Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Diabetes & Metabolic Syndrome: Clinical Research & Reviews 15 (2021) 643–644

Contents lists available at ScienceDirect

Diabetes & Metabolic Syndrome: Clinical Research & Reviews

journal homepage: www.elsevier.com/locate/dsx

Letter to the Editor

Letter to the editor in response to the article: "Vitamin D concentrations and COVID-19 infection in UK biobank" (Hastie et al.)

UK Biobank analyses concluded that COVID-19 risk, particularly higher risk in ethnic minorities were not explained by vitamin D [1–10]. Hastie et al. dismissed previous critiques, responding that their analyses were " … as powerful as any to date" [8]. However, the reported statistical significance and high precision are illusory; these papers used unreliable data and contained grave errors: mislabelled data, flawed models, low power and high bias.

Only 449 Covid-19 test-positive cases were available, containing just 31 Black and 19 Asian individuals; plus 1,025 test negatives [1]. The Covid-19-negative set ("controls") was artificially inflated by adding all 347,124 untested individuals [1]. At that time, only those hospitalized (~8.2% of cases) were tested [10,11]. Therefore, the "Covid-19-negative" control set likely contained nine times as many positives as the "test-positive" set, including pre-hospitalisations, some in care homes, and milder cases [11]. Moreover, because COVID-19 risk is zero in the absence of SARS-CoV-2 exposure, the vast majority of the so-called controls was meaningless noise [12].

This data inflation led to serious errors: overfit, over-adjusted, and unnecessarily adjusted models [13,14]. Too many model variables in logistic regressions introduces bias, obscures effects and reduces precision [15,16]. Estimation efficiency deteriorates with each added covariate and reduces statistical significance, which can lead to important associations being declared insignificant [17]. Controlling for more variables does not necessarily reduce confounding; in fact, adding variables amplifies bias faster than it reduces confounding [17]. Selection criteria based on a priori theoretical or biological relationships should have been used to judiciously construct models [18].

These mistakes were compounded by using vitamin D levels and confounder variables (including self-reported subjective indexes) measured 10–14 years ago [1,9,10]. The authors claimed vitamin D levels remain stable over time, appearing to confuse the correlation coefficient, R, with explained variance, R^2 [1]. Indeed, studies they referenced demonstrated levels are not stable over many years, particularly among 25(OH)D-deficient individuals [1,5,19,20] - nor are blood pressure, pulse, and body mass index [19]. Biobank data explained only ~16% of variance in 2020 vitamin D values [19].

Categorising continuous variables is inadvisable in regressions, even for precise measures; categorising unreliable data amplifies errors by up to ten times [21]. A much larger sample size could increase power [22], but inestimably large and insurmountable bias issues would remain [23,24].

The reported conclusions were unjustified and incorrect. The data set was 1,474, not 348,598; misused statistical methods led to misleading results; and the UK Biobank data are too old to be appropriate for investigating this subject. A more detailed critique is available [25].

Declaration of competing interest

The authors received no funding for this work.
The authors have no conflicts of interest to disclose.

References

[1] Hastie CE, Mackay DF, Ho F, Celis-Morales CA, Katikireddi SV, Niedzwiedz CL, et al. Vitamin D concentrations and COVID-19 infection in UK biobank. Diabetes & metabolic syndrome. Clin Res Rev 2020;14:561–5. https://doi.org/10.1016/j.dsx.2020.04.050.

[2] Hastie CE, Mackay DF, Ho F, Celis-Morales CA, Katikireddi SV, Niedzwiedz CL, et al. Corrigendum to "vitamin D concentrations and COVID-19 infection in UK biobank" [diabetes metabol syndr: clin res rev 2020 14 (4) 561-5] Diabetes Metab Syndr 2020;14. 1315–6.

[3] Grant WB, McDonnell SL. Letter in response to the article: vitamin D concentrations and COVID-19 infection in UK biobank (Hastie et al.). Diabetes & Metabolic Syndrome: Clin Res Rev 2020;14:893–4. https://doi.org/10.1016/j.dsx.2020.05.046.

[4] Roy AS, Matson M, Herlekar R. Response to 'vitamin D concentrations and COVID-19 infection in UK biobank: Diabetes & Metabolic Syndrome. Clin Res Rev 2020;14:777. https://doi.org/10.1016/j.dsx.2020.05.049.

[5] Benskin LL. A basic review of the preliminary evidence that COVID-19 risk and severity is increased in vitamin D deficiency. Front Public Health 2020;8:513. https://doi.org/10.3389/fpubh.2020.00513.

[6] Boucher BJ. Vitamin D status as a predictor of Covid-19 risk in Black, Asian and other ethnic minority groups in the UK. Diabetes Metabol Res Rev 2020;36:e3375. https://doi.org/10.1002/dmrr.3375.

