Identification of key factors related to digital health observational study adherence and retention by data-driven approaches: an exploratory secondary analysis of two prospective longitudinal studies

Peter J Cho, Iredia M Olaye*, Md Mobashir Hasan Shandhi*, Eric J Daza, Luca Foschini, Jessilyn P Dunn*

Lancet Digit Health 2025; 7: e23–34

See **Comment** page e4

*Contributed equally

Biomedical Engineering Department (P J Cho BA, M M H Shandhi PhD, Prof J P Dunn PhD) **and Biostatistics and Bioinformatics Department** (Prof J P Dunn), **Duke University, Durham, NC, USA; Evidation Health, San Mateo, CA, USA** (I M Olaye PhD)**; Stats-of-1, Palo Alto, CA, USA** (E J Daza DrPH)**; Sage Bionetworks, Seattle, WA, USA** (L Foschini PhD)

Correspondence to: Dr Jessilyn P Dunn, Biomedical Engineering Department, Duke University, Durham, NC 27708, USA
jessilyn.dunn@duke.edu

Summary

Background Longitudinal digital health studies combine passively collected information from digital devices, such as commercial wearable devices, and actively contributed data, such as surveys, from participants. Although the use of smartphones and access to the internet supports the development of these studies, challenges exist in collecting representative data due to low adherence and retention. We aimed to identify key factors related to adherence and retention in digital health studies and develop a methodology to identify factors that are associated with and might affect study participant engagement.

Methods In this exploratory secondary analysis, we used data from two separate prospective longitudinal digital health studies, conducted among adult participants (age ≥18 years) during the COVID-19 pandemic by the BIG IDEAs Laboratory (BIL) at Duke University (Durham, NC, USA; April 2, 2020 to May 25, 2021) and Evidation Health (San Mateo, CA, USA; April 4 to Aug 31, 2020). Prospective daily or weekly surveys were administered for up to 15 months in the BIL study and daily surveys were administered for 5 months in the Evidation Health study. We defined metrics related to adherence to assess how participants engage with longitudinal digital health studies and developed models to infer how demographic factors and the day of survey delivery might be associated with these metrics. We defined retention as the time until a participant drops out of the study. For the purpose of clustering analysis, we defined three metrics of survey adherence: (1) total number of surveys completed, (2) participation regularity (ie, frequency of filling out surveys consecutively), and (3) time of activity (ie, engagement pattern relative to enrolment time). We assessed these metrics and explored differences by age, sex, race, and day of survey delivery. We analysed the data by unsupervised clustering, survival analysis, and recurrent event analysis with multistate modelling, with analyses restricted to individuals who provided data on age, sex, and race.

Findings In the BIL study, 5784 unique participants with the required demographic data completed 388 600 unique daily surveys (mean 67 [SD 90] surveys per participant). In the Evidation Health study, 89 479 unique participants with the required demographic data completed 2 080 992 unique daily surveys (23 [32] surveys per participant). Participants were grouped into adherence clusters based on the three metrics of adherence, and we identified statistically discernible differences in age, race, and sex between clusters. Most of the individuals aged 18–29 years were observed in the clusters with low or medium adherence, whereas the oldest age group (≥60 years) was generally more represented in clusters with high adherence than younger age groups. For retention, survival analysis indicated that 18–29 years was the age group with the highest risk of exiting the study at any given point in time (BIL study, hazard ratio [HR] for 18–29 years *vs* ≥60 years, 1·69 [95% CI 1·53–1·86; p<0·0001]; Evidation Health study, HR 1·50 [1·47–1·53; p<0·0001]). Sex and race were not discernible predictors of retention in the BIL study. In the Evidation Health study, male participants (*vs* female participants; HR 0·96 [0·94–0·98]; p<0·0001) and White participants (*vs* Asian participants; HR 0·96 [0·93–0·98; p=0·0004) had a lower risk of study exit, and Other race participants (*vs* Asian participants) had a higher risk of study exit (HR 1·10 [1·06–1·14; p<0·0001]). Recurrent event analysis confirmed age as the factor most associated with adherence; for the 18–29 years age group (*vs* ≥60 years group), the transition intensity from an active to inactive state per day in the BIL study was 1·661 (95% CI 1·606–1·718) and in the Evidation Health study was 1·108 (1·094–1·121). Participation patterns were variable by race and sex between the studies.

Interpretation Our analyses revealed that age was consistently associated with adherence and retention, with younger participants having lower adherence and higher dropout rates than older participants. Unsupervised clustering and survival analyses are established methods in this field, whereas the use of recurrent event analysis, was, to our knowledge, the first instance of the application of this method to remote digital health data. These methods can help to understand participant engagement in digital health studies, supporting targeted measures to improve adherence and retention.

Funding US National Science Foundation, US National Institutes of Health, Microsoft AI for Health, Duke Clinical and Translational Science Institute, North Carolina Biotechnology Center, Duke MEDx, Duke Bass Connections, Duke Margolis Center for Health Policy, and Duke Office of Information Technology.

Introduction

Smartphones, smartwatches, and activity trackers are ubiquitous in the USA, with 85% of the US population owning smartphones and 31% owning smartwatches in 2021.[1–3] These technologies provide an opportunity to capture continuous information regarding physiology and behaviour in real-world settings and have specific health-tracking benefits.[4–6] Given their low cost and high accessibility, many digital health technologies can address health-care inequities by reaching traditionally underserved populations. However, not all people will benefit equally unless challenges in equitable technology, study design, and implementation are addressed.

A substantial challenge is the under-representation of underserved groups in digital health research, including older adults, minoritised racial and ethnic groups (such as Black, Hispanic, and Indigenous populations), individuals with limited English proficiency, and those with lower socioeconomic status. These groups often have low access to digital health tools, low rates of technology adoption, and unique health challenges that might not be well represented in current digital health datasets.[7,8] This lack of representation limits the generalisability of learnings from these studies and might widen existing disparities.[9,10] Such unbalanced study populations are particularly problematic for machine learning-based technologies, which require representative data to ensure equitable functionality to avoid biased or incorrect results for particular groups.[10] For example, underserved groups have had low adherence in remote longitudinal digital health studies that aim to validate new technologies.[11] In the QUASAR and Cloudy with a Chance of Pain studies, there were challenges in maintaining engagement among older adults, Black and Hispanic populations, and those with chronic conditions, with factors such as usability of the technology, participant burden, and limited familiarity with digital platforms contributing to this disengagment.[12,13] Studies can also be limited by low adherence when participants take a break from filling out surveys or from wearing a smartwatch or other wearable devices regularly. The source of the low adherence is particularly challenging to identify when adherence characteristics vary substantially between specific demographic factors.[14]

Another challenge related to adherence is the retention rate of study participants. To tackle the issue of enrolment and retention, a growing body of literature is providing

Research in context

Evidence before this study

Primary methods of characterising study participant retention and adherence in digital health studies have focused on survival analysis and exploratory data analysis. We searched PubMed from Jan 1, 1990 (the year when "digital health" was first cited), to March 1, 2023, for studies in English, using combinations of words or terms that included "adherence", "retention", "recurrent event analysis", OR "digital health". The search identified a range of studies exploring participant retention and adherence in digital health. Most of these studies used survival analysis to estimate time-to-dropout and exploratory data analysis to describe adherence trends. Previous studies identified key factors associated with adherence and retention, including participant age, gender or sex, socioeconomic status, study incentives, and digital literacy. However, there has been limited focus on modelling the recurrent nature of engagement behaviours in digital health, which our study aimed to address by applying recurrent event analysis. In addition, previous studies have not assessed both retention and adherence.

