



ARTICLE OPEN



Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants

Abhishek Pratap^{1,2} , Elias Chaibub Neto¹, Phil Snyder¹, Carl Stepnowsky^{3,4}, Noémie Elhadad⁵, Daniel Grant⁶, Matthew H. Mohebbi⁷, Sean Mooney², Christine Suver¹, John Wilbanks¹, Lara Mangravite¹, Patrick J. Heagerty⁸, Pat Areán⁹ and Larsson Omberg ¹

Digital technologies such as smartphones are transforming the way scientists conduct biomedical research. Several remotely conducted studies have recruited thousands of participants over a span of a few months allowing researchers to collect real-world data at scale and at a fraction of the cost of traditional research. Unfortunately, remote studies have been hampered by substantial participant attrition, calling into question the representativeness of the collected data including generalizability of outcomes. We report the findings regarding recruitment and retention from eight remote digital health studies conducted between 2014–2019 that provided individual-level study-app usage data from more than 100,000 participants completing nearly 3.5 million remote health evaluations over cumulative participation of 850,000 days. Median participant retention across eight studies varied widely from 2–26 days (median across all studies = 5.5 days). Survival analysis revealed several factors significantly associated with increase in participant retention time, including (i) referral by a clinician to the study (increase of 40 days in median retention time); (ii) compensation for participation (increase of 22 days, 1 study); (iii) having the clinical condition of interest in the study (increase of 7 days compared with controls); and (iv) older age (increase of 4 days). Additionally, four distinct patterns of daily app usage behavior were identified by unsupervised clustering, which were also associated with participant demographics. Most studies were not able to recruit a sample that was representative of the race/ethnicity or geographical diversity of the US. Together these findings can help inform recruitment and retention strategies to enable equitable participation of populations in future digital health research.

npj Digital Medicine (2020)3:21; <https://doi.org/10.1038/s41746-020-0224-8>

INTRODUCTION

Traditional in-person clinical trials serve as the cornerstone of modern healthcare advancement. While a pivotal source of evidence generation for advancing clinical knowledge, in-person trials are also costly and time-consuming, typically running for at least 3–5 years from conception to completion, at a cost of millions of dollars per study. These timelines have often meant that promising treatments take years to get to dissemination and uptake, which can create unnecessary delays in advancing clinical practice. Additionally, clinical research suffers from several other challenges^{1,2} including (1) recruiting sufficiently large and diverse cohorts quickly, and (2) tracking day-to-day fluctuations in disease severity that often go undetected in study-related intermittent protocolized in-clinic evaluations.^{3,4} Scientists have recently turned to digital technology^{5,6} to address these challenges, hoping to collect real-world evidence⁷ from large and diverse populations to track long-term health outcomes and variations in disease trajectories at a fraction of the cost of traditional research.⁸

The global penetration⁹ and high-frequency usage of smartphones (up to 4 h daily^{10,11}) offer researchers a potentially cost-effective means to recruit a large number of participants into health research across the US (and the world).^{12,13} In the last 5 years, investigators have conducted several large-scale studies^{14–22} that deployed interventions^{23,24} and operationally conduct clinical trials^{25–27} using mobile technologies. These studies are able to recruit at-scale because participants can be identified and consented²⁸ to participate in the study without ever having

stepped foot in a research lab, with significantly lower costs than conventional clinical trials.^{23,24} Mobile technologies also allow investigators an opportunity to collect data in real-time based on people's daily lived experiences of the disease, that is, real-world data.⁷ Rather than retrospectively asking people to recall their health over the past week or month, researchers using mobile technologies can assess participants frequently including outside clinic and at important points in time without having to rely on recall that is known to have bias.²⁹ While these recent studies show the utility of mobile technology, challenges in participant diversity and long-term participant retention still remain a significant problem.³⁰

Digital studies continue to suffer from participant retention problems that also plagued internet-based studies^{31,32} in the early 2000s.^{33–35} However, our understanding of factors impacting retention in remote research remains to be limited. High levels of user attrition combined with variations in long-term app usage may result in the creation of a study cohort that does not represent the characteristics of the initially recruited study population with regards to demographics and disease status. This has called into question the reliability and utility of the collected data from these studies.³⁶ Of note, the representativeness of remote study cohort (e.g., demographics, geographical diversity, etc) may vary based on the study design and inclusion criteria. Many large-scale digital health studies enroll participants from a general population, where anyone eligible with or without target disease of interest can self-select to join the study. Such strategies

¹Sage Bionetworks, Seattle, WA, USA. ²Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA. ³University of California, San Diego, CA, USA. ⁴American Sleep Apnea Association, Washington, DC, USA. ⁵Columbia University, New York, NY, USA. ⁶Novartis Pharmaceutical Corporation, East Hanover, NJ, USA. ⁷GoodRx, Santa Monica, CA, USA. ⁸Department of Biostatistics, University of Washington, Seattle, WA, USA. ⁹Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA, USA. ✉email: apratap@sagebionetworks.org

may be prone to selection and ascertainment biases.³⁶ Similarly, cohorts in studies targeting a population with a specific clinical condition of interest may need to be evaluated in the context of the clinical and demographic characteristics of the underlying population with that disease. Evaluation of participant recruitment and retention from large-scale remotely collected data could help detect confounding characteristics that may be present and which have been shown to severely impact the generalizability of the resulting statistical inference.^{36,37} Participant retention may also be partially dependent on the engagement strategies used in remote research. While most studies assume participants will remain in a study for altruistic reasons,³⁸ other studies provide compensation for participant time,³⁹ or leverage partnerships with local community organizations, clinical registries, and clinicians to encourage participation.^{23,24} Although monetary incentives are known to increase participation in research,⁴⁰ we know little about the relative impact of demographics and different recruitment and engagement strategies on participant retention, especially in remote health research.

The purpose of this study is to document the drivers of retention and long-term study mobile application usage in remote research. To investigate these questions, we have compiled user-engagement data from eight digital health studies that enrolled more than 100,000 participants from throughout the US between 2014–2019. These studies assessed different disease areas including asthma, endometriosis, heart disease, depression, sleep health, and neurological diseases. While some studies enrolled participants from the general population (i.e., with and without disease of interest) others targeted a specific subpopulation with the clinical condition of interest. For analysis, we have combined individual participant data across these studies. Analysis of the pooled data considers overall summaries of demographic or operational characteristics while accounting for study heterogeneity in retentions (see Methods for further details on individual studies and analytical approach). The remote assessments in these studies consisted of a combination of longitudinal subjective surveys and objective sensor-based tasks including passive data⁴¹ collection. The diversity of the collected data allows for a broad investigation of different participant characteristics and engagement strategies that may be associated with higher retention in the collected real-world data.

RESULTS

Participant characteristics

The combined user-activity data from eight digital health studies resulted in a pool of 109,914 participants who together completed ~3.5 million tasks on more than 850,000 days (Table 1). The demographics of participants across studies (Table 2) varied widely in part due to study-specific eligibility criteria, which may

impact the underlying characteristics of the recruited population. Except for three studies (Brighten, Phendo, and Start) that aimed to recruit people with a specific clinical condition of interest, the rest of the five studies enrolled people from the general population with and without the target disease of interest. The majority of participants were between 17–39 years (Median percent of subject age across studies 17–39 = 65.2%, Range = 37.4–91.5%) with those 60 years and older being the least represented (Median age percent greater than 60 = 6%, Range = 0–23.3%). The study samples also had a larger proportion of Females (Median = 56.9%, Range = 18–100%). A majority of recruited participants were Non-Hispanic Whites (Median = 75.3%, Range = 60.1–81.3%) followed by Hispanic/Latinos (Median = 8.21%, Range = 4.79–14.29%) and African-American/Blacks (Median = 3.45%, Range = 2–10.9%). The race/ethnic and geographical diversity of the present sample showed a marked difference from the general population of the US. Minority groups were under-represented in the present sample with Hispanic/Latinos and African-America/Black showing a substantial difference of –8.1% and –9.2% respectively compared to the 2010 US census data s (Table 2, Fig. 1b). Similarly, the median proportion of recruited participants per state also showed notable differences from the state's population proportion of the US (Fig. 1a).

