JECH Online First, published on February 12, 2016 as 10.1136/jech-2015-206698

Challenges in reproducing results from publicly available data: an example of sexual orientation and cardiovascular disease risk

Nichole Austin,[1] Sam Harper,[1] Jay S Kaufman,[1] Ghassan B Hamra[2]

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jech-2015-206698).

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[2]Department of Environmental and Occupational Health, School of Public Health, Drexel University, Philadelphia, Pennsylvania, USA

Correspondence to
Dr Sam Harper, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec, Canada H3A 1A2; sam.harper@mcgill.ca

Received 24 September 2015
Revised 8 January 2016
Accepted 23 January 2016

ABSTRACT

Background Replication is a vital part of the research process and has recently received considerable attention. Analyses using publicly available data should, if adequately described, be reproducible without assistance from the original investigators. Using data from the US National Health and Nutrition Examination Survey (NHANES), a recent study reported a statistically significant difference in cardiovascular disease risk comparing subgroups of sexual minority men. We attempted to reproduce these findings and assessed whether the results were robust to alternative analytic strategies and assumptions.

Methods We used the exclusion criteria and coding strategy described in the original paper to construct our analytical data set. Sampling weights were constructed in accordance with NHANES analytical guidelines. We estimated crude and covariate-adjusted associations between sexual orientation and vascular age using the regression models specified in the original report. We also conducted a series of sensitivity analyses to improve on the original findings.

Results Our replication attempt was partially successful: we replicated the general trends reported in the original analysis, but not identical effect estimates. Importantly, we identified a potential misapplication of the Framingham Risk Score; correcting for this increased the probability that the reported null hypothesis test was a type I error.

Conclusions This paper supports the recent calls for greater transparency and improved reporting in research. Even with a publicly available and well-documented data source, we were unable to exactly replicate another study's original findings. Our sensitivity analyses revealed key issues in the original analysis and demonstrate the scientific importance of research replication.

INTRODUCTION

The production of reproducible research findings is a hallmark of the scientific method. Replication has been described by this journal as a potentially effective way to reduce false-positive findings,[1] but a number of high profile studies suggest that many, if not most, results are not replicable.[2 3] There are many reasons for non-reproducible findings, including genuine heterogeneity. Recently, there have been numerous calls to increase the emphasis on reproducibility in research generally, and in epidemiology in particular, but such practices are currently not incentivised, nor always possible, given inadequate and selective methodological reporting practices.[3 4]

A number of factors have been cited as barriers to replication, including publication bias and selective outcome reporting. Reporting guidelines such as CONSORT (for clinical trials) and STROBE (for observational studies) are meant to encourage clear and transparent reporting of research results, but such guidelines are infrequently followed.[5] Data sharing facilitates analytical replication,[6] but the difficulty of replication is well documented in biomedical research even when data and protocols are available.[3] In epidemiological research, where results are often reported as universal, qualitative (if not exact) replication should be the expectation.

A recent study by Farmer *et al*[7] in this journal used data from the publicly available US National Health and Nutrition Examination Survey (NHANES) to explore the association between sexual orientation and the risk of cardiovascular disease (CVD) in men. Their results found that, after adjusting for drug use and education, bisexual men were at increased risk for CVD compared with heterosexuals, while heterosexually identified men reporting at least one male sexual partner over the course of their lifetime (homosexually experienced heterosexuals, or HEHs) were at decreased risk of CVD. The authors concluded that greater attention should be paid to the mechanisms by which CVD risk is conferred to sexual minorities, with particular emphasis on the potentially heterogeneous mechanisms in specific subgroups of sexual minority men (SMM).

In this paper, we aim to reproduce the findings reported by Farmer *et al*. Given the publicly available nature of the NHANES data and the methods described by Farmer *et al*, we believed these findings should be replicable without assistance from the original authors. We also extend their analysis and perform several additional sensitivity analyses.

