20] and *positive* selection for health – the so-called 'healthy migrant effect'.[10,11] However, De Wet et al.'s[7] study focussed exclusively on households in just eight of Johannesburg's poorest Wards, and used a composite measure of chronic ill-health which might have underestimated the prevalence of morbidity amongst younger, migrant-headed households. While the multivariable analytical models De Wet et al.[7] used adjusted for the age of the 'household head' and 'length of residence' in the current dwelling (which should have attenuated the impact of younger, migrant-headed households on the relation between housing quality and health), all of these models suffered from under-adjustment for measured/available confounders, and from inappropriate adjustment for likely mediators; while the use of a single model (in which all covariates are entered simultaneously) seriously undermines the interpretation of coefficient estimates for all but the last covariate, as a result of the so-called 'Table 2 fallacy'.[21]

The four aims of the present study were therefore to establish whether De Wet et al.'s[7] findings: (i) could be replicated using a measure of self-reported health status ('health-limited work/social activities') likely to be more sensitive to the health of younger, healthier individuals; (ii) were the result of incomplete/inappropriate adjustment for potential confounders/mediators; (iii) are also evident amongst households in a range of different urban and peri-urban contexts; and (iv) might be attenuated following adjustment for a more precise measure of migrant status.

Methods

The Gauteng City Regional Observatory's third Quality of Life household survey

The analyses described in the present study drew on data collected in the third Quality of Life (QoL-III) household survey undertaken by the Gauteng City Regional Observatory (GCRO), completed in July 2013 using a questionnaire containing more than 300 items within separate 'modules' (including those on housing and health).[12] These items included two relevant to the outcome of interest in the present study, worded: "How often, if ever, does your health status prevent you from doing your daily work?"; and "How often, if ever, does your health status prevent you from taking part in your usual social activities (physical and emotional problems)?" – both of which had four pre-coded answers: "Always"; "Some of the time"; "Hardly ever"; or "Never". For the analyses conducted in the present study, the answers to both of these items were combined into a single variable ('health-limited work/social activities') and re-coded as: 'good' (where the answer "Never" was given for *both* 'work' *and* 'social activities') vs. 'poor' (where the answers "Hardly ever"; "Some of the time"; or "Always" was given for *either* 'work' *or* 'social activities', *or* both).

The QoL-III questionnaire also contained suitable items relevant to the twelve discrete individual- and household-level characteristics that De Wet et al.[7] explored as potential determinants of health. For most of these the QoL-III items were directly comparable to the measures used by De Wet et al.[7] including: respondent age and gender; educational attainment; length of residence in the current dwelling; employment; housing tenure (rented vs. owned); number of persons per household; household services (toilet, water and electricity); and social participation. Finally, the QoL-III questionnaire contained two items relevant to the classification of dwellings as 'formal' vs. 'informal' – the first comprising an objective assessment based on interviewer observation; the second embedded within the answer categories offered when respondents were asked about their housing tenure. While the numbers of rooms/dwelling and the quality of services available was lower for dwellings classified as 'informal' using either the objective or respondent-derived items (see Table S1; online Supplementary Material) the present study used the interviewers' classification of 'formal'/'informal' on the grounds that this was more likely to provide a consistent and objective assessment of this key analytical characteristic.

In the absence of a strictly comparable item on 'social support' (which, in De Wet et al.'s[7] study, drew on an item worded "at least one person to count upon in the event of a problem"), the QoL-III item on 'social trust' ("Generally speaking, do you think that most people in your community can be trusted or that you need to be very careful when dealing with people in your community?") was used. Likewise, in the absence of the Household Food Security Access Scale,[7,13] a broadly comparable item in the QoL-III survey ("In the last year, has there ever been a time when you or any other adult in this household had to skip a meal because there was not enough money to buy food?") was used. Finally, detailed information on migration (both 'internal', from elsewhere in South Africa; and 'transnational', from outside South Africa) collected by a series of items within the QoL-III questionnaire was used to create a dedicated 'migrant status' variable capable of distinguishing between 'internal' and 'transnational' migrants, and between those 'internal' migrants who had moved to Gauteng from South Africa's most and least urbanised Provinces.[11,14]

Study Design

To establish whether De Wet et al.'s[7] 'poor housing, good health' paradox applied not only to residents in eight of the "most deprived" Wards within the City of Johannesburg[15] but also to: residents of the City's poorest households; and residents of households throughout the City, and Gauteng Province - four separate samples of the n=27,490 households included in the GCRO's QoL-III survey were identified: Sample 1 comprising households in the same eight Johannesburg Wards as those surveyed by De Wet et al.[7] (Ivory Park, Diepsloot, Riverlea, Doornkop, Phiri/Senoane, Alexandra, Jeppestown and Orange Farm);

Sample 2 comprising Johannesburg households in the lowest tertile of total/combined household income (≤ZAR3,200 per month; equivalent to approximately US$270 per month at average 2013 exchange rates); Sample 3 comprising households throughout Johannesburg (irrespective of Ward or income); and Sample 4 comprising households throughout the Province of Gauteng (i.e. including Johannesburg and elsewhere).

