How methodological pitfalls have created widespread misunderstanding about long COVID

Tracy Beth Høeg 1,2, Shamez Ladhani 3,4, Vinay Prasad 1

Introduction
High rates of long COVID or post-acute sequelae of COVID-19 (PASC) continue to be reported in academic journals and subsequently filtered to the public. For instance, the Centers for Disease Control and Prevention (CDC) recently stated ‘nearly one in five American adults who have had COVID-19 still have long Covid’.1 Many scientific publications overestimate PASC prevalence because of overly broad definitions, lack of control groups, inappropriate control groups, and other methodological flaws. This problem is further compounded by inclusion of poorly conducted studies into systematic reviews and meta-analyses that overstate the risk. This is led to the public by the media and social media, raising undue concern and anxiety. This paper aims to discuss these estimation errors and why epidemiologic research on long COVID has been misleading.

The problem with current case definitions
For the purposes of this paper, we define long COVID as a syndrome or individual symptoms which are direct sequelae of the virus, SARS-CoV-2, and last at least 12 weeks. Some post-COVID sequelae such as post-ICU syndrome, and post-pneumonia respiratory compromise are common to many upper respiratory viruses. While post-infectious conditions common to other respiratory illnesses may be included in estimates of prevalence of lasting symptoms, we propose future research avoid the umbrella term ‘long COVID’ and instead more narrowly define certain post-COVID syndromes or symptoms (such as anosmia) which may be specific to the SARS-CoV-2 virus.

Existing PASC case definitions from four international health organisations are shown in table 1. None of them requires a causal link between the SARS-CoV-2 infection, meaning any new symptoms after confirmed or suspected SARS-CoV-2 infection, regardless of their aetiology, could be considered consistent with long COVID. In general, in the scientific literature, imprecise definitions have resulted in more than 200 symptoms being associated with the condition termed long COVID.2

While all four definitions require antecedent infection with SARS-CoV-2, a recent review of PASC definitions found, of all studies of long COVID interventions, only 54% required laboratory-confirmed infection.3 Some argue that confirmation of SARS-CoV-2 infection was not always possible, particularly early in the pandemic; however, these studies also did not use serology to confirm prior infection, which can be done at any time. Failing to confirm prior SARS-CoV-2 infection is particularly relevant given that one French study, for example, found that self-reporting of persistent symptoms was more strongly associated with the belief in having been infected than with having had laboratory-confirmed SARS-CoV-2 infection.4

Another important failing of the term ‘long COVID’, interchangeably used with PASC, is that it connotes a permanent or long-term condition, such as epilepsy after bacterial meningitis, for example. However, there is good evidence post-infectious symptoms after COVID-19 improve over time even if some symptoms may take longer to improve than others.5,6

Moreover, many studies include a broad range of symptoms without any evidence of a causal link to SARS-CoV-2 infection. One UK study noted that 40% of patients with PASC first reported any symptoms ≥90 days after infection, which would not have been included as PASC if persistent or contiguous symptoms had been part of the definition.7 Three of four working definitions of PASC require symptoms to be persistent or continuous, while the CDC definition allows any symptom lasting at least 4 weeks after SARS-CoV-2 infection.8 This paper aims to discuss these estimation errors and why epidemiologic research on long COVID has been misleading.

Key messages

⇒ The existing epidemiological research on long COVID has suffered from overly broad case definitions and a striking absence of control groups, which have led to distortion of risk.

⇒ The unintended consequences of this may include, but are not limited to, increased societal anxiety and healthcare spending, a failure to diagnose other treatable conditions misdiagnosed as long COVID and diversion of funds and attention from those who truly suffer from chronic conditions secondary to COVID-19.

⇒ Future research should include properly matched control groups, sufficient follow-up time after infection and internationally-established diagnostic or inclusion and exclusion criteria.
infection. The CDC’s definition is likely to create misclassification bias by making it more likely that a temporally unrelated symptom or condition after SARS-CoV-2 infection is improperly labelled long COVID.

**Lack of a control group**

Given long COVID’s current broad definition, researchers have a most basic obligation to compare the nature and prevalence of reported symptoms among cases to a control population, which ideally would be similar to the cases in demographics, underlying health, geography and time. However, one recent systematic review identified control groups in only 22/194 (11%) of long COVID studies. In this particular review, around 45% of those with COVID-19 had one unresolved symptom 4 months after diagnosis, but this review did not estimate the prevalence among the uninfected in the 22 studies with a control group.