METHODS

Data source

We obtained five cycles of NHANES data from 2001–2002 to 2009–2010 for the variables specified in the Farmer *et al* study and pooled these individual data sets. As per NHANES analytic guidelines, we created the appropriate 10-year sampling weight for combining surveys.[8]

Exclusion criteria

We followed the authors' exclusion strategy to identify the analytic sample. In total, 7571 men responded to the sexual orientation question, of which 162 were excluded due to responding either 'something else' (n=26), 'not sure?' (n=82),

To cite: Austin N, Harper S, Kaufman JS, *et al*. *J Epidemiol Community Health* Published Online First: [*please include* Day Month Year] doi:10.1136/jech-2015-206698

'refused' (n=13) or 'don't know' (n=41). A further 312 respondents were excluded from the sample due to self-reported prior cardiovascular events. Our final sample consisted of 7097 participants who self-identified as heterosexual, bisexual or homosexual, and who had not experienced a prior cardiovascular event.

Exposure definition

Following Farmer *et al*, we defined sexual orientation in two ways: by self-identity (the 'identity only' definition), and by self-identity and/or history of same-sex activity (the 'identity/behaviour' definition). We created a binary variable denoting participant's sexual behaviour, coded as 1 for men who either self-identified as gay or bisexual, or self-identified as heterosexually identified but reported at least one same-sex partner in their lifetime, and 0 for men self-identified as heterosexual and reported no lifetime same-sex partners. Our final sample contained 128 homosexual, 104 bisexual and 6865 heterosexually identified men. Among the heterosexually identified men, 135 reported lifetime same-sex activity and were thus considered as SMM in analyses using the identity/behaviour definition of sexuality.

Outcome definition: the Framingham Risk Score algorithm

The outcome of interest was CVD risk, which Farmer *et al* operationalised by using the ratio of each participant's 'vascular age' to their chronological age. Vascular age is defined as a participant's projected age based on their cardiovascular risk factors. The Framingham Risk Score (FRS) was used to obtain CVD risk and vascular age estimates. The risk score for CVD can be calculated with either a point system or a parametric formula.[9] The point system assigns values to established cardiovascular risk factors based on a participant's age and sex. The participant's cumulative score is then used to determine his/her vascular age and the probability of experiencing a cardiovascular event in the following 10-year period. The point system is derived from Cox proportional hazards models based on a sample of Framingham cohort study participants aged 30–74. In the formula-based approach, the coefficients yielded by these models can be used directly to calculate risk.

The authors did not explicitly state their approach, but they appear to have used the point system given their example calculation, which includes a range (rather than an exact value) of untreated systolic blood pressure. Although the two approaches typically produce similar results, the point system is regarded as less precise.[10] In the interest of replication, we calculated risk and vascular age using the point system. We also repeated our analyses using the formula-based approach; calculation details and results are provided in the online supplementary material.

Covariates

We retained the same covariates as Farmer *et al* and attempted to code them as described in the original analysis, though there were a few anomalies. Two of the income categories overlapped in Farmer *et al*'s table 1; to account for this, we adjusted the maximum value of the third category ($35 000–$49 999) to $44 999. Additionally, some NHANES respondents classified their income as simply 'over $20 000'. We treated this response as non-informative, based on the category structure employed by Farmer *et al*, and considered these responses missing. We also coded the lowest age category to include participants up to and including the age of 29; it was unclear how these participants were classified in the original analysis. Finally, we were not certain how Farmer *et al* calculated systolic blood pressure as NHANES reports up to four readings. Our systolic blood pressure variable reflects the participant-specific average across all available readings; this approach is in accordance with the American Heart Association's clinical guidelines and produces more accurate estimates than a single reading.[11]

Missing data

Aside from the main exposure, Farmer *et al* did not discuss how they handled missing values in the analysis. Most covariates had missing data, but the per cent missing never exceeded the NHANES-suggested threshold for complete case analysis (10%).[i] We therefore assumed the authors employed a complete case approach and we mirrored this in our own analysis.