Participant retention

As is the nature of these studies, participants were required to complete all health assessments and other study-related tasks (e.g., treatments) through a mobile application (app) throughout the length of the study. The median time participants engaged in the study in the first 12 weeks was 5.5 days of which in-app tasks were performed on 2 days (Table 2). Higher proportions of active tasks were completed by participants during the evening (4–8 PM) and night (8–12 Midnight) hours (Fig. 2a). Across the studies, the median retention time varied significantly ($P < 1e-16$) between 2 and 12 days with the Brighten study being an outlier with higher median retention of 26 days (Fig. 2b). A notable increase in median retention time was seen for sub-cohorts that continue to engage with the study apps after day one and beyond (Fig. 2c). For example, the median retention increased by 25 days for the sub-cohort that was engaged for the first 8 days. The participant retention also showed a significant association with participant characteristics. While older participants (60 years and above) were the smallest proportion of the sample, they remained in the study for a significantly longer duration (Median = 7 days, $P < 1e-16$) compared to the majority younger sample (17–49 years) (Fig. 2d). Participants declared gender showed no significant difference in retention ($P = 0.3$). People with clinical conditions of interest to the study (e.g., heart disease, depression, multiple sclerosis) remained in the studies for a significantly longer time (Median = 13 days, $P < 1e-16$) compared to participants that were recruited

Table 1. Summary of user-engagement data compiled from eight digital health studies.

Study	Disease focus/study type	Study period	Number of participants	Total participant days	Active tasks completed
Start	Antidepressant Efficacy–Observational	Aug, 2015–Feb, 2018	42,704	280,489	1,219,656
MyHeartCounts	Cardiovascular Health–Observational	Mar, 2015–Oct, 2015	26,902	165,455	305,821
SleepHealth	Sleep Apnea–Observational	Jul, 2015–Jun, 2019	12,914	99,696	401,628
mPower	Parkinson's–Observational	Mar, 2015–Jun, 2019	12,236	104,797	568,685
Phendo	Endometriosis–Observational	Dec, 2016–Jul, 2019	7,802	81,938	735,778
Asthma	Asthma–Observational	Mar, 2015–Dec, 2016	5,875	77,815	175,699
Brighten	Depression–Randomized Control Trial	Jul, 2014–Aug, 2015	876	34,987	45,951
ElevateMS	Multiple Sclerosis–Observational	Aug, 2017–Jul, 2019	605	11,211	31,568
			109,914	856,388	3,484,786

Table 2. Summary of select participant demographics and study-app usage across the eight digital health studies.

	Asthma	Brighten	ElevateMS	mPower	MyHeartCounts	Phendo	SleepHealth	Start	Overall (median)
<i>Age group</i>									
<i>N</i>	2512	875	569	6810	1555	7484	12392	42690	
18–29 (%)	43.31	50.06	10.9	31.5	25.08	55.38	32.79	55.72	38
30–39 (%)	27.83	25.14	26.54	18.37	32.67	36.09	28.72	24.14	27.2
40–49 (%)	14.41	14.74	28.47	13.19	16.27	8.23	20.77	12.38	14.6
50–59 (%)	9.08	6.97	22.14	13.61	12.09	0.25	11	5.26	10
60 + (%)	5.37	3.09	11.95	23.33	13.89	0.04	6.72	2.51	6
<i>Sex</i>									
<i>N</i>	2509	875	329	6916	6976	7532	12558	42704	
Female (%)	39.58	77.83	74.16	28.93	18.94	100	29.14	75.86	56.9
<i>Race</i>									
<i>N</i>	3274	875	334	6884	4703	7530	5311	—	
Non-Hispanic White (%)	68.69	60.11	80.84	75.32	77.95	81.29	74.13	—	75.3
Hispanic/Latinos (%)	13.29	14.29	4.79	8.21	6.97	5.67	12.82	—	8.21
African-American/Black (%)	4.95	10.86	6.89	2.05	3.1	2.71	3.45	—	3.45
Asian (%)	4.98	8.23	2.99	8.4	7.72	2.79	5.87	—	5.9
Hawaiian or other Pacific Islander (%)	0.89	0.57	0	0.28	0.32	0	0.23	—	0.3
AIAN (%)	0.43	0.46	0	0.65	0.53	0.74	0.28	—	0.5
Other (%)	6.78	5.49	4.49	5.1	3.4	6.8	3.22	—	5.1
Duration in Study (Median ± IQR)	12 ± 38	26 ± 82	7 ± 45	4 ± 21	9 ± 19	4 ± 25	2 ± 8	2 ± 16	5.5
Days active tasks performed (Median ± IQR)	4 ± 12	14 ± 58	2 ± 8	2 ± 4	4 ± 7	2 ± 6	2 ± 4	2 ± 4	2

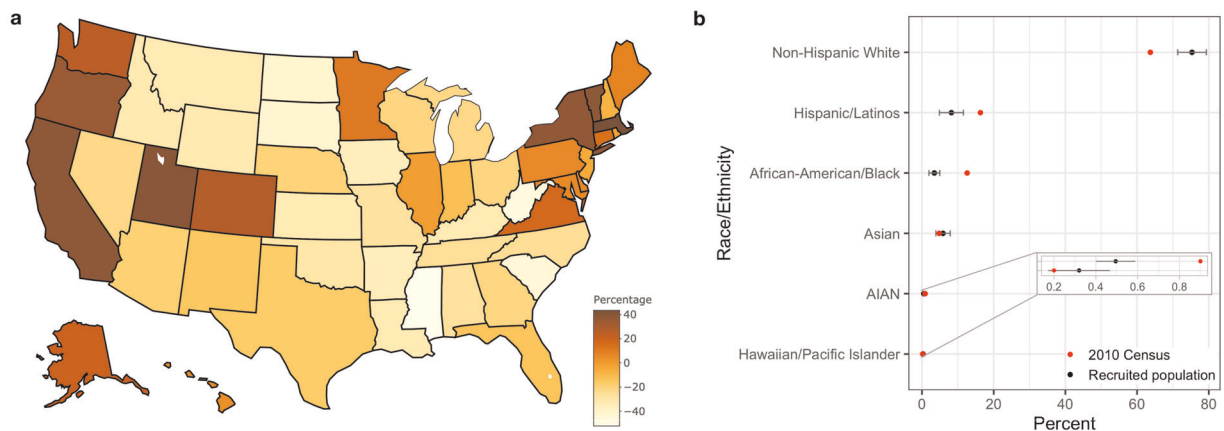


Fig. 1 Geographical and race/ethnic diversity of the recruited participants. **a** Map of US showing the proportion (median across the studies) of recruited participants relative to state's population proportion of the US and **b** Race/Ethnicity proportion of recruited participants compared to 2010 census data. The median value across the studies is shown by the black point with error bars indicating the interquartile range.

as non-disease controls (Median = 6 days) (Fig. 2e). Median retention time also showed a marked and significant increase of 40 days ($P < 1e-16$) for participants that were referred by a clinician to join one of the two studies (mPower and ElevateMS) (Median = 44 days) compared to participants who self-selected to join the same study (Median = 4 days) (Fig. 2f). See Supplementary Tables 1–6 for a further breakdown of survival analysis results. Sensitivity analysis by including participants with missing age showed no impact on the association of age with participant retention. However, participants with missing demographics showed variation in retention compared to participants who shared their demographics (Supplementary Fig. 1). This could be

related to different time points at which demographic related questions were administered in individual studies.