Statistical analysis

Descriptive analyses, summarising the frequencies and percentages of demographic, economic, household, psychosocial and health variables across all n=27,940 households included in the GCRO's QoL-III survey, and across households with complete data in Samples 1-4, were used to assess the comparability of households in each of these samples. The two multivariable logistic regression models used by De Wet et al.[7] to examine the relationship between 'informal' housing and health were then replicated using data from households in the same eight Wards within the City of Johannesburg (i.e. Sample 1). In each of these models the different covariate adjustment sets selected from the 12 available/measured covariates were entered *simultaneously* as potential predictors/determinants of self-reported health status (i.e. exactly as conducted by De Wet et al.[7] – an approach that undermines the causal interpretation of coefficient estimates for all but the last covariate as a result of the 'Table 2 fallacy'.[21] Additional multivariable logistic regression analyses were therefore performed with each of the 12 key variables entered *sequentially*, using data from Sample 1 and the three remaining samples (i.e. Samples 2, 3 and 4). This (sequential) approach permits interpretation of the (adjusted) relationships between each of available covariates and self-reported health, because a separate model is used when each covariate is specified as the 'exposure' of interest to ensure that these models only include adjustment for other covariates that act as potential confounders for that 'exposure'.[21] The identification of potential confounders in each of these models was determined using a forward saturated directed acyclic graph (DAG; see Figure 1), which drew on the likely temporal relationships between the 12 covariates examined (as well as migrant status), such that variables denoting events or characteristics occurring at any given point in time were assumed to act as putative, *probabilistic causes* of any subsequent variables (subject to empirical confirmation). [16,17,22] In this way, Figure 1 represents the authors' theoretical assumptions regarding the likely, and dominant, temporal causal sequence of variables as a reflection of the underlying (but unknown and unknowable) 'data generating mechanism' which, for example, recognises that the causal relationship between employment and health *can* operate in either direction but *in this instance* was considered most likely to reflect the impact of *established* employment on self-reported health (as assessed *on the day of the interview*).[22] A key advantage of using DAGs in this fashion is that it facilitates the selection of appropriate covariate

adjustment sets (i.e. those containing potential confounders, but not likely mediators)[22] which, in the present study, was verified using the online tool www.dagitty.net.[18,19]

Results

Table 1 compares the demographic, economic, household, psychosocial and self-reported health status of all n=27,940 respondents to the GCRO's QoL-III survey with those providing complete data on self-reported health and each of the covariates from households in the four Samples examined in the present study. Of these, n=1,494 were from households in eight of Johannesburg's poorest Wards (i.e. the same eight Wards examined by De Wet et al. 2011[7]); n=3,059 were from Johannesburg households with the lowest tertile of household income (Sample 2); n=8,263 were from households throughout Johannesburg (regardless of Ward or income; Sample 3); and n=24,727 were from households across Gauteng Province (i.e. within Johannesburg and beyond; Sample 4). A comparison of all n=27,940 QoL-III respondents with those in Sample 4 (i.e. all QoL-III survey respondents providing complete data) suggests that these were very similar in terms of the distribution of self-reported health and each of the covariates examined. As such, it is unlikely that the slightly lower number of respondents with complete data on these variables (i.e. n=24,727 or 88,5% of the n=27940 QoL-III respondents) were unrepresentative of the sample intended by the GCRO.

Respondents from Samples 1, 2 and 3 displayed very similar rates of: residential mobility (only 11.2-13.1% having moved to their current dwelling within the past 18 months); social participation (58.0-59.6% participating in at least one club/organization); and social trust (only 15.8-18.1% responding that 'most people can be trusted') to those sampled from Gauteng Province as a whole (i.e. Sample 4). However, in many other respects these four Samples displayed considerable variation in sociodemographic, economic and social characteristics. For example, respondents in Sample 1 were: the youngest (18-25yrs: 23.9%) and least old (>45yrs: 27.1%) of those in any of the four Samples. Sample 1 also had the highest proportion of respondents who were internal migrants (44.1%) and the lowest proportion who were transnational migrants (9.0%). Sample 1 respondents also had the highest rates of unemployment (62.8%); the second highest prevalence of 'skipping a meal' in the past year (19.1%); yet the second highest prevalence of reporting 'good' health (53.9%). However, the dwellings occupied by respondents in Sample 1 were most likely to: be owned (68.7%) and 'informal' (24.6%); house four or more people (51.7%); and have access to two or fewer key services (i.e. water, electricity or a toilet). In contrast, respondents in Sample 2 were the oldest (>45yrs: 35.5%) and least young (18-25yrs: 17.1%) of those in any of the four Samples examined;

and were most likely to be female (62.2%); and least likely to have completed Grade 8 or beyond (79.3%). Sample 2 also had the highest percentage of transnational migrants (14.0%); the highest prevalence of respondents who reported 'skipping a meal' in the past year (19.3%); and the lowest prevalence of self-reported 'good' health (47.4%). The dwellings occupied by Sample 2 respondents were least likely to be owned (55.1%); though they were also least likely to house four or more people (44.1%).

Respondents from households from Sample 3 were: least likely to be female (55.9%); most likely to have completed Grade 8, or above (87.8%); least likely to be internal migrants (34.2%); least likely to be unemployed (53.3%); and least likely to report having 'skipped a meal' in the past year (13.0%). The dwellings occupied by respondents in Sample 3 were: least likely to be 'informal' (12.9%); and most likely to have access to two or more key household services (87.1%). However, these respondents also had the second lowest prevalence of self-reported 'Good health' (51.6%). In comparison, respondents from Sample 4 had the highest self-reported 'good' health (55.0%) of all four Samples examined; and were least likely to be migrants (i.e. non-Gauteng 'residents': 55.0%).