Statistical analysis

Our analyses consist of the replication attempt and a series of sensitivity analyses. We estimated crude and covariate-adjusted associations between sexual orientation and vascular age using linear regression, which was the same approach employed by Farmer *et al*. Analyses were conducted using Stata V.13 (StataCorp, College Station, Texas, USA) and we accounted for the survey design and weighting structure described in the NHANES analytic guidelines.

Modifications to the original analysis

We performed two sensitivity analyses to assess the robustness of the findings of Farmer *et al* to assumptions about vascular age among young people and missing data on sexual orientation, respectively. The FRS for CVD has been validated for individuals aged 30–74[8] but is not defined for individuals under 30.[12 13] Most online assessment tools, including the self-assessment tool on the Framingham website (http://bit.ly/1l4qDg7), will not calculate risk for a participant outside of the validated parameters. In addition, the cited source for risk calculation[8] provided no information on point-based calculation in participants under 30, so it was not possible to infer how these participants (n=2109, 29.7% of the original sample) were treated in the authors' analysis.

In an attempt to replicate their results, we initially programmed the point system to include participants under 30 by simply extending the point structure, which assigned zero points for age to men from 30 to 35, so that any participant under 30 would also receive zero points for age. However, because predicted risks and vascular ages have not been validated for those <30 years of age, we also re-estimated the authors' models in the subpopulation of individuals aged 30 and over.

Second, we used a simulation strategy to assign individuals Farmer *et al* considered 'missing' on sexual orientation to one of the four main exposure categories. Of the 7571 men who were asked about their sexual orientation, 162 did not explicitly self-define as heterosexual, gay or bisexual. Twelve of these participants reported prior CVD and were excluded from analysis. Of the remaining 150 men, 13 (8.6%) refused to respond and the remaining participants responded 'don't know' (n=37, 24.7%), 'something else' (n=24, 16%) or 'not sure' (n=76, 50.7%). The modified sample, including the participants with non-informative responses and excluding all participants with prior CVD (n=324), contained 7247 men, 2045 of whom were under 30. We randomly reassigned these individuals to one of the four main exposure categories and re-estimated the adjusted associations. We bootstrapped this process 10 000 times and

