Research and Applications

A digital health research platform for community engagement, recruitment, and retention of sexual and gender minority adults in a national longitudinal cohort study——The PRIDE Study

Mitchell R. Lunn,[1,2] Micah Lubensky,[2,3] Carolyn Hunt,[2,3] Annesa Flentje,[2,3] Matthew R. Capriotti,[2,4] Chollada Sooksaman,[5] Todd Harnett,[5] Del Currie,[6] Chris Neal,[6] and Juno Obedin-Maliver[2,7]

[1]Division of Nephrology, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA, [2]The PRIDE Study/PRIDEnet, Stanford University School of Medicine, Palo Alto, California, USA, [3]Department of Community Health Systems, School of Nursing, University of California, San Francisco, San Francisco, California, USA, [4]Department of Psychology, San Jose State University, San Jose, California, USA, [5]THREAD Research, Tustin, California, USA, [6]Analog Republic, London, UK and [7]Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, California, USA

Corresponding Author: Mitchell R. Lunn, MD, MAS, Division of Nephrology, Department of Medicine, Stanford University School of Medicine, 1701 Page Mill Road, Palo Alto, CA 94304, USA (lunn@stanford.edu)

ABSTRACT

Objective: Sexual and gender minority (SGM) people are underrepresented in research. We sought to create a digital research platform to engage, recruit, and retain SGM people in a national, longitudinal, dynamic, cohort study (The PRIDE Study) of SGM health.

Materials and Methods: We partnered with design and development firms and engaged SGM community members to build a secure, cloud-based, containerized, microservices-based, feature-rich, research platform. We created PRIDEnet, a national network of individuals and organizations that actively engaged SGM communities in all stages of health research. The PRIDE Study participants were recruited via in-person outreach, communications to PRIDEnet constituents, social media advertising, and word-of-mouth. Participants completed surveys to report demographic as well as physical, mental, and social health data.

Results: We built a secure digital research platform with engaging functionality that engaged SGM people and recruited and retained 13 731 diverse individuals in 2 years. A sizeable sample of 3813 gender minority people (32.8% of cohort) were recruited despite representing only approximately 0.6% of the population. Participants engaged with the platform and completed comprehensive annual surveys— including questions about sensitive and stigmatizing topics— to create a data resource and join a cohort for ongoing SGM health research.

Discussion: With an appealing digital platform, recruitment and engagement in online-only longitudinal cohort studies are possible. Participant engagement with meaningful, bidirectional relationships creates stakeholders and enables study cocreation. Research about effective tactics to engage, recruit, and maintain active participation from all communities is needed.

Conclusion: This digital research platform successfully recruited and engaged diverse SGM participants in The PRIDE Study. A similar approach may be successful in partnership with other underrepresented and vulnerable populations.

INTRODUCTION

Sexual and gender minority (SGM) people— those who identify as lesbian, gay, 2-spirit, bisexual, and transgender as well as those whose sexual orientation, gender identity and expressions, or reproductive development varies from traditional, societal, cultural, or physiological norms[1]— are an underserved, vulnerable, and understudied population. The National Institutes of Health has recognized SGM people as a "health disparity population for research"[2] because they experience numerous health and health care inequities including high smoking rates,[3] worse mental health outcomes,[4–7] high prevalence of certain infectious diseases,[8,9] and low utilization of preventative care services.[10–13]

SGM people are underrepresented in biomedical research for 2 primary reasons: (1) stigma and discrimination drive SGM people away from the health care system, and (2) there is limited SGM health-related data collection. SGM people report mistreatment from health care professionals including verbal and physical abuse.[14] Approximately 1-quarter (23%) of transgender people did not seek care due to fear of disrespect or mistreatment.[15] Structural discrimination is widespread; only 12 states having nondiscrimination laws to protect SGM people from being denied health insurance coverage,[16] and 21 states having laws prohibiting employment discrimination based on sexual orientation or gender identity.[17]

In addition to discrimination, there are limited SGM-related national data for researchers and epidemiologists to characterize SGM health. Despite comprising approximately 4–6% of the United States population,[18] SGM people are largely invisible to the federal government and in federal health surveillance surveys/systems. While growing in visibility in some arenas, sexual orientation and gender identity (SOGI) are not collected in the decennial United States Census or the ongoing American Community Survey,[19] which thereby inhibits our ability to describe SGM communities in terms of age, race/ethnicity, geography, etc. In 2011, the Institute of Medicine (now National Academy of Medicine) released a report on SGM health and found that "the relative lack of population-based data presents the greatest challenge to describing the health status and health-related needs of LGBT people."[20] As a result of this, sexual orientation was added to the National Health Interview Survey in 2013; SOGI questions are available in an optional module for state-based implementation in the Behavioral Risk Factor Surveillance System. Outside of limited federal efforts and poor/inconsistent collection of SOGI data in electronic health records,[21] there are limited mechanisms to comprehensively describe the health of diverse SGM populations.

Longitudinal cohorts are valuable for epidemiologic studies. They are, however, expensive and difficult to start, grow, and maintain.[22] The largest and longest longitudinal cohort studies,such as the Framingham Heart Study[23] and the Nurses' Health Study,[24] cost hundreds of millions of dollars and involved significant in-person physical examination and biospecimen collection. Newer longitudinal cohort studies, such as the UK Biobank[25] or the *All of Us* Research Program,[26] are larger (ie, 500 000 to 1 million participants), cost billions of dollars, and are conducted primarily online except for a single in-person physical examination and biospecimen collection visit. With notable technological advancements and rapidly decreasing costs of digital technologies, exclusively online longitudinal cohort studies may increase the efficiency and effectiveness of clinical research.

Participation of SGM people in longitudinal cohort studies would be enhanced through an exclusively online longitudinal cohort study. An online longitudinal cohort study of SGM people may be particularly relevant and successful because they are avid users of the internet and social media to get health information,[27] meet partners,[28,29] build community, and participate in research.[30–32] Digital interactions provide safety by avoiding interpersonal interactions that may be fraught with discrimination in one's local community. Although online health research frequently involves completing surveys, popular online survey software packages used in academic settings have limited functionality with data validation/verification, cohort engagement, facile reporting, integrations with third-party services, and other services that nurture a longstanding relationships with participants within the context of high-quality data collection.

In order to better characterize SGM physical, mental, and social health, a longitudinal cohort of SGM people for observational and interventional studies was needed. However, with limited existing SGM population-based data to develop sampling frames and with reported SGM mistreatment in health care, we desired to employ a community-engaged approach in order to provide the required trust, safety, and convenience to participants. We hypothesized that an online-only platform would be a valuable tool to engage and recruit a diverse national cohort of SGM adults. To this end, we designed and developed a robust digital research platform to support an online-only, community-engaged, longitudinal, dynamic (ie, continually enrolling), cohort study of SGM people distributed across the country entitled "The Population Research in Identities and Disparities for Equality (PRIDE) Study". In this article, we detail The PRIDE Study's digital research platform as a tool to engage, recruit, and retain SGM adults in an engaged, research-ready cohort.