Another systematic review reported a PASC prevalence of 25% in children but, again, did not consider symptom prevalence among controls, citing ‘heterogeneity in the definition.’ The same authors also reported a PASC prevalence of 80% in adults in a 2020 systematic review. Not only did they not compare cases with controls, but they also included studies with a short median follow-up of only 1 month, studies that did not specify length of follow-up and studies that included abnormal laboratory results as ‘symptoms.’ Lack of control groups, convenience sampling and heterogeneity of follow-up time have made drawing conclusions from systematic reviews challenging. If systematic reviews include studies with major methodological limitations, they should refrain from providing prevalence estimates which are likely to be less accurate and with wider confidence intervals than well conducted individual studies.

A more recent publication from Norway of children and young people aged 12–25 used a modified Delphi definition for long COVID (table 1) and found a strikingly high point prevalence of those meeting the case definition of post–COVID-19 condition and controls (the latter being SARS-CoV-2 seronegative) of 48.5% among SARS-CoV-2–positive cases and 47.1% in the control group, which was not significantly different. This study demonstrates why it is critical to have a control group when the definition of a condition is vague and includes numerous common symptoms, particularly when alternative causes could not be entirely ruled out, as described in the study (Figure S1, Selvakumar et al29) and to us (J Selvakumar, personal communication, 25 April 2023).

**Inappropriately-matched controls**

Not only should control groups be included, but they should also be properly matched to cases, ideally by age, sex, geography, socioeconomic status and, if possible, underlying health and health behaviours. The CDC, for example, estimated 38% of case-patients experienced an incident condition within a year of COVID-19 diagnosis documented in the electronic health record compared with 16% of controls. However, they failed to acknowledge that those who are diagnosed with COVID-19 in healthcare settings tend to be less healthy at baseline than those who do not seek COVID-19 testing in the healthcare system, which could have biased the estimate by including more severe cases in the post–COVID group and less severe in the controls. Additionally, the study did not describe how participants were matched and provided no information about underlying health, age or socioeconomic status of cases or controls. Researchers should also, to the best of their ability, ensure cases have been infected and controls have not, but in this study there was no attempt to link the timing of ongoing symptoms with SARS-CoV-2 infection among cases, or to rule out a history of SARS-CoV-2 infection in the controls.

As another example, the US Veterans Affairs (VA) research has produced misleading results because those who received a diagnosis of COVID-19 through the VA (as opposed to being asymptomatic or mildly asymptomatic and testing at home or not testing at all) have fundamentally different health status than controls. The authors themselves described the cases as being predominantly white, male, older, more obese, on multiple regular medications and having poorer underlying health than the general population; thus, it was expected they would also have very high rates of multiple symptoms and outpatient encounters post–COVID-19.

**Control groups created using test-negative design**

Having a SARS-CoV-2 negative control group with upper respiratory symptoms may provide better context for understanding the risks and prevalence of PASC compared with other respiratory viral illnesses. Theoretically, this can be achieved with a test-negative design. However, this design is prone to bias as

<table>
<thead>
<tr>
<th>Table 1 International health organisations’ various definitions of long COVID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organisation</strong></td>
</tr>
<tr>
<td>Centers for Disease Control and Prevention (CDC)</td>
</tr>
<tr>
<td>World Health Organization (WHO)</td>
</tr>
<tr>
<td>National Institute for Health and Care Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN) and Royal College of General Practitioners (RCGP)</td>
</tr>
<tr>
<td>Delphi definition for children and young people</td>
</tr>
</tbody>
</table>
test-positive individuals are not the same as test-negatives and this can affect results in both directions.

For example, a prospective Swiss study using the test-negative design found that those testing SARS-CoV-2 PCR positive during the omicron period were more likely to live with children, be employed and be younger than test-negative controls. Even so, the difference in PASC prevalence at 12 weeks between cases (11.7%) and test-negative controls (10.4%) was small (1.3%). Most importantly, the only significant differences in symptom prevalence were loss of taste and smell, and insomnia; the latter could easily be explained by confounding due to demographic differences between cases and controls. This study, however, suffered from misclassification bias by only considering ‘symptoms with new onset after the test date’ and, therefore, would have simultaneously missed symptoms that were continuous from the first COVID-19 symptoms and, instead, included new, potentially unrelated symptoms that developed within the 12 week post-diagnosis period.

**Sampling bias**

Sampling bias occurs when certain members of a population have a higher probability of being included in a study sample than others. This type of bias can lead to a non-representative sample, which may limit the generalisability of a study’s findings.