Participant daily engagement patterns

The subgroup of participants who remain engaged with study apps for a minimum of 7 days, showed four distinct longitudinal engagement patterns (Fig. 3b) with the dedicated users in cluster C1, high utilizers (C2), moderate users (C3) and sporadic users (C4). The participants who did not participate for at least 7 days were placed in a separate group (abandoners, C5*) (See Supplementary Fig. 2 and Methods for cluster size determination and exclusion criteria details). The engagement and demographic characteristics

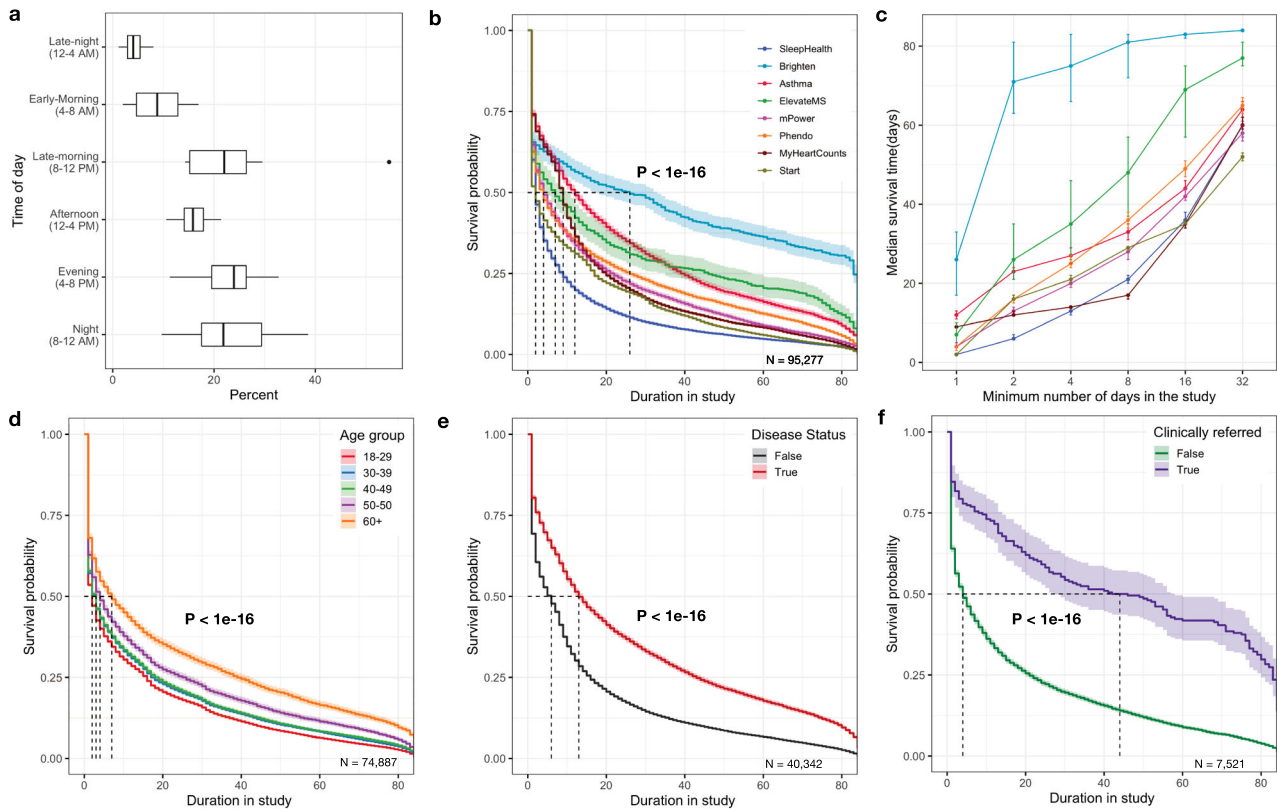


Fig. 2 Factors impacting participant retention in digital health studies. **a** Proportion of active tasks ($N = 3.3$ million) completed by participants based on their local time of day. The centerline of the boxplot shows the median value across the studies and upper and lower whisker corresponding to outlier point (1.5 times the interquartile range). **b** Kaplan Meir survival curve showing significant differences ($P < 1e-16$) in user retention across the apps. Brighten App where monetary incentives were given to participants showed the longest retention time (Median = 26 days, 95% CI = 17–33) followed by Asthma (Median = 12 days, 95% CI = 11–13), MyHeartCounts (Median = 9 days, 95% CI = 9–9), ElevateMS (Median = 7 days, 95% CI = 5–10), mPower (Median = 5 days, 95% CI = 4–5), Phendo (Median = 4 days, 95% CI = 3–4), Start (Median = 2 days, 95% CI = 2–2) and SleepHealth (Median = 2 days, 95% CI = 2–2). **c** Lift curve showing the change in median survival time (with 95% CI indicated by error bars) based on the minimum number of days (1–32) a subset of participants continued to use the study apps, Kaplan-Meier survival curve showing significant differences in user retention across **d** Age group, with 60 years and older using the apps for longest duration (Median = 7 days, 95% CI = 6–8, $P < 1e-16$) followed by 50–59 years (Median = 4 days, 95% CI = 4–5) and 17–49 years (Median = 2–3 days, 95% CI = 2–3). **e** Disease status; participants reporting having a disease stayed active longer ($N_{50} = 13$ days, 95% CI = 13–14) compared to people without disease ($N_{50} = 6$ days, 95% CI = 5–6) and finally **f** Clinical referral; Two studies (mPower and ElevateMS), had a subpopulation, that were referred to the study by clinicians and showed significantly ($P < 1e-16$) longer app usage period (Median = 44 days, 95% CI = 27–58) compared to self-referred participants with disease ($N_{50} = 4$ days, 95% CI = 4–4). For all survival curves the shaded region shows the 95% confidence limits based on the survival model fit.

across these five groups (C1–5*) varied significantly. Cluster 1 and 2 showed the highest daily app usage (Median app usage in the first 84 days = 96.4% and 63.1%, respectively) but also had the smallest overall proportion of participants (Median = 9.5%) with the exception of Brighten where 23.7% of study participants were in the dedicated users cluster C1. While daily app usage declined significantly for both moderate and sporadic clusters (C3–21.4% and C4–22.6%), the median number of days between app usage was significantly higher for participants in the sporadic C4 cluster (Median = 5 days) compared to cluster C3 (Median = 2 days). The majority of participants (median 54.6%) across the apps were linked to the abandoner group (C5*) with the median app usage of just 1 day (Fig. 4a, b). Furthermore, distinct demographic characteristics emerged across these five groups. Higher engagement clusters (C1–2) showed significant differences ($P = 1.38e-12$) in proportion of adults 60 years and above (Median range = 15.1–17.2% across studies) compared to lower engagement clusters C3–5* (Median range = 5.1–11.7% across studies) (Fig. 4c). Minority groups such as Hispanic/Latinos, Asians, and African-American/Black, on the other hand, were represented in higher proportions in the clusters (C3–5*) ($P = 4.12e-10$) with the least

engagement (Fig. 4d) (See Supplementary Table 8 for further details).

DISCUSSION

Our findings are based on one of the largest and most diverse engagement dataset compiled to date. We identified two major challenges with remote data collection: (1) more than half of the participants discontinued participation within the first week of a study and the rates at which people discontinued was drastically different based on age, disease status, clinical referral, and use of monetary incentives, and (2) most studies were not able to recruit a sample that was representative of the Race/ethnicity or geographical diversity of the US. Although these findings raise questions about the reliability and validity of data collected in this manner, they also shed light on potential solutions to overcome biases in populations using a combination of different recruitment and engagement strategies.