Added value of this study

The overall aim of this work was to identify factors in digital health study design via a repurposed methodology that could be targeted to improve representation of study data. We found that age was inversely related to survey adherence and study retention, which is consistent with the literature. Furthermore, this study is the first to apply recurrent event analysis to model study engagement in digital health studies.

Implications of all the available evidence

Designers of prospective digital health studies need to understand how participants engage to improve data collection methods. Modelling methods that are appropriate for longitudinal data with varying amounts of missingness can guide the development of strategies to improve study engagement and garner representative datasets. The evidence suggests that demographic factors, including age, socioeconomic status, and digital literacy, can affect retention and adherence rates. Young adults often exhibit higher initial engagement in digital health studies than older adults, but might have lower adherence over time. To address this pattern, studies might benefit from tailored engagement strategies that resonate with specific age groups and demographic characteristics, such as age-appropriate incentives for young adults or technical support for older participants. Recognising these differences can help in designing inclusive digital health studies that maintain high retention and adherence across diverse participant groups.

guidance on how to increase diversity among participants in digital health studies. Strategies include forming partnerships with community organisations (eg, faith-based organisations and cultural centres) for recruitment and offering complimentary wearable devices or internet services to have participants remain in studies.[7,15]

Low adherence and retention can adversely affect the external validity of research results. Statistical approaches have characterised retention and adherence,[16,17] but innovation is needed for longitudinal studies. For example, in survival analysis, an event only occurs once, and although digital health studies often employ such methods to evaluate time-to-event for adherence, in reality, these methods fail to involve multiple recurring events (eg, not responding to a survey more than once), elaborated on previously in the literature[18] and in a preprint paper.[19] Standard survival analysis methods (eg, Kaplan–Meier curves and Cox regression models) describe hazard rates for the occurrence of single final events, but cannot model competing events and transitions.[19,20] Thus, traditional methods of survival analysis are not best suited to conceptualising or analysing adherence in digital health studies.

To address these limitations, we sought to develop and apply alternative methods that could better capture the nuances of recurring participation behaviours in digital health studies, and identify factors that affect these behaviours (figure 1). In the current study, we characterised participant retention and survey adherence patterns in two large-scale, longitudinal digital health studies, specifically in the context of influenza-like illness surveillance (which is commonly studied at a population level for long periods of time),[21–23] and explored whether and how demographic factors and the day of survey delivery contribute to survey adherence.

Methods

Data collection and preprocessing

The data utilised for this study were derived from a partnership between Evidation Health, a technology company located in San Mateo, CA, USA, which generates insights from real-world behavioural data via intensive longitudinal methods, and the BIG IDEAs Laboratory (BIL) at Duke University in Durham, NC, USA. Evidation Health and the BIL each did a separate prospective longitudinal digital health study during the COVID-19 pandemic, collecting wearable physiological and behavioural data and electronic patient-reported diagnoses and symptoms. The studies have been reported on previously in the literature[6,24] and a preprint paper.[25]

Evidation Health launched the Daily Surveillance of COVID-19 Symptoms and Experience, also referred to as the COVID2020 study, between April 4 and Aug 31, 2020, under institutional review board protocol number 2020-0320 approved by Solutions IRB (San Mateo, CA, USA). The COVID2020 study was advertised to people who were users of the Evidation Health app. Adults (age ≥18 years) living in the USA were recruited. For 5 months, COVID-19-related data were collected prospectively on a daily basis by wearable activity trackers, health apps, and surveys via the Evidation Health app. The rationale of the study was to improve understanding of the geographical disparities of self-reported respiratory disease, influenza-like illness, and, primarily, COVID-19, and their effect on daily life.[24] Additional information on the study design, eligibility criteria, and data collection is provided in appendix 1 (p 1). The enrolment and daily surveys are provided in appendix 2.

See **Online** for appendix 1

See **Online** for appendix 2

The BIL study, referred to as CovIdentify, was launched between April 2, 2020 and May 25, 2021, under institutional review board protocol number 2020-0412 approved by the Institutional Review Board of Duke University. The aim of the study was to collect commercial wearable device data before, during, and after COVID-19 infections and utilise the data to develop an intelligent testing allocation algorithm.[6] The study recruited adult participants (age >18 years) through social media campaigns, flyers, and word-of-mouth via community centres. After enrolment, participants were given the option to donate 12 months of retrospective and 12 months of prospective wearable data and to fill out daily electronic surveys of symptoms. Initially, participants were sent daily surveys for 1 month and then weekly surveys for 2 months. The survey strategy evolved concomitantly with the pandemic based on an intermediate exploratory data analysis and was converted into daily surveys for 12 months (an additional 12 months for participants enrolled before the change) beginning on May 19, 2020. Additional information on the study design, eligibility criteria, and data collection is provided in appendix 1 (p 1). The enrolment and follow-up surveys are provided in appendix 3. In both studies, participants provided consent through remote methods to participate in the study and ethical approvals were accepted by the corresponding review boards.

See **Online** for appendix 3

Since the BIL study had different survey intervals (ie, weekly and daily), we downsampled the Evidation Health data to the weekly interval (single case approach: if the participant completed at least one survey within a given week, they were considered to have data for that week) so that we could have comparable datasets. We separated survey data into either weekly or daily intervals to identify patterns at differing frequencies. This binary aggregation generated the following four datasets: BIL daily, BIL weekly, Evidation Health daily, and Evidation Health weekly. To handle the different frequencies with which surveys were sent out from the BIL, we analysed at both daily and weekly intervals. We applied the same downsampling to the BIL daily data when conducting the weekly interval analysis. We excluded participants with only weekly data from the daily analysis. We focus on daily intervals in this paper and present weekly interval results from the clustering analysis in appendix 1 (p 3). Further details on the preprocessing of survey data are provided in appendix 1 (p 1).

For the **Evidation Health app** see https://evidation.com/blog/what-is-evidation

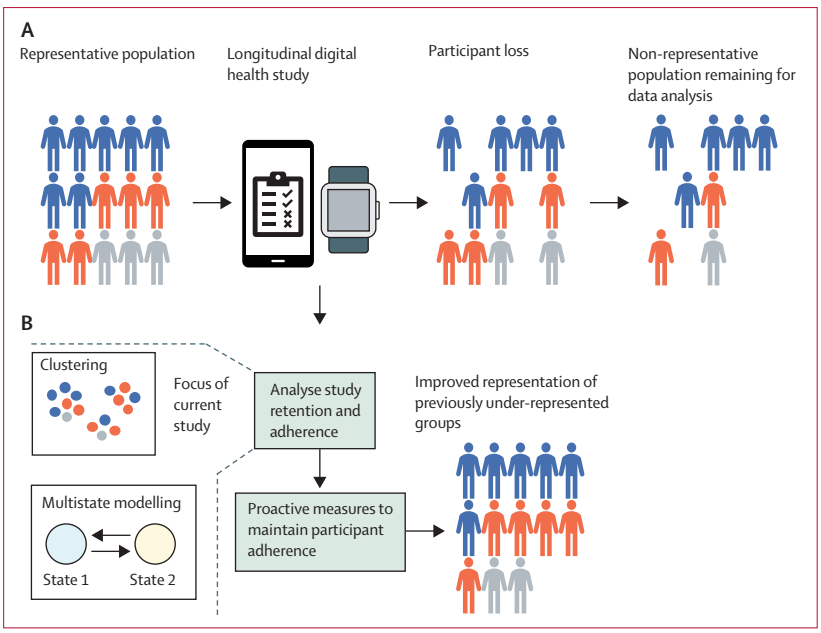


Figure 1: **Proposed method to achieve target demographics from enrolment through to final data analysis**
(A) Graphical representation of the demographics of participants enrolling in a conventional longitudinal digital health study and the subsequent study population generating the data for analysis. (B) By analysing participant engagement during the study, research teams can undertake proactive measures to mitigate challenges with participant adherence and retention.