[i]http://www.cdc.gov/nchs/tutorials/NHANES/Preparing/CleanRecode/Info1.htm

Table 1 Comparison of demographic characteristics by SMM category

Variable	Farmer *et al*				Replication			
	Hetero* (n=6713)	Gay† (n=128)	Bi‡ (n=102)	HEH§ (n=135)	Hetero* (n=6730)	Gay† (n=128)	Bi‡ (n=104)	HEH§ (n=135)
Age, years (%)								
≤29	27.2	18.1	23.1	15.9	27.2	18.1	22.9	15.9
30–39	24.5	34.7	29.5	29.3	24.5	34.7	29.2	29.3
40–49	27.2	28.6	22.8	28.7	27.2	28.6	23.1	28.7
50+	21.1	18.5	24.6	26.1	21.1	18.5	24.8	26.1
Race (%)								
White	69.5	73.3	66.8	70.4	69.4	73.3	66.6	70.4
Black	10.7	7.4	13.0	9.7	10.7	7.4	12.8	9.7
Mexican	9.9	6.8	9.9	8.9	9.9	6.8	9.8	8.9
Other Hispanic	4.8	3.9	8.0	7.3	4.8	3.9	8.5	7.3
Other	5.2	8.7	2.3	3.9	5.2	8.7	2.2	3.9
Education (%)								
<High school	16.1	2.5	17.7	9.3	16.1	2.5	17.5	9.3
High school	26.8	7.7	23.9	17.5	26.8	7.7	23.7	17.5
Some college	30.8	29.0	36.4	44.7	30.8	29.0	37.1	44.7
College grad+	26.3	60.9	22.0	28.5	26.3	60.9	21.7	28.5
Income (%)								
<25 000	16.5	14.4	30.6	23.7	16.6	14.4	30.8	23.7
25 000–34 999	10.0	7.3	6.9	4.6	10.0	7.3	6.8	4.6
35 000–44 999	10.0	3.4	13.0	7.6	9.7	3.4	13.5	7.6
45 000–54 999	10.5	16.3	10.7	11.0	10.5	16.3	10.6	11.0
55 000+	53.3	58.6	38.8	53.0	53.2	58.6	38.4	53.0
Smoking (%)								
Non-smoker	49.3	50.6	45.2	41.2	49.3	50.6	45.2	41.2
Former smoker	21.6	19.4	14.2	29.0	21.6	19.4	14.1	29.0
Current smoker	29.1	30.0	40.6	29.8	29.1	30.0	40.8	29.8
Diabetes (%)	4.0	0.4	15.4	5.5	4.2	0.4	17.3	2.9
Antihypertensive medication use (%)	12.8	7.8	23.0	8.5	12.8	7.8	23.2	8.5
Family history of CVD (%)	7.3	9.4	7.2	11.3	12.3	15.1	19.7	21.2
Body mass index (kg/m^2)								
Normal/under (<25)	58.0	63.6	54.8	63.0	29.3	41.0	28.0	34.8
Over (≥25, <30)	24.2	21.1	19.8	21.9	39.5	35.1	33.1	38.3
Obese (≥30)	17.8	15.3	25.4	15.1	31.2	23.9	38.8	27.0
Total cholesterol (mg/dL)	199.4	200.3	197.3	208.2	199.4	200.3	197.3	208.2
HDL cholesterol (mg/dL)	47.1	47.3	46.4	47.8	47.1	46.4	47.4	48.2
Systolic blood pressure (mm Hg)	122.0	119.0	122.8	116.5	121.3	118.2	122.6	117.9
History of drug use (%)	25.8	31.3	42.5	46.5	25.8	31.3	42.7	46.5
Alcohol use								
Risky (>4/day or >14 week or any day with 5+)	39.1	31.4	40.6	37.7	33.2	20.0	41.5	24.2
Social	55.8	65.5	55.7	62.0	60.7	76.9	54.0	75.1
Infrequent	5.1	3.1	3.7	0.3	6.1	3.1	4.5	0.7

*Self-identified heterosexual.
†Self-identified gay/homosexual.
‡Self-identified bisexual.
§Self-identified heterosexual with at least one lifetime same-sex sexual partner (HEH).
CVD, cardiovascular disease; HDL, high-density lipoprotein; HEH, homosexually experienced heterosexual; SMM, sexual minority men.

used the empirical distribution of the results to calculate coefficients and 95% CIs.

Reproducible research statement

The raw data and statistical code for reproducing the results in this study are available with unrestricted open access from the Dataverse: https://dataverse.harvard.edu/dataverse/samharper

RESULTS

Our results are presented in two sections: the first describes our attempt to replicate the findings reported by Farmer *et al*, and the second presents our sensitivity analyses, which are intended to expand on the original findings. For brevity, we present our replication of Farmer *et al*'s four-category 'identity/behaviour' analyses. A summary table of replicated elements and replication results based on the binary 'identity only' definition of sexuality are provided in the online supplementary material.

Replication

Table 1 provides a summary of demographic attributes by sexual identity/behaviour category. Despite close adherence to the

Table 2 Vascular age ratio comparison by sexual orientation category (point-based, all ages)