One solution could be the use of a flexible randomized withdrawal design.⁴² Temporal retention analysis (Fig. 2c) shows that a run-in period could be introduced in the research design, wherein participants who are not active in the study app in the

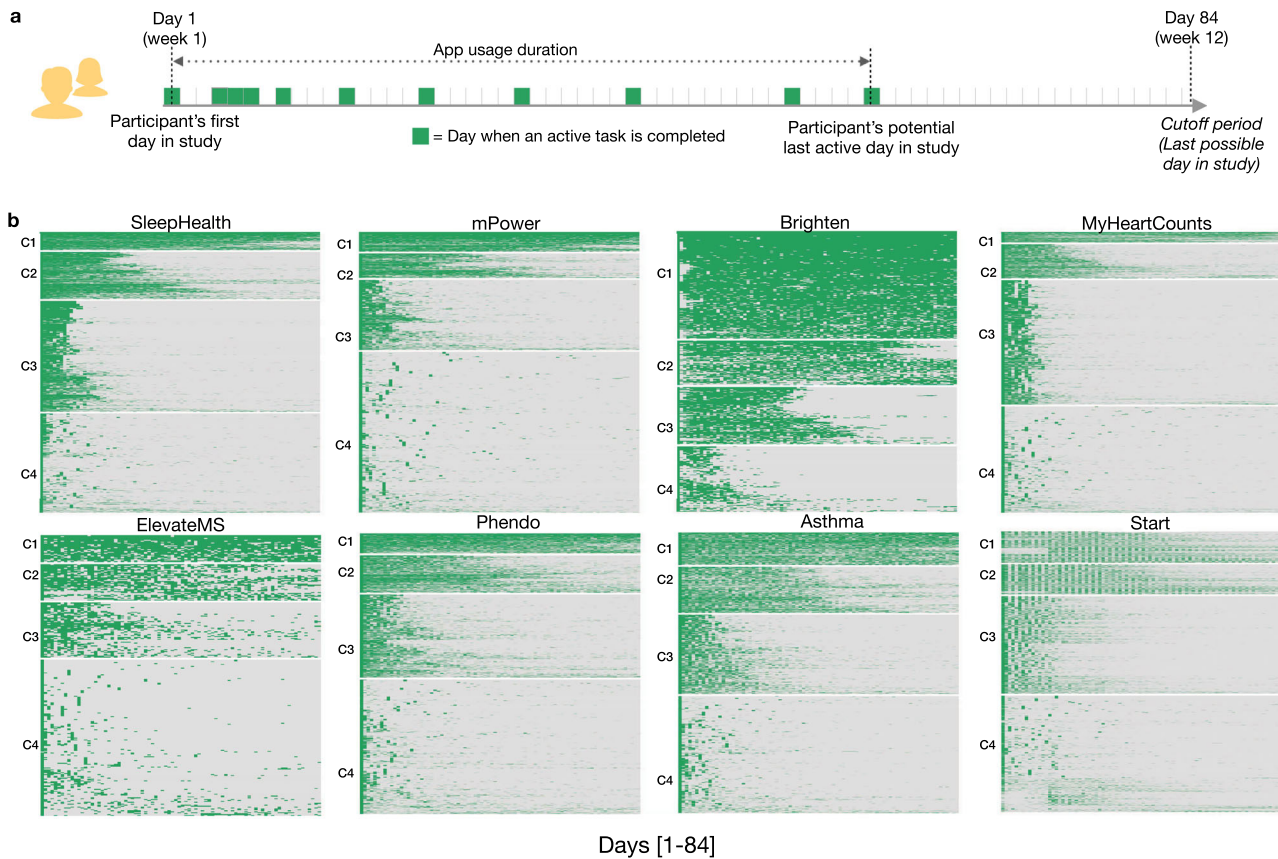


Fig. 3 Daily participant engagement patterns in digital health studies. **a** Schematic representation of an individual's in-app activity for the first 84 days. The participant app usage time is determined based on the number of days between the first and last day they perform an active task (indicated by the green box) in the app. Days active in the study is the total number of days a participant performs at least one active task (indicated by the number of green boxes). **b** Heatmaps showing participants in-app activity across the apps for the first 12 weeks (84 days), grouped into four broad clusters using unsupervised k-means clustering. The optimum number of clusters was determined by minimizing the within-cluster variation across different cluster sizes between 1–10. Seven out of eight studies indicated four clusters to be an optimum number using the elbow method. The heatmaps are arranged by the highest (C1) to the lowest user-engagement cluster (C4).

first week or two of the study can be excluded after enrollment but before the start of the actual randomized study. The resulting smaller but more engaged cohort will help increase the statistical power of the study but does not fix the potential for non-representative participant bias.⁴³

Another solution is to rely on monetary incentives to enhance engagement. Although only one study paid participants, the significant increase in retention and the largest proportion of frequent app users indicate the utility of the fair-share compensation model^{1,44,45} in remote research. Such “pay-for-participation” model could be utilized by studies that require long-term and frequent remote participation. Researchers conducting case-control studies should also plan to further enrich and engage the population without the disease. Studies run the risk of not collecting sufficient data from controls to perform case-control analysis with participants without disease seen to be dropping out significantly early. Similarly, more efforts^{46–48} are needed to retain the younger population that although demonstrates large enrollment, also features a majority that drops out on day one.

Distinct patterns in daily app usage behavior, also shown previously,⁴⁹ further strengthen the evidence of unequal technology utilization in remote research. The majority of the participants found in the abandoners group (C5*) who dropped out of the study on day 1, may also reflect initial patterns in willingness to participate in research, in a way that cannot be captured by recruitment in traditional research. Put another way, although there is significant dropout in remote trials, these early dropouts may be able to yield very useful information about differences in

people who are willing to participate in research and those who are not willing to participate. For decades clinical research has been criticized for its potential bias because people who participate in research may be very different from people who do not participate in research.^{50–52} Although researchers will not have longitudinal data from those who discontinue participation early, the information collected during onboarding can be used to assess potential biases in the final sample and may inform future targeted retention strategies.

Only 1 in 10 participants were in the high app use clusters (C1–2), and these clusters tended to be largely Non-Hispanic whites and older adults. Minority and younger populations, on the other hand, were represented more in the clusters with the lowest daily app usage (Fig. 4d). The largest impact on participant retention (>10 times) in the present sample was associated with clinician referral for participating in a remote study. This referral can be very light touch in nature, for example in the ElevateMS study, it consisted solely of clinicians handing patients a flyer with information about the study during a regular clinic visit. This finding is understandable, given recent research⁵³ showing that the majority of Americans trust medical doctors.

For most studies, the recruited sample was also inadequately diverse highlighting a persistent digital divide⁵⁴ and continued challenges in the recruitment of racial and ethnic underserved communities.⁵⁵ Additionally, the underrepresentation of States in the southern, rural and midwest regions indicates that areas of the US that often bear a disproportionate burden of disease⁵⁶ are under-represented in digital research.^{56–58} This recruitment bias