Exploratory data analysis

For our analysis, participant retention was defined as the duration of time until the participant drops out of the study, and survey adherence was defined by how the participant engages with regular surveys (ie, survey filling patterns). Adherence is a crucial metric, particularly in longitudinal studies that seek to detect irregular events over long periods (eg, COVID-19 infection).[16,26] For the purpose of describing the clustering, we defined three dimensions or metrics for assessing adherence: (1) total number of surveys completed; (2) participation regularity (ie, frequency of filling out surveys consecutively); and (3) time of activity (ie, engagement pattern relative to enrolment time; appendix 1 p 7).

Unsupervised learning methods were used to provide an overview of survey adherence and visualise the survey filling patterns, via clustering of adherence based on the three dimensions. We identified clusters of participants by k-means clustering, using methods similar to those used in previous studies.[16,17] We used the elbow method to identify the number of clusters (appendix 1 pp 2, 4). Missing observations (no survey was filled) were labelled "false" and valid observations (survey was filled) as "true". We used heatmaps to visualise the survey filling patterns and clusters, and assessed the adherence patterns (high, medium, or low; and early or late) relative to each other.

To understand whether there were common factors that distinguished the adherence dimensions, we explored whether any identified clusters corresponded with specific demographic factors (age, sex, and race). We did χ^2 tests of independence to detect if variable distributions were statistically different between clusters, and estimated Cramér's V, which is a measurement of association magnitude (ie, the categorical variable analogue to Pearson's χ^2 statistic for continuous variables), to assess the strength of the relationship between cluster membership and each of the categorical variables[27] (appendix 1 p 2). Cramér's V values were interpreted as follows: 0–0·1, weak association; >0·1–0·3, moderate association; >0·3–0·5, strong association; and >0·5, very strong association.[27] Post-hoc analysis was conducted by comparing each cluster against one another to identify which cluster contributed to the χ^2 tests.

We subsequently used time-to-event methods to analyse study retention and survey filling patterns. After an initial literature search and exploration of digital health studies that sought to assess adherence and retention (see Research in context panel and appendix 1 pp 6–7), we selected two common methods: survival analysis and recurrent event analysis. Survival analysis is a common tool applied to clinical data for estimating the average time to an event; in the health and clinical sciences, the event is usually death or disease progression.[19,20] We used survival analysis to estimate the probability of participants exiting the study before its completion and to identify factors that might affect study participant retention. For our analyses, we relabelled completed and missing surveys to create three different states at any given survey timepoint: (1) the participant completes the survey, indicating they are active; this was labelled as "ACT"; (2) the participant misses (ie, does not complete) the survey, but completes at least one subsequent survey before the study ends, indicating that they are temporarily inactive; this was labelled as "INACT"; (3) the participant misses the survey and all subsequent surveys during the rest of the study period, indicating they exited the study; this was labelled as "EXIT".

A participant can transition back and forth between ACT and INACT states, but cannot transition to an ACT state from an EXIT state. A missing survey response was initially labelled INACT/EXIT. If a subsequent ACT state was observed, then all preceding INACT/EXIT states between that ACT state and the preceding ACT state (or the first study date, whichever occurred later) were relabelled as INACT. Otherwise, if no subsequent ACT state was observed by the end of the study period, then all preceding INACT/EXIT states occurring after the preceding ACT state (or the first study date, whichever occurred later) were relabelled as EXIT. For the unsupervised clustering analysis, we did not consider transitions but used these labels to cluster information.

For the survival analysis, we defined the failure event as the EXIT state. This definition meant that a participant who had completed surveys until the end of the study was considered to have survived (ie, been retained) for the entire period. We thereby defined retention in terms of

survival as the proportion of participants surviving (retained) on any given day. Overall retention was estimated after enrolment, along with 95% CIs. The patterns of retention were summarised with Kaplan–Meier curves[28] across demographic factors of age, sex, and race as explanatory variables. We also analysed retention as a function of the demographic factors via survival analysis using a Cox regression.[29] In the regression analysis, hazard ratios (HRs) were calculated to assess the effect of the explanatory variables on the hazard of the event (ie, study exit) occurring. 95% CIs for HRs were calculated assuming a normal distribution.

The same Kaplan–Meier survival analysis was done with the failure event as the INACT state. For this analysis, we defined the failure event as the first instance of an INACT state within 30 days of enrolment. A participant with only ACT states was considered to have survived (been retained) for this entire 30-day period (ie, the participant completed the daily survey every day of the 30-day period).

To model survey filling behaviour, we did recurrent event analysis using multistate modelling, a statistical method commonly observed in clinical studies, whereby patients can transition between different health states.[19,30] We generated metrics from this model and differentiated transition states for participants based on time-independent demographic factors (age, sex, and race, which were assumed not to change throughout the study). To meet the assumptions of multistate modelling, we assumed a two-state model, in which participants would be in either the ACT or INACT state. We did not include the EXIT state as it would require foreknowledge that a participant would not return to the study at a future point in time, and this knowledge violates the future-state independence of a multistate model.

Adherence metrics similar to those generated from the unsupervised clustering were generated from the multistate modelling. The adherence metrics were: (1) mean sojourn time (ie, the estimated average amount of time in a state) and (2) total length of stay (ie, the estimated total time in a state given a start and end time for a 90-day period). We likened the mean sojourn time to the participation regularity metric. Both metrics were descriptive in nature and not predictive. The equations for these metrics are included in appendix 1 (p 7). Additionally, our models estimated transition intensities, which represented the conditional likelihoods of each transition per day given covariate values (ie, the probability of an individual moving from ACT to INACT). A transition intensity of 1 indicated no difference in the likelihood of transition versus a reference group given the presence of the covariate, whereas a value greater than or less than 1 indicated a higher or lower likelihood of transitioning, respectively. We first explored age given that we had identified relationships between age and retention in the survival analysis, and next explored sex, race, and day of survey delivery (ie, weekday *vs* weekend). For the BIL study, we also explored the performance of the models using the presence of symptom reports (ie, if the participant reported symptoms on a given date) as input variables. Prevalence plots to represent participants transitioning from the ACT to EXIT state (theoretically moving from ACT to INACT to EXIT) by age, sex, and race were generated similar to Kaplan–Meier curves but included the expected proportion of individuals at a specific point in time given the demographics, calculated from multistate modelling. We applied the multistate modelling package msm in R (version 4.1.2).[30] The clustering and survival analyses were done in Python (version 3.8.5). For all of our models, we restricted the analyses to individuals who provided data on all three demographic factors of age, sex, and race.

In this paper, we report on statistical discernibility rather than statistical significance to address the significance fallacy, whereby statistical significance is often conflated with clinical or scientific importance.[31] Statistical discernibility provides an alternative lexicon that better reflects the evidence and avoids over-interpreting p values as rigid thresholds for meaningfulness. Our reporting of statistical discernibility is explained fully in appendix 1 (p 2). We nonetheless report thresholds and levels of statistical significance because these metrics give statistical context and are necessary for constructing confidence intervals. The threshold for statistical significance was a p value of less than 0·05.

