Developing Inclusive and Meaningful Surveys for Long-COVID Impact: Findings from an Australian Usability Study

Dr Kate Anderson[1], Dr Jacob Sheahan[2], Mr Nelson Lay[2]
[1]School of Computing Technologies, RMIT University
[2]School of Health and Biomedical Sciences, RMIT University

Introduction

Our team is undertaking a national survey to better understand the impacts of Long COVID in Australia. The survey combines several patient-reported measures including:

- COVID-19 Yorkshire Rehabilitation Scale (C19-YRS)
- EuroQoL – 5 Dimensions Quality of Life Scale (EQ-5D)
- International Physical Activity Questionnaire (IPAQ)
- Custom questions co-developed with our Lived Experience Advisory Network.

This poster presents key findings from our survey user testing phase that can guide the design of future Long-COVID evaluations.

Background

While tools like the C19-YRS have shown strong psychometric properties, limited research has examined their *usability.* Survey usability issues such as wording, recall demands, cultural appropriateness, and emotional burden can impact both data quality and participant wellbeing (e.g., Nguyen et al., 2021; Burgess et al., 2022). User testing offers a systematic way to identify these barriers through direct observation and feedback (Hornbæk, 2006, Hughes, 2018), ensuring that survey tools are not only statistically valid but also practical, accessible, and sensitive to the needs of people with Long COVID.

Methods

Draft versions of the survey were iteratively tested with 14 adults living with Long COVID to address common usability challenges that undermine data quality, including item wording, interpretation, accessibility, recall, and cognitive or emotional workload. Testers included 9 women, 4 men, and 1 nonbinary person, located across VIC, NSW, QLD, SA. Four testers identified as LGBTIQ+. Ages ranged from 36 – 67 years, and testers had lived with Long-COVID for 19 months to 4 years.

Testing sessions lasted approximately 90-minutes and were completed online (via Microsoft Teams). The sessions used a 'Think Aloud' method where participants verbalised their thoughts and feelings in real time. This was followed by a retrospective, semi-structured cognitive interview where participants could review and clarify their reactions to the survey.

A thematic analysis is currently underway. Deductive coding draws on two complementary frameworks:

- Tourangeau's four-stage model of survey response (comprehension, retrieval, judgment, and response) provides a structured lens for identifying where items may introduce cognitive burden or misinterpretation (Tourangeau, 2020).
- Nielsen's Usability Heuristics (Nielsen, 2024) examines design features that affect usability, including clarity of wording, consistency, error prevention, and support for accessibility.

Analysis is supplemented with inductive coding to capture emergent sociocultural, linguistic, and emotional considerations, and developing insights are periodically reviewed with the LEAN and survey development team.

References

Hornbæk, K. (2006). *International journal of human-computer studies*, *64*(2), 79-102. https://doi.org/10.1016/j.ijhcs.2005.06.002
Hughes, D. J. (2018). *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 751-779. https://doi.org/10.1002/9781118489772.ch24
Nielsen, J. (2024). *Nielsen Norman Group.* 10 Usability Heuristics for User Interface Design - NN/g
Tourangeau, R., Maitland, A., Steiger, D., & Yan, T. (2020). *Advances in questionnaire design, development, evaluation and testing*, 47-73.

Emerging Findings: **Insights for Long COVID Survey Development and Interpretation**

IMPACT: It's important to people with Long-COVID that research captures both their *worst experiences* and their *current state*.

→ If surveys only ask about the last 7 days, participants may distort answers so their worst impacts are still represented.

"I feel like I'm being slightly swayed toward lower scores on this side [pre-COVID] and higher scores on this side [current], because I have a vested interest in the outcome of the research, and I'm not the researcher, so I don't have to do like bracketing and good stuff like that".

"I want to tell you how bad it was. Like, I really want to tell you how awful it's been."

SEVERITY: Severity ratings are strongly shaped by context: comparisons to others, coping strategies, and co-morbid conditions.

→ People often down-rated themselves because they knew others were worse off, or because they had learned to carefully manage activity.

"**Everything can be done with enough planning and enough assistance and a wheelchair. "**

"I can't go lower [on the scale] because I know how immense the health suffering of others are, and I'm so much better than that."

SERVICES: Questions about service use/experience, support, and economic impacts are complex to model within standard questions.

→ People distinguished between *accessing*, *actively using*, and *being satisfied* with services and supports. Service experience varied across providers, and unaffordable services were important to capture.

→ Financial situations and impacts fluctuated with time, circumstance, and Government intervention.

"I feel like I'm saying it's a bad doctor, but it's not. It's just hugely under resourced and lack of information so he has given me conflicting advice. It's a system problem, not a doctor problem."

MEDICATIONS: Testers used a mix of off-label prescriptions, over-the-counter drugs, and pre-existing treatments. The relevance of medications may be unclear.

→ Medication and supplement use won't be captured reliably through administrative data (PBS/MBS) alone.

→ Side effects from medications influenced the overall picture of Long-COVID symptoms and could be difficult to disentangle.

"There's no long COVID pill… Most people would rather provide that information and know that what they've said to you is accurate. ".

Physical activity is inconsistently interpreted. Standard definitions (vigorous/moderate) often didn't match participants' reality, especially with post-exertional malaise (PEM). Anchoring activity questions to personal baselines or capability profiles would be more meaningful.

→ Asking people to characterise their baseline/lowest/current exercise level (e.g., physical activity scale or profile tool) before asking about frequency or duration of exercise at that level, may be more relevant and more sensitive to change across the severity levels and diverse baseline profiles.

Survey burden is high, but participants expressed willingness to persist if they felt heard and supported.

→ Topics including weight, financial cost, suicide/self-harm, and social/family connection can be sensitive or triggering. Transparency about *why* sensitive topics are included can help.

→ Support can include provision of timely, relevant, and actionable support links, and messages of encouragement from others with Long COVID.

→ Participants may need reminders to take a break during long surveys to prevent emotional and physical exhaustion. Flexibility in completion format and return time is valued.

Next Steps: **Co-Designing a Roadmap for Trauma-Informed, Accessible Long COVID Survey Research.**

If you are involved in survey design and would like to be part of this work or hear more about our findings, please reach out to **longcovid.project@rmit.edu.au**

Long COVID Assessment Management Practice (LAMP)