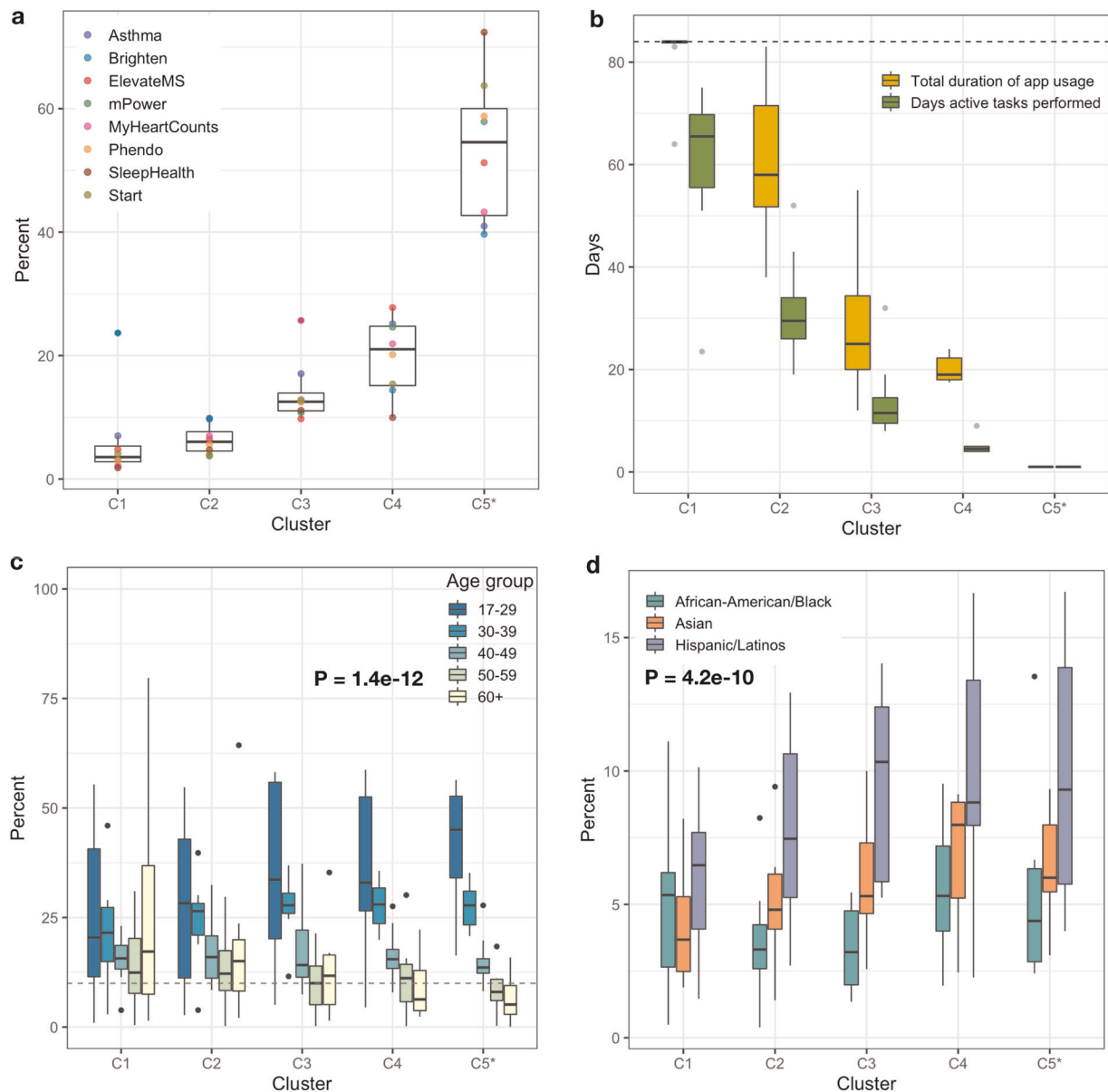


Fig. 4 Comparison of characteristics across participant engagement clusters. **a** Proportion of participants in each cluster across the study apps, **b** Participants total app usage duration (between 1–84 days) and the number of days participants completed tasks in the study apps, **c** Significant differences [$F(4,163) = 18.5$, $P = 1.4e-12$] in the age groups of participants across five clusters and **d** Significant differences [$F(2,81) = 28.5$, $P = 4.1e-10$] in proportion of minority population present in the five clusters. C5* cluster contains the participants that used the apps for less than a week and were removed from the clustering; however, they were added back for accurate proportional comparison of participants in each cluster. The centerline of the boxplots shows the median value across the studies for each cluster and upper and lower whisker corresponds to the outlier point that is at least 1.5 times the interquartile range.

could impact future studies that aim to collect data for health conditions that are more prevalent among certain demographic⁵⁹ and associated with geographic groups.⁶⁰ Recent research⁶¹ has also shown that participants' willingness to join remote research studies and share data are tied to their trust in the scientific team conducting the study including the institutional affiliations of researchers. Using different recruitment strategies^{46–48} including targeted online ads in regions known to have a larger proportion of the minority groups, partnerships with local community organizations, clinics and universities may help improve the penetration of remote research and improve diversity in the recruited sample. The ongoing "All of Us" research program that includes remote digital data collection has shown the feasibility of using a multifaceted approach to recruit a diverse sample with a

majority of the cohort coming from communities under-represented in biomedical research.⁶² Additionally, simple techniques, such as stratified recruitment that is customized based on the continual monitoring of the enrolling cohort demographics, can help enrich for a target population.

Finally, communication in digital health research may benefit from adopting the diffusion of innovations approach^{63,64} that has been applied successfully in healthcare settings to change behavior including the adoption of new technologies.^{65–67} Research study enrollments, advertisements including in-app communication and return of information to participants,⁶⁸ could be tailored to fit three distinct personality types (trendsetters, majority, and laggards). While trendsetters will adopt innovations early, they are a minority (15%) compared to the majority (greater

than two-thirds of the population) who will adopt a new behavior after hearing about its real-benefits, utility and believe it is the status quo. On the other hand, laggards (15%) are highly resistant to change and hard to reach online and as a result, will require more targeted and local outreach efforts.

These results should also be viewed within the context of limitations related to integrating diverse user-engagement data across digital health studies that targeted different disease areas with varying underlying disease characteristics and severity. While we did adjust for potential study level heterogeneity, we were not able to account for within-study differences such as variations in participants' disease severity and any other study-specific temporal changes. For example, the user experience and burden could have changed or improved over time based on changes in the study protocol or other technical fixes in the app. The variations in participant recruitment were not fully documented across studies so not be analyzed and accounted for fully except for clinical referrals in two of the studies. Furthermore, the comparison of participant race/ethnicity and geographical diversity to a general US population was meant to assess the representativeness of the study that is aimed at recruiting from a general population and not necessarily targeted towards a specific clinical condition of interest. Researchers should also prioritize to collect demographic data such as age, gender, race/ethnicity, participant state during onboarding which help characterize user attrition in future studies. While sensitivity analysis showed the main findings from user retention analysis do not change by including participants with missing data, however, missing demographic characteristics remains a significant challenge for digital health (See Supplementary Table 7). Finally, obtaining raw user-level engagement data from digital health studies that is well annotated and computable remains a significant challenge. The present findings are based on eight select US-based digital health studies and thus may not be generalizable broadly. We do, however, hope that this work will help motivate digital health researchers to share user-level engagement data to help guide a larger systematic analysis of participant behavior in digital health studies.

Despite these limitations, the present investigation to the best of our knowledge is the largest cross-study analysis of participant retention in remote digital health studies using individual-level data. While the technology has enabled researchers to reach and recruit participants for conducting large-scale health research in short periods of time, more needs to be done to ensure equitable access and long-term utilization by participants across different populations. The low retention in "fully remote, app-based" health research may also need to be seen in the broad context of the mobile app industry where similar user attrition is reported.⁶⁹ Attrition in remote research may also be impacted by study burden³⁰ as frequent remote assessments can compete with users' everyday priorities and perceived value proposition for completing a study task that may not be linked to an immediate monetary incentive. Using co-design techniques⁷⁰ for developing study apps involving researchers and participants could help guide the development of most parsimonious research protocols that fit into the daily lives of people and are still sufficiently comprehensive for researchers.

In the present diverse sample of user-activity data, several cohort characteristics, such as age, disease status, clinical referral, monetary benefits, etc, have emerged as key drivers for higher retention. These characteristics may also guide the development of new data-driven engagement strategies^{71,72} such as tailored just-in-time interventions⁷³ targeting sub-populations that are most likely to dropout early from remote research. Left unchecked the ongoing bias in participant recruitment combined with inequitable long-term participation in large-scale "digital cohorts" can severely impact the generalizability^{36,37} and undermine the

promise of digital health in collecting representational real-world data.

