Obtained funding: Chou.
Administrative, technical, or material support: Chou.
Study supervision: Rajkomar, Chou, Cui, Dean.

Conflict of Interest Disclosures: All authors are employed by and own stock in Google. In addition, as part of a broad-based equity portfolio intending to mirror the US and International equities markets (eg, MSCI All Country World, Russell 3000), Jeff Dean holds individual stock positions in many public companies in the health care and pharmacological sectors, and also has investments in managed funds that also invest in such companies, as well as limited partner and direct venture investments in private companies operating in these sectors. All other health care–related investments are managed by independent third parties (institutional managers) with whom Jeff Dean has no direct contact and over whom Jeff Dean has no control. The authors have a patent pending for the machine learning tool described in this study. No other conflicts are reported.

Additional Contributions: We thank Kathryn Rough, PhD, and Mila Hardt, PhD, for helpful discussions on the manuscript; Mike Pearson, MBA, Ken Su, MBA, MBH, and Kasumi Widner, MS, for data collection; Diana Jaunzeikare, BA, Chris Co, PhD, Daniel Tse, MD, and Nina Gonzalez, MD, for labeling; Linh Tran, PhD, Nan Du, PhD, Yu-hui Chen, PhD, Yonghui Wu, PhD, Kyle Scholz, BS, Izhak Shafran, PhD, Patrick Nguyen, PhD, Chung-cheng Chiu, PhD, Zhifeng Chen, PhD, for helpful discussions on modeling; and Rebecca Rolfe, MSc, for illustrations. All individuals work at Google. They were not compensated outside of their normal duties for their contributions.

1. Verghese A, Shah NH, Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA*. 2018;319(1):19-20. doi:10.1001/jama.2017.19198

2. Chiu C-C, Tripathi A, Chou K, et al. Speech Recognition for Medical Conversations. In: Interspeech 2018. ISCA: ISCA; 2018. https://www.isca-speech.org/archive/Interspeech_2018/abstracts/0040.html. Accessed December 8, 2018.

3. Sutskever I, Vinyals O, Le QV. Sequence to Sequence Learning with Neural Networks. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ, eds. Advances in Neural Information Processing Systems. vol 27. 2014. http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf. Accessed December 8, 2018.

4. Cho K, van Merriënboer B, Gülçehre Ç, et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014:1724-1734.

5. Kannan A, Chen K, Jaunzeikare D, Rajkomar A. Semi-supervised Learning for Information Extraction from Dialogue. In: Interspeech 2018. ISCA: ISCA; 2018. https://www.isca-speech.org/archive/Interspeech_2018/abstracts/1318.html. Accessed December 8, 2018.

Characteristics of Digital Health Studies Registered in ClinicalTrials.gov

Digital health is the application of software or hardware, often using mobile smartphone or sensor technologies to improve patient or population health and health care delivery.[1] In contrast to drugs and traditional medical devices, which have strict regulatory guidelines on safety and efficacy, the clinical evidence generation for digital health tools may be motivated by other factors, including adoption, utilization, and value, that may influence study design and quality. The landscape of clinical evidence underlying digital health interventions has not been well characterized.[2,3] We sought to evaluate the characteristics of digital health studies registered in ClinicalTrials.gov.

Invited Commentary
page 840

Methods | We performed a cross-sectional analysis of digital health studies in ClinicalTrials.gov.[4,5] To identify studies evaluating mobile-, web-, and electronic-based tools as well as digital medical devices, we searched ClinicalTrials.gov on January 22, 2017, using the Medical Subject Heading concepts (*mobile health*, *mHealth*, *ehealth*, *telehealth*, and *telemedicine*) and commonly used lay terms (*digital health*, *consumer health*, *mobile application*, and *wireless technology*). Variables were exported as structured fields when downloaded from ClinicalTrials.gov.[6] A single reviewer (C.E.C.) verified studies for inclusion, removed duplicates, and assigned each study to 1 of 13 clinical areas determined by iterative qualitative clustering against commonly accepted medicine domains. Descriptive statistics were calculated for key study characteristics, with additional stratification by study type (interventional vs observational, randomization status). We used the χ^2 test to compare proportions, and $P < .05$ was considered to be statistically significant.