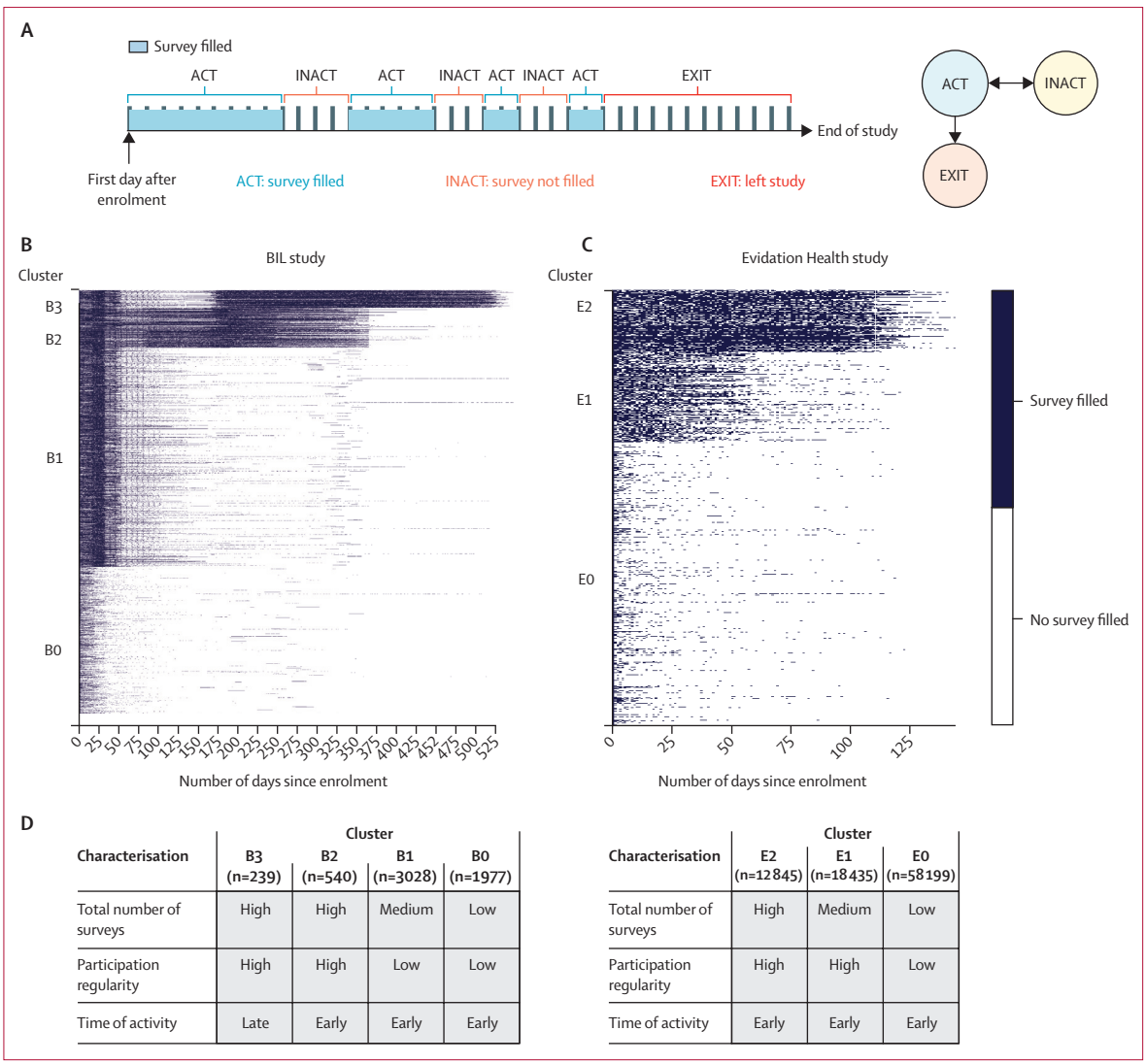
Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

	BIL study	Evidation Health study
Total participants*	5784	89 479
Sex		
Female	3521 (60·9%)	73 891 (82·6%)
Male	2263 (39·1%)	15 588 (17·4%)
Race		
White	5249 (90·8%)	75 048 (83·9%)
Black	270 (4·7%)	4943 (5·5%)
Asian	198 (3·4%)	4805 (5·4%)
Other	67 (1·2%)	4683 (5·2%)
Age, years		
18–29	498 (8·6%)	22 635 (25·3%)
30–39	1029 (17·8%)	30 604 (34·2%)
40–49	1239 (21·4%)	17 573 (19·6%)
50–59	1443 (24·9%)	13 231 (14·8%)
≥60	1575 (27·2%)	5436 (6·1%)

Percentages do not always add to 100% due to rounding. Participants were placed into a specific age range to create a categorical feature. BIL=Duke University BIG IDEAs Laboratory. *Participants who completed enrolment surveys with demographic information on sex, age, and race.

Table: **Demographic information for the BIL and Evidation Health studies**



D

Characteristation	Cluster B3 (n=239)	B2 (n=540)	B1 (n=3028)	B0 (n=1977)
Total number of surveys	High	High	Medium	Low
Participation regularity	High	High	Low	Low
Time of activity	Late	Early	Early	Early

Characteristation	Cluster E2 (n=12845)	E1 (n=18435)	E0 (n=58199)
Total number of surveys	High	Medium	Low
Participation regularity	High	High	Low
Time of activity	Early	Early	Early

Figure 2: **Adherence to digitally delivered surveys**
(A) An example participant's survey filling patterns over the course of a study. ACT represents the days surveys were completed, INACT represents the days surveys were not completed, and EXIT represents the days after the participant left the study; grey dashes represent days. The network model on the right shows possible paths between states as represented by the arrowhead direction. Once a participant exits the study, there is no path to re-enter. (B) Heatmap showing clustering of daily survey filling patterns in the BIL study. Due to the study design changing from a weekly to daily survey method, the number of days extended beyond a 12-month period. (C) Heatmap showing clustering of daily survey filling patterns in the Evidation Health study. In the heatmaps, each column represents one day and each row represents one participant. Unclustered maps are presented in appendix 1 (p 5). (D) Adherence characterisation of the different clusters for both studies (left, BIL study; right, Evidation Health study). Participants within each cluster were ordered by their study ID number. BIL=Duke University BIG IDEAs Laboratory.

Results

In the BIL CovIdentify study, 7348 adults (age >18 years) electronically consented to the study. Demographic information (sex, age, and race) and enrolment survey data were available for 5784 participants (table). These 5784 unique participants completed 388 600 unique daily surveys (mean 67 [SD 90] surveys per participant) throughout the study. In the Evidation Health COVID2020 study, 96 804 adults (age ≥18 years) living in the USA electronically consented to the study. Demographic information (sex, age, and race) and enrolment survey data were available for 89 479 participants (table). These 89 479 unique participants completed 2 080 992 unique daily surveys (mean 23 [32] surveys per participant) throughout the study. The overall data collection period in the BIL study was longer than that of the Evidation Health study (546 days *vs* 149 days, respectively).