METHODS

Data acquisition

The user-engagement data were compiled by combining data from four studies that published annotated, accessible, and computable user-level data^{16,19,74,75} under qualified researcher program⁷⁶ as well as new data from four other digital health studies (SleepHealth,⁷⁷ Start,⁷⁸ Phendo,⁷⁹ and ElevateMS⁸⁰) that were contributed by collaborators. These eight studies aimed at assessing different diseases ranging from Parkinson's (mPower), asthma (Asthma health), heart condition (MyHeartCounts), sleep health (SleepHealth), multiple sclerosis (ElevateMS), endometriosis (Phendo) to depression (Brighten and Start). Except for three studies (Brighten, Phendo, and Start) that aimed to recruit people with a specific clinical condition, the other five studies enrolled people from the general population. Anyone with and without the target disease that met the other study eligibility criteria could join. The studies recruited participants from throughout the US between 2014–2019 using a combination of different approaches including placing ads on social media, publicizing or launching the study at a large gathering, partnerships with patient advocacy groups, clinics, and through word of mouth. In all studies, participants were enrolled fully remotely either through a study website or directly through the study app and provided electronic consent⁸¹ to participating in the study. Ethical oversight of the eight remote health studies included in the analysis was conducted by the respective institution's Institutional Review Board/Ethics Boards: Brighten (University of California, San Francisco), SleepHealth (University of California, San Diego), Phendo (Columbia University), Start (GoodRx), mPower (Sage Bionetworks), elevateMS (Sage Bionetworks), Asthma (Mt. Sinai), MyHeartCounts (Stanford University). The present retention analysis study used existing de-identified data only and qualifies for exemption status per OHRP guidelines.⁸²

The studies were launched at different time points during the 2014–2019 period, including three studies mPower, MyHeartCounts, and Asthma being launched with the public release of ResearchKit framework⁸³ released by Apple in March 2015. The studies were also active for different time periods including significant differences in the minimum time participants were expected to participate in the studies remotely. While Brighten and ElevateMS had a fixed 12-week participation period, other studies allowed participants to remain active for as long as they desired. Given this variation in the expected participation period across the studies, we selected the minimum common time period of the first 12 weeks (84 days) of each participant's activity in each study for retention analysis. Finally, with the exception of Brighten study which was a randomized interventional clinical trial and enrolled depressed cohort offering them monetary incentives for participation, the rest of the seven studies were observational and did not offer any direct incentives for participation. The studies also collected different real-world data ranging from frequent subjective assessments, objective sensor-based tasks to continual passive data⁴¹ collection.

Data harmonization

User-activity data across all the apps were harmonized to allow for inter-app comparison of user-engagement metrics. All in-app surveys and sensor-based tasks (e.g., Finger tapping on the screen) were classified as "active tasks" data type. The data gathered without explicit user action such as daily step count (Apple's health kit API⁸⁴), daily local weather patterns were classified as "passive" data type and were not used for assessing active user-engagement. The frequency at which the active tasks were administered in the study apps were aligned based on the information available in the corresponding study publication or obtained directly from the data contributing team in case the data were not publicly available. Furthermore, there were significant differences in the baseline demographics that were collected by each app. A minimal subset of four demographic characteristics (age, gender, race, state) was used for participant recruitment and retention analysis. A subset of five studies (mPower, ElevateMS, SleepHealth, Asthma, MyHeartCounts) had enrolled participants with and without disease status and were used to assess retention differences between people with (case) and without (control) disease. Two studies (mPower and ElevateMS) had a subset of participants that were referred to use the same study app by their care providers. For this smaller but unique subgroup, we compared the retention differences between clinically referred participants to self-referred participants.

Statistical analysis

We used three key metrics to assess participant retention and long-term engagement. (1) Duration in the study: the total duration, a study participant remained active in the study i.e., the number of days between the first and last active task completed by the participant, during the first 84 days of each participant's time in the study. (2) Days active in the study: the number of days a participant performed any active task in the app within the first 84 days. (3) User-activity streak: a binary-encoded vector representing the 84 days of potential app participation for each participant (Fig. 3a) where the position of the vector indicates the participant's day in the study and is set to 1 (green box) if at least one active task is performed on that day or else is 0 (white). User-activity streak metric was used to assess sub-populations that show similar longitudinal engagement patterns over a 3 month period.

Participant retention analysis (survival analysis⁸⁵) was conducted using the total duration of time in the study as the outcome metric to compare the retention differences across studies, sex, age group, disease status, and clinical referral for study-app usage. The duration of each participant in the study was assessed based on "active task" completion i.e., tasks that require active user input (e.g., a survey or a sensor-based active task). With the underlying user-level engagement data available across selected eight studies, we used an individualized pooled data analysis (IPDA) approach⁸⁶ to compare participant retention. IPDA has shown to yield more reliable inference compared to pooling estimates from published studies.⁸⁶ Log-rank test⁸⁷ was used to compare significant differences in participant retention between different comparator groups of interest. In order to adjust for potential study level heterogeneity, we used a stratified version of the log-rank test. Kaplan-Meier⁸⁸ plots were used to summarize the effect of the main variable of interest by pooling the data across studies where applicable. Two approaches were used to evaluate participant retention using survival analysis. (1) No censoring (most conservative)—if the last active task completed by participant fell within the pre-specified study period of the first 84 days, we considered it to be a true event i.e., participant leaving the study (considered "dead" for survival analysis). (2) Right-censoring⁸⁸—to assess the sensitivity of our findings using approach 1, we relaxed the determination of true event (participant leaving the study) in the first 12 weeks to be based on the first 20 weeks of app activity (additional 8 weeks). For example, if a participant completes last task in an app on day 40 (within the first 84 days) and then additionally completes more active task/s between week 13–20 he/she was still considered alive (no event) during the first 84 days (12 weeks) of the study and therefore "right-censored" for survival analysis. Given that age and gender had a varying degree of missingness across studies; additional analysis comparing the retention differences between the two sub-groups that provided the demographics and that opted out was done to assess the sensitivity of missing data on main findings.

Unsupervised k-means clustering method was used to investigate the longitudinal participant engagement behavior within each study using the user-activity streak metric (described above). The number of optimum clusters (between 1–10) in each study was determined using the elbow method⁸⁹ that aims to minimize the within-cluster variation. Enrichment of demographic characteristics in each cluster was assessed using a one-way analysis of variance. Since the goal of this unsupervised clustering of user-activity streaks was to investigate the patterns in longitudinal participant engagement; we filtered out individuals who remained in the study for less than 7 days from clustering analysis. However, for post hoc comparisons of demographics across the clusters, the initially left-out participants were put in a separate group (C5*). The state-wise proportions of recruited participants in each app were compared to the 2018 US state population estimates using the data obtained from the US census bureau.⁹⁰ To eliminate potential bias related to marketing and advertising of the launch of Apple's Research kit platform on March 09, 2015, participants who joined and left the mPower, MyHeartCounts, Asthma studies within the first week of Research Kit launch ($N = 14,573$) were taken out from the user retention analysis. We initially considered using Cox proportional hazards model⁹¹ to test for the significance of variable of interest on user retention within each study accounting for other study-specific covariates. However, because the assumption of proportional hazards (tested using scaled Schoenfeld residuals) was not supported for some studies, these analyses were not further pursued. All statistical analyses were performed using R.⁹²

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Aggregated data for all studies are included as part of tables in the main manuscript or supplementary tables. Additionally, individual-level user-engagement data for all studies are available under controlled access through the Synapse (<https://doi.org/10.7303/syn20715364>).

CODE AVAILABILITY

The complete code used for the analysis is available through a GitHub code repository (https://github.com/Sage-Bionetworks/digitalHealth_RetentionAnalysis_PublicRelease).

Received: 1 October 2019; Accepted: 17 January 2020;