During the early stages of the pandemic, when SARS-CoV-2 testing was not widely available, studies were more likely to include a non-representative sample of SARS-CoV-2-positive patients by including fewer patients with mild or no symptoms. On the other hand, studies that employed SARS-CoV-2 antibody seroprevalence to identify cases and controls instead of relying on rt-PCR or rapid testing are less prone to this bias. Two studies used this methodology and found no significant difference in the prevalence of long COVID between cases and controls. Future studies should also take into account that seroconversion to anti-nucleocapsid antibodies was more than 90% prior to vaccination, but appears to be lower at only around 40% after vaccination. Seroprevalence will also be of limited value in populations with repeated infections given the long half-life of anti-nucleocapsid antibodies of around 283 days.

Study results may also be biased towards poorer health outcomes if participants are recruited by advertising the study as pertaining to COVID-19 recovery or long COVID. People who are experiencing lasting symptoms after COVID-19 may be motivated to participate, potentially because they believe doing so may provide insight into their own condition or help others experiencing similar symptoms. For example, this type of sampling bias due to self-selection was described as a possible limitation in the Zurich SARS-CoV-2 cohort study of adults. This study found one in four patients did not report feeling fully recovered 6–8 months after their COVID diagnosis. The authors suggested ‘individuals who were more concerned with their health or experiencing symptoms related to post-COVID-19 syndrome may have been ‘more likely to participate.’ If present, this self-selection may have provided a non-representative sample with more symptomatic participants after COVID-19 and overestimated the prevalence of lasting symptoms. Recruiting patients without advertising the study as pertaining to long COVID may help reduce this bias. However, without information on symptoms among non-participants, the presence and effect of this bias are difficult to ascertain. This is of particular concern if the study participation rate is low; in the Swiss study the participation rate was only around one in three. Furthermore, beyond the potential sampling bias, the Swiss study did not include a control group, which could have provided important context about ongoing symptom prevalence among uninfected people.

A subsequent adult cohort study from the same group of Swiss researchers, which also looked at lasting symptoms following infection with the alpha strain had a similarly low participation rate of 35%. They found those who agreed to participate did have a slightly higher rate of symptomatic infection (86% vs 79.5%) than non-participants, suggestive of sampling bias. This subsequent study also included an uninfected comparator group obtained via a Swiss seroprevalence study. At 24 months, there was an adjusted 17% difference in self-reported symptoms between the infected and the comparator group. Beyond sampling bias, differences in underlying health, age, education and employment status between the infected and uninfected comparator group may have been impossible to fully adjust for, which the authors concede. This again highlights the importance of appropriately matched controls when attempting to define the nature and prevalence of long COVID. However, that ‘taste and smell alterations’ were essentially absent in the uninfected and still present in around 10% 6 months post-infection, with an odds ratio of 26 between infected and controls, speaks strongly to this being a real lasting symptom of COVID-19, at least from the alpha variant.

The most well-designed studies provide reassuring estimates

In the UK, national surveys conducted by the Office for National Statistics (ONS) continue to report a 2.9% prevalence of self-reported long COVID in adults and children. Yet, when a control group was included with age, sex, health and socio-demographically matched controls, the prevalence of any of 12 common symptoms was 5.0% at 12–16 weeks after infection compared with 3.4% in a control group without a positive SARS-CoV-2 test, demonstrating the relative commonness of these symptoms in the population at any given time. There was no significant difference in symptom prevalence between cases and controls among people younger than 50 years, though the analysis was only able to detect a 3% difference between groups. Notably, too, this national study was performed prior to the omicron variant, which has been associated with significantly lower prevalence of persistent symptoms compared with previous variants, with one UK study estimating 0·24–0·50 odds of long COVID with the omicron versus the delta variant.

Supporting these findings, a well designed Swiss study used antibody seroconversion during the study period to confirm SARS-CoV-2 infection in children. In randomly assigned school classes at the end of 2020, they found essentially the same prevalence of lasting symptoms among 12–16 year olds who had been infected compared with those who had not been. Specifically they found 9% of antibody-positive children had at least one symptom after 4 weeks compared with 10% of those without antibodies. Though the study was small, the authors should be commended for including the most representative group of children exposed to SARS-CoV-2 in a study which excluded both biases such as testing and health-seeking behaviours and avoided over-representation of severe or hospitalised cases. This study highlights the importance of well conducted studies, even with small sample sizes; these can be more informative than systematic reviews that include studies with serious methodological shortcomings.