	Farmer *et al*			Replication		
	Ratio	Difference	95% CI for difference	Ratio	Difference	95% CI for difference
Unadjusted						
Heterosexual	1.20	Ref	Ref	1.18	Ref	Ref
Gay	1.11	−0.09	(−0.14 to −0.04)	1.10	−0.08	(−0.12 to −0.03)
Bisexual	1.29	0.08	(0.01 to 0.15)	1.27	0.09	(0.01 to 0.17)
HEH*	1.14	−0.07	(−0.12 to −0.02)	1.13	−0.05	(−0.09 to −0.00)
Adjusted†						
Heterosexual	1.09	Ref	Ref	1.07	Ref	Ref
Gay	1.05	−0.04	(−0.09 to 0.003)	1.04	−0.03	(−0.07 to 0.02)
Bisexual	1.16	0.07	(0.00 to 0.13)	1.14	0.08	(0.00 to 0.15)
HEH*	1.02	−0.07	(−0.12 to −0.02)	1.02	−0.04	(−0.08 to −0.00)

*Homosexually experienced heterosexuals.
†Adjusted for history of hard drug use and education.

exclusion strategy described by Farmer *et al*, we could not reproduce their analytical sample exactly (n=7097 vs n=7078). Our covariate distribution was generally close to the original findings, with a few notable exceptions. Our estimates of body mass index (BMI) categories, family history of CVD and alcohol use were different from the estimates in the published paper and we ultimately could not reconcile the two sets of estimates. In particular, we questioned Farmer *et al*'s BMI distribution: the prevalence of adult male obesity in 1999–2000 was 27.5%, and obesity rates among men increased monotonically throughout the last decade.[14 15] It seems unlikely that the prevalence of obesity for heterosexual men and SMM would be 17.8% and 17.7%, respectively, for the 2001–2009 period. Our estimates (31.2%, 28.6%) appear closer to published national trends. It should be noted that none of the highly discrepant covariates were components of the Framingham CVD risk algorithm, so these inconsistencies were unlikely to directly affect regression estimates.

Table 2 shows regression-based measures of vascular age for Farmer *et al* and our replication models. Because some participants had missing values on necessary components of the risk score, the mean was calculated based on the 6519 participants for whom risk and vascular age were available, representing 92% of the initial sample. Our closest replication of the authors' adjusted findings used the highest level of education (college +) as the reference group; our reported values therefore represent age ratios for men with a college education who did not report drug use. In keeping with Farmer *et al*, we also report the difference (and 95% CI) in vascular age comparing sexual minority and heterosexual men.

Our results suggested that the average heterosexual participant's vascular age, adjusting for education and drug use, was 1.07 times higher than his chronological age, which was slightly lower than Farmer's estimate of 1.09. In their unadjusted analysis, Farmer *et al* reported a significant increase in CVD risk among bisexual men, as well as decreased CVD risk among homosexual and HEHs, compared with heterosexuals. The estimates for bisexuals and HEH remained statistically significant after controlling for education and hard drug use. We found similar estimates in our replication, though the lower and upper bounds (respectively) for the adjusted bisexual and HEH estimates remained very close to the null.

Sensitivity analyses

We build on Farmer *et al*'s findings in two ways: first, by imposing an age restriction in accordance with the sample used to validate the Framingham CVD algorithm, and then by attempting to include the men who provided a non-informative response to the sexual orientation question. Results from sensitivity analyses using the simplified 'identity only' definition of sexuality and the formula-based calculation of CVD risk are provided in the online supplementary material.

Table 3 Age ratio calculation comparison by sexual minority men category: age restriction versus original

	Ages 30–69			Ages 18–69		
	Ratio	Difference	95% CI for difference	Ratio	Difference	95% CI for difference
Unadjusted						
Heterosexual	1.12	Ref	Ref	1.18	Ref	Ref
Gay	1.07	−0.05	(−0.10 to 0.00)	1.10	−0.08	(−0.12 to −0.03)
Bisexual	1.20	0.08	(0.01 to 0.16)	1.27	0.09	(0.01 to 0.17)
HEH*	1.10	−0.01	(−0.06 to 0.03)	1.13	−0.05	(−0.09 to −0.00)
Adjusted†						
Heterosexual	1.04	Ref	Ref	1.07	Ref	Ref
Gay	1.03	−0.01	(−0.06 to 0.04)	1.04	−0.03	(−0.07 to 0.02)
Bisexual	1.12	0.08	(0.00 to 0.15)	1.14	0.08	(0.00 to 0.15)
HEH*	1.03	−0.01	(−0.05 to 0.03)	1.02	−0.04	(−0.08 to −0.00)

*Homosexually experienced heterosexuals.
†Adjusted for history of hard drug use and education.

Age restriction

Restricting the original sample to ages 30 and above resulted in a modified sample size of 5078 participants. We used the age-restricted sample to calculate identity/behaviour age ratio estimates (table 3). Most age ratios and differences were substantially reduced following age restriction. The adjusted heterosexual age ratio (previously reported as 1.07) decreased considerably to 1.04, suggesting that the average heterosexual participant's vascular age in the modified sample was actually 4%, rather than nearly 7%, higher than their chronological age. The adjusted estimates for bisexual and HEH participants had larger null hypothesis p values than those reported in the original publication.

The shift in estimates was likely due to the influence of age ratios for men under 30 in the original analysis: using point-based calculation in our replication sample, the mean age ratio among men under 30 was 1.35, suggesting that the average vascular age was 35% higher than chronological age in this age group. This is clearly an overestimate. Because concrete values of vascular age obtained with the point-based system have a minimum of 30 (participants receiving <0 points receive a vascular age of '<30', which is not possible to code), the difference in vascular and chronological age was inflated in our replication analysis. Given our initial proximity to Farmer *et al*'s findings, this issue likely affected the original results meaningfully.

Simulation: non-informative respondents

Our simulation strategy randomly reassigned participants who provided a non-informative response to the sexual orientation question (and who had subsequently been excluded from analysis) to one of the four main exposure categories and re-estimated the adjusted associations (figure 1). Including these participants in the analytical sample increased the precision of the bisexual estimate and pushed the lower bound away from the null; the associated point estimate was relatively robust to model specification. This lends some support to the findings of increased CVD risk among bisexual men originally reported by Farmer *et al*. The HEH estimate, however, was less robust: following inclusion of non-informative respondents, the HEH estimate and CI closely resembled that originally reported by Farmer *et al*, but age restriction (both with and without non-informatives) shifted estimates towards the null.

DISCUSSION

Farmer *et al* reported a significantly increased risk for CVD in bisexual men and a significantly decreased risk for CVD in HEH men, compared with heterosexuals. Our replication attempt was reasonably successful: we were able to identify the same general trends, but we were unable to reproduce the exact effect estimates. Given that the data source is publicly available and well documented, the original findings should have been replicable: the fact that they were not supports the recent calls for greater transparency and improved reporting in research.[1–6] Although we expended considerable effort in our replication attempt, there is certainly the possibility that our own analyses contain errors; towards that end, we have posted the raw data and statistical code for our analysis in a public repository (https://dataverse.harvard.edu/dataverse/samharper).

The ambiguity surrounding the authors' description of the FRS calculation in men under 30 may have affected our estimates, and was likely a barrier to exact replication. Inappropriate use of the FRS has been documented.[16] Our sensitivity analyses suggested that age restriction is particularly important when using the point-based version of the

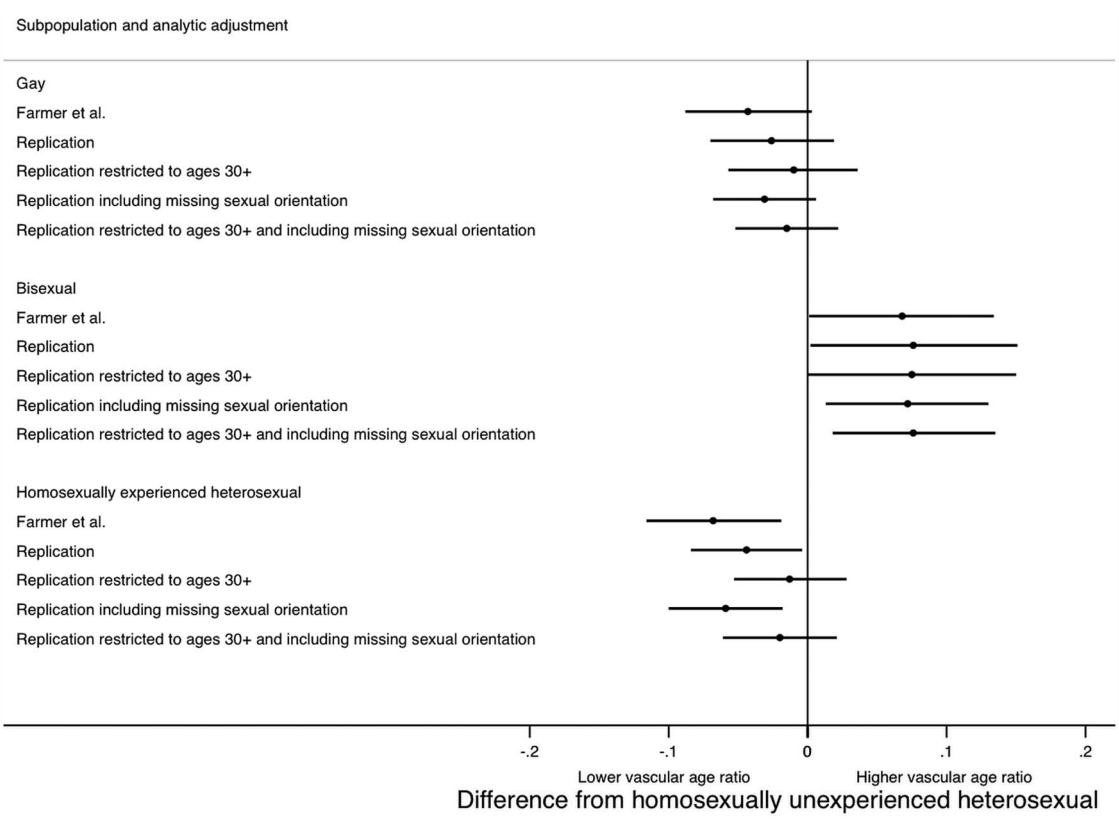

Figure 1 Associations between sexual orientation and the ratio of vascular to chronological age (adjusted for education and history of hard drug use) under different analytical strategies.

Framingham CVD algorithm. Age-restricted models had larger null hypothesis p values than those reported in the Farmer study. Supplementary findings (see online supplementary material) further suggested that the formula-based calculation, rather than point-based, may be a more robust approach of estimating CVD risk, as the substantial shift between original and age-restricted estimates was less pronounced when the formula was used.

We noted two other concerns in our replication attempt. First, the NHANES analytic guidelines caution that, "the sample size for the non-heterosexual subgroups is limited and do not meet the minimum sample size requirements for statistical reliability".[ii] The guidelines also note a high rate of non-response (and non-informative response) to the sexual orientation question. Although our sensitivity analyses did not reveal substantial changes in most estimates when non-informative participants were incorporated, use of this survey item as a marker of sexual orientation may be inadvisable. Second, we question the confounder selection process. Farmer *et al* used a 10% change in estimate criterion to identify drug use and education as model covariates, but one could easily argue that neither of these are determinants of sexual orientation; they are perhaps best conceptualised as mediators in the relationship between sexual orientation and CVD risk.[17] Future work in this area should carefully consider appropriate causal mechanisms, and the conceptualisation of sexual orientation as an exposure, prior to model fitting.

Ideally, research reports should be written that adhere to the replication standard: that is, the methods are described in sufficient detail, so that a third party might replicate the findings without any additional input from the authors.[18] Our interests here were prosaic: we wanted to see whether, given a publicly available data source and a straightforward descriptive analysis, we could replicate prior authors' works without their assistance. Although our attempt was moderately successful and yielded the same qualitative conclusions as the original study, our sensitivity analyses revealed potential issues in the original work and a possibly different take-home message regarding CVD risk in SMM. We hope to elucidate both the utility and importance of replication, and the need for rigorously testing assumptions, particularly when data are readily available for reanalysis. While errors are inevitable in any analysis (including our own), replication and improved transparency make these errors easier to identify.

What is already known on this subject

Replication remains underutilised in epidemiology, but when data are publicly available and methods are well described, findings should be reproducible without assistance from the original authors. In this study, we attempted to replicate the findings of a recent study on cardiovascular disease risk in men.

What this study adds

This paper supports the recent calls for greater transparency and improved reporting in research. It also describes a potential misapplication of the Framingham Risk Score, which has an important effect on estimates of cardiovascular disease risk.

[ii]http://www.cdc.gov/nchs/nhanes/nhanes2009-2010/SXQ_F.htm.

Contributors GBH proposed this study. NA compiled the data, and NA and SH conducted the analyses, drafted the initial manuscript, and prepared the public-use data set. All authors (NA, SH, JSK and GBH) collaborated on model interpretation and contributed to the final version of this manuscript.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 de Jonge P, Conradi HJ, Thombs BD, *et al*. Prevention of false positive findings in observational studies: registration will not work but replication might. *J Epidemiol Community Health* 2011;65:95–6.
2 Baggerly KA, Berry DA. Reproducible research. *Amstat News* 2011;403:16–17.
3 Ioannidis JP, Greenland S, Hlatky MA, *et al*. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166–75.
4 Chan AW, Song F, Vickers A, *et al*. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257–66.
5 Glasziou P, Altman DG, Bossuyt P, *et al*. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267–76.
6 Hernán MA, Wilcox AJ. Epidemiology, data sharing, and the challenge of scientific replication. *Epidemiology* 2009;20:167–8.
7 Farmer GW, Bucholz KK, Flick LH, *et al*. CVD risk among men participating in The National Health and Nutrition Examination Survey (NHANES) from 2001 to 2010: differences by sexual minority status. *J Epidemiol Community Health* 2013;67:772–8.
8 Johnson CL, Paulose-Ram R, Ogden CL, *et al*. National health and nutrition examination survey: analytic guidelines, 1999–2010. *Vital Health Stat 2* 2013;161:1–24.
9 D'Agostino RB Sr, Vasan RS, Pencina MJ, *et al*. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008;117:743–53.
10 Marma AK, Lloyd-Jones DM. Systematic examination of the updated Framingham heart study general cardiovascular risk profile. *Circulation* 2009;120:384–90.
11 Pickering TG, Hall JE, Appel LJ, *et al*. Recommendations for blood pressure measurement in humans and experimental animals: part 1: blood pressure measurement in humans: a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. *Hypertension* 2005;45:142–61.
12 Berry JD, Lloyd-Jones DM, Garside DB, *et al*. Framingham risk score and prediction of coronary heart disease death in young men. *Am Heart J* 2007;154:80–6.
13 Sheridan S, Pignone M, Mulrow C. Framingham-based tools to calculate the global risk of coronary heart disease: a systematic review of tools for clinicians. *J Gen Intern Med* 2003;18:1039–52.
14 Ogden CL, Carroll MD, Kit BK, *et al*. Prevalence of obesity in the United States, 2009–2010. *NCHS Data Brief* 2012;82:1–8.
15 Flegal KM, Carroll MD, Ogden CL, *et al*. Prevalence and trends in obesity among US adults, 1999–2008. *JAMA* 2010;303:235–41.
16 Tzoulaki I, Seretis A, Ntzani EE, *et al*. Mapping the expanded often inappropriate use of the Framingham Risk Score in the medical literature. *J Clin Epidemiol* 2014;67:571–7.
17 Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009;20:488–95.
18 King G. Replication, replication. *Political Sci Polit* 1995;28:444–52.