Table 1. Digital Health Studies Registered in ClinicalTrials.gov

Study Type	No. (%) of Studies
All (N = 1783)	
Interventional	1570 (88.1)
Observational	213 (11.9)
Study allocation (n = 1776)	
Randomized	1257 (70.8)
Nonrandomized	519 (29.2)
Recruitment status[a]	
Not yet recruiting	218 (12.2)
Recruiting or enrolling	535 (30.0)
Active, not recruiting	176 (9.9)
Completed	692 (38.9)
Withdrawn or terminated	56 (3.1)
Unknown	103 (5.8)
Interventional (n = 1570)	
Intervention model (n = 1563)	
Parallel assignment	1147 (73.4)
Crossover assignment	88 (5.6)
Factorial assignment	43 (2.8)
Single group assignment	282 (18.0)
Masking (n = 1561)	
Double-blind	107 (6.9)
Single-blind[a]	417 (26.7)
Open-label or no masking	1031 (66.0)
Observational (n = 213)	
Observational model (n = 180)	
Case-control	26 (14.4)
Case-only	39 (21.7)
Cohort	100 (55.6)
Other	15 (8.3)
Time perspective (n = 199)	
Prospective	157 (78.9)
Retrospective	19 (9.5)
Cross-sectional	21 (10.5)
Completed, Suspended, Withdrawn, Terminated (n = 751)	
Study results	
Available	85 (11.3)
Not available	666 (88.7)

[a] Randomization status unknown for 7 studies.
[b] Patient, principal investigator, or assessor.

