Predicting panel attrition in longitudinal HRQoL surveys during the COVID-19 pandemic in the US

Tianzhou Yu, MPH[1]; Jiafan Chen, MS[1]; Ning Yan Gu, PhD[2]; Joel W. Hay, PhD[3]; Cynthia L. Gong, PharmD, PhD[3,4]

[1]Department of Pharmaceutical and Health Economics, School of Pharmacy, University of Southern California, Los Angeles, CA; [2]School of Nursing and Health Professions, University of San Francisco, San Francisco, CA; [3]Leonard D. Schaeffer Center for Health Policy and Economics, School of Pharmacy, University of Southern California, Los Angeles, CA; [4]Fetal & Neonatal Institute, Division of Neonatology, Children's Hospital Los Angeles, Department of Pediatrics, Keck School of Medicine, University of Southern California

Background

The COVID-19 pandemic has led to an explosion of research conducted via online platforms, which provide an effective and efficient strategy to collect data on COVID-19 within a short time period.[1] However, online longitudinal surveys may be subject to potential biases due to sample attrition.[2] This study was designed to identify potential predictors of attrition using a longitudinal quality of life panel survey collected during the COVID-19 pandemic.

Methods

We collected three waves of survey data from April 2020 to March 2021. Wave 1 data were collected from Apr 1st to May 6th, 2020 (n=2,734). Wave 2 started on July 4th and ended on Sept 4th, 2020 (n=2,454). Wave 3 collected data from Jan 10th to Mar 15th, 2021 (n=2,252). We used Amazon's Mechanical Turk (MTurk) platform to field the survey. Respondents aged 18 years or older registered as "workers" in MTurk are eligible to participate in our survey. There were no other exclusion criteria. Age, gender and race were stratified like the general US population to ensure sample generalizability. Participants were compensated €1.50 for each survey. Informed consent was obtained at the beginning of the survey.

We collected information on various sample characteristics including respondents' demographics, COVID-19 status, HRQoL, and behavior, employment, and productivity changes related to COVID-19. HRQoL was measured primarily using the EuroQol EQ-5D-5L (EQ-5D) and the Veterans Rand 12-Item Health Survey (VR-12), supplemented by questions adopted from several other questionnaires (e.g., PHQ-4, CD-RISC 2, PROMIS, BRFSS, etc.).

Missing responses were imputed based on the sample. We employed standard logistic regression with stepwise selection to identify the most parsimonious model that predicted the attrition in wave 2 and wave 3. Age, sex, race, and ethnicity were fixed in the regression model based on background knowledge.

Reference

1. De Man J, Campbell L, Tabana H, Wouters E. The pandemic of online research in times of COVID-19. BMJ Open. 2021;11(2):e043866. Published 2021 Feb 23. doi:10.1136/bmjopen-2020-043866
2. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [published correction appears in doi:10.2196/jmir.2042]. J Med Internet Res. 2004;6(3):e34. Published 2004 Sep 29. doi:10.2196/jmir.6.3.e34

Table 1. Selected factors associated with panel attrition (Odds ratio & 95% CI)

Predictor in the previous wave	Attrition at wave 2	Attrition at wave 3
Age group (ref: 18-24)		
25-34	0.596* (0.423-0.839)	0.636 (0.375-1.079)
35-44	0.553* (0.388-0.786)	0.378* (0.219-0.654)
45-54	0.426* (0.291-0.625)	0.313* (0.175-0.560)
55-64	0.735 (0.508-1.064)	0.241* (0.133-0.437)
65+	0.620* (0.409-0.939)	0.078* (0.038-0.159)
Gender (ref: Male)		
Female	1.064 (0.903-1.255)	1.076 (0.872-1.328)
Race (ref: White)		
American Indian or Alaska Native	1.259 (0.447-3.548)	0.673 (0.181-2.506)
Asian	0.925 (0.668-1.281)	0.573* (0.381-0.861)
Black or African American	0.906 (0.660-1.244)	0.907 (0.618-1.332)
Multiple races	1.630* (1.297-2.048)	0.653 (0.351-1.217)
Native Hawaiian or Pacific Islander	2.674 (0.254-28.129)	<.001 (N/A b/c sample size)
Ethnicity (ref: Non-Hispanic)		
Hispanic	1.698* (1.266-2.279)	2.097* (1.344-3.273)
Medical Care deferred due to COVID-19	-	1.612* (1.238-2.101)
Diagnosed with COVID-19	-	5.026* (2.026-12.473)
Experienced COVID-19-like symptoms not serious enough to require hospitalization	1.384* (1.052-1.821)	-
Support social distancing policy	-	1.047* (1.004-1.092)
Employment change (ref: No change)		
Work form home	-	1.327* (1.067-1.651)
Lost job	-	2.028* (1.035-3.975)
Laid off temporarily	-	0.577* (0.337-0.988)
Work deemed essential	1.223* (1.020-1.466)	-
COVID-19 impact on productivity	1.061* (1.028-1.095)	-
Self-rated survey difficulty	1.075* (1.026-1.126)	1.125* (1.061-1.193)
EQ-5D-5L VAS score	-	1.011* (1.005-1.017)
VR-12 Q2a, Moderate activity (ref: Not limited at all)		
Limited a little	1.355* (1.088-1.688)	-
Limited at lot	0.885 (0.608-1.288)	-

Table 1. Cont'd

Predictor in the previous wave	Attrition at wave 2	Attrition at wave 3
VR-12 Limited in work or activities due to physical health (ref: None of the time)		
Some of the time	-	2.234* (1.524-3.276)
A little of the time	-	1.502* (1.107-2.039)
Most of the time	-	2.221* (1.367-3.609)
All of the time	-	1.404 (0.748-2.636)

Results

Compared to the general US population, our wave 1 sample was slightly older, more likely to be single, and had higher education level. Wave 2 participants were younger, less likely to be female, but the race and ethnicity composition was comparable. A variety of variables were associated with attrition. Some were consistently predictive of attrition (e.g., younger age, survey difficulty, and being Hispanic). Some were always associated with attrition, but the association varied (e.g., race). The remaining variables' association with attrition suggested a context-specific relationship.

Conclusions & Limitations

A multitude of factors can affect attrition in online longitudinal surveys. Our findings suggested that researchers should look into more than commonly identified demographic factors to evaluate attrition in their own study. Moreover, we established a framework of 3 types of attrition predictors, each with different level of context-dependence. This may help future research to evaluate potential factors that may lead to non-random attrition and therefore improve study design and obtain more appropriate data interpretation.

Limitations: 1) Participants from the MTurk platform may have mixed external validity to the general US population. 2) Wave 2 sample included participants from wave 1, who were less likely to attrite, given their continued participation in wave 2. 3) The gap between wave 2 and wave 3 was longer than that between wave 1 and wave 2, which could lead to increased attrition rates.

Contact

Tianzhou Yu, MPH, tyu09191@usc.edu