[7] Hosack T, Baktash V, Mandal AKJ, Missouris CG. Prognostic implications of vitamin D in patients with COVID-19. Eur J Nutr 2020. https://doi.org/10.1007/s00394-020-02429-4.

[8] Hastie CE, Pell JP, Sattar N. Reply to: prognostic implications of vitamin D in patients with COVID-19. Eur J Nutr 2020. https://doi.org/10.1007/s00394-020-02430-x.

[9] Hastie CE, Pell JP, Sattar N. Vitamin D and COVID-19 infection and mortality in UK Biobank. Eur J Nutr 2020;1–4. https://doi.org/10.1007/s00394-020-02372-4.

[10] Raisi-Estabragh Z, McCracken C, Bethell MS, Cooper J, Cooper C, Caulfield MJ, et al. Greater risk of severe COVID-19 in non-White ethnicities is not explained by cardiometabolic, socioeconomic, or behavioural factors, or by 25(OH)-vitamin D status: study of 1,326 cases from the UK Biobank. MedRxiv 2020. https://doi.org/10.1101/2020.06.01.20118943. 2020.06.01.20118943.

[11] Stewart K. LANCET: comprehensive COVID-19 hospitalization and death rate estimates. Today's Practitioner; 2020. https://todayspractitioner.com/covid-19/lancet-comprehensive-covid-19-hospitalization-and-death-rate-estimates/. [Accessed 14 November 2020].

[12] Official UK Coronavirus Dashboard. n.d. https://coronavirus.data.gov.uk/details/casess. [Accessed 27 January 2021].

[13] Day NE, Byar DP, Green SB. Overadjustment in case-control studies. Am J Epidemiol 1980;112:696–706. https://doi.org/10.1093/oxfordjournals.aje.a113042.

[14] Greenland S, Pearce N. Statistical foundations for model-based adjustments. Annu Rev Publ Health 2015;36:89–108. https://doi.org/10.1146/annurev-publhealth-031914-122559.

[15] Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. International Statistical Review/Revue

https://doi.org/10.1016/j.dsx.2021.02.016

Internationale de Statistique 1991;59:227–40. https://doi.org/10.2307/1403444.
[16] Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology 2009;20:488–95. https://doi.org/10.1097/EDE.0b013e3181a819a1.
[17] Pearl J. Invited commentary: understanding bias amplification. Am J Epidemiol 2011;174:1223–7. https://doi.org/10.1093/aje/kwr352.
[18] VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol 2019;34:211–9. https://doi.org/10.1007/s10654-019-00494-6.
[19] Jorde R, Sneve M, Hutchinson M, Emaus N, Figenschau Y, Grimnes G. Tracking of serum 25-hydroxyvitamin D levels during 14 Years in a population-based study and during 12 Months in an intervention study. Am J Epidemiol 2010;171:903–8. https://doi.org/10.1093/aje/kwq005.
[20] Meng JE, Hovey KM, Wactawski-Wende J, Andrews CA, LaMonte MJ, Horst RL, et al. Intraindividual variation in plasma 25-hydroxyvitamin D measures 5 Years apart among postmenopausal women. Cancer Epidemiol Biomark Prev 2012;21:916–24. https://doi.org/10.1158/1055- 9965.EPI-12-0026.
[21] Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34:1370–6. https://doi.org/10.1093/ije/dyi184.
[22] McKeown-Eyssen GE, Tibshirani R. Implications of measurement error in exposure for the sample sizes of case-control studies. Am J Epidemiol 1994;139:415–21. https://doi.org/10.1093/oxfordjournals.aje.a117014.
[23] Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. BMJ 2010;340:c2289. https://doi.org/10.1136/bmj.c2289.
[24] van Smeden M, Lash TL, Groenwold RHH. Reflection on modern methods: five myths about measurement error in epidemiological research. Int J Epidemiol 2020;49:338–47. https://doi.org/10.1093/ije/dyz251.
[25] Davies G, Mazess R, Benskin L. Serious statistical Flaws in Biobank analyses. 2020. https://www.researchgate.net/publication/346922274_Serious_Statistical_Flaws_in_Biobank_Analyses.

Gareth Davies*
Imperial College (alumnus), London, UK

Richard B. Mazess
Emeritus Professor of Medical Physics, University of Wisconsin, Madison, WI, USA

Linda L. Benskin
Independent Researcher and Ferris Mfg. Corp. (Makers of PolyMem Dressings), Austin, TX, USA

* Corresponding author. Sencillo Research, 125 Harbut Road, London, SW11 2RD, USA.
E-mail address: gareth@sencillo.co.uk (G. Davies).