Table 2. Study Characteristics Stratified by Study Design Characteristics

Study Characteristic	No. (%) of Studies			χ^2 *P* Value	No. (%) of Studies		χ^2 *P* Value
	All (N = 1783)	Observational (n = 213)	Interventional (n = 1570)		Nonrandomized (n = 519)	Randomized (n = 1257)	
Start year							
Before 2011	244 (13.7)	27 (12.7)	215 (13.7)		53 (8.6)	185 (13)	
2011	97 (5.4)	14 (6.6)	83 (5.3)		31 (5.6)	66 (5.1)	
2012	119 (6.7)	13 (6.1)	106 (6.8)		37 (6.2)	82 (6.3)	
2013	200 (11.2)	27 (12.7)	173 (11)		53(10)	147 (12)	
2014	259 (14.5)	35 (16.4)	224 (14.3)	.26	81 (16)	178 (15)	.26
2015	334 (18.7)	41 (19.2)	293 (16.7)		100 (20)	234 (20)	
2016	415 (23.3)	47 (22)	368 (23.4)		132 (28)	282 (23)	
2017	108 (6.1)	9 (8.3)	101 (6.4)		30 (5.4)	79 (6.5)	
Not stated	7 (0.4)	NA	7 (0.4)		2	4	
Target age[a]							
Child (<18 y)	374 (20.1)	45 (21.1)	329 (21)	.95	117 (22.5)	253 (20.1)	.25
Adult (18-65 y)	1680 (94.2)	204 (95.8)	1476 (94)	.30	487 (93.8)	1186 (94.3)	.67
Senior (≥66 y)	1302 (73)	174 (81.7)	1128 (71.8)	.002	387 (74.5)	908 (72.2)	.31
Target sex							
Men only	34 (1.9)	2 (0.9)	32 (2.0)		7 (1.4)	27 (2.1)	
Women only	139 (9.0)	10 (4.7)	129 (8.22)	.005	33 (6.4)	106 (8.4)	.04
Both	1607 (89)	199 (93.4)	1408 (89.7)		477 (91.9)	1124 (89.4)	
Not stated	3 (0.1)	2 (0.9)	1 (0.06)		NA	NA	
Clinical area							
Autoimmune	84 (4.7)	5 (2.3)	79 (5.0)		20 (3.9)	64 (5.1)	
Cardiometabolic	382 (21.4)	44 (20.7)	338 (21.5)		94 (18.1)	287 (2.9)	
Hematology-oncology	107 (6)	10 (4.7)	97 (6.2)		38 (7.3)	69 (5.5)	
Infectious disease	77 (4.3)	5 (2.4)	72 (4.6)		13 (2.5)	64 (5.1)	
Mental health	216 (12.1)	15 (7.0)	201 (12.8)		51 (9.8)	163 (13)	
Musculoskeletal or pain	54 (3)	6 (2.8)	48 (3.1)		12 (2.3)	41 (3.3)	
Neurology	114 (6.4)	29 (13.7)	85 (5.4)	<.001	58 (11.2)	56 (4.5)	<.001
Obstetrics-gynecology	63 (3.5)	10 (4.7)	53 (3.4)		20 (3.9)	43 (3.4)	
Pulmonary	113 (6.3)	15 (7.0)	98 (6.2)		37 (7.1)	76 (6.1)	
Renal	24 (1.4)	3 (1.4)	21 (1.3)		8 (1.5)	16 (1.3)	
Substance abuse	112 (6.3)	9 (4.2)	103 (6.6)		21 (4.1)	91 (7.2)	
Surgery	32 (1.8)	7 (3.3)	25 (1.6)		14 (2.7)	17 (1.4)	
Wellness	183 (10.2)	4 (2.2)	179 (11.4)		35 (6.7)	146 (11.6)	
Other	222 (12.4)	51 (23.9)	171 (10.9)		98 (18.9)	124 (9.9)	
Funding[b]							
NIH or federal	369 (20.7)	31 (14.6)	338 (21.5)	.02	72 (13.9)	294 (23.4)	<.001
Commercial or industry	214 (12.0)	48 (22.5)	166 (0.6)	<.001	88 (17.0)	125 (9.9)	<.001
Other	1602 (91)	184 (86.4)	1418 (90.3)	.07	467 (89.9)	1131 (70.8)	>.99
Trial size							
0-100	829 (46.5)	102 (47.9)	727 (46.3)		317 (61.1)	508 (40.4)	
101-1000	812 (45)	82 (38.5)	730 (46.5)	.002	154 (29.7)	657 (52.3)	<.001
>1000	142 (8.0)	29 (13.7)	113 (7.2)		48 (9.3)	92 (7.3)	

Abbreviations: NA, not applicable; NIH, National Institutes of Health.

[a] Some studies listed multiple funding sources and target audiences.

Results | We identified 1783 studies that met our inclusion criteria (from the top-level search of 3833 and after deduplication); 1570 studies (88.1%) were interventional, and 1257 (70.8%) were randomized. Among interventional studies, 107 (6.9%) were double-blinded and 417 (26.7%) single-blinded. Among observational studies, the most common study designs were case-control (26 [14.4%]), case-only (39 [21.7%]), and cohort (100 [55.6%]) (**Table 1**).

Most studies consisted of adults or elderly individuals; 374 (20.1%) enrolled children. The most common clinical areas

were cardiometabolic (382 [21.4%]), mental health (216 [12.1%]), and wellness (183 [10.2%]). Funding sources included federal and National Institutes of Health (369 [20.7%]) and industry (214 [12.0%]). Median enrollment was 120 (interquartile range, 50-300), although study sample size varied from fewer than 100 individuals (829 [46.5%]) to more than 1000 (142 [8.0%]) (**Table 2**). A higher proportion of publicly funded studies were interventional (338 [21.5%]) or randomized (294 [23.4%]), whereas a higher proportion of industry-funded studies were observational (48 [22.5%]) or nonrandomized (88 [17.0%]) (Table 2). Overall, 692 of 1783 studies (38.9%) had completed recruitment, and 85 completed or terminated studies (11.3%) had reported results. After multivariate adjustment, federally funded trials were more likely to have reported results (odds ratio, 4.9; 95% CI, 3.1-7.8; $P < .001$).