OBJECTIVES

In this article, we describe the development of a cloud-based digital research infrastructure to effectively conduct a community-engaged longitudinal cohort study. Specifically, we sought to (1) develop a secure digital research platform with engagement, recruitment, retention, and reporting features to recruit and support a national, longitudinal, cohort study of diverse SGM adults distributed across the country; (2) develop a containerized, microservices-based platform that enables rapid implementation of new features or requirements; and (3) develop a facile system to develop and deploy PRIDE Ancillary Studies (additional research studies administered to the entire cohort or a subset). The platform we developed may be deployed on any computing infrastructure and may allow collaborative, community-engaged research with other groups including underserved/vulnerable populations.

SYSTEM DESCRIPTION

PRIDE research platform design and development

From June 2015 through April 2017, we conducted a pilot phase of The PRIDE Study.[33] In the pilot, we received participant feedback that influenced the specifications of this digital research platform

including editing demographic questions to make them more participant-centered, reordering of survey items to ensure recognition of identities, and providing the ability to edit profile information, as names and identities can change within SGM communities.

We partnered with a technology design and project management firm (THREAD Research; Tustin, CA) to gather and implement our business and functional requirements of the PRIDE digital research platform. In-house designers at THREAD Research proposed the user interface for an optimal user experience, which was reviewed, refined, and approved by the research team. THREAD Research managed the project and performed quality assurance (QA) checks, bug remediation, user acceptance testing, and live QA review. Software development (ie, coding), database configuration, third-party integrations, cloud infrastructure configuration, and load testing was performed by THREAD Research's trusted partner, Analog Republic (United Kingdom). Platform development used the scrum software development framework.[34] In addition to members from The PRIDE Study team, the development team included a digital producer, director of user experience, director of client services, and a senior QA specialist from THREAD Research as well as the director of solutions, project manager, director of development, senior software developer, and QA engineer from Analog Republic.

The PRIDE digital research platform was coded in PHP, HTML, CSS3, and JavaScript using responsive web design to ensure effective page rendering on all screen sizes regardless of operating system. CSS employed Block Element Modifier and Inverted Triangle CSS methodologies. Google Tag Manager was used to enable web analytics data collection using Google Analytics (Mountain View, CA). User acceptability testing was performed with SGM community members in iterative cycles using loop11 online user testing software (loop11.com). SGM community feedback received through e-mail, toll-free telephone, and Zendesk-processed support tickets was welcomed, evaluated, and implemented when feasible to improve the experience of The PRIDE Study participants.

The PRIDE Study informational and enrollment website (pridestudy.org)

A comprehensive public-facing informational website is an invaluable community engagement and recruitment method for a digital study. We therefore created pridestudy.org to provide potential participants and other interested parties information about The PRIDE Study, its goals, study participation commitment, and the study team as well as frequently asked questions and study contact information. When data are available, the site will also be used to disseminate results back to SGM communities in traditional (eg, scientific manuscripts, scientific slide/poster presentations) and nontraditional (eg, infographics, short summary videos) ways. Website content was editable via the content management system accessible to PRIDE digital research platform administrators. SGM community involvement during user acceptability testing improved the website's functionality (eg, revised layout, address validation step) and acceptability by suggesting new website images of SGM people from an SGM photographer. The website provides information for collaborating researchers about accessing The PRIDE Study data via an Ancillary Study proposal (see "Data Access").

Visitors interested in joining in The PRIDE Study can immediately begin the eligibility screening and enrollment process, whereas those who are not ready or eligible to enroll in The PRIDE Study can provide their e-mail address (and optionally, their first name, last name, and ZIP code) to be added to The PRIDE Study's general interest list. Visitor-provided information is added directly to EveryAction (customer relationship manager) to receive monthly study newsletters and other digital engagement communications.

PRIDE research platform infrastructure

The PRIDE digital research platform was built as a containerized platform with a microservices-based architecture (Figure 1). Technical documentation is provided in Supplementary Appendix A.

All incoming traffic was secure sockets layer (SSL) terminated at the ingress load balancer that distributed the incoming traffic across compute instances in the cluster to increase the number of concurrent users and application reliability.

We used containers to separate the application from the actual operating system in which it runs. Use of containers allowed facile packaging of application code, associated libraries, and additional dependencies into a portable package for deployment on any computing instance without the need for an operating system; it also allowed rapid debugging and easy replication for scalability across compute instances. The Docker-based (docker.com; San Francisco, CA) containers were managed with Kubernetes (kubernetes.io) open-source container orchestration software. When the PRIDE digital research platform was running on Amazon Web Services, Kubernetes was provided by running Rancher (rancher.com; Cupertino, CA) as there was no AWS-managed Kubernetes service. On Google Cloud Platform (GCP), containers were orchestrated using GCP's managed Kubernetes solution, Google Kubernetes Engine.

Kong (konghq.com; San Francisco, CA) was deployed as an application programming interface gateway in order to manage data across the PRIDE digital research platform microservices (Table 1). Kong is built upon the Nginx reverse proxy HTTP server. Each microservice was a separate container. In addition to microservices, the platform used third-party services and integrations to add features that improve the user experience (Table 2).

Four datastores were used by the PRIDE digital research platform. Kong used PostgreSQL (postgresql.org) to store its configuration. All microservices, including Participants, used Oracle's MySQL (mysql.com; Redwood Shores, CA). Device data was stored in a nonrelational (NoSQL) database provided by MongoDB (mongodb.com; Palo Alto, CA). Redis (redis.io; Mountain View, CA) was used for session store data. Total design and development costs were ~$390 000.

Cloud computing services

The PRIDE digital research platform opened on May 1, 2017 using UCSF's preferred cloud computing vendor, AWS. On February 1, 2019, we moved to Stanford University School of Medicine and migrated to GCP. The required GCP compute, database, and storage resources are in Table 3. Staging and production environments were in the same Kubernetes node pool to avoid paying to run low-activity compute instances exclusively for staging. Kubernetes cluster autoscaling was activated to automatically expand and delete additional node pools as activity demands. Database-associated (ie, MySQL, PostgreSQL) storage autoscaling was activated to ensure no database failure due to limited storage. MongoDB nonrelational database service for device data was provided by mLab (mlab.com; San Francisco, CA). Monthly recurring costs for cloud computing and third-party integrations in Table 2 were ~$875.