Published online: 17 February 2020

REFERENCES

- Fogel, D. B. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp. Clin. Trials Commun.* **11**, 156–164 (2018).
- Briel, M. et al. A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *J. Clin. Epidemiol.* **80**, 8–15 (2016).
- Bradshaw, J., Saling, M., Hopwood, M., Anderson, V. & Brodtmann, A. Fluctuating cognition in dementia with Lewy bodies and Alzheimer's disease is qualitatively distinct. *J. Neurol. Neurosurg. Psychiatry* **75**, 382–387 (2004).
- Snyder, M. & Zhou, W. Big data and health. *Lancet Digital Health.* [https://doi.org/10.1016/s2589-7500\(19\)30109-8](https://doi.org/10.1016/s2589-7500(19)30109-8) (2019).
- Sim, I. Mobile devices and health. *N. Engl. J. Med.* **381**, 956–968 (2019).
- Steinhilb, S. R., Muse, E. D. & Topol, E. J. The emerging field of mobile health. *Sci. Transl. Med.* **7**, 283rv3 (2015).
- ElZarrad, M. K., Khair ElZarrad, M. & Corrigan-Curay, J. The US Food and Drug Administration's Real-World Evidence Framework: A Commitment for Engagement and Transparency on Real-World Evidence. *Clin. Pharmacol. Therapeutics.* <https://doi.org/10.1002/cpt.1389> (2019).
- May, M. Clinical trial costs go under the microscope. *Nat. Med.* <https://doi.org/10.1038/d41591-019-00008-7> (2019).
- Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. *Pew Research Center's Global Attitudes Project.* <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/> (2019).
- Turner, B. A. Smartphone Addiction & Cell Phone Usage Statistics in 2018. *BankMyCell.* <https://www.bankmycell.com/blog/smartphone-addiction/> (2018).
- Global time spent on social media daily 2018 | Statista. *Statista.* <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>.
- Steinhilb, S. R. & Topol, E. J. Digital medicine, on its way to being just plain medicine. *npj Digit Med.* **1**, 20175 (2018).
- Perry, B. et al. Use of mobile devices to measure outcomes in clinical research, 2010–2016: a systematic literature review. *Digit Biomark.* **2**, 11–30 (2018).
- Trister, A., Dorsey, E. & Friend, S. Smartphones as new tools in the management and understanding of Parkinson's disease. *npj Parkinson's Dis.* **2**, 16006 (2016). <https://doi.org/10.1038/npjparkd.2016.6>.
- Dorsey, E. R. et al. The use of smartphones for health research. *Acad. Med.* **92**, 157–160 (2017).
- Bot, B. M. et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci. Data* **3**, 160011 (2016).
- Webster, D., Suver, C., Doerr, M. et al. The Mole Mapper Study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. *Sci. Data* **4**, 170005 (2017). <https://doi.org/10.1038/sdata.2017.5>.
- Crouthamel, M. et al. Using a ResearchKit smartphone app to collect rheumatoid arthritis symptoms from real-world participants: feasibility study. *JMIR mHealth uHealth* **6**, e1177 (2018).
- McConnell, M. V. et al. Feasibility of obtaining measures of lifestyle from a Smartphone App: the MyHeart Counts Cardiovascular Health Study. *JAMA Cardiol.* **2**, 67–76 (2017).
- Chan, Y.-F. Y. et al. The Asthma Mobile Health Study, a large-scale clinical observational study using ResearchKit. *Nat. Biotechnol.* **35**, 354–362 (2017).
- McKillop, M., Mamykina, L. & Elhadad, N. Designing in the Dark: Eliciting Self-tracking Dimensions for Understanding Enigmatic Disease. in *Proc 2018 CHI Conference on Human Factors in Computing Systems - CHI '18* 1–15 (ACM Press, 2018).
- Waalens, J. et al. Real world usage characteristics of a novel mobile health self-monitoring device: results from the Scanadu Consumer Health Outcomes (SCOUT) Study. *PLoS ONE* **14**, e0215468 (2019).

23. Pratap, A. et al. Using mobile apps to assess and treat depression in Hispanic and Latino Populations: fully remote randomized clinical trial. *J. Med. Internet Res.* **20**, e10130 (2018).
24. Anguera, J. A., Jordan, J. T., Castaneda, D., Gazzaley, A. & Areán, P. A. Conducting a fully mobile and randomised clinical trial for depression: access, engagement and expense. *BMJ Innov.* **2**, 14–21 (2016).
25. Smalley, E. Clinical trials go virtual, big pharma dives in. *Nat. Biotechnol.* **36**, 561–562 (2018).
26. Virtual Clinical Trials Challenges and Opportunities: Proceedings of a Workshop: Health and Medicine Division. (2019). <http://www.nationalacademies.org/hmd/Reports/2019/virtual-clinical-trials-challenges-and-opportunities-pw.aspx>.
27. Orri, M., Lipset, C. H., Jacobs, B. P., Costello, A. J. & Cummings, S. R. Web-based trial to evaluate the efficacy and safety of tolterodine ER 4 mg in participants with overactive bladder: REMOTE trial. *Contemp. Clin. Trials* **38**, 190–197 (2014).
28. Moore, S. et al. Consent processes for mobile app mediated research: systematic review. *JMIR Mhealth Uhealth* **5**, e126 (2017).
29. Althubaiti, A. Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* **9**, 211–217 (2016).
30. Druce, K. L., Dixon, W. G. & McBeth, J. Maximizing engagement in mobile health studies: lessons learned and future directions. *Rheum. Dis. Clin. North Am.* **45**, 159–172 (2019).
31. Christensen, H., Griffiths, K. M., Korten, A. E., Brittliffe, K. & Groves, C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. *J. Med. Internet Res.* **6**, e46 (2004).
32. Christensen, H., Griffiths, K. M. & Jorm, A. F. Delivering interventions for depression by using the internet: randomised controlled trial. *BMJ* **328**, 265 (2004).
33. Eysenbach, G. The law of attrition. *J. Med. Internet Res.* **7**, e11 (2005).
34. Christensen, H. & Mackinnon, A. The law of attrition revisited. *J. Med. Internet Res.* **8**, e20 (2006).
35. Eysenbach, G. The law of attrition revisited—Author's -reply. *J. Med. Internet Res.* **8**, e21 (2006).
36. Kaplan, R. M., Chambers, D. A. & Glasgow, R. E. Big data and large sample size: a cautionary note on the potential for bias. *Clin. Transl. Sci.* **7**, 342–346 (2014).
37. Neto, E. C. et al. A Permutation Approach to Assess Confounding in Machine Learning Applications for Digital Health. in *Proc 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19*, <https://doi.org/10.1145/3292500.3330903> (2019).
38. Dainesi, S. M. & Goldbaum, M. Reasons behind the participation in biomedical research: a brief review. *Rev. Bras. Epidemiol.* **17**, 842–851 (2014).
39. Balachandran, M. Way To Health. (2019). <https://www.waytohealth.org/>.
40. Bentley, J. P. & Thacker, P. G. The influence of risk and monetary payment on the research participation decision making process. *J. Med. Ethics* **30**, 293–298 (2004).
41. Laport-López, F., Serrano, E., Bajo, J. & Campbell, A. T. A review of mobile sensing systems, applications, and opportunities. *Knowl. Inf. Syst.* (2019). <https://doi.org/10.1007/s10115-019-01346-1>.
42. Laursen, D. R. T., Paludan-Müller, A. S. & Hróbjartsson, A. Randomized clinical trials with run-in periods: frequency, characteristics and reporting. *Clin. Epidemiol.* **11**, 169–184 (2019).
43. Pablos-Méndez, A., Barr, R. G. & Shea, S. Run-in periods in randomized trials: implications for the application of results in clinical practice. *JAMA* **279**, 222–225 (1998).
44. Dickert, N. & Grady, C. What's the price of a research subject? Approaches to payment for research participation. *N. Engl. J. Med.* **341**, 198–203 (1999).
45. Gelinas, L. et al. A framework for ethical payment to research participants. *N. Engl. J. Med.* **378**, 766–771 (2018).
46. A guide to actively involving young people in research—INVOLVE. (2004). <https://www.invo.org.uk/posttypepublication/a-guide-to-actively-involving-young-people-in-research/>.
47. Nicholson, L. M. et al. Recruitment and retention strategies in longitudinal clinical studies with low-income populations. *Contemp. Clin. Trials* **32**, 353–362 (2011).
48. Nicholson, L. M., Schwirian, P. M. & Groner, J. A. Recruitment and retention strategies in clinical studies with low-income and minority populations: progress from 2004–2014. *Contemp. Clin. Trials* **45**, 34–40 (2015).
49. Druce, K. L. et al. Recruitment and ongoing engagement in a UK smartphone study examining the association between weather and pain: cohort study. *JMIR MHealth UHealth* **5**, e168 (2017).
50. Liu, H.-E. & Li, M.-C. Factors influencing the willingness to participate in medical research: a nationwide survey in Taiwan. *PeerJ* **6**, e4874 (2018).
51. Shavers, V. L., Lynch, C. F. & Burmeister, L. F. Factors that influence African-Americans' willingness to participate in medical research studies. *Cancer* **91**, 233–236 (2001).
52. Trauth, J. M., Musa, D., Siminoff, L., Jewell, I. K. & Ricci, E. Public attitudes regarding willingness to participate in medical research studies. *J. Health Soc. Policy* **12**, 23–43 (2000).
53. 5 key findings about public trust in scientists in the U.S. *Pew Research Center*. (2019). <https://www.pewresearch.org/fact-tank/2019/08/05/5-key-findings-about-public-trust-in-scientists-in-the-u-s/>.
54. Nebeker, C., Murray, K., Holub, C., Haughton, J. & Arredondo, E. M. Acceptance of mobile health in communities underrepresented in biomedical research: barriers and ethical considerations for scientists. *JMIR Mhealth Uhealth* **5**, e87 (2017).
55. Christopher Gibbons, M. Use of health information technology among racial and ethnic underserved communities. *Perspect. Health Inf. Manag.* **8**, 1f (2011).
56. Massoud, M. R., Rashad Massoud, M. & Kimble, L. Faculty of 1000 evaluation for Global burden of diseases, injuries, and risk factors for young people's health during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *F1000 - Post-Publ. Peer Rev. Biomed. Lit.* (2016). <https://doi.org/10.3410/f.726353640.793525182>.
57. Oh, S. S. et al. Diversity in clinical and biomedical research: a promise yet to be fulfilled. *PLoS Med.* **12**, e1001918 (2015).
58. Kondo, N. et al. Income inequality, mortality, and self rated health: meta-analysis of multilevel studies. *BMJ* **339**, b4471 (2009).
59. Krieger, N., Chen, J. T., Waterman, P. D., Rehkopf, D. H. & Subramanian, S. V. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures—The Public Health Disparities Geocoding Project. *Am. J. Public Health* **93**, 1655–1671 (2003).
60. Roth, G. A. et al. Trends and patterns of geographic variation in cardiovascular mortality among US Counties, 1980–2014. *JAMA* **317**, 1976–1992 (2017).
61. Pratap, A. et al. Contemporary views of research participant willingness to participate and share digital data in biomedical research. *JAMA Netw. Open* **2**, e1915717 (2019).
62. All of Us Research Program Investigators et al. The 'All of Us' Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
63. Rogers, E. M., Singhal, A. & Quinlan, M. M. in *An Integrated Approach to Communication Theory and Research*. Taylor & Francis Group. (2019). <https://doi.org/10.4324/9780203710753-35>.
64. Karni, E. & Safra, Z. in *Risk, Decision and Rationality*. Taylor & Francis Group. (1988). https://doi.org/10.1007/978-94-009-4019-2_11.
65. Greenhalgh, T. et al. Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. *BMJ* **337**, a1786 (2008).
66. Ward, R. The application of technology acceptance and diffusion of innovation models in healthcare informatics. *Health Policy Technol.* **2**, 222–228 (2013).
67. Emani, S. et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. *J. Med. Internet Res.* **14**, e150 (2012).
68. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy & Committee on the Return of Individual-Specific Research Results Generated in Research Laboratories. *Returning Individual Research Results to Participants: Guidance for a New Research Paradigm*. (National Academies Press, 2018).
69. Mobile Apps: What's A Good Retention Rate? (2018). <http://info.localytics.com/blog/mobile-apps-whats-a-good-retention-rate>.
70. Yardley, L., Morrison, L., Bradbury, K. & Muller, I. The person-based approach to intervention development: application to digital health-related behavior change interventions. *J. Med. Internet Res.* **17**, e30 (2015).
71. O'Connor, S. et al. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. *BMC Medical Informatics and Decision Making* **16** (2016).
72. Pagoto, S. & Bennett, G. G. How behavioral science can advance digital health. *Transl. Behav. Med.* **3**, 271–276 (2013).
73. Nahum-Shani, I. et al. Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* **52**, 446–462 (2018).
74. Chan, Y.-F. Y. et al. The asthma mobile health study, smartphone data collected using ResearchKit. *Sci. Data* **5**, 180096 (2018).
75. Hershman, S. G. et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart Counts Study. *Sci. Data* **6**, 24 (2019).
76. Wilbanks, J. & Friend, S. H. First, design for data sharing. *Nat. Biotechnol.* **34**, 377–379 (2016).
77. SleepHealth Study and Mobile App - SleepHealth. *SleepHealth*. (2019). <https://www.sleephealth.org/sleephealthapp/>.
78. Website. (2019). <https://www.goodrx.com/>.
79. Phendo. (2019). <http://citizenendo.org/phendo/>.
80. elevateMS. (2019). <http://www.elevatems.org/>.
81. Coiera, E. e-Consent: the design and implementation of consumer consent mechanisms in an electronic environment. *J. Am. Med. Inform. Assoc.* **11**, 129–140 (2003).
82. Electronic Code of Federal Regulations (eCFR). *Electronic Code of Federal Regulations (eCFR)*. (2019). <https://www.ecfr.gov/>.
83. ResearchKit and CareKit. *Apple*. (2019). <http://www.apple.com/researchkit/>.