We explored whether distinct patterns of daily survey adherence existed via unsupervised clustering. We identified three clusters of adherence for the Evidation Health study (E0–2) and four clusters of adherence for the BIL study (B0–3; figure 2; weekly patterns in appendix 1 p 3). Generally, the four BIL clusters ranged from low to high adherence based on the three metrics of

adherence (total number of surveys completed, participation regularity, and time of activity). Participants in group B0 (n=1977) had a low total number of surveys completed (mean 54 [SD 75]), low participation regularity, and early time of activity. Participants in B1 (n=3028) had a medium total number of surveys completed (123 [89]), low participation regularity, and early time of activity. Participants in B2 (n=540) had a high total number of surveys completed (509 [35]), high participation regularity, and early time of activity. Participants in B3 (n=239) had a high total number of surveys completed (319 [68]), high participation regularity, and late time of activity. The Evidation Health clusters could be characterised as: E0 (n=58 199), participants with a low total number of surveys completed (mean 5 [SD 5]), low participation regularity, and early time of activity; E1 (n=18 435), participants with a medium total number of surveys completed (35 [13]), high participation regularity, and early time of activity; and E2 (n=12 845), participants with a high total number of surveys completed (90 [19]), high participation regularity, and early time of activity.

In both studies, the highest numbers of participants were present in the clusters with a low or medium total number of surveys completed, low participation regularity, and early time of activity (B0, B1, and E0; figure 2).

To further test whether demographic factors were associated with the unsupervised clustering, we explored if the clusters varied by race, sex, and age (figure 3). Most of the individuals aged 18–29 years were observed in the clusters with low or medium adherence (245 [48·7%] of 503 individuals aged 18–29 years in cluster B0 and 228 [45·3%] of 503 in cluster B1; 17 516 [77·6%] of 22 581 in cluster E0; appendix 1 pp 7–9). We observed that the oldest age group (≥60 years) was generally more represented in clusters with high adherence (B2, B3, and E2) than younger age groups (figure 3). We explored whether any demographic differences were statistically discernible (ie, statistically significant). χ^2 tests of independence were done to assess the relationship between cluster labels and the demographic factors of sex, race, and age (appendix 1 pp 7–9). For both studies, we observed statistically discernible relationships between all demographic factors and cluster membership. This gave us a foundation to examine which associations were scientifically meaningful.

Since the p value only informs whether an association exists, but not its extent, we sought to quantify the magnitude of the associations by applying Cramér's V. For the BIL study, the correlations were weak between cluster membership and race, sex, and age, with Cramér's V values of 0·086, 0·072, and 0·099 respectively (appendix 1 pp 7–8). For the Evidation Health dataset, the correlations were weak between cluster membership and race and sex, with values of 0·037 and 0·024, respectively. However, the correlation between cluster membership and age was moderate, with a value of 0·151 (appendix 1 p 8).

To identify whether there was a relationship between clusters and specific demographic factors, we did a post-hoc analysis which sought to establish the source of the difference present in the χ^2 tests. Results for the BIL study showed that the B0 cluster (low total number of surveys completed and low participation regularity, and early time of activity) was discernibly different from clusters B1, B2, and B3 for race. Specifically, the B0 cluster had a higher proportion of Black or Other race participants and a lower proportion of White participants than the other clusters (appendix 1 pp 7–8). Sex and age were discernibly different across all clusters except for B0 versus B1 (which was not discernible on sex) and B1 versus B3 (not discernible on age; appendix 1 pp 15–16). For the Evidation Health study, the demographic compositions of the clusters were discernibly different except for E1 versus E2 (which was not discernible on sex; appendix 1 p 16).

Retention is another key aspect of digital health studies as it measures the duration an individual remains within a study. We did a Kaplan–Meier analysis to estimate the probability of participants exiting the study before its completion and to identify factors that might affect study participant retention. We modelled the time to exit (the time between enrolment and exit; defining the failure event as the EXIT state) using a Kaplan–Meier curve with age, race, and sex as explanatory variables (figure 4). Both the BIL and Evidation Health studies showed a steady decline in the retention probability, with both studies having a higher dropout rate at the study's outset versus

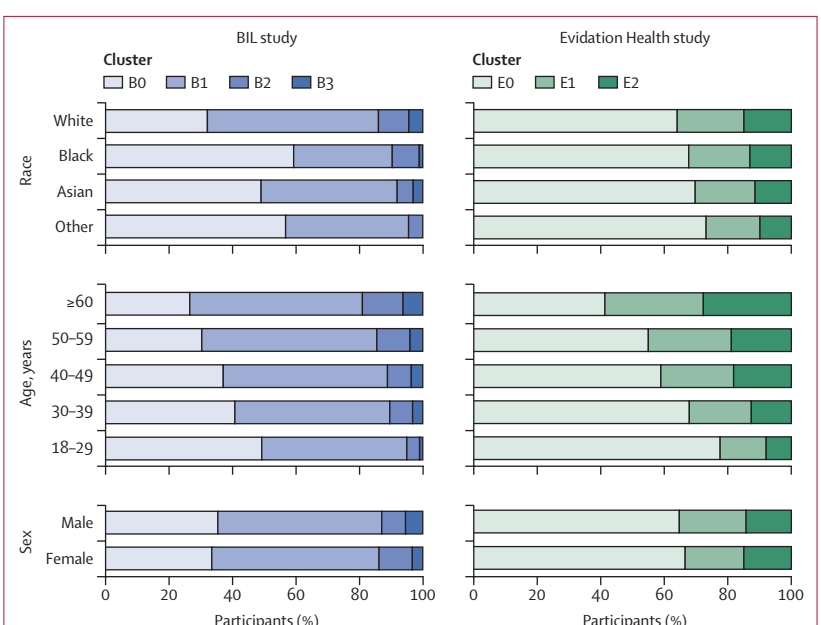


Figure 3: **Study participants clustered on daily survey filling patterns and stratified by demographic factors**
Plots show the percentages of each survey filling pattern cluster for participants in each study separated by race, age, and sex. BIL=Duke University BIG IDEAs Laboratory.