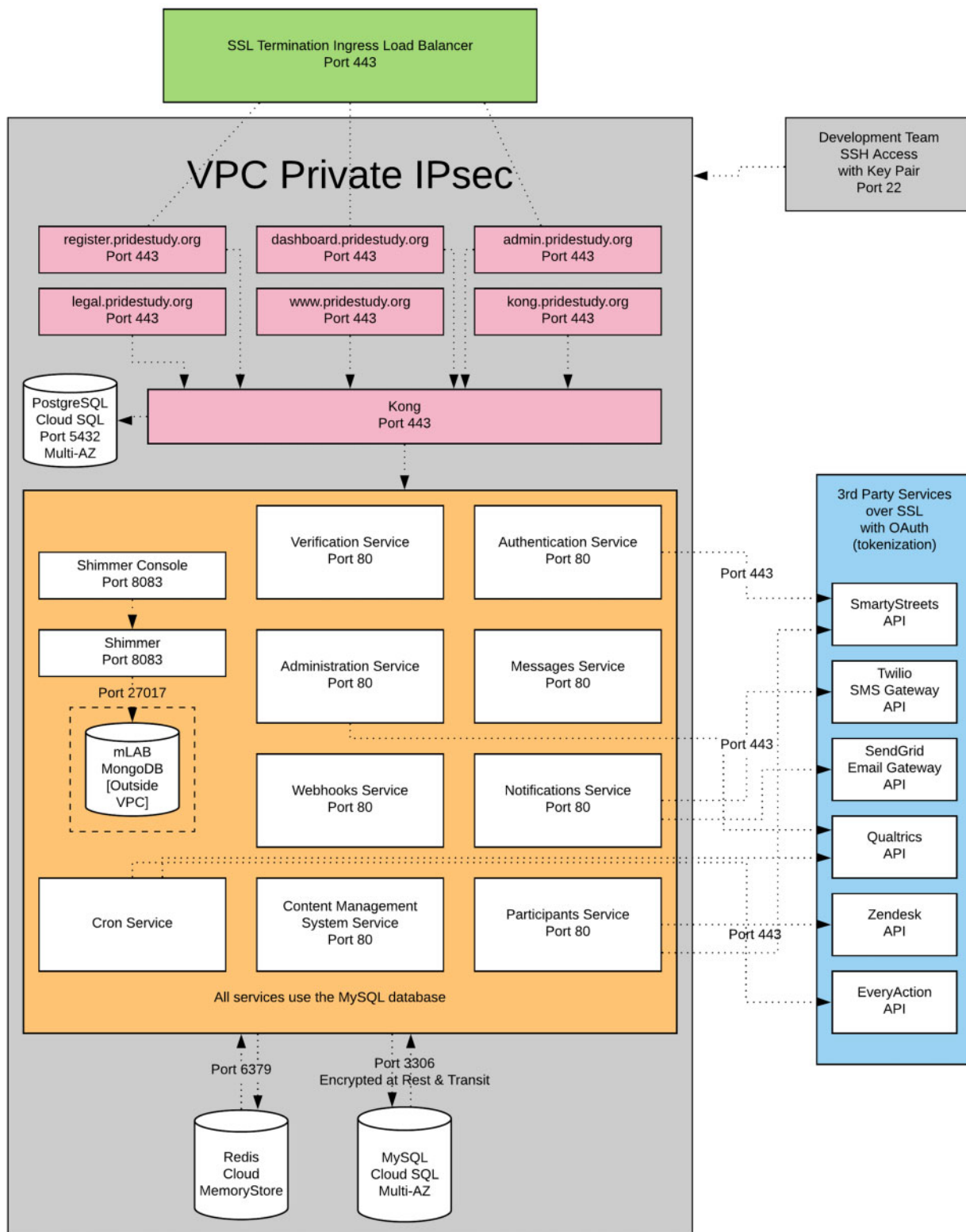## PRIDE Digital Research Architecture

Figure 1. PRIDE digital research platform architecture diagram. Abbreviations: API, application programming interface; AZ, availability zone; SMS, short message service; SQL, structured query language; SSH, secure shell; SSL, secure sockets layer; VPC, virtual private cloud.

Data security, regulatory, and compliance

All microservices in the PRIDE digital research platform were within a private subnet within a GCP virtual private cloud using internet protocol-secured traffic. All connections external to the VPC were SSL-encrypted. All connections with third-party application pro-gramming interfaces employed tokenization-based authentication and authorization using the open authorization standard and were SSL-encrypted. All data were stored in redundant, fault-tolerant databases; the MySQL database (with participant-level data) was encrypted at rest. Short message service (SMS)-based 2-factor authentication using Twilio's Authy service (twilio.com/authy) was required for all administrator access.

In addition to GCP's robust infrastructure security,[35] The PRIDE digital research platform employs technical best practices to keep data secure and limit access including limiting access to participant-level data to a need-to-know basis and only by administrators with proper permissions. The platform is compliant with the requirements of the Health Insurance Portability and Accountability Act, and GCP entered into a business associate agreement with Stanford University School of Medicine. GCP also maintains Federal Risk and Authorization Management Program (FedRAMP; fedramp.gov) authority to operate.[36] The platform also uses the Stanford-provided centralized logging service called Splunk (uit.stanford.edu/service/splunk) to log administrative activity and data access, which would assist a forensic investigation in the event of platform compromise.

The Stanford University Information Security Office and Privacy Office evaluated the PRIDE digital research platform to ensure all applicable security and privacy laws, regulations, and university policy were appropriately followed in the collection, storage, and use of high-risk data (eg, health information, social security numbers). This study was approved by the Institutional Review Board at the University of California, San Francisco (#16-21213) and at Stanford University (#48707).

Table 1. Microservices used in the PRIDE digital research platform

Microservice	Function
Administration	Manages all administrator-level platform functions
Authentication	Participant and administrator identity management
Content Management System	Manages and stores content for pridestudy.org
Cron	Schedules routine or repeating tasks (eg, reports, notifications)
Messages	Participant and administrator in-platform message service
Notifications	Manages participant notifications sent via e-mail and text message
Participants	Manages all participant-level data
Shimmer	Aggregates consumer health device data via OAuth connections
Verification	Manages verification of participant e-mail addresses and/or mobile telephone numbers
Webhooks	Receives all platform webhooks for Google Cloud, SendGrid, Twilio, and Zendesk

Table 2. Integrations and third-party services used in the PRIDE digital research platform

Service	Purpose
EveryAction (everyaction.com)	Customer relationship manager
Open mHealth Shimmer (getshimmer.co)	Device data aggregator
PaperTrail (papertrailapp.com)	Log management
Pingdom (pingdom.com)	Server/endpoint uptime monitoring
Qualtrics (qualtrics.com)	Survey design and administration
Sentry (sentry.io)	Error tracking and debugging
SendGrid (sendgrid.com)	Transactional e-mail gateway
SmartyStreets (smartystreets.com)	Address validation and geocoding
Twilio (twilio.com)	Short message service (SMS) message gateway
Zendesk (zendesk.com)	Customer service/help desk support ticket service

The PRIDE digital research platform: administrator experience

Administrators can manage participant accounts including viewing and editing all participant-provided information. Participant groups (eg, current smokers ages 18–39 interested in quitting) can be created to facilitate sending messages (to the participants' "My Messages") or surveys (created in Qualtrics) to specific participants rather than the entire cohort. Surveys can have a prerequisite if completion of a specific survey is required before another survey is accessible. Platform-wide (global) and survey-specific consent management customizes the process to deploy new consents and reconsent only specific participants. Personalized participant notifications sent by e-mail (via SendGrid) or text message (via Twilio) when certain administrator-defined criteria (eg, new survey assigned, birthday card) are met. Administrators also customize the notification content and frequency. Full control over all website content is available via a custom content management system. Other systems log platform activity and promptly notify administrators during platform downtime and errors. Complete details about