Discussion | We characterized the digital health clinical research landscape. Although the number of registered studies increased by a mean of 29% per year from 2011 to 2017, many were small. Federally funded studies were more likely to use interventional designs and randomization. However, few studies have reported findings to date, even among studies marked completed or terminated.

Our use of the ClinicalTrials.gov database has some limitations, most notably that submission of digital health trials, unlike that for drugs and devices, remains voluntary.[5] Although most stakeholders and sponsors generally require that prospective interventional trials be reported to ClinicalTrials.gov, these standards do not always apply to observational studies. Selection bias could lead to overestimation of the proportion of studies that are randomized across the full landscape and mean trial size, particularly if small pilot and nonregulated validation studies are underascertained. Because these data were extracted in 2017, ongoing assessment of the state of digital health studies is warranted.

Whether results will drive substantial clinical adoption is unknown because small studies, even if randomized, are unlikely to be significantly powered to demonstrate meaningful treatment effects. Although the pipeline of digital health studies appears to be promising, these factors could limit their ability to yield a high level of evidence, demonstrate value, or motivate stakeholder adoption.

Connie E. Chen, MD
Robert A. Harrington, MD
Sumbul A. Desai, MD
Kenneth W. Mahaffey, MD
Mintu P. Turakhia, MD, MAS

Author Affiliations: Department of Medicine, Stanford University School of Medicine, Stanford, California (Chen, Harrington, Desai, Mahaffey, Turakhia); Center for Digital Health, Stanford University School of Medicine, Stanford, California (Harrington, Desai, Turakhia); Stanford Center for Clinical Research, Department of Medicine, Stanford University School of Medicine, Stanford, California (Mahaffey); Veterans Affairs Palo Alto Health Care System, Palo Alto, California (Turakhia).

Accepted for Publication: October 28, 2018.

Corresponding Author: Mintu P. Turakhia, MD, MAS, Veterans Affairs Palo Alto Health Care System, 3801 Miranda Ave, 111C, Palo Alto, CA 94304 (mintu@stanford.edu).

Published Online: February 25, 2019. doi:10.1001/jamainternmed.2018.7235

Author Contributions: Dr Chen had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Chen, Harrington, Desai, Turakhia.
Acquisition, analysis, or interpretation of data: Chen, Mahaffey, Turakhia.
Drafting of the manuscript: Chen, Turakhia.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Chen, Turakhia.
Administrative, technical, or material support: Harrington, Turakhia.
Supervision: Harrington, Desai, Turakhia.

Conflict of Interest Disclosures: Dr. Chen reported being employed by Lyra Health and serving as a consultant for and having equity in Vida Health. Dr Harrington reported serving as a consultant for Adverse Events, Amgen, Bayer, Gilead, Merck & Co, Vida Health, and WebMD; performing data safety monitoring for AstraZeneca, BMS, and Janssen; having equity in Element Science and MyoKardia; having a fiduciary role in Scanadu, Signal Path, the American Heart Association, College of the Holy Cross, and Stanford Healthcare; and receiving research grants from CSL Behring, GlaxoSmithKline, Merck & Co, Novartis, Portola, Sanofi, and The Medicines Company. Dr Desai reported being employed by Apple. Dr. Mahaffey reported receiving research grants from Afferent, Amgen, Apple Inc, AstraZeneca, Cardiva Medical Inc, Daiichi Sankyo, Ferring, Google (Verily), Johnson & Johnson, Luitpold, Medtronic, Merck & Co, the National Institutes of Health, Novartis, Sanofi, St. Jude, and Tenax; serving as a consultant for Abbott, Ablynx, Baim Institute, Boehringer Ingelheim, Bristol-Myers Squibb, Cardiometabolic Health Congress, Elsevier, GlaxoSmithKline, Medergy, Medscape, Mitsubishi, MyoKardia, Novo Nordisk, Oculeve, Portola, Radiometer, Springer Publishing, Theravance, University of California San Francisco, and WebMD; and having equity in BioPrint Fitness. Dr Turakhia reported receiving research grants from Medtronic Inc, Research, Janssen Pharmaceuticals, AstraZeneca, the American Heart Association, and Amazon; serving as a consultant for Medtronic Inc and Abbott; and having equity in AliveCor. No other disclosures were reported.