84. Apple Inc. HealthKit - Apple Developer. (2019). <https://developer.apple.com/healthkit/>.
85. Bewick, V., Cheek, L. & Ball, J. Statistics review 12: survival analysis. *Crit. Care* **8**, 389 (2004).
86. Blettner, M. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int. J. Epidemiol.* **28**, 1–9 (1999).
87. Bland, J. M. & Altman, D. G. The logrank test. *BMJ* **328**, 1073 (2004).
88. Rich, J. T. et al. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol. Head. Neck Surg.* **143**, 331–336 (2010).
89. Parsons, S. Introduction to Machine Learning by Ethem Alpaydin, MIT Press, 0-262-01211-1, 400 pp., \$50.00/£32.95. The Knowledge Engineering Review. 432–433 (2005).
90. US Census Bureau. 2018 National and State Population Estimates. <https://www.census.gov/newsroom/press-kits/2018/pop-estimates-national-state.html>.
91. Kumar, D. & Klefsjö, B. Proportional hazards model: a review. *Reliab. Eng. Syst. Saf.* **44**, 177–188 (1994).
92. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2013). <http://www.R-project.org/>.

ACKNOWLEDGEMENTS

This work was supported in part by various funding agencies that included the National Institute of Mental Health (MH100466), Robert Wood Johnson Foundation (RWJF - 73205), National Library of Medicine (R01LM013043), National Center for Advancing Translational Sciences (1UL1TR002319-01) and American Sleep Apnea Foundation, Washington, DC. The funding agencies did not play a role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit this article for publication. We also would like to thank Phendo, SleepHealth, and GoodRx groups for sharing user-engagement data from their respective digital health studies for the retention analysis. We also acknowledge and thank researchers from previously completed digital health studies for making the user-engagement data available to the research community under qualified researcher program. Specifically, mPower study data were contributed by users of the Parkinson mPower mobile application as part of the mPower study developed by Sage Bionetworks and described in Synapse [<https://doi.org/10.7303/syn4993293>]. MyHeartCounts study data were contributed by users of the My Heart Counts Mobile Health Application study developed by Stanford University and described in Synapse [<https://doi.org/10.7303/syn11269541>]. Asthma study data were contributed by users of the asthma app as part of the Asthma Mobile Health Application study developed by The Icahn School of Medicine at Mount Sinai and described in Synapse [<https://doi.org/10.7303/syn8361748>]. We would like to thank and acknowledge all study participants for contributing their time and effort to participate in these studies.

AUTHOR CONTRIBUTIONS

A.P. conceptualized the study, wrote the initial analytical plan, carried out the analysis, and wrote the first draft of the paper. A.P. and P.S. integrated and harmonized user-engagement data across the studies. P.A.A. and A.P. managed the data-sharing agreements with data contributors (N.E., M.M., C.S. and D.G.). E.C.N., P.H. and L.O. made significant contributions to the data analysis. P.A.A., L.M. and L.O. helped interpret the analytical findings and provided feedback on the initial manuscript draft. A.P. and P.A.A. made major revisions to the first draft. C.S. and J.W. implemented the data governance and sharing model to enable the sharing of raw user-engagement data with the research community. L.O. and P.A.A. jointly oversaw the study. All authors assisted with the revisions of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41746-020-0224-8>.

Correspondence and requests for materials should be addressed to A.P.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020