**Implications for current practice and future research**

Our analysis indicates that, in addition to including appropriately-matched controls, there is a need for better case definitions and
Box 1  Recommended criteria for epidemiological research of long COVID

- Avoid misclassification bias: Include clear case definitions, with every attempt to avoid improper attribution of non-specific common and non-pathological symptoms to SARS-CoV-2 infection. Establish long COVID as a diagnosis of exclusion.
- Avoid selection bias: Include representative cases and controls to allow extrapolation of findings to the general population.
- Avoid detection bias: Monitor symptoms and signs through longitudinal studies rather than cross-sectional studies.
- Avoid confounding by underlying health: Include properly matched controls when establishing incidence and prevalence. Account for pre-infection physical and mental health status of cases and controls.
- Avoid information bias: Require diagnostic evidence of SARS-CoV-2 infection in cases and lack of infection in controls.
- Avoid sampling bias: Include a representative sample of participants who do not differ from non-participants in terms of severity or duration of symptoms.
- Avoid mischaracterization: Collect data over a longer time period to describe the different courses and progression of different symptoms over time, given that most symptoms improve with time.
- Reduce diagnostic ambiguity: Attempt to identify specific symptoms or syndromes that emerge clearly linked to SARS-CoV-2 infection and are absent or substantially less frequent in controls. Create names and diagnostic criteria for specific post-COVID symptoms and syndromes for future study.

more stringent PASC criteria, which should include continuous symptoms after confirmed SARS-CoV-2 infection and take into consideration baseline characteristics, including physical and mental health, which may contribute to an individual’s post-COVID experience.

When limiting studies to those with acceptable PASC definitions and appropriate controls, we find little to no difference in the prevalence of reported persistent symptoms in children by 4 weeks or in adults younger than 50 years by 12 weeks post-infection compared with controls. It is noteworthy that the findings of the highest-quality research stand in contrast to much of what is reported in the media. Such high-quality studies can and should be used to reassure the public about the risks of PASC.

Importantly, however, even large-scale population-based studies are currently unable to rule out or estimate rarer post-infectious symptoms associated with SARS-CoV-2 infection, some of which may be debilitating. For a symptom or syndrome to be truly defined as post-COVID, it needs to be specific to— or at least a characteristic feature of—SARS-CoV-2 infection (such as anosmia). It may in the future be preferable to have different names for specific sequelae which are found to arise after SARS-CoV-2 infection, such as post-COVID-19 anosmia, rather than using the umbrella term ‘long COVID’. We also need better studies comparing the prevalence of well described post-infectious syndromes associated with other respiratory viruses, especially influenza, such as shortness of breath after severe pneumonia or debilitation and fatigue after intensive care admission.

In summary, the results of well designed population-based studies of long COVID in adults and children have been reassuring. However, taken together, the existing literature is replete with studies with critical biases that clinicians and researchers alike should be aware of. To this end, we have listed common pitfalls identified in long COVID research in box 1.

Ultimately, biomedicine must seek to aid all people who are suffering. In order to do so, the best scientific methods and analysis must be applied. Inappropriate definitions and flawed methods do not serve those whom medicine seeks to help. Improving standards of evidence generation is the ideal method to take long COVID seriously, improve outcomes, and avoid the risks of misdiagnosis and inappropriate treatment.

Twitter Tracy Beth Høeg @tracybethhoeg, Shamez Ladhani @ShamezLadhani and Vinay Prasad @VPrasadMDMPH

Contributors This article was conceived of by VP and TBH, stemming from concerns about potential negative societal effects of numerous epidemiological investigations which had, with inappropriate methodology, likely overestimated the prevalence of long COVID. In the planning stages of this article, SL was contacted about being a co-author because of his expertise in the subject matter. TBH is a physician and PhD epidemiologist who has led or been senior author on multiple investigations on COVID-19 transmission and risk-benefit analyses, particularly in children and young adults. VP is a professor of epidemiology and biostatistics who specialises in evaluating medical and epidemiological evidence, has published over 450 papers, with two dozen ongoing or forthcoming on COVID-19. SL has developed, led and contributed to more than 100 publications on COVID-19, including on long COVID, with particular focus on paediatric COVID-19 and post-COVID conditions. TBH is the guarantor of the article.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Tracy Beth Høeg http://orcid.org/0000-0002-2341-6573
Shamez Ladhani http://orcid.org/0000-0002-0856-2476
Vinay Prasad http://orcid.org/0000-0002-6110-8221
References