1. Turakhia MP, Desai SA, Harrington RA. The outlook of digital health for cardiovascular medicine: challenges but also extraordinary opportunities. *JAMA Cardiol*. 2016;1(7):743-744. doi:10.1001/jamacardio.2016.2661

2. Ahern DK. Challenges and opportunities of ehealth research. *Am J Prev Med*. 2007;32(5)(suppl):S75-S82. doi:10.1016/j.amepre.2007.01.016

3. O'Neil A, Cocker F, Rarau P, et al. Using digital interventions to improve the cardiometabolic health of populations: a meta-review of reporting quality. *J Am Med Inform Assoc*. 2017;24(4):867-879. doi:10.1093/jamia/ocw166

4. National Institutes of Health. ClinicalTrials.gov. https://clinicaltrials.gov/. Accessed June 19, 2017.

5. National Institutes of Health. ClinicalTrials.gov FDAAA 801 requirements. https://clinicaltrials.gov/ct2/search/advanced?cond=&term=&state1=&cntry1=. Accessed June 19, 2017.

6. National Institutes of Health. ClinicalTrials.gov advanced search. https://clinicaltrials.gov/ct2/search/advanced?cond=&term=&state1=&cntry1=. Accessed June 19, 2017.

Invited Commentary

Evaluating Digital Health Tools—Prospective, Experimental, and Real World

Digital health software tools (DHSTs) are becoming increasingly available to patients, health care systems, and other key stakeholders seeking to enhance patient-centered care with innovative apps, sensors, algorithms, and data visualization approaches. Broadly speaking, DHSTs seek to improve care by providing more informed treatment recommendations, clarifying and refining diagnoses, optimizing workflows and efficiency, and facilitating access to and use of complex health care data. However, even as these tools increase in use and popularity, real challenges exist to identify which DHSTs are appropriate for which patients or which clinical settings, how to integrate them into

Related article page 838

clinical and other care, and which DHSTs are safe and effective in practice.[1]

In this issue of *JAMA Internal Medicine,* Chen and colleagues[2] present an important and interesting examination of the number and type of DHST studies registered in ClinicalTrials.gov,[1] a registry of clinical trials operated by the US National Library of Medicine at the National Institutes of Health. Their results are encouraging, suggesting that use of experimental studies of DHSTs has been increasing and that at least 70% were enrolling or completed by the end of 2017. The scope of these studies matches important trends in DHSTs, with a focus on mental health and cardiovascular illness, mapping to consumer trends more broadly. This focus is likely in part to be consumer market driven, as indicated by the large proportion of funding originating from nonfederal sources. It is also encouraging that most studies are using experimental designs for their evaluations and therefore might be more likely to produce scientifically valid results.

Nevertheless, DHST evaluations remain a challenge because DHSTs rapidly evolve (as software, sensors, and analytics advance) and are often deeply dependent on context. That is, the technology's effectiveness is often determined by the way in which it is provided to patients or practitioners, how it is supported (or taught), and how the DHST is added to clinical work or daily life.[3] As a result, randomized clinical trials of DHSTs may be more prone to limitations in terms of generalizability than traditional devices and drugs. The authors of this article did not examine the protocols in terms of the context or process of adapting the DHST into use, but understanding whether and to what degree these trials described more implementation science–focused factors (such as how and whether early modifications were required for the DHST or sensor and whether training or technical assistance was required) would be valuable information. It may also be reasonable for DHST trials to prespecify that there will be post hoc analysis focused on why certain groups used the DHST (or not) and whether use as intended vs as actually observed was associated with differences in outcomes. In this way, trials of DHSTs may be more akin to program evaluation of quality improvement studies than traditional drug or device studies.