Table 3. Google cloud platform services used in the PRIDE digital research platform

Service	Quantity	Instance Configuration	Purpose
Kubernetes Engine	1 pool with 4 nodes	n1-standard-1 per node (1 vCPU, 3.75 GB memory) with 100 GB boot disk per node	Microservice containers, managed by Kong
Cloud SQL (MySQL)	1	db-n1-standard-2 (2 vCPUs, 7.5 GB memory, 10 GB SSD storage)	Primary PRIDE datastore
Cloud SQL (PostgreSQL)	1	db-n1-standard-2 (2 vCPUs, 7.5 GB memory, 10 GB SSD storage)	Kong datastore
Cloud MemoryStore (Redis)	1	1 GB	Session store
Cloud Storage	10	N/A	Asset and object storage

Abbreviations: GB: gigabyte; SQL: structured query language; SSD: solid-state drive; vCPU: virtual central processing unit (ie, core).
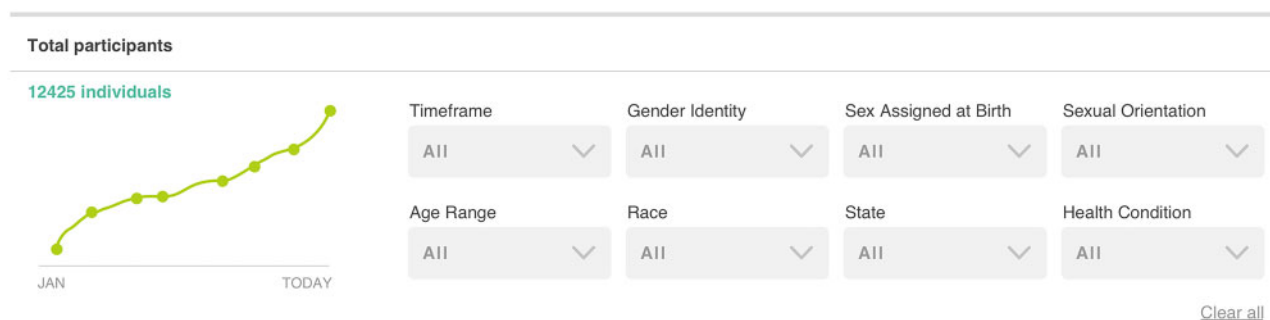
Figure 2. Real-time cohort statistics displayed on participant and administrator dashboards.

administrator-level functionality are available in Supplementary Appendix B.

In addition to the real-time cohort counts (Figure 2), administrators receive a nightly e-mail with overall cohort numbers: unverified (ie, participants who did not verify their e-mail address or mobile telephone number), verified, withdrawn (ie, participants who withdrew themselves), and banned (ie, participants who were removed from the study). Individually password-protected data reports in comma-separated values format are generated nightly and are downloadable by administrators with specific permissions (Table 4).

The PRIDE digital research platform: participant experience

Participant enrollment

Individuals interested in enrolling in The PRIDE Study are presented with eligibility screening questions and, if eligible, the Stanford University-approved informed consent for electronic affirmation. Consented participants proceed to account creation using either an e-mail address or mobile telephone number. (Participants without either of these can register using a mailing address for an offline experience.) Participants choose to verify their e-mail address or mobile telephone number. E-mail verification occurs by visiting a verification URL sent to their e-mail address; mobile telephone number verification occurs by entering a 6-digit code sent by text message. To complete enrollment, participants are asked to activate short message service (SMS)-based 2-factor authentication. Individuals who try to exit the enrollment process are shown a modal that collects their e-mail addresses for later follow-up (Figure 3).

Authenticated PRIDE Study participant dashboard

After logging in, consented PRIDE Study participants are taken to their dashboard, which shows pending activities (eg, "Tell us about yourself," "Complete your medical history," "Enter your medications," etc.), available surveys, and cohort-level statistics (Supplementary Figure S1). Available surveys display an administrator-provided survey title, brief description, and estimated completion time. Participants who do not complete a survey in a single session can resume incomplete surveys from the dashboard.

Participants can access "My Messages," "My Profile," "My Health," "My Devices," and "Account Settings." In "My Messages," participants receive messages from study administrators. In "My Profile," participants provided comprehensive demographic information, social security number, contact information (including 2 e-mail addresses, 2 telephone numbers, and a mailing address), communication preferences, and backup contact information to

Table 4. Available data reports in the PRIDE digital research platform

Report Name	Contents
Demographics	Demographics by participant
Personally Identifiable Information	Contact information, social security number by participant
Health Data	Medical and surgical/procedural histories, medication lists by participant
Hospitalization Data	List of responses to quarterly hospitalization assessment
Under 18 Inquiries	List of age-ineligible participants who requested notification on their 18th birthday
Lapsed Users for 90 Days	List of participants who have not logged in within the past 90 days
Lapsed Users for 180 Days	List of participants who have not logged in within the past 180 days
US Mail Registrants	List of participants who elected to register using their mailing address (instead of e-mail or mobile telephone number)

locate participants who were lost to follow-up and to get additional information on participants who die. Mailing addresses were instantly corrected, validated, and geocoded using SmartyStreets. In "My Health," participants provided their medical history and surgical history (particularly gender-affirming surgeries) by selecting from a pick-list of common conditions and procedures (Supplementary Figure S2). Participants also provided basic information about their sexual histories. Participants selected their current medications using an auto-complete interface based on the US Food and Drug Administration's National Drug Code Directory (updatable/uploadable by an administrator). In "My Devices," participants can authorize OAuth connections to their Fitbit and Withings accounts; data are pulled into the PRIDE database daily. Participants can access their signed consents and can change their password, 2-factor authentication settings, and communication preferences in "Account Settings" (Supplementary Figure S3). A "Help Desk" enabled participants to submit requests for assistance (approximately 40–50 tickets per month), provide feedback, and suggest new features. Requests automatically created a support ticket in Zendesk that included participant data (eg, name, participant ID number) for easy participant record lookup and rapid response to ensure a high-quality customer service experience.