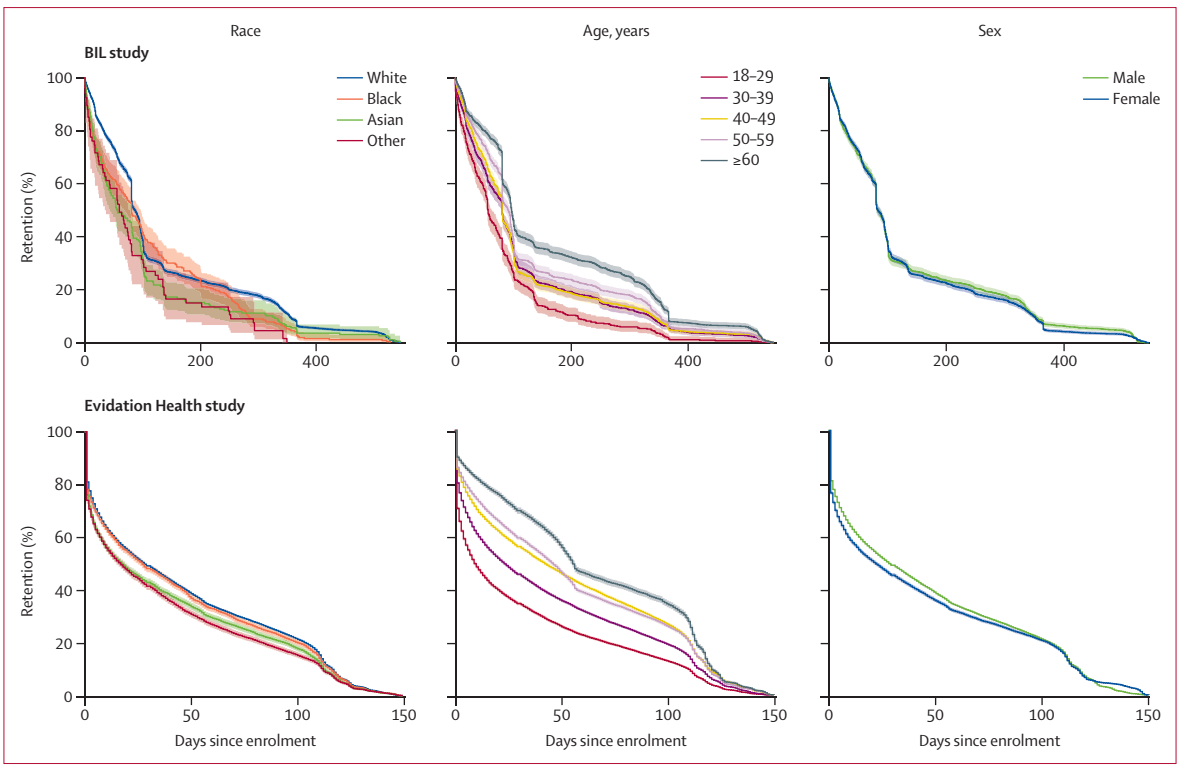


Figure 4: **Kaplan–Meier curves by age, race, and sex**
Kaplan–Meier curves showing the changing retention probability over time in demographic groups in each study, with the event defined as participants who had exited the study (EXIT state). Numbers at risk, numbers censored, and log-rank p values are provided in appendix 1 (pp 13–15). BIL=Duke University BIG IDEAs Laboratory.

later timepoints. When counting a single missed survey (entering the INACT state) as the failure event, we observed sharp declines in the numbers of participants early in the studies (appendix 1 pp 5–6).

We characterised the association between retention probability and demographic factors using Cox regression (appendix 1 pp 16–17). We observed that as age increased, the retention probability also increased. For the BIL study, the retention probability was lowest for the 18–29 years age group (*vs* ≥60 years group), with an HR of 1·69 (95% CI 1·53–1·86; p<0·0001) for study exit at any given point in time. The same trend was observed for the Evidation Health study with an HR of 1·50 (1·47–1·53; p<0·0001) for the 18–29 years age group. For other demographic factors, sex was not a statistically discernible predictor of study exit in the BIL study (HR for male participants *vs* female participants (0·98 [0·93–1·03]; p=0·50), but male participants (*vs* female participants) in the Evidation Health study had a lower risk for study exit (HR 0·96 [0·94–0·98]; p<0·0001). Comparisons of race showed no associations with study exit in the BIL study. In the Evidation Health study, participants identifying as Other race (*vs* Asian race) had a higher risk of study exit (HR 1·10 [1·06–1·14; p<0·0001]), and those identifying as White (*vs* Asian) had a lower risk of study exit (HR 0·96 [0·93–0·98; p=0·0004]).

The previous methodology used in this study assessed adherence qualitatively and did not account for epochs of different behaviour types. Therefore, we did multistate modelling analysis in which we sought to identify how covariates (age, sex, and race) might relate to adherence based on maximum likelihood estimation. We also included the time-based covariate of day of survey delivery (weekday *vs* weekend) to model the difference in behaviour given demographic factors. We decided upon the two relevant states of ACT and INACT. In the multistate model, for the BIL study we observed a similar pattern of adherence concerning age to that in the survival analysis: as age increased, the likelihood of entering the INACT state from the ACT state generally decreased. However, this pattern was not apparent for the Evidation Health study (figure 5; appendix 1 pp 11–13). Although the two digital health studies we examined were separate, the transition intensities for age were generally similar for the ACT to INACT transition, but not for the INACT to ACT transition. Specifically, for the 18–29 years age group (*vs* ≥60 years group), the transition intensity from ACT to INACT in the BIL study was 1·661 (95% CI 1·606–1·718) and in the Evidation Health study was 1·108 (1·094–1·121), whereas the transition intensities from INACT to ACT were 1·220 (1·179–1·262) and 0·597 (0·590–0·605), respectively. Results were

similar when comparing the race categories, with most transition intensities having overlapping 95% CIs for the BIL study. For the White racial group in both the BIL and Evidation Health studies, 95% CIs for transitions (ACT to INACT and INACT to ACT) did not overlap with those of other racial groups, indicating some separation in transition intensities. When comparing results by sex, male participants showed higher transition intensities for ACT to INACT transitions than INACT to ACT transitions in both studies (figure 5; appendix 1 pp 11–13). As for survey timing, our inclusion of weekends versus weekdays revealed some separation in transition intensities. For example, in the Evidation Health study, ACT to INACT transitions were lower on weekends than weekdays, while INACT to ACT transitions were higher on weekends than weekdays, indicated by 95% CIs that did not cross the null value of 1. Conversely, in the BIL study, ACT to INACT transitions were higher on weekends than weekdays, and INACT to ACT transitions were lower on weekends than weekdays (figure 5; appendix 1 pp 11, 13). Use of BIL symptom reports as input variables indicated that the presence of symptoms caused a lowering of transition intensities for both possible transitions versus the absence of symptoms (appendix 1 p 12). However, the role of symptom reports in modulating individual-level transitions warrants further investigation.

To characterise adherence from the multistate model, we defined two separate descriptive metrics: (1) participation regularity (mean sojourn time); and (2) total length of stay, defined for a 90-day window. For the BIL and Evidation Health studies, younger participants (age 18–29 years *vs* ≥60 years) had a lower participation regularity in the ACT state, shorter total length of stay in the ACT state, and longer total length of stay in the INACT state (appendix 1 pp 9–11). There were some differences between the patterns observed in the BIL and Evidation Health studies with regard to race. In the BIL study, individuals who were White or Asian had higher participation regularity in the ACT state and a longer total length of stay in the ACT state than individuals who were Black or Other race. In the Evidation Health study, individuals who were White or Other race had higher participation regularity in the ACT state than individuals who were Black or Asian, with similar total lengths of stay between the racial groups (appendix 1 pp 9–10). In the BIL study, male participants had higher participation regularity in both the ACT and INACT states than female participants. In the Evidation Health study, female participants had higher participation regularity in the ACT state than male participants. In both studies, total lengths of stay in the ACT and INACT states were generally similar between male and female participants (appendix 1 pp 9–10). Weekdays in the BIL dataset were associated with longer ACT state adherence, whereas weekends were associated with more time in the INACT state. By contrast, the Evidation Health dataset showed higher adherence to the ACT state during weekends, with higher participation regularity in the ACT state than for weekdays (appendix 1 pp 10–11). Prevalence plots representing the percentage of participants transitioning from the ACT state to the EXIT state over time by age, sex, and race were similar to the Kaplan–Meier curves (appendix 1 p 6).