The pipeline of DHST studies found at ClinicalTrials.gov appears to be primed, but it is hard to tell how these 2000 or so studies compare with the entire universe of DHSTs, which we suspect is larger. The increased use and rapid evolution of widely used DHSTs have prompted the US Food and Drug Administration (FDA) to begin a software precertification program[4] wherein software vendors, particularly those producing software as a medical device[5] or software that produces or supports diagnostic or therapeutic decisions, will be reviewed for the quality of their software production and quality-assurance activities at a company level. Companies' individual products will then be reviewed for safety, effectiveness, and usability as part of initial regulatory approval; part of this evaluation will include review of preapproval clinical outcomes data. Because of the rapidity of the evolution of digital tools, however, ongoing evaluation of DHSTs as part of precertification will take place after marking and FDA regulatory approval. The relevance of this approach to the article by Chen et al[2] is that future studies will likely still include experimental trials of the sort that they evaluated but will more likely be dominated by observational studies and real-world evidence.[6] These studies may consist of software and digital tools that may have undergone regulatory approval but most of which will have not. In most cases, there will be little to no phase 1 equivalent data to precede any studies.

As a result, most of our understanding of safety or effectiveness of DHSTs is likely to be made based on real-world evidence and performance data. The role of trials and trial registration in the future is not irrelevant but may need refocusing and repurposing. There are likely disease states, clinical use cases, or patient groups, such as situations in which the disease in question is serious or unstable or when the decision prompted by the DHST takes place in a risky clinical situation, for which clinical trials are required to ensure safety and effectiveness before a tool can be considered for wide use or expanded indications. In other cases, real-world performance data may in turn prompt a prospective clinical trial. In both situations, trial registration, reporting of context and digital nuance, and a commitment to empirical study will remain critically important to ensure that innovation is on target, rapid, and safe.

Andrew D. Auerbach, MD, MPH

Author Affiliations: Division of Hospital Medicine, University of California, San Francisco; Center for Digital Health Innovation, University of California, San Francisco.

Corresponding Author: Andrew D. Auerbach, MD, MPH, Division of Hospital Medicine, University of California, San Francisco, UCSF Box 0131, San Francisco, CA 94143-0131 (andrew.auerbach@ucsf.edu).

Published Online: February 25, 2019. doi:10.1001/jamainternmed.2018.7229

Conflict of Interest Disclosures: Dr Auerbach reported receiving grants from the US Food and Drug Administration–Centers of Excellence in Regulatory Science and Innovation, the Centers for Disease Control and Prevention, and Patient-Centered Outcomes Research Institute and other support from Society of Hospital Medicine and UpToDate outside the submitted work.

1. Auerbach A, Neinstein A, Khanna R. Balancing innovation and safety when integrating digital tools into health care. *Ann Intern Med*. 2018;168(10):733-734.

2. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of digital health studies registered in ClinicalTrials.gov [published online February 25, 2019]. *JAMA Intern Med*. doi:10.1001/jamainternmed.2018.7235

3. Meeks DW, Takian A, Sittig DF, Singh H, Barber N. Exploring the sociotechnical intersection of patient safety and electronic health record implementation. *J Am Med Inform Assoc*. 2014;21(e1):e28-e34. doi:10.1136/amiajnl-2013-001762

4. US Food and Drug Administration. Software Precertification Program: Excellence Model. 2018. https://www.fda.gov/medicaldevices/digitalhealth/digitalhealthprecertprogram/default.htm. Accessed February 13, 2018.

5. International Medical Device Regulators Forum. Software as a Medical Device (SaMD): Key Definitions. 2013:9. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf. Accessed February 13, 2018.

6. National Academies of Sciences. The National Academies Collection: reports funded by National Institutes of Health. In: *Real-World Evidence Generation and Evaluation of Therapeutics: Proceedings of a Workshop.* Washington, DC: National Academies Press; 2017.