These differences highlight substantial heterogeneity within and between each of the Samples examined in the present study, with differences in living conditions and in sociodemographic and economic characteristics likely to create complex relationships with self-reported health. Indeed, since age, sex, migration status, educational attainment and employment can all play a role in both the quality of housing available to, and the quality of health experienced by, respondents in the present study (see Figure 1), all of these covariates are likely to act as potential confounders in any relationship between housing and health. At the same time, because the type of housing (i.e. 'formal' vs. 'informal') is likely to determine the numbers of people housed therein, the household services available, and the contexts within which social participation and trust develop, these characteristics are all likely to fall on the causal pathway between type of housing and self-reported health (i.e. as 'mediators' within Figure 1). Since the exclusion of potential confounders, and the inclusion of likely mediators, in the covariate adjustment sets used by multivariable statistical models can both lead to biased causal inference,[17] it is important to specify these adjustment sets accordingly.

This does not seem to have been the case in the two multivariable models used by De Wet et al.[7] to examine the relationship between 'formal'/'informal' housing and health. Both of these failed to adjust for at least one of the available/measured confounders (education and/or housing tenure), and both adjusted for at least one likely mediator (household services; social participation and trust). To assess the potential impact of these covariate adjustment sets on the relationship observed between type of housing ('formal'

8

vs. 'informal') and health, these have been applied in multivariable analyses of data from Sample 1 (comprising respondents living in the same eight Wards as those surveyed by De Wet et al.[7]) and compared to appropriately adjusted models in Table 2 (see Models 1, 2 and 3, respectively). These analyses suggest there is no relationship between 'formal' housing and health-limited work/social activities (OR:1.03; 95%CI:0.77,1.39) when using the first of De Wet et al.'s[7] models, which lacked adjustment for educational attainment (a potential confounder) but included adjustment for household services (a likely mediator; see Table 2, Model 1). When using the second of De Wet et al.'s[7] models, which failed to adjust for two potential confounders (educational attainment and housing tenure), but adjusted for two likely mediators (social participation and social trust), there is a strong relationship between 'formal' housing and health-limited work/social activities (OR:1.33; 95%CI:1.03,1.72; see Table 2, Model 2). However, this relationship is strengthened further in the revised model that includes adjustment for educational attainment (and all other available/measured confounders) but does not adjust for any likely mediators (OR:1.39; 95%CI:1.07,1.79; see Table 2, Model 3). As such, Model 3 confirms that 'informal' housing is associated with 'good' health amongst respondents from households in the eight Wards examined by De Wet et al. 2011,[7] and that this relationship: (i) applies to both composite measures of chronic health (as used by De Wet et al.[7]) and self-reported health (as used in the present study); and (i) does not appear to be an artefact of inappropriate statistical adjustment.

No such relationship between 'informal' housing and 'good' self-reported is evident in comparable analyses of data from respondents in Sample 2 or Sample 4 (see Table 3). However, there was a relationship between 'formal' housing and health-limited work/social activities amongst respondents in Sample 3, although this was less strong than that observed in Sample 1 (OR:1.15; 95%CI:1.00,1.31). Nonetheless, this relationship, and that observed amongst respondents in Sample 1, was somewhat attenuated after additional adjustment for migrant status (Sample 1: OR:1.26; 95%CI:0.97,1.65; see Table 2, Model 4; Sample 3: OR:1.07; 95%CI:0.93,1.22; see Table 3). This suggests that the relationship between 'informal' housing and 'good' health observed in Samples 1 and 3 partly reflects confounding for migrant status (as a result of the greater proportion of migrants living in 'informal' housing, and the better health of migrants as a result of the 'healthy migrant' effect) – as De Wet et al.[7] proposed.

Finally, as a result of the sequential nature of the multivariable models summarised in Table 2 (Models 3 and 4) and Table 3, it is also possible to assess the role that each of the covariates might play in health-limited work/social activities (i.e. as 'exposures' of interest), after adjustment for other covariates identified as potential confounders using the DAG presented in Figure 1. From the adjusted effect estimates provided for each of these covariates it is clear that many operate in a similar fashion regardless

of the Sample examined. For example: older, female respondents were more likely to report health-limited work/social activities; as were those participating in at least one club/organisation and those reporting that they needed to "to be very careful when dealing with people" in their community. In contrast, respondents who were (internal or transnational) migrants, those who had recently moved to their current dwelling (within the past 18 months), and those who were employed, were all *less* likely to report health-limited work/social activities. And while household tenure ('owned' vs. 'rented') and the number of people in the household appeared unrelated to self-reported health, higher educational attainment was associated with 'Poor' health amongst respondents from Johannesburg's poorest households, but with 'good' health amongst respondents throughout Gauteng as a whole; and respondents who reported they had 'skipped a meal' in the past year were more likely to report health-limited work/social activities, but only in Gauteng as a whole. To some extent these relationships help explain the 'poor housing, good health' paradox observed in Samples 1 and 3, as these had a high prevalence of internal and transnational migrants, respectively, and both migrant groups were less likely to report health-limited work/social activities than Gauteng residents.

Discussion

The results of the analyses summarised in Table 2 clearly indicate that the 'poor-housing, good health' paradox was still evident when using a more generic (and potentially sensitive) measure of health and, in this respect, De Wet et al.'s[7] paradox appears insensitive to/unaffected by the health status measure each study used. Separate sensitivity analyses (see Tables S2 and S3; online Supplementary Material) also confirmed that these findings were largely unchanged after repeating these using each of the two different components of self-reported health (i.e. health-limited work; and health-limited social activities). At the same time, Table 2 demonstrates that inappropriate covariate adjustment in the statistical models used by De Wet et al.[7] (see Table 2, Models 1-2) was not responsible for their apparent paradox, since this was also found following DAG-led adjustment for all potential confounders (but not for likely mediators; see Table 2, Model 3).