Participants experience several features developed for increasing data completeness and timeliness. Upon completion of the various sections within "My Profile" and "My Health," a modal with
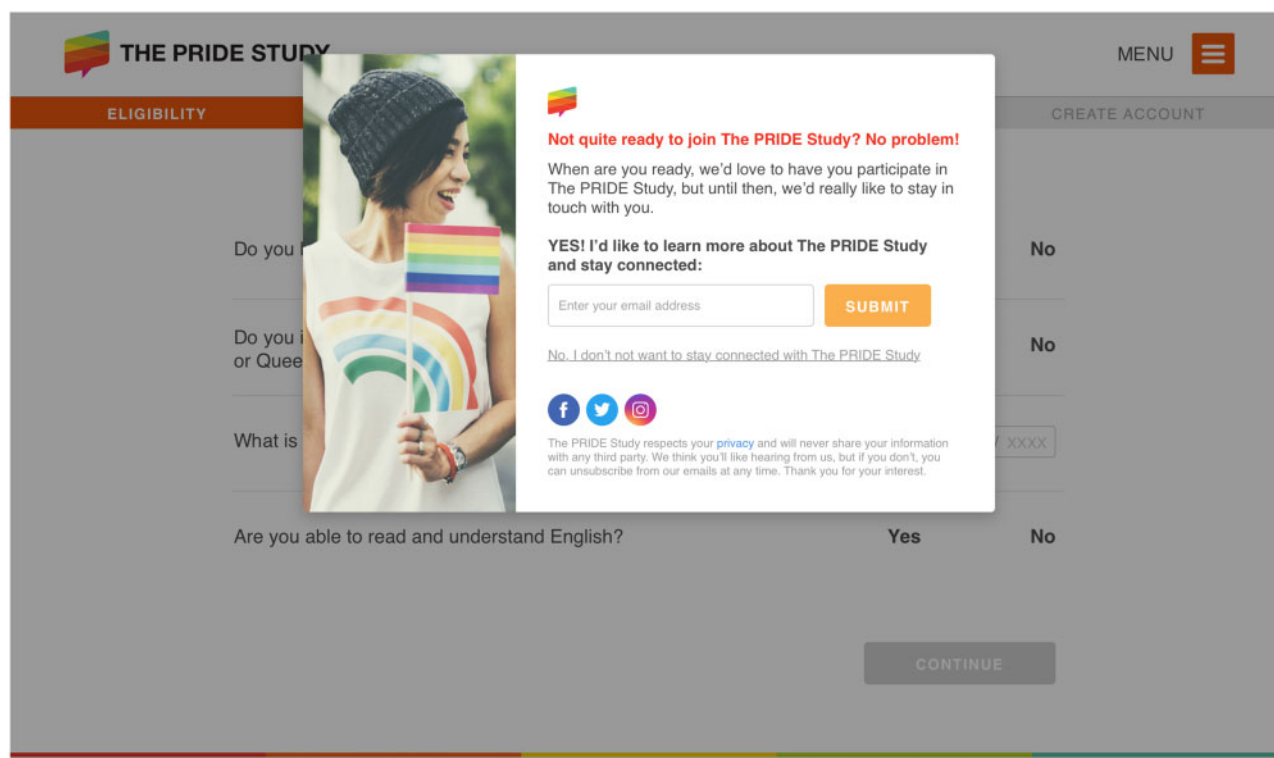
Figure 3. Modal capturing e-mail address of individuals who did not join The PRIDE Study.

congratulatory language and fun imagery inspires continued data completion (Figure 4). Upon registration, personalized e-mail- or text message-based notifications (depending on the participant's preferences) communicate to The PRIDE Study participants when new surveys were available, when surveys were incomplete, when they had not logged in within 3 months, and on their birthday (see "Notifications" in Supplementary Appendix B). Every 6 months, a modal appears upon login to remind participants to update "My Profile" and "My Health" (Supplementary Figure S4). Every 3 months, PRIDE Study participants receive an e-mail or text message (depending on the participant's preferences) asking if they had been hospitalized in the prior 3 months.

The customized dashboard displays the real-time cohort proportion with the same specific (ie, gender identity, sexual orientation, race) demographic characteristics as the participant. Additionally, real-time cohort counts are available based on any combination of 7 participant attributes: gender identity, sex assigned at birth, sexual orientation, age range, race, state, and health conditions. The timeframe (eg, last 7 days, last month, last 3 months, last year) can be selected to see the change over time (Figure 2).

Participants can access The PRIDEnet Blog (blog.pridestudy.org), where we post community-friendly research summaries, share The PRIDE Study developments, and disseminate study results. Participants can easily share their participation in The PRIDE Study with prepopulated messages on Facebook and Twitter via dedicated buttons.

MATERIALS AND METHODS

Community engagement
The PRIDE Study is a community-engaged research study that strives to engage participants at each step in the research process: research question generation, study design, recruitment, participation, data analysis and interpretation, and results dissemination. In order to operationalize this philosophy, a community engagement structure was needed. We created PRIDEnet, a national network of individuals and organizations that actively engaged SGM communities in all stages of health research. PRIDEnet includes a 41-member (as of April 2019) national Community Partner Consortium composed of trusted SGM-serving health clinics, community centers, and professional/advocacy organizations; a 12-member Participant Advisory Committee that provides study guidance and oversight; and 8 PRIDEnet ambassadors that work through their established networks of influence to engage SGM communities. All are committed partners with us in improving the health and well-being of SGM communities. Built on decades of work by activists, health advocates, service providers, and researchers, PRIDEnet reflects the voices and views of the people whose health is being studied.

The PRIDE Study participant recruitment and enrollment
Participant eligibility screening and enrollment occurred exclusively online; participant recruitment efforts focused therefore on driving traffic to pridestudy.org. The PRIDE Study team recruited primarily by conducting outreach at SGM conferences and events, word-of-mouth within SGM health researcher networks, distributing The PRIDE Study-branded promotional items (eg, pens, water bottles, first aid kits), and social media advertising. PRIDEnet recruited primarily by digital communications (eg, blog posts, newsletters)[37,38] and by distributing The PRIDE Study-branded promotional items to their constituents. PRIDEnet Community Partner Consortium members had the option to create an organization-branded PRIDE Study

Figure 4. Example modal with congratulatory imagery to foster data completion. Simulated datum (ie, a nonreal participant) is shown in the figure.

landing page with a friendly URL (ie, pridestudy.org/*organization_-name*) and customizable text, images, and video to educate their constituents about The PRIDE Study and engage them to enroll. All website traffic— including participant enrollment through the 4-step funnel (eligibility screening, informed consent, data privacy information, and account creation)— was tracked using Google Analytics. Participant-provided demographics were collected after enrollment via "My Profile."

The PRIDE Study annual questionnaires

Annual questionnaires (launched every June) are the primary research instrument in The PRIDE Study. Each annual questionnaire (AQ) contains 5 blocks (Introduction, Mental Health, Physical Health, Social Health, and Miscellaneous); the order in which the middle 3 blocks are presented to the participant is random. The AQ is comprehensive and assesses diagnoses (including behavioral health), surgeries and procedures, cancer screening, vaccinations, substance use, sexual behavior and satisfaction, traumatic experiences (including sexual assault), experiences of stigma and discrimination, identity formation, acceptance from self and others about SGM status, suicidal history,

health insurance, social supports, resilience, health behaviors (exercise, smoking, sleep, sexually-transmitted infection prevention, etc.), family formation and structure, and many others. (Complete surveys are available at pridestudy.org/collaborate.) Branching logic hides questions that are irrelevant for a participant in order to create a more engaging experience. The 2017 AQ contained a maximum of 670 questions and took approximately 30–40 minutes to complete. In response to community inquiry, the 2018 AQ was made more comprehensive to cover the previously mentioned topics. Therefore, in order to minimize survey burden in 2018, we launched a 2018 AQ (maximum of 732 questions, ∼35–45 minutes) and a 2018 AQ Supplement (maximum of 214 questions, ∼10–15 minutes). In 2019, participants will complete an entry questionnaire once to report on past health experiences (maximum of ∼400 questions, ∼20–25 minutes) and the AQ to update this information annually (maximum of ∼550 questions, ∼25–35 minutes).

