

Citizen science in the time of COVID-19

Linda J Birkin,¹ Eleftheria Vasileiou,² Helen Ruth Stagg ³

The COVID-19 pandemic has been a masterclass in the need for accurate, population-wide, frequently updated and rapidly analysed digital data sources in global health.

This issue of *Thorax* presents two studies using data from the smartphone Zoe COVID Symptom Study app, which has been collecting voluntarily self-reported information from consenting participants ≥ 18 years on COVID in the UK since its launch in March 2020. As of 16 December 2020, 4 481 148 individuals had registered to use the app across the UK, USA and Sweden.¹ While there are other citizen science studies running in the UK to track COVID symptoms (eg, FluSurvey/Influenzanet and TrackTogether), the Zoe COVID app is by far the most extensive of these in coverage and participation numbers, and has contributed, for example, to identification of anosmia as a key symptom of COVID-19 in general (in May 2020),² and delirium as a key symptom in older people (October 2020).³

In this issue, Hopkinson *et al* sought to examine the impact of current smoking on the development of COVID-19, using data from March to April 2020.⁴ Among individuals who did not think they had previously had COVID-19, self-declared current smokers reported slightly increased odds for presenting the classic triad (cough, fever and breathlessness) of symptoms than non-smokers (OR: 1.14 (95% CI: 1.10 to 1.18)), in results adjusted for age, sex and body mass index. This association did not hold when analyses were constricted to individuals tested for SARS-CoV-2 (OR: 0.73 (95% CI: 0.65 to 0.81)), including after adjustment for potential confounding due to being a healthcare worker (such individuals were early targets of testing and had a lower prevalence of smoking). (Of note, collider bias due to selected sampling has been reported as a potential cause of apparent 'protective' effects of smoking in early

COVID-19 studies.⁵) Smokers who tested positive did, however, report more than double the odds of being hospitalised due to their COVID-19 (OR: 2.11 (95% CI: 1.41 to 3.11)), an association which remained when individuals who reported comorbidities were removed from the analysis, or the model was adjusted for healthcare worker status. Although establishment of an association between being a current smoker and COVID-19 disease remains uncertain within this study, the association with hospitalisation is concerning and deserves further exploration in models adjusting for socioeconomic status, ethnicity and other confounding factors. Notably, a series of systematic reviews published in recent months have demonstrated an association between smoking and mortality as well as disease severity, for example, Dorjee *et al*.⁶

Bowyer *et al* investigated the geographical distribution of COVID-19 and its association with deprived areas, using self-reported data from March to April 2020 across the UK.⁷ Higher predicted COVID-19 incidence was observed in urban and more deprived areas compared with rural and less deprived areas, respectively, adjusted by air pollution, primary care centres per area, household density, urbanicity, age, sex and spatial autocorrelations. This association is in line with other studies, for example, from the USA.⁸ People living in deprived areas are more likely to have unstable and short-term employment that cannot facilitate home-based remote working, and issues taking sick leave. Further instability has been added into the job market due to the economic downturn. Additionally, people living in deprived areas are more likely to live in high geographical density higher-occupancy housing, and have greater prevalence of key comorbidities. These factors are relevant for initial COVID-19 disease, and many will also be pertinent for the impact of long COVID-19. The role of deprivation in the pandemic speaks to the general need for government policies to reduce inequalities, as well as holistic planning across different sectors for future pandemics—all government departments should be committed to pandemic preparedness, not simply those for healthcare and public health.

As the above studies demonstrate, citizen science—the active and voluntary

participation of the public in research; typically collecting data that are impractical to record otherwise, often over short timescales or large geographical areas—has been invaluable in the pandemic. In the context of the Zoe app, it has not only allowed the monitoring of the spread of COVID-19 across the UK when access to testing was limited and slow (and people were perhaps not aware their symptoms might be relevant), but also to answer particular research questions. Even with the recognised limitations in app-collected and self-reported data, a wider picture of likely infection levels was provided than was otherwise possible.⁹ The Zoe COVID app data has broadly agreed with the government's ranking of most affected areas, confirming the validity of the approach and providing especial benefit in places with limited routine testing. Additionally, depending on country-specific and community-specific factors, citizen science projects undertaken by independent institutions can be met with greater public trust than data collection by the government. In either respect, trust between data provider and data collector is paramount, including on issues of data protection and ownership

Tracking disease spread is an increasingly common usage for citizen science within ecological fields,¹⁰ and similar principles apply for COVID-19, where effective responses have hinged on knowing where it is and how fast it is spreading, as quickly as possible. Citizen science studies within the pandemic have been greatly enhanced by smartphones and the availability of the internet, allowing for broad and fast information capture and active reporting back to participants. Biological testing requires distribution, application and processing, whereas self-reporting of symptoms takes a few minutes.