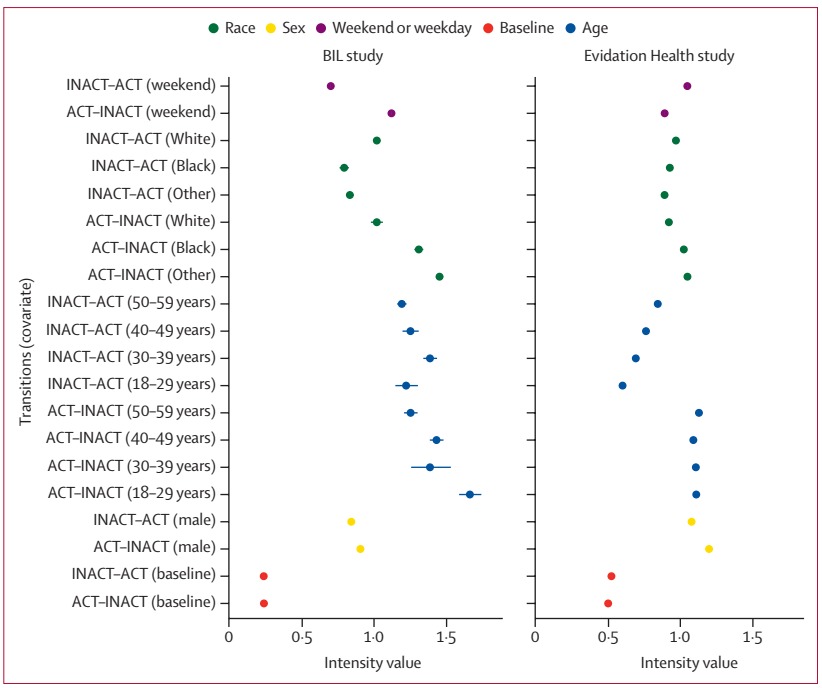


Figure 5: **Multistate model transition intensities by race, age, sex, and day of survey delivery**
Transition intensities ordered by INACT to ACT (INACT–ACT) and ACT to INACT (ACT–INACT) for different covariates in the multistate model for each study, where ACT represents an active state and INACT represents an inactive state. The BIL model also included symptom reports as input variables (appendix 1 p 12). The transition intensity value represents the rate at which individuals move from one state to another per day given the covariate. The selected covariates are shown in parentheses; the baseline model had the following covariates: Asian, female, age ≥60 years, and weekday (and no symptoms reported, BIL study only). Errors bars represent 95% CIs. BIL=Duke University BIG IDEAs Laboratory.

Discussion

The overall aim of this work was to identify factors in digital health study design that can influence data representation and to test a different methodological approach for identifying such factors. We found that age was inversely related to adherence and retention, which is consistent with the literature,[16,17] indicating a need to improve engagement among young adults.

A major contribution of this work is the methodology used to analyse adherence and retention. Typical methods such as survival analysis or multivariate regression[20,32] do not comprehensively represent participation in repeated tasks over time which are common in digital health studies.[17] Although these methods explain variation in participant dropout across multiple factors, they might misrepresent how participants engaged with the study. To address this possibility, we applied multistate modelling to capture engagement dynamics over time. The two-state

multistate model closely resembled conventional Cox regression in identifying factors, such as age, that influenced adherence and retention, with both methods showing similar effect sizes. For instance, HRs from the Cox regression for study exit (eg, BIL study, 1·69 for age 18–29 years *vs* ≥60 years; Evidation Health study, 1·50) aligned with multistate modelling transition intensities for the ACT to INACT state (eg, BIL study, 1·661 for age 18–29 years *vs* ≥60 years; Evidation Health study, 1·108).

Multistate modelling offers additional insights through metrics such as mean sojourn time and length of stay, indicating the duration that participants are in active or inactive states. These measures help to predict dropout points and identify low engagement periods. For example, knowing that younger participants have shorter mean sojourn times (ie, lower participation regularity) in the active state than older participants can guide proactive retention methods, such as personalised reminders, incentives, or flexible study protocols that can accommodate their schedules and preferences. The differences in demographic characteristics across clusters indicate that study participants with particular demographic characteristics might be at increased risk of low adherence, and study teams might address this issue by developing targeted solutions to aim to balance demographic characteristics across clusters of adherence types. Understanding the mean length of stay in an inactive state can help to time re-engagement efforts, such as follow-up communications, to bring participants back into the study. Recurrent event analysis offers a nuanced view of participant behaviour. Although recurrent event analysis has been applied in other fields, this is the first instance, to our knowledge, of applying this method to assess adherence in digital health studies.

Our work also highlights the need for a common lexicon to describe and measure adherence and retention.[14,17,32] Terms such as compliance, adherence, and engagement are often used interchangeably and the definitions vary by study. Thus, we clearly and quantitatively defined adherence as a participant's survey filling patterns and retention as the duration of time before a participant drops out of the study.

Our study should be contextualised within the broader landscape of digital health research, particularly with regard to past experiences in influenza-like illness surveillance. Studies such as those by Dalton and colleagues,[21] Baltrusaitis and colleagues,[22] and Bajardi and colleagues[23] provide valuable insights into factors influencing follow-up participation and the representativeness of participants in web-based participatory surveillance systems. From these studies, it is evident that addressing usability, motivation, trust, and demographic considerations can substantially enhance adherence and retention in digital health studies. Incorporating these strategies into study design can improve the representativeness and reliability of data collected from web-based surveillance systems.

We note that a major limitation in comparing the two datasets was the differing study designs. Although the two studies were similar in the resulting data, some differences were not captured by the surveys. These differences include the user experience of completing surveys via apps versus a website portal, the method with which participants found out about the study, and the public's perception of the institution hosting the study. In addition, the BIL CovIdentify study was initially designed to be done for 3 months but was run for an additional year based on the evolution of the COVID-19 pandemic. This change might be the source of the B3 cluster and the return of many participants to renew their participation in the study.

Another important limitation of our study is the absence of continual monitoring of symptoms and diagnostic data, which prevented analysis of the effect of illness timing and severity on study engagement. For instance, people who feel unwell might be likely to cease participation. Due to the secondary nature of the study, although the surveys collected time-varying covariates such as illness and symptoms, we could not include them for both studies and for all analyses due to a high amount of missingness of these data. The ability to collect continuous information and to account for this potential source of bias in data missingness should be explored in future work, and we recommend considering all the potential factors that can affect adherence over time.

Furthermore, we note that neither of the digital health studies were originally designed to address the research questions posed in this paper, and we did secondary analyses to answer questions regarding participant survey adherence. We recommend that for future longitudinal studies, increased consideration be given to participant engagement patterns as part of the study design. This consideration will allow for analyses to isolate factors that directly contribute to how a participant interacts with a study. Furthermore, although studies might initially enrol representative populations, an improved understanding of the engagement patterns can aid in the resulting data quality. For example, for studies that collect data from devices, participants' wear time can be affected by the device's battery life and how comfortable the device is to wear. Not considering such factors in the choice of the device could result in insufficient data collection.

Finally, although this study assessed participant engagement, it did not address the issue of which populations were reached. Communities most affected by COVID-19, including Black or African American and Hispanic or Latinx populations,[33] had low representation in our study (Hispanic representation was <5% in the BIL study and about 8% in the Evidation Health study; data not shown), which is a common issue encountered in studies with a bring-your-own-devices strategy, in which participants can donate data for the research

studies from their personally owned devices.[7] We recognise that the initial survey populations were not representative and encourage future researchers to improve the reach of their studies.

The generalisable methods developed in this study might help research groups to better understand the behaviour patterns of their study populations and guide the development of targeted strategies for improving representation. Future research could explore the interaction of different demographic factors and other individual characteristics, such as socioeconomic, geographical, and biopsychosocial factors, and intersectional characteristics, in shaping engagement patterns in digital health studies.

Contributors

PJC, MMHS, and IMO conceived the study and the design. JPD and LF oversaw the analysis and writing. PJC, MMHS, and IMO were involved in data collection and curation. EJD assisted in modelling and analysis. PJC, MMHS, and JPD working within the BIL at Duke University had access to the CovIdentify data, and IMO, EJD, and LF working at Evidation Health at the time of data collection had access to the COVID2020 study data. PJC and MMHS also had access to the COVID2020 survey data. PJC and MMHS accessed and verified both datasets. PJC performed the data analysis. PJC and EJD validated the analysis. PJC, MMHS, IMO, EJD, and JPD interpreted the data. PJC, MMHS, and IMO drafted the manuscript. All authors revised and edited the manuscript critically for important intellectual content. All authors had final responsibility for the decision to submit for publication.