The PRIDE Study participant retention

Because participants may forget about their participation in The PRIDE Study and become lost to follow-up, we developed several

Table 5. The PRIDE study participant sociodemographics (as of April 30, 2019)

Characteristic	N (%)
Age, years (N = 13 731)	Median 30.7
	IQR 25.2–40.9
Age (N = 13 731)	
18–19 years	350 (2.6)
20–24 years	2946 (21.5)
25–29 years	3196 (23.3)
30–34 years	2147 (15.6)
35–39 years	1457 (10.6)
40–49 years	1610 (11.7)
50–59 years	1143 (8.3)
60–69 years	690 (5.0)
>= 70 years	192 (1.4)
Gender Identity (N = 11 639)[a]	
Genderqueer	1931 (16.6)
Man	3831 (32.9)
Transgender man	1133 (9.7)
Transgender woman	560 (4.8)
Woman	5165 (44.4)
Another gender identity	1238 (10.6)
Sex Assigned at Birth (N = 10 941)	
Female	7094 (64.8)
Male	3847 (35.2)
Sexual Orientation (N = 11 630)[b]	
Asexual	1018 (8.8)
Bisexual	3194 (27.5)
Gay	4002 (34.4)
Lesbian	2804 (24.1)
Pansexual	2018 (17.4)
Queer	4154 (35.7)
Questioning	392 (3.4)
Same-Gender Loving	693 (6.0)
Straight	245 (2.1)
Another sexual orientation	398 (3.4)
Gender Minority[c] (N = 11 639)	3813 (32.8)
Sexual Minority[d] (N = 11 630)	11 476 (98.7)
Sexual and Gender Minority (N = 11 623)	3677 (31.6)
Race (N = 11 546)[e]	
African-American	471 (4.1)
American Indian or Alaska Native	400 (3.5)
Asian	501 (4.3)
Native Hawaiian or Pacific Islander	55 (0.5)
Middle Eastern/North African	21 (0.2)
White	10 589 (91.7)
Another race	448 (3.9)
Hispanic/Latino/Spanish Ethnicity (N = 11 593)	978 (8.4)
Born in the US (N = 11 616)	10 995 (94.6)
Education (N = 11 587)	
No schooling	7 (0.1)
Less than high school	103 (0.9)
High school graduate or equivalent	712 (6.1)
Trade/Technical/Vocational training	190 (1.6)
Some college	2487 (21.5)
2-year degree	614 (5.3)
4-year college degree	3840 (33.1)
Graduate degree (Masters/Doctoral/Professional)	3634 (31.4)

(continued)

Table 5. continued

Characteristic	N (%)
Region[f] (N = 7165)	
Northeast	1301 (18.2)
Midwest	1489 (20.8)
South	2058 (28.7)
West	2317 (32.3)

[a]Items sum to more than 100% because multiple selections were permitted; 1893 (16.3%) participants selected 2 or more gender identities.

[b]Items sum to more than 100% because multiple selections were permitted; 4607 (39.6%) participants selected 2 or more sexual orientations.

[c]Gender minority individuals were those whose current gender identity differed from that most consistent with their sex assigned at birth.

[d]Sexual minority individuals were those who did not exclusively choose straight/heterosexual as their sexual orientation.

[e]Items sum to more than 100% because multiple selections were permitted; 840 (7.3%) participants selected 2 or more races.

[f]Region determined by participant-entered ZIP code.

Abbreviation: IQR, interquartile range.

than 6 months have elapsed since their last login, participants are presented with a modal recommending they update their information to ensure accuracy (Supplementary Figure S4). PRIDE Study participants receive an e-mail or text message every 3 months asking if they had been hospitalized in the prior 3 months, which brings them back to their dashboard. E-mail and text message notifications about the newly released questionnaire (including AQ and Ancillary Studies) also brings participants back into their accounts. Finally, *ad hoc* incentive campaigns (eg, entry into a prize drawing if surveys are completed by a specified date) increase activity on the PRIDE digital research platform. Calculation of longitudinal (year-after-year) AQ completion will begin in June 2019.

RESULTS

The PRIDE Study website (pridestudy.org) engagement

Between September 20, 2017 and February 4, 2019, pridestudy.org hosted 104 679 sessions for 69 122 users with 60.2% using a computer, 3.7% using a tablet, and 36.1% using a mobile device. Of these sessions, most (76.6%) were direct traffic to pridestudy.org. Approximately 9.3% originated from social media with 82.2% of them from Facebook and 14.3% from Twitter. Like social media, approximately 9.2% resulted from organic search. Among the 9133 eligible for The PRIDE Study in this period, 8317 (91.1%) people signed consent. Among those, 5742 (69.0%) created an account.

The PRIDE Study enrollment

Between May 1, 2017 and April 30, 2019, 13 932 individuals consented to join The PRIDE Study. Of them, 192 participants have withdrawn their consent for reasons including loss of interest, loss of commitment to the research, lack of trust in the setting of the current political environment, and death. A total of 9 participant accounts were removed by the study staff because either they were duplicates, or they were participants who, after registering with a false date-of-birth, modified their date-of-birth to be less than 18 years old. Demographic information for the 13 731 nonwithdrawn/removed participants is in Table 5.

methods to increase retention. The "Inactive Participant" automated notification sends an e-mail and/or text message to participants at administrator-set intervals with a "We've missed you" message and an invitation to see what is new with The PRIDE Study. If more

The PRIDE Study participant engagement and retention

During the period in which Google Analytics collected data (September 20, 2017– April 30, 2019), there were 74 802 sessions with 52.7% using a computer, 4.4% using a tablet, and 43.0% using a mobile device. Among the 35 403 sessions using tablets and mobile devices, 67.8% were Apple devices. The average session length was 5 minutes, 36 seconds with 18.9% of the sessions lasting longer than 10 minutes in duration. The bounce rate (proportion of visitors who only view 1 page before leaving the platform) was 28.96%.

In examining survey response data, 7208 responses (65.8%) were received from 10 952 eligible participants during a 12-month window for the 2017 AQ. For the 2018 AQ, 6574 responses (47.9%) have been received from 13 731 eligible participants during an 11-month (June 2018–April 2019) window. During the same period, 5134 responses (37.4%) to the 2018 AQ Supplement were received.

DISCUSSION

We created a containerized, comprehensive, feature-rich, digital platform to support community-engaged longitudinal and cross-sectional digital research studies. We created PRIDEnet, a network of dedicated SGM organizations and advocates, to build relationships and keep participants connected, engaged, and informed. We subsequently used the platform to recruit a national sample of more than 13 700 SGM adults in 24 months for longitudinal participation in The PRIDE Study.

Because SGM people experience discrimination in health care, creating long-lasting, meaningful, bidirectional relationships with participants is critical. The PRIDE Study's initial iPhone app-based pilot phase[33] generated valuable community-provided insights that influenced the development of this digital research platform. These insights included having a platform that is accessible from all devices regardless of screen size, enabling participants to learn about the other participants in a way that protects individual privacy, being transparent about how community members are involved in The PRIDE Study governance, and conducting research on topics important to SGM communities (manuscript in preparation).