Citizen science comes with recognised biases. App users are not representative of the general population, generally showing clear differences in age, gender, educational level and income. The subject area of a voluntary study tends to attract people with pre-existing reasons to be interested in that field: more health conscious groups in this case; and other areas of collider biases. It is possible to compensate for some of these biases in analysis, but results must be interpreted with this in mind and the implications of a lack of generalisability and poor statistical certainty in poorly represented groups be kept at the forefront of all policy and resource-prioritisation decisions so as to

¹Independent researcher, Nottingham, UK

²Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK

³Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, UK

Correspondence to Dr Helen Ruth Stagg, Centre for Global Health Research, Usher Institute, MacKenzie House, University of Edinburgh, Edinburgh EH8 9AG, UK; helen.stagg@ed.ac.uk

avoid further exacerbating pre-existing inequalities.

Like nothing before in human history, the COVID-19 pandemic has called for rapid, interdisciplinary, scientific studies and innovative methods of data collection, as professionals worldwide have worked to contain the virus. Within this, citizen engagement has been critical for effective pandemic responses, both in terms of engagement with rules and regulations, and understanding how responsiveness has been limited by the speed of possible scientific progress. Citizen science has been a key element in our engagement toolbox and provides a template for future work in global health. Moving forwards, it is critical that the expertise gained during this pandemic is not lost; for example, the need for hibernated projects that can be rapidly activated to provide data to support public health surveillance and research questions when a potential pandemic is first detected.

Contributors All authors conceived of and designed the work, drafted it and critically revised it for content, approval the final version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding HRS is funded by the Medical Research Council, UK [MR/R008345/1].

Competing interests HRS is an advisor to the Scottish Parliament's COVID-19 committee and is an author on a paper using the Zoe app data.

Patient consent for publication Not required.

Provenance and peer review Commissioned; externally peer reviewed.

This article is made freely available for use in accordance with BMJ's website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

LJB and EV are joint first authors.



To cite Birkin LJ, Vasileiou E, Stagg HR. *Thorax* 2021;**76**:636–637.

Accepted 10 February 2021
Published Online First 2 March 2021



- ▶ <http://dx.doi.org/10.1136/thoraxjnl-2020-215119>
- ▶ <http://dx.doi.org/10.1136/thoraxjnl-2020-216422>

Thorax 2021;**76**:636–637.
doi:10.1136/thoraxjnl-2020-216673

ORCID iD

Helen Ruth Stagg <http://orcid.org/0000-0003-4022-3447>

REFERENCES

- 1 ZOE. COVID symptom study. Available: <https://covid.joinzoe.com/> [Accessed 10 Dec 2020].
- 2 Menni C, Valdes AM, Freidin MB, *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26:1037–40.
- 3 Zazzara MB, Penfold RS, Roberts AL, *et al.* Probable delirium is a presenting symptom of COVID-19 in frail, older adults: a cohort study of 322 hospitalised and 535 community-based older adults. *Age Ageing* 2021;50:40–8.
- 4 Hopkinson NS, Rossi N, El-Sayed Moustafa J, *et al.* Current smoking and COVID-19 risk: results from a population symptom APP in over 2.4 million people. *Thorax* 2021;76:714–22.
- 5 Griffith GJ, Morris TT, Tudball MJ, *et al.* Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020;11:5749.
- 6 Dorjee K, Kim H, Bonomo E, *et al.* Prevalence and predictors of death and severe disease in patients hospitalized due to COVID-19: a comprehensive systematic review and meta-analysis of 77 studies and 38,000 patients. *PLoS One* 2020;15:e0243191.
- 7 Bowyer RCE, Varsavsky T, Thompson EJ, *et al.* Geo-social gradients in predicted COVID-19 prevalence in Great Britain: results from 1 960 242 users of the COVID-19 symptoms study APP. *Thorax* 2021;76:723–5.
- 8 Chang S, Pierson E, Koh PW, *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 2021;589:82–7.
- 9 Varsavsky T, Graham MS, Canas LS, *et al.* Detecting COVID-19 infection hotspots in England using large-scale self-reported data from a mobile application: a prospective, observational study. *Lancet Public Health* 2021;6:e21–9.
- 10 Meentemeyer RK, Dorning MA, Vogler JB, *et al.* Citizen science helps predict risk of emerging infectious disease. *Front Ecol Environ* 2015;13:189–94.