Declaration of interests

At the time of writing, JPD received a US National Institutes of Health (NIH) grant (number R01 DK133531) and consulting fees as a consultant for Gold Track. JPD has received honoraria for lectures and support for attending meetings from the American College of Cardiology, the Academy of Managed Care Pharmacy, Shionogi, the American College of Sports Medicine, the International Society for the Measurement of Physical Behaviour, Harvard University, Dartmouth College, the Foundation for the National Institutes of Health, and the Chan Zuckerberg Initiative. JPD was a Scientific Advisor for Veri and was an Associate Editor for *npj Digital Medicine*. EJD and LF were employees of Evidation Health at the time of data collection and writing. LF is the co-founder of Evidation Health and holds stocks in Evidation Health. IMO is a consultant at Evidation Health. PJC and MMHS declare no competing interests.

Data sharing

The de-identified BIL dataset generated during the current study has been published on Physionet under the title CovIdentify Dataset (https://doi.org/10.13026/ncq1-vp79). The data can be requested via the PhysioNet Credentialed Health Data Use Agreement. The de-identified Evidation Health COVID2020 dataset is available on the Synapse Platform (https://www.synapse.org/Synapse:syn22891469/wiki/605716) run by Sage Bionetworks. Before accessing any study data, partners and collaborators will have to create a Synapse account, agree to the Synapse terms and privacy policy, and agree to the Evidation Health data use agreement, which prohibits re-identification and limits data use to only those activities and research efforts covered by this agreement.

Acknowledgments

The BIL study was supported in part by a US National Science Foundation (NSF) grant (number 2200047; Predictive Intelligence for Pandemic Prevention programme), an NSF Faculty Early Career Development Program grant (number 2339669), an NIH grant (number R01 DK133531), Microsoft AI for Health, the Duke Clinical and Translational Science Institute (grant number UL1TR002553), North Carolina Biotechnology Center (2020-FLG-3884), Duke MEDx, Duke Bass Connections, the Duke Margolis Center for Health Policy, and Duke Office of Information Technology. MMHS is supported by an American Heart Association Postdoctoral Fellowship (number 23POST1025599). The Evidation Health study was done by Evidation Health. EJD thanks Chris Barker for their helpful feedback on analysis approaches via the American Statistical Association's Statistical Consulting Section Community Discussion online forum. We thank all the participants who took part in our studies.

References

1 Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Jan 9, 2020. Pew Research Center. https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ (accessed Sept 16, 2020).

2 Pew Research Center. Mobile fact sheet. Jan 31, 2024. https://www.pewresearch.org/internet/fact-sheet/mobile/ (accessed June 14, 2024).

3 Laricchia F. Wearable devices ownership in the United States in 2021, by age. March 10, 2023. https://www.statista.com/statistics/1310364/wearable-devices-ownership-age-us/ (accessed May 25, 2023).

4 Li X, Dunn J, Salins D, et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol* 2017; **15:** e2001402.

5 Gadaleta M, Radin JM, Baca-Motes K, et al. Passive detection of COVID-19 with wearable sensors and explainable machine learning algorithms. *NPJ Digit Med* 2021; **4:** 1–10.

6 Shandhi MMH, Cho PJ, Roghanizad AR, et al. A method for intelligent allocation of diagnostic testing by leveraging data from commercial wearable devices: a case study on COVID-19. *NPJ Digit Med* 2022; **5:** 130.

7 Cho PJ, Yi J, Ho E, et al. Demographic imbalances resulting from the bring-your-own-device study design. *JMIR Mhealth Uhealth* 2022; **10:** e29510.

8 Mumtaz H, Riaz MH, Wajid H, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. *Front Digit Health* 2023; **5:** 1203945.

9 Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. *Gerontologist* 2019; **59:** 6–14.

10 Kostick-Quenet KM, Cohen IG, Gerke S, et al. Mitigating racial bias in machine learning. *J Law Med Ethics* 2022; **50:** 92–100.

11 Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. *Rheum Dis Clin North Am* 2019; **45:** 159–72.

12 D'Angelo S, Gilio M, D'Attino RM, et al. Observational study on the quality of life of Italian axial spondyloarthritis patients (QUASAR): baseline data. *Clin Exp Rheumatol* 2019; 37: 748–55.

13 Reade S, Spencer K, Sergeant JC, et al. Cloudy with a Chance of Pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. *JMIR Mhealth Uhealth* 2017; **5:** e37.

14 Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. *J Med Internet Res* 2022; **24:** e35120.

15 Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health* 2021; **3:** e260–65.

16 Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100 000 participants. *NPJ Digit Med* 2020; **3:** 21.

17 Zhang Y, Pratap A, Folarin AA, et al. Long-term participant retention and engagement patterns in an app and wearable-based multinational remote digital depression study. *NPJ Digit Med* 2023; **6:** 25.

18 Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Wiley & Sons, 2002.

19 Groha S, Schmon SM, Gusev A. A general framework for survival analysis and multi-state modelling. *arXiv* 2021; published online Feb 15. https://doi.org/10.48550/arXiv.2006.04893 (preprint).

20 George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol* 2014; **21:** 686–94.

21 Dalton C, Carlson S, Butler M, et al. Insights from flutracking: thirteen tips to growing a web-based participatory surveillance system. *JMIR Public Health Surveill* 2017; **3:** e48.

22 Baltrusaitis K, Santillana M, Crawley AW, Chunara R, Smolinski M, Brownstein JS. Determinants of participants' follow-up and characterization of representativeness in Flu Near You, a participatory disease surveillance system. *JMIR Public Health Surveill* 2017; **3:** e18.

23 Bajardi P, Vespignani A, Funk S, et al. Determinants of follow-up participation in the internet-based European influenza surveillance platform Influenzanet. *J Med Internet Res* 2014; **16:** e78.

24 Shapiro A, Marinsek N, Clay I, et al. Characterizing COVID-19 and influenza illnesses in the real world via person-generated health data. *Patterns (N Y)* 2020; **2:** 100188.

25 Nestor B, Hunter J, Kainkaryam R, et al. Dear watch, should I get a COVID-19 test? Designing deployable machine learning for wearables. *medRxiv* 2021; published online May 17. https://doi.org/10.1101/2021.05.11.21257052 (preprint).

26 Pathiravasan CH, Zhang Y, Trinquart L, et al. Adherence of mobile app-based surveys and comparison with traditional surveys: eCohort study. *J Med Internet Res* 2021; **23:** e24773.

27 Cramér H. Mathematical methods of statistics. Princeton University Press, 1999.

28 Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958; **53:** 457–81.

29 Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol* 1972; **34:** 187–202.

30 Jackson C. msm: multi-state Markov and hidden Markov models in continuous time. Jan 4, 2011. https://CRAN.R-project.org/package=msm (accessed July 7, 2022).

31 Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat* 2016; **70:** 129–33.

32 Daniore P, Nittas V, von Wyl V. Enrollment and retention of participants in remote digital health studies: scoping review and framework proposal. *J Med Internet Res* 2022; **24:** e39910.

33 Boehmer TK, Koumans EH, Skillen EL, et al. Racial and ethnic disparities in outpatient treatment of COVID-19—United States, January–July 2022. *MMWR Morb Mortal Wkly Rep* 2022; **71:** 1359–65.