However, the analyses summarised in Table 3 indicate that the 'poor housing, good health' paradox appears context-specific. For the three pairs of analyses presented in Table 3 there was only evidence of a relationship between 'informal' (i.e. poor) housing and health-limited work/social activities amongst households in Sample 3. This relationship was not observed in analyses of data from Sample 2 or Sample 4. Moreover, even amongst households in Sample 3, the relationship between 'good housing' and 'poor

health' was far less pronounced (OR: 1.15; 95%CI: 1.00,1.31) than that observed amongst households within 8 of Johannesburg's poorest Wards (Sample 1; OR: 1.39; 95%CI: 1.07,1.79).

As such, it appears that the 'poor housing, good health' paradox observed by De Wet et al.[7] is restricted to those neighbourhoods where the determinants of *poor* housing also convey/reflect *good* health. An important factor in this, as suggested by De Wet et al.[7] appears to be migrant status, since both of the contexts in which the 'poor housing, good health' paradox was evident in the present study, had a high prevalence of internal (Sample 1) and/or transnational migrants (Sample 3); and additional adjustment for 'migrant status' attenuated the relationship between informal ('poor') housing and self-reported ('good') health in both Samples 1 and 3 (see Table 2, Model 4; and Table 3, respectively). Such attenuation is consistent with a disproportionate number of migrants inhabiting (poorer quality) informal housing yet retaining better self-reported health due, even if only in part, to the legacy of health selection at migration (the 'healthy migrant' effect).[10,11] This constitutes an important consideration when evaluating alternative public health policies to support the health and housing needs of migrants and non-migrants alike in the heterogeneous and rapidly growing neighbourhoods of South Africa's largest urban conurbation. However, it is also important to acknowledge that the relationship between good (self-reported) health and poor (informal) housing remained strong in those contexts where this was found, *even after* adjustment for migrant status and length of residence in the current dwelling – suggesting that imprecision in the classification of migrant status or other health-related characteristics are responsible for much of the paradoxical relationship observed.

Compliance with ethical standards

References

1. Mitchell A, Macció J. *Housing and wellbeing: evidence from the informal settlements of Buenos Aires*. Documento de Trabajo No. 52. Facultad de Ciencias Económicas. Departamento de Investigación

"Francisco Valsecchi"; 2015. Available from:
http://bibliotecadigital.uca.edu.ar/repositorio/investigacion/housing-wellbeing-mitchell-maccio.pdf; 2015

2. Satterthwaite D. The links between poverty and the environment in urban areas of Africa, Asia, and Latin America. *Ann Am Acad Polit SS* 2003;**590:**73-92. doi: 10.1177/0002716203257095

3. Shaw M. Housing and public health. *Ann Rev Public Health* 2004;**25:**397-418. doi: 0.1146/annurev.publhealth.25.101802.123036

4. Moore M, Gould P, Keary BS. Global urbanization and impact on health. *Int J Hyg Envir Heal* 2003;**206:**269-278. doi: 10.1078/1438-4639-00223

5. Thomson H, Petticrew M, Morrison D. Health effects of housing improvement: systematic review of intervention studies. *Brit Med J* 2001;**323:**187-90. doi: 10.1136/bmj.323.7306.187

6. Thomson H, Thomas S, Sellstron E, Petticrew M. The health impacts of housing improvement: a systematic review of intervention studies from 1887 to 2007. *Am J Publ Health* 2009;**99:**S681-692. doi: 10.2105/AJPH.2008.143909

7. De Wet T, Plagerson S, Harpham T, Mathee A. Poor housing, good health: a comparison of formal and informal housing in Johannesburg, South Africa. *Int J Publ Health* 2011;**56:**625–633. doi: 10.1007/s00038-011-0269-1

8. Christopher AJ. The slow pace of desegregation in South African cities, 1996-2001. Urban Stud 2005;**42:**2305-2320. doi: 10.1080/00420980500332122

9. Turok I. *Urbanisation and development in South Africa: Economic imperatives, spatial distortions and strategic responses*. Urbanisation and Emerging Population Issues Working Paper 8. Human Settlements Group, International Institute for Environment and Development (IIED), London; 2012. ISBN: 978-1-84369-890-6. Available from: https://pdfs.semanticscholar.org/4285/c890b1bd847ce44902ff6bf0af04dbf5a2b7.pdf?_ga=2.23561 018.1064485027.1553642227-250884056.1549389125; 2012.

10. Pablos-Méndez A. Mortality among Hispanics. *J Am Med Ass* 1994;**271:**1237-1238. doi: 10.1001/jama.1994.03510400023017

11. Ellison GTH, De Wet T. Poverty, disability and self-reported health amongst residents and migrants in Gauteng, South Africa. Ann Hum Biol 2016;**43:**131-143. doi: 10.3109/03014460.2016.1147597

12. Gauteng City Regional Observatory (GCRO). *The Gauteng City-Region*. Available at: www.gcro.ac.za; 2015.

13. Coates J, Swindale A, Bilinsky P. *Household food insecurity access scale (HFIAS) for measurement of household food access: indicator guide.* Food and nutrition technical assistance project. Academy

for Educational Development, Washington; 2006. Available at: http://www.fao.org/fileadmin/user_upload/eufao-fsi4dm/doc-training/hfias.pdf; 2006

14. Prinsloo DA. *Urbanisation and the impact on future shopping centre development in Africa and South Africa*; 2014. Available at: http://www.urbanstudies.co.za/wp-content/uploads/2014/08/2662-SAC-Urbanisation-Report-Sept-2014-T3-2.pdf; 2014.