In conjunction with PRIDEnet's robust community engagement efforts, The PRIDE Study successfully recruited participants who were diverse in terms of age, sexual orientation, gender identity, and geography. The recruitment of 3813 (32.8%) gender minority people is particularly notable given that gender minorities represent only an estimated 0.6% of the US population.[39] The addition of less frequently searched gender identities (eg, transgender woman, genderqueer) and sexual orientations (eg, asexual, pansexual, queer) and the ability to select multiple identities highlight the heterogeneity with SGM identities. Our sample, however, was less diverse than desired in terms of race and ethnicity; this is consistent with other SGM studies.[40,41] Only 8.4% of The PRIDE Study reported a Hispanic/Latino/Spanish ethnicity compared with 16.3% of the US population.[42] Adjustments in communication assets (including images of those underrepresented in biomedical research), targeted campaigns, and high-touch relationship-building with racial/ethnic community partners may be needed to gain the trust of these SGM MC subcommunities that have historically been underserved and stigmatized. The PRIDE Study cohort was also highly educated with nearly 65% having a 4-year college degree or higher compared to 32.2% of US adult population.[43] Future efforts will ensure The PRIDE Study is accessible to diverse reading levels, and additional media (such as short, informational videos) will be used to educate about participation in The PRIDE Study.

Annual Questionnaire response rates of ~48%–66% may be slightly lower than other longitudinal cohort studies of SGM people for several reasons.[44,45] The larger number of participants in The PRIDE Study makes more-frequent, personalized check-ins with participants challenging. Smaller cohorts benefited from participant–researcher relationships with in-person recruitment, follow-up, and interviews.[46,47] Prior exclusively online, longitudinal, SGM cohort studies with shorter follow-up (ie, 6 months) resulted in fewer participants lost to follow-up.[48] Whereas the sensitive nature of some AQ questions may deter participation, completing surveys online in the participants' own environments—as opposed to an in-person clinical research center interview- and/or paper survey-based data collection—provides safety for participants and may limit social desirability bias.[49]

As The PRIDE Study evolves and new technologies emerge, there are multiple areas ripe for additional platform development. Some areas include participant-level biospecimen collection and storage tracking, a digital signature service to collect legally binding signatures on health record release forms, and linkage to electronic health records and other data sources including direct-to-consumer genetic testing results (eg, 23andme, Veritas). Electronic identity verification of participants using a credit history-based question set or biometric facial recognition may help ensure that the true individual is authorizing access to sensitive information in digital studies without face-to-face encounters. Finally, identifying tactics to maintain participant engagement (including survey completion rates) in a digital-only experience is critical to longitudinal studies. Gamification and innovative methods to return study results to participants may be effective at optimizing retention.

CONCLUSION

We created a digital research platform to support the development of a nationwide, community-engaged, longitudinal cohort study (The PRIDE Study) of SGM people. The PRIDE Study, as a data resource for SGM health researchers, will improve our understanding of SGM physical, mental, and social health. With the continual evolution of digital health research technologies, digital research platforms, such as the 1 described here, may be successful approaches to engaging, recruiting, and retaining individuals from underrepresented and vulnerable populations into clinical research studies that document and improve the health of their communities.

FUNDING

AUTHOR CONTRIBUTORS

All authors have fulfilled the criteria for authorship established by the International Committee of Medical Journal Editors and approved submission of the manuscript. MRL and JOM made substantial contributions to the conception and design of the study and secured study-specific funding. MRL drafted the manuscript. ML and CH made important intellectual contributions to the study design and platform design as experts in SGM community engagement. AF and MRC made important intellectual contributions to the study design as experts in SGM mental and social health. CS, TH, DC, and CN made important intellectual contributions to platform design/development and supervised their teams for platform development. All coauthors participated in revising the manuscript critically, made important intellectual contributions, and approved the final version to be published.

DATA ACCESS

Members of the sexual and gender minority (SGM) communities have experienced significant stigma and discrimination from society including the medical and research communities. We are ethically bound to upholding the principle of nonmaleficence; we promise our participants to not let any data (including deidentified) fall into the hands of people who may use it to publish stigmatizing results about the SGM communities. As such, we have a developed an Ancillary Study process in which investigators interested in using our data submit a brief application which is reviewed by both a Research Advisory Committee (composed of scientists) and Participant Advisory Committee (composed of participants) to affirm appropriate data use. Details about the Ancillary Study process are available at pridestudy.org/collaborate or by contacting us at support@pridestudy.org or 855-421-9991 (toll-free).

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Sexual & Gender Minority Research Office, National Institutes of Health. https://dpcpsi.nih.gov/sgmro Accessed March 23, 2019.
2. Pérez-Stable E. *Sexual and Gender Minorities Formally Designated as a Health Disparity Population for Research Purposes*. NIMHD Director's Message. https://www.nimhd.nih.gov/about/directors-corner/messages/message_10-06-16.html Accessed March 23, 2019.
3. Boyd CJ, Veliz PT, Stephenson R, Hughes TL, McCabe SE. Severity of alcohol, tobacco, and drug use disorders among sexual minority individuals and their "not sure" counterparts. *LGBT Health* 2019; 6(1): 15–22.
4. Kidd SA, Howison M, Pilling M, Ross LE, McKenzie K. Severe mental illness in LGBT populations: a scoping review. *Psychiatr Serv* 2016; 67(7): 779–83.
5. Crissman HP, Stroumsa D, Kobernik EK, Berger MB. Gender and frequent mental distress: comparing transgender and non-transgender individuals' self-rated mental health. *J Womens Health* 2019; 28(2): 143–51.
6. Yarns BC, Abrams JM, Meeks TW, Sewell DD. The mental health of older LGBT adults. *Curr Psychiatry Rep* 2016; 18(6): 1-11. doi: 10.1007/s11920-016-0697-y.
7. Steele LS, Daley A, Curling D, et al. LGBT identity, untreated depression, and unmet need for mental health services by sexual minority women and trans-identified people. *J Womens Health* 2017; 26(2): 116–27.
8. Lutz AR. Screening for asymptomatic extragenital gonorrhea and chlamydia in men who have sex with men: significance, recommendations, and options for overcoming barriers to testing. *LGBT Health* 2015; 2(1): 27–34.
9. Centers for Disease Control and Prevention. *HIV Surveillance Report 2016*. 2017. https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2016-vol-28.pdf Accessed March 25, 2019.
10. Qureshi RI, Zha P, Kim S, et al. Health care needs and care utilization among lesbian, gay, bisexual, and transgender populations in New Jersey. *J Homosex* 2018; 65(2): 167–80.
11. Shiu C, Kim H-J, Fredriksen-Goldsen K. Health care engagement among LGBT older adults: the role of depression diagnosis and symptomatology. *Gerontologist* 2017; 57(suppl 1): S105–14.
12. Floyd SR, Pierce DM, Geraci SA. Preventive and primary care for lesbian, gay and bisexual patients. *Am J Med Sci* 2016; 352(6): 637–43.
13. Whitehead J, Shaver J, Stephenson R. Outness, stigma, and primary health care utilization among rural LGBT populations. *Plos One* 2016; 11(1): e0146139.
14. Lambda Legal. *When Health Care Isn't Caring: Lambda Legal's Survey of Discrimination against LGBT People and People with HIV*. New York: Lambda Legal; 2010.
15. James SE, Herman JL, Rankin S, Keisling M, Mottet L, Anafi M. *The Report of the 2015 US Transgender Survey*. Washington, DC: National Center for Transgender Equality; 2016. https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf. Accessed March 23, 2019.
16. Movement Advancement Project. *Healthcare Laws and Policies*. http://www.lgbtmap.org/equality-maps/healthcare_laws_and_policies Accessed March 23, 2019.
17. Movement Advancement Project. *Non-Discrimination Laws*. http://www.lgbtmap.org/equality-maps/non_discrimination_laws Accessed March 23, 2019.
18. Newport F. *In US, Estimate of LGBT Population Rises to 4.5%*. Gallup. https://news.gallup.com/poll/234863/estimate-lgbt-population-rises.aspx Accessed March 23, 2019.
19. Thompson JH. *Director's Blog: Planned Subjects for the 2020 Census and the ACS*. The United States Census Bureau. https://www.census.gov/newsroom/blogs/director/2017/03/planned_subjects_2020.html Accessed March 23, 2019.
20. Committee on Lesbian, Gay, Bisexual, and Transgender Health Issues and Research Gaps and Opportunities, Board on the Health of Select Populations, Institute of Medicine of the National Academies. *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding*. Institute of Medicine. http://www.nationalacademies.org/hmd/Reports/2011/The-Health-of-Lesbian-Gay-Bisexual-and-Transgender-People.aspx Accessed November 12, 2017.
21. Grasso C, McDowell MJ, Goldhammer H, Keuroghlian AS. Planning and implementing sexual orientation and gender identity data collection in electronic health records. *J Am Med Inform Assoc* 2019; 26(1): 66–70.
22. Toledano MB, Smith RB, Brook JP, Douglass M, Elliott P. How to establish and follow up a large prospective cohort study in the 21st century—lessons from UK COSMOS. *Plos One* 2015; 10(7): e0131521.

23. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham study. *Am J Public Health Nations Health* 1951; 41(3): 279–86.

24. Belanger CF, Hennekens CH, Rosner B, Speizer FE. The Nurses' health study. *Am J Nurs* 1978; 78(6): 1039–40.

25. Bycroft C, Freeman C, Petkova D, *et al*. The UK biobank resource with deep phenotyping and genomic data. *Nature* 2018; 562(7726): 203.

26. National Institutes of Health. *All of Us Research Program*. https://allofus.nih.gov/ Accessed March 23, 2019.

27. Mustanski B, Greene GJ, Ryan D, Whitton SW. Feasibility, acceptability, and initial efficacy of an online sexual health promotion program for LGBT youth: the Queer Sex Ed intervention. *J Sex Res* 2015; 52(2): 220–30.

28. Roth Y. Zero feet away: the digital geography of gay social media. *J Homosex* 2016; 63(3): 437–42.

29. Smith LW, Guy R, Degenhardt L, *et al*. Meeting sexual partners through internet sites and smartphone apps in Australia: national representative study. *J Med Internet Res* 2018; 20(12): e10683.

30. Muessig KE, Pike EC, Fowler B, *et al*. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. *AIDS Patient Care STDS* 2013; 27(4): 211–22.

31. Fleming JB, Hill YN, Burns MN. Usability of a culturally informed mHealth intervention for symptoms of anxiety and depression: feedback from young sexual minority men. *JMIR Hum Factors* 2017; 4(3): e22.

32. Smiley SL, Elmasry H, Hooper MW, Niaura RS, Hamilton AB, Milburn NG. Feasibility of ecological momentary assessment of daily sexting and substance use among young adult African American gay and bisexual men: a pilot study. *JMIR Res Protoc* 2017; 6(2): e9.

33. Lunn MR, Capriotti MR, Flentje A, *et al*. Using mobile technology to engage sexual and gender minorities in clinical research. *Plos One* 2019; 14(5). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0216282.

34. Schwaber K, Sutherland J. The Scrum Guide. Scrum.Org. https://www.scrumguides.org/scrum-guide.html Accessed March 24, 2019.

35. Google Cloud. *Google Infrastructure Security Design Overview*. Google; 2018. https://cloud.google.com/security/infrastructure/design/ Accessed March 23, 2019.

36. FedRAMP, United States General Services Administration. *FedRAMP Marketplace Dashboard*. https://marketplace.fedramp.gov/#/products? sort= productName&productNameSearch=google Accessed March 23, 2019.

37. DelloStritto L. Bisexuals DO Exist! And They're Joining The PRIDE Study. 2018. https://biresource.org/bisexuals-do-exist-and-theyre-joining-the-pride-study/ Accessed April 28, 2019.

38. Murchison G. *UCSF Team Launches Major Study of LGBTQ Health*. 2017. https://www.hrc.org/blog/ucsf-team-launches-major-study-of-lgbtq-health Accessed April 28, 2019.

39. Flores AR, Herman JL, Gates GJ, Brown TNT. *How Many Adults Identify as Transgender in the United States*? Los Angeles, CA: The Williams Institute; 2016. https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf Accessed April 28, 2019.

40. Grov C, Cain D, Whitfield TH, *et al*. Recruiting a US national sample of HIV-negative gay and bisexual men to complete at-home self-administered HIV/STI testing and surveys: challenges and opportunities. *Sex Res Soc Policy Berkeley* 2016; 13(1): 1–21.

41. Sullivan PS, Khosropour CM, Luisi N, *et al*. Bias in online recruitment and retention of racial and ethnic minority men who have sex with men. *J Med Internet Res* 2011; 13(2): e38.

42. Humes KR, Jones NA, Ramirez RR. *Overview of Race and Hispanic Origin: 2010*. United States Census Bureau; 2011. https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf Accessed April 28, 2019.

43. US Census Bureau, Current Population Survey, 2018 Annual Social and Economic Supplement. *Educational Attainment in the United States: 2018*. U.S. Census Bureau; 2018. https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html Accessed April 28, 2019.

44. Hughes TL, Wilsnack SC, Szalacha LA, *et al*. Age and racial/ethnic differences in drinking and drinking-related problems in a community sample of lesbians. *J Stud Alcohol* 2006; 67(4): 579–90.

45. Millar BM, Starks TJ, Rendina HJ, Parsons JT. Three reasons to consider the role of tiredness in sexual risk-taking among gay and bisexual men. *Arch Sex Behav* 2019; 48(1): 383–95.

46. Martos AJ, Wilson PA, Gordon AR, Lightfoot M, Meyer IH. "Like finding a unicorn": healthcare preferences among lesbian, gay, and bisexual people in the United States. *Soc Sci Med* 2018; 208: 126–33.

47. Fredriksen-Goldsen KI, Kim H-J. The science of conducting research with LGBT older adults- an introduction to aging with pride: