



Yates, T., Griffith, G. J., & Morris, T. T. (2021). Re: Human cytomegalovirus and risk of incident cardiovascular disease in UK Biobank. *Journal of Infectious Diseases*.
<https://doi.org/10.1093/infdis/jiab571>

Version created as part of publication process; publisher's layout; not normally made publicly available

License (if available):
CC BY

Link to published version (if available):
[10.1093/infdis/jiab571](https://doi.org/10.1093/infdis/jiab571)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Oxford University Press at [10.1093/infdis/jiab571](https://doi.org/10.1093/infdis/jiab571)/ Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Re: Human cytomegalovirus and risk of incident cardiovascular disease in UK Biobank

Tom A. Yates ^{1*}, Gareth J. Griffith ^{2,3}, Tim T. Morris ^{2,3}

1. Department of Infectious Disease, Faculty of Medicine, Imperial College London, UK
2. Medical Research Council Integrative Epidemiology Unit at the University of Bristol, UK
3. Population Health Sciences, Bristol Medical School, University of Bristol, UK

* Dr Tom Yates, t.yates@imperial.ac.uk

Accepted Manuscript

© The Author(s) 2021. Published by Oxford University Press for the Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Hamilton *et al*'s analysis of the association between cytomegalovirus (CMV) infection and incident cardiovascular disease (CVD) in 8,531 UK Biobank (UKBB) participants is stronger than most observational studies of the CMV–CVD association[1]. There was careful adjustment for age, sex, socioeconomic position and cardiovascular risk factors. Neither inflammatory markers nor prior CVD were included as covariates, which would have been inappropriate. However, whilst the authors acknowledge that sections of the population are underrepresented in UKBB, the potential impact of this on their main results is not discussed.

Non-random sample selection can lead to spurious in-sample associations, a form of collider bias. A schematic example is presented in Figure 1A, where individuals with high CVD risk or high CMV exposure are less likely to be sampled. This induces a negative correlation between CVD and CMV in the sample that does not hold in the broader population. To evaluate the extent to which collider bias might explain Hamilton *et al*'s null findings, we compared CVD risk and CMV exposure in UKBB participants to that in more representative population samples.

UKBB is a highly selected sample. Of 9 million 40-69 year olds invited to participate, only 5.5% enrolled[2]. The sample is enriched for individuals of higher socioeconomic position and White ethnicities (94.6 vs 91.3%)[2]. UKBB participants are less likely to be obese, drink less alcohol and are less likely to smoke when compared to more population representative respondents to the Health Survey for England (HSE)[2]. All-cause mortality rates among UKBB participants aged 70-74 years are 46% and 56% lower in men and women, respectively, than in the general population of England and Wales[2].

UKBB participants have lower self-reported CVD prevalence at enrolment[2], lower exposure to CVD risk factors[2,3], and CVD mortality rates that are 2.9 fold lower than individuals of a comparable age in HSE[3]. Previous analyses have shown that these selection effects can attenuate risk factor–CVD associations. For example, for hypertension, the hazard ratio for CVD mortality is 1.9 in UKBB and 2.6 in HSE[3].

As CMV is more prevalent in those of lower socioeconomic position[4,5], it is plausible that UKBB participants have lower CMV exposure than the general population. CMV seroprevalence appears lower in UKBB than in participants from a 2002 serosurvey using blood samples from individuals accessing NHS care in England and Wales[6]. Age-specific CMV seroprevalence in the serosurvey was only presented graphically, preventing quantification of these differences.

Comparing the age-specific CMV seroprevalence reported by Hamilton *et al*[1] with that among pregnant women in the Born in Bradford (BiB) birth cohort[7] suggests markedly lower CMV prevalence in UKBB. CMV seropositivity is more prevalent in older individuals, which is partly a cohort effect, with most incident infections occurring in individuals under the age of 40 years[6]. Seroprevalence in 40-49 year olds in UKBB is 48.0%, similar to that among pregnant White British women in BiB (48.6%), despite the latter having a mean age of 26.6 years[1,7]. CMV is also more prevalent in people of non-White ethnicity[4,5,7]. Seroprevalence among UK born pregnant women of South Asian ethnicity in BiB was 89.3% (mean age 27.4 years). Note that CMV seroprevalence in UKBB is similar in men (56.0%) and women (59.3%)[1].

It appears that UKBB participants have lower CVD and CMV prevalence than the general population. The potential impact of selection effects into UKBB on observed associations can be simulated. For

example, selecting 5% of individuals with lower risk of both characteristics can produce a null in-sample association if the two characteristics have a Pearson's correlation coefficient of only 0.2 in the broader population (Figure 1B). Here, both characteristics are treated as continuous variables, e.g., CVD incidence and lifetime exposure to CMV. Code to reproduce this analysis under varying selection mechanisms is available at <https://github.com/timtmorris/collider-bias>.

Collider bias is a common and underappreciated problem in health research. Its potential to impact analyses in UKBB and other unrepresentative samples is well described[3,8,9]. This bias may be particularly consequential when investigating the impact of ubiquitous exposures on common diseases. Here, modest differences in attributable risk imply a substantial burden of disease.

The best solution to this problem would be to invest in making participation in research accessible to a broader section of the community. This includes access to healthcare services that routinely contribute data to, e.g., Hospital Episodes Statistics. Where this has not been done, researchers should conduct their analyses in other less selected datasets or, where data exist on the determinants of inclusion, weight the sample and adjust for selection. This is an area of active methodological research.[10]

Figure 1. A) Schematic demonstrating an in-sample negative correlation (filled circles and dashed line) that does not exist in the underlying population (all circles) may be generated by selectively including individuals with lower levels of both exposure and outcome. B) Simulated data showing how in-sample associations (dashed line) may not reflect population level associations (line) if selected individuals (blue circles) are a specific subset of the population (red circles).

Accepted Manuscript

Funding: TAY is funded via an NIHR Academic Clinical Fellowship and acknowledges support from the NIHR Imperial Biomedical Research Centre. The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_00011/1, MC_UU_00011/3].

Declaration of interest: TAY is currently applying for funding to undertake research on the indirect consequences of CMV infection. GJG and TTM have no relevant conflicts of interest to declare.

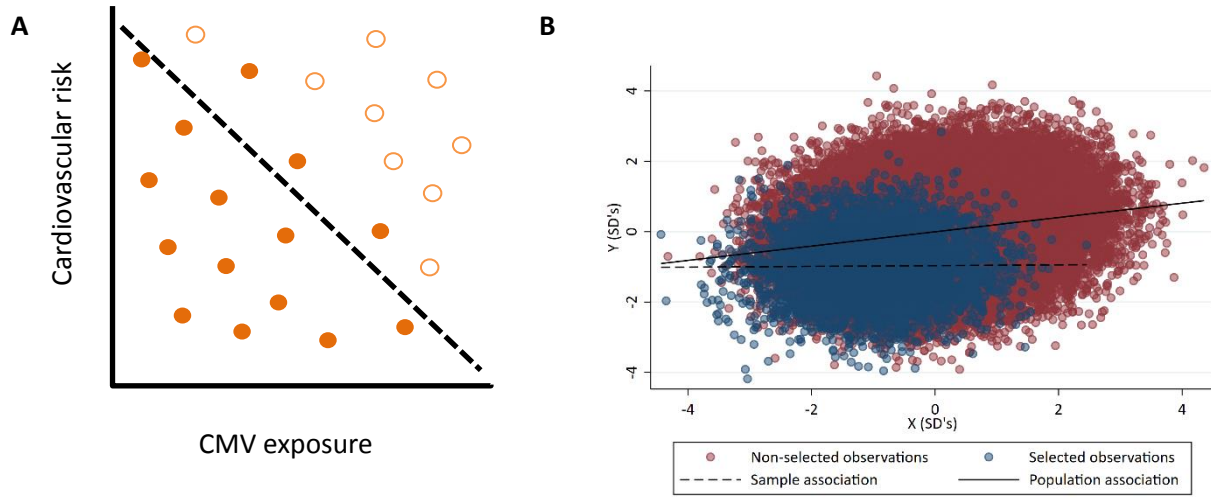
Accepted Manuscript

1. Hamilton EM, Allen NE, Mentzer AJ, Littlejohns TJ. Human cytomegalovirus and risk of incident cardiovascular disease in UK Biobank. *J Infect Dis.* **2021**; .
2. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. *Am J Epidemiol.* **2017**; 186(9):1026–1034.
3. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ.* **2020**; 368:1–8.
4. Cannon MJ, Schmid DS, Hyde TB. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol.* **2010**; 20(4):202–213.
5. Winter JR, Taylor GS, Thomas OG, Jackson C, Lewis JEA, Stagg HR. Factors associated with cytomegalovirus serostatus in young people in England: a cross-sectional study. *BMC Infect Dis. BMC Infectious Diseases*; **2020**; 20(1):1–9.
6. Vyse AJ, Hesketh LM, Pebody RG. The burden of infection with cytomegalovirus in England and Wales: How many women are infected in pregnancy? *Epidemiol Infect.* **2009**; 137(4):526–533.
7. Pembrey L, Raynor P, Griffiths P, Chaytor S, Wright J, Hall AJ. Seroprevalence of cytomegalovirus, Epstein Barr virus and varicella zoster virus among pregnant women in Bradford: A cohort study. *PLoS One.* **2013**; 8(11):1–8.
8. Munafò MR, Tilling K, Taylor AE, Evans DM, Smith GD. Collider scope: When selection bias can substantially influence observed associations. *Int J Epidemiol.* **2018**; 47(1):226–235.

9. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun.* **2020**; 11(1):1–12.
10. Griffith GJ, Hemani G, Herbert A, et al. We should be cautious about associations of patient characteristics with COVID-19 outcomes that are identified in hospitalised patients [Internet]. *Heal. Data Res. UK.* 2021 [cited 2021 Sep 21]. Available from: <https://www.hdruk.org/news/we-should-be-cautious-about-associations-of-patient-characteristics-with-covid-19-outcomes-that-are-identified-in-hospitalised-patients/>

Accepted Manuscript

Figure 1



Accepted Manuscript