15. Noble M, Babita M, Barnes H, Dibben C, Magasela W, Noble S, Ntshongwana P, Phillips H, Rama S, Roberts B, Wright G, Zungu S. *The provincial indices of multiple deprivation for South Africa 2001*. Technical Report. University of Oxford, UK; 2006. Available at: http://www.statssa.gov.za/census/census_2001/PMID/PIMDTechnicalReport2006.pdf; 2006.

16. Tennant PWG, Arnold K, Berrie L, Ellison GTH, Gilthorpe MS. *Advanced modelling strategies: Challenges and pitfalls in robust causal inference with observational data.* Leeds Institute for Data Analytics (LIDA), Leeds; 2017. ISBN: 978-1-5272-1208-4

17. Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GTH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol* 2016;**45:**1887-1894.

18. Textor J, Hardt J, Knüppel S. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiol* 2011;**22:**745.

19. Textor J. *Drawing and analyzing causal DAGs with DAGitty*. Available at: http://www.dagitty.net/manual-2.x.pdf; 2015.

20. Robinson D, Reeve K. *Neighbourhood experiences of new immigration: reflections from the evidence base*. York: Joseph Rowntree Foundation, York; 2006. ISBN: 13: 978 1 85935 441 4. Available at: https://pdfs.semanticscholar.org/b654/b8e6c73a71056d14717e8ae53573b9726213.pdf; 2006.

21. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol* 2013; **177:** 292-8.

22. Tennant PWG, Harrison WJ, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, Keeble C, Ranker LR, Textor J, Tomova GD, Gilthorpe MS, Ellison GTH. Use of directed acyclic graphs (DAGs) in applied health research: review and recommendations. *medRxiv* 2020; doi: 10.1101/2019.12.20.19015511

Figure 1. Hypothesised temporal and causal relationships between the demographic, economic, household and psychosocial variables relevant to the association between housing quality and health in Johannesburg and Gauteng (represented in the form of a Directed Acyclic Graph and published online: dagitty.net/mJgAmkZ).[18,19] The nodes (variables) along the base of the DAG (numbered C_1-C_{12}) indicate where latent (unmeasured) confounders may lie.
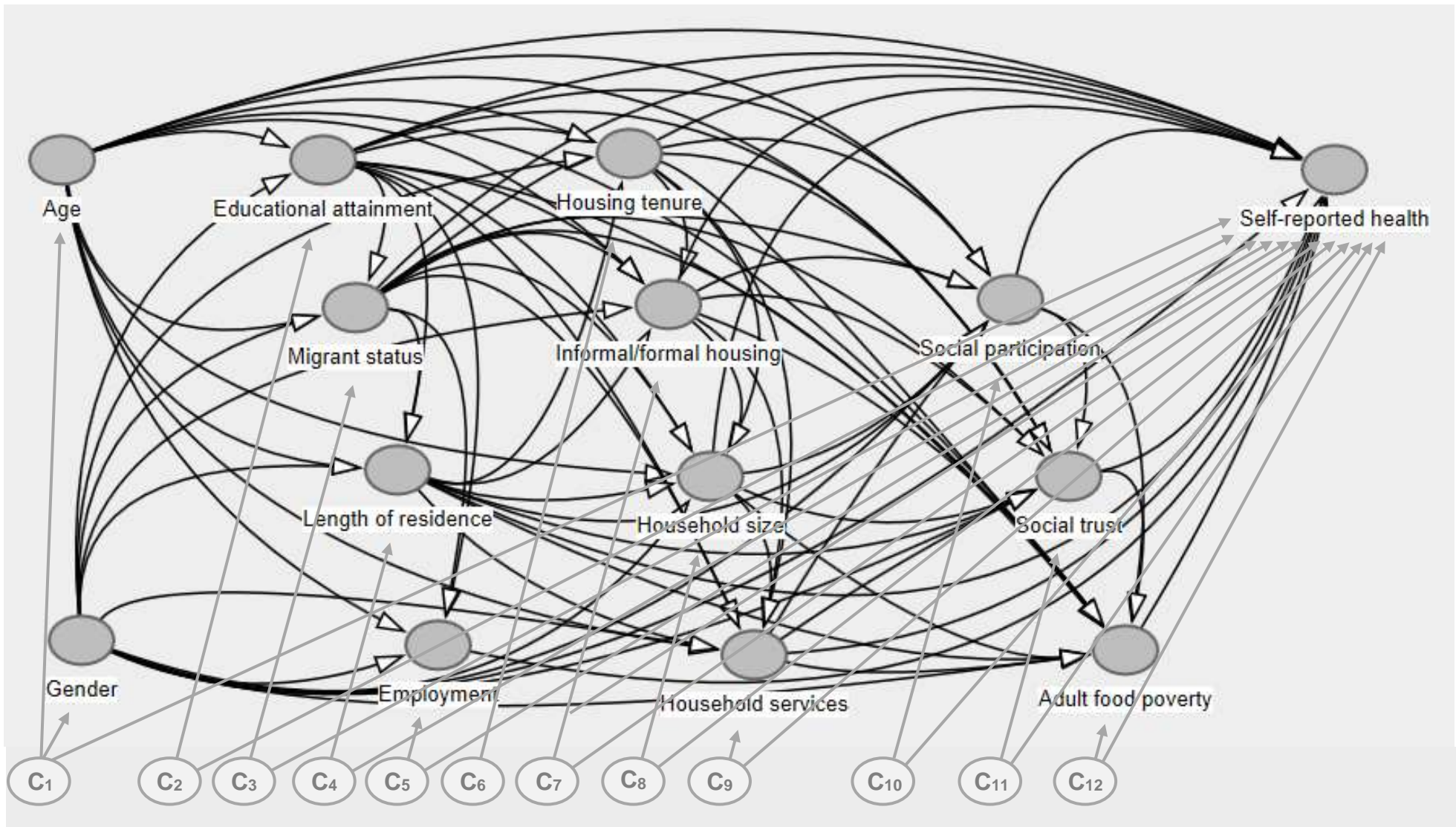
Table 1. A comparison of the demographic, economic, household, psychosocial and health characteristics of all respondents to the GCRO QoLIII Survey (including respondents with missing data for one or more variable); and respondents with complete data on all variables living: in eight of the poorest Wards within the City of Johannesburg (Sample 1); in households with the lowest tertile of household income within the City of Johannesburg (Sample 2); throughout the City of Johannesburg (Sample 3); and throughout Gauteng Province (Sample 4).

Sample:	GCRO QoLIII	Sample 1	Sample 2	Sample 3	Sample 4
Households included in the model:	Gauteng Province	City of Johannesburg	City of Johannesburg	City Johannesburg	Gauteng Province
	all income strata	8 of the poorest Wards	lowest income tertile	all income strata	all income strata
	n=27,490	n=1,494	n=3,059	n=8,263	n=24,727
Missingness:	Incomplete data	Complete data	Complete data	Complete data	Complete data
Characteristic	n (%)	n (%)	n (%)	n (%)	n (%)
Age					
18-25yrs	5,478 (20.0)	357 (23.9)	522 (17.1)	1,608 (19.5)	5,011 (20.3)
26-35yrs	7,461 (27.2)	417 (27.9)	829 (27.1)	2,306 (27.9)	6,681 (27.0)
36-45yrs	5,768 (21.0)	315 (21.1)	625 (20.4)	1,825 (22.1)	5,190 (21.0)
46-60yrs	5,573 (20.3)	289 (19.3)	644 (21.1)	1,662 (20.1)	5,002 (20.2)
>60yrs	3,174 (11.6)	116 (7.8)	439 (14.4)	862 (10.4)	2,843 (11.5)
Missing	36	-	-	-	-
Gender					
Male	11,835 (43.1)	603 (40.4)	1,156 (37.8)	3,644 (44.1)	10,558 (42.7)
Female	15,655 (57.0)	891 (59.6)	1,903 (62.2)	4,619 (55.9)	14,169 (57.3)
Missing	0	-	-	-	-
Educational attainment					
Primary or lower (<Grade 8)	4,016 (14.8)	242 (16.2)	634 (20.7)	1,012 (12.3)	3,553 (14.4)
Secondary or higher (≥Grade 8)	23,142 (85.2)	1,252 (83.8)	2,425 (79.3)	7,251 (87.8)	21,174 (85.6)
Missing	332	-	-	-	-
Migrant status					
Gauteng resident	14,805 (53.9)	701 (46.9)	1,525 (49.9)	4,382 (53.0)	13,598 (55.0)
Internal (Urban)	2,458 (8.9)	136 (9.1)	265 (8.7)	644 (7.8)	2,163 (8.8)
Internal (Rural)	7,511 (27.3)	523 (35.0)	840 (27.5)	2,180 (26.4)	6,649 (26.9)
Transnational	2,716 (9.9)	134 (9.0)	429 (14.0)	1,057 (12.8)	2,317 (9.4)
Missing	0	-	-	-	-
Length of residence in current dwelling					
≥18 months	23,802 (86.6)	1,326 (88.8)	2,676 (87.5)	7,221 (87.4)	21,480 (86.9)
<18 months	3,688 (13.4)	168 (11.2)	383 (12.5)	1,042 (12.6)	3,247 (13.1)

Missing	0	-	-	-	-
Employment					
Unemployed	15,815 (57.5)	938 (62.8)	1,843 (60.3)	4,407 (53.3)	14,318 (57.9)
Employed	11,675 (42.5)	556 (37.2)	1,216 (39.8)	3,856 (46.7)	10,409 (42.1)
Missing	0	-	-	-	-
Housing tenure					
Rented	9,233 (33.6)	468 (31.3)	1,374 (44.9)	3,311 (40.1)	8,000 (32.4)
Owned	18,257 (66.4)	1,026 (68.7)	1,685 (55.1)	4,952 (59.9)	16,727 (67.7)
Missing	0	-	-	-	-
Housing type					
Informal	4,055 (15.1)	367 (24.6)	604 (19.8)	1,067 (12.9)	3,750 (15.2)
Formal	22,875 (84.9)	1,127 (75.4)	2,455 (80.3)	7,196 (87.1)	20,977 (84.8)
Missing	560	-	-	-	-
Number of people in the household					
1-3 people	13,910 (50.6)	722 (48.3)	1,709 (55.9)	4,129 (50.0)	12,280 (49.7)
Four or more people	13,580 (49.4)	772 (51.7)	1,350 (44.1)	4,134 (50.0)	12,447 (50.3)
Missing	0	-	-	-	-
Household services					
Two or less	4,172 (15.2)	307 (20.6)	609 (19.9)	1,066 (12.9)	3,565 (14.4)
Water, electricity and toilet	23,318 (84.8)	1,187 (79.5)	2,450 (80.1)	7,197 (87.1)	21,162 (85.6)
Missing	0	-	-	-	-
Social participation					
No participation	11,090 (40.3)	603 (40.4)	1,284 (42.0)	3,447 (41.7)	9,818 (39.7)
Participation in at least one club/organisation	16,400 (59.7)	891 (59.6)	1,775 (58.0)	4,816 (58.3)	14,909 (60.3)
Missing	0	-	-	-	-
Social trust					
Most people can be trusted	4,636 (18.1)	245 (16.4)	483 (15.8)	1,457 (17.6)	4,470 (18.1)
Need to be very careful	20,939 (81.9)	1,249 (83.6)	2,576 (84.2)	6,806 (82.4)	20,257 (81.9)
Missing/Don't know	1,915	-	-	-	-
Household adult food poverty					
Not skipped a meal in the past year	23,400 (85.1)	1,209 (80.9)	2,468 (80.7)	7,191 (87.0)	21,042 (85.1)
Skipped a meal in the past year	4,090 (14.9)	285 (19.1)	591 (19.3)	1,072 (13.0)	3,685 (14.9)
Missing	0	-	-	-	-
Self-reported health-limited work and/or social activities					
None	15,170 (55.2)	805 (53.9)	1,451 (47.4)	4,262 (51.6)	13,601 (55.0)

16

One or both	12,320 (44.8)	689 (46.1)	1,608 (52.6)	4,001 (48.4)	11,126 (45.0)
Missing	0	-	-	-	-

Table 2. Multivariable logistic regression models exploring the relationship between a range of demographic, economic, household, psychosocial characteristics and health-limited work and/or social activities amongst n=1,494 households in eight of the poorest Wards of the City of Johannesburg (after De Wet et al.[7]; Sample 1). Models 1-2 mimic the analytical models used by De Wet et al.[7], in which all covariates were included simultaneously in a single step; while Models 3 and 4 includes covariates entered one-by-one in a sequential fashion[1] as determined by the DAG (see Figure 1; Model 4 including additional adjustment for migrant status).

Covariates entered: Adjustment for migrant status: Characteristic (referent)	Model 1 Simultaneously No OR(95%CI)	Model 2 Simultaneously No OR(95%CI)	Model 3 Sequentially No OR(95%CI)[1]	Model 4 Sequentially Yes OR(95%CI)[1]
Age (18-25yrs)				
26-35yrs	1.33 (0.98,1.81)[3]	1.31 (0.97,1.78)[3]	1.32 (0.98,1.77)	1.32 (0.98,1.77)
36-45yrs	1.90 (1.37,2.63)[3]	1.91 (1.38,2.64)[3]	1.96 (1.43,2.68)	1.96 (1.43,2.68)
46-60yrs	2.37 (1.70,3.30)[3]	2.38 (1.71,3.33)[3]	2.57 (1.86,3.55)	2.57 (1.86,3.55)
>60yrs	5.38 (3.32,8.70)[3]	5.46 (3.37,8.84)[3]	5.81 (3.62,9.32)	5.81 (3.62,9.32)
Gender (Male)				
Female	1.62 (1.29,2.02)[3]	1.56 (1.25,1.95)[3]	1.58 (1.28,1.97)	1.58 (1.28,1.97)
Educational attainment (Primary or lower; <Grade 8)				
Secondary or higher	-	-	1.00 (0.73,1.37)	1.00 (0.73,1.37)
Migrant status (Gauteng resident)				
Internal (most urban)	-	-	-	0.92 (0.63,1.36)
Internal (least urban)	-	-	-	0.68 (0.54,0.86)
Transnational	-	-	-	0.54 (0.36,0.81)
Length of residence in current dwelling (≥18 months)				
<18 months	0.92 (0.65,1.30)[2,3]	0.88 (0.63,1.25)[2,3]	0.87 (0.62,1.22)	0.98 (0.69,1.39)
Employment (Unemployed)				
Employed	1.06 (0.84,1.34)[2,3]	1.05 (0.83,1.33)[2,3]	1.08 (0.85,1.36)	1.11 (0.88,1.40)
Housing tenure (Rented)				
Owned	1.10 (0.87,1.40)[2,3]	-	1.12 (0.89,1.42)	1.05 (0.82,1.33)
Housing type (Informal)				
Formal	1.03 (0.77,1.39)[2,3]	1.33 (1.03,1.72)[2,3]	1.39 (1.07,1.79)	1.26 (0.97,1.65)
Number of people in the household (1-3 people)				
Four or more people	-	-	0.97 (0.77,1.21)	0.91 (0.72,1.14)
Household services (Two or less)				
Water, electricity and toilet	1.78 (1.29,2.45)[2]	-	1.81 (1.31,2.50)	1.72 (1.24,2.38)

Social participation (No participation)				
Participation in at least one club or organisation	-	1.40 (1.12,1.75)[2,3]	1.38 (1.11,1.73)	1.35 (1.08,1.70)
Social trust (Most people can be trusted)				
Need to be very careful	-	1.48 (1.10,1.98)[2]	1.47 (1.09,1.97)	1.47 (1.09,1.98)
Household adult food poverty (Not skipped a meal in the past year)				
Skipped a meal	-	-	0.80 (0.60,1.06)	0.79 (0.60,1.05)

[1]Age and gender were adjusted for one another; and all subsequent covariates were adjusted for age and gender, and any preceding covariates (as in Figure 1).

[2]Under-adjusted (given the availability of data on potential confounders that have not been included in the model).

[3]Inappropriately adjusted (given the inclusion of likely mediators in the model).

Table 3. Sequential multivariable logistic regression models exploring the relationship between a range of demographic, economic, household, psychosocial characteristics and health-limited work and/or social activities amongst respondents from households: with the lowest tertile of total household income within the City of Johannesburg (Sample 2); throughout the City of Johannesburg, regardless of Ward or income (Sample 3); and throughout Gauteng Province (Sample 4). Covariates were entered additively, one-by-one in a sequential fashion[1] as determined by the DAG (see Figure 1), before and after additional adjustment for migrant status.

Households included:	Sample 2 City of Johannesburg lowest tertile of household income n=3,059		Sample 3 City of Johannesburg all income strata n=8,263		Sample 4 Gauteng Province all income strata n=24,727	
Covariates entered:	Sequentially	Sequentially	Sequentially	Sequentially	Sequentially	Sequentially
Adjustment for migrant status:	No	Yes	No	Yes	No	Yes
Characteristic (referent)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)	OR(95%CI)
Age (18-25yrs)[1]						
26-35yrs	1.11 (0.89,1.38)	1.11 (0.89,1.38)	1.12 (0.99,1.28)	1.12 (0.99,1.28)	1.08 (1.00,1.16)	1.08 (1.00,1.16)
36-45yrs	1.51 (1.19,1.91)	1.51 (1.19,1.91)	1.34 (1.17,1.54)	1.34 (1.17,1.54)	1.38 (1.27,1.49)	1.38 (1.27,1.49)
46-60yrs	2.06 (1.63,2.60)	2.06 (1.63,2.60)	1.78 (1.55,2.05)	1.78 (1.55,2.05)	2.06 (1.90,2.23)	2.06 (1.90,2.23)
>60yrs	5.21 (3.91,6.93)	5.21 (3.91,6.93)	4.04 (3.38,4.84)	4.04 (3.38,4.84)	4.54 (4.11,5.02)	4.54 (4.11,5.02)
Gender (Male)[1]						
Female	1.40 (1.21,1.63)	1.40 (1.21,1.63)	1.39 (1.27,1.51)	1.39 (1.27,1.51)	1.41 (1.34,1.49)	1.41 (1.34,1.49)
Educational attainment (Primary or lower; <Grade 8)[2]						
≥Grade 8	1.24 (1.01,1.51)	1.24 (1.01,1.51)	0.96 (0.83,1.11)	0.96 (0.83,1.11)	0.82 (0.75,0.88)	0.82 (0.75,0.88)
Migrant status (Gauteng resident)[2]						
Internal (most urban)	-	0.70 (0.54,0.92)	-	0.79 (0.66,0.93)	-	0.92 (0.84,1.01)
Internal (least urban)	-	0.57 (0.48,0.68)	-	0.71 (0.64,0.79)	-	0.89 (0.84,0.95)
Transnational	-	0.56 (0.45,0.71)	-	0.61 (0.53,0.70)	-	0.80 (0.72,0.87)
Length of residence in current dwelling (≥18 months)[2]						
<18 months	0.76 (0.61,0.96)	0.88 (0.70,1.11)	0.73 (0.63,0.83)	0.79 (0.69,0.91)	0.85 (0.78,0.92)	0.88 (0.81,0.95)
Employment (Unemployed)[2]						
Employed	0.86 (0.73,1.01)	0.94 (0.80,1.11)	0.84 (0.76,0.92)	0.87 (0.79,0.96)	0.84 (0.80,0.89)	0.85 (0.81,0.90)
Housing tenure (Rented)[2]						
Owned	0.99 (0.84,1.16)	0.92 (0.78,1.08)	0.97 (0.88,1.06)	0.89 (0.81,0.98)	1.01 (0.95,1.07)	0.98 (0.92,1.04)
House structure (Informal)[2]						
Formal	1.02 (0.85,1.23)	0.91 (0.75,1.10)	1.15 (1.00,1.31)	1.07 (0.93,1.22)	0.98 (0.91,1.06)	0.96 (0.89,1.03)
Number of people in the household (1-3 people)[2]						

Four or more people	1.14 (0.97,1.34)	1.06 (0.90,1.25)	1.05 (0.96,1.16)	1.00 (0.91,1.11)	1.02 (0.96,1.07)	1.00 (0.95,1.06)
Household services (Two or less)[2]						
Water, electricity and toilet	1.32 (1.06,1.64)	1.22 (0.98,1.52)	1.47 (1.25,1.71)	1.38 (1.18,1.62)	1.14 (1.04,1.25)	1.12 (1.03,1.23)
Social participation (No participation)[2]						
Participation in at least one club or organisation	1.25 (1.07,1.46)	1.21 (1.03,1.41)	1.32 (1.20,1.45)	1.28 (1.17,1.41)	1.34 (1.27,1.42)	1.34 (1.26,1.41)
Social trust (Most people can be trusted)[2]						
Need to be very careful	1.41 (1.15,1.73)	1.40 (1.14,1.72)	1.77 (1.57,2.00)	1.77 (1.57,2.00)	1.18 (1.11,1.27)	1.18 (1.10,1.27)
Household adult food poverty (Not skipped a meal in the past year)[2]						
Skipped a meal	0.96 (0.79,1.16)	0.93 (0.77,1.13)	0.93 (0.81,1.06)	0.91 (0.80,1.05)	1.16 (1.08,1.25)	1.16 (1.07,1.**24**)

[1]Age and gender were adjusted for one another; [2]All characteristics were adjusted for age and gender, and any preceding covariates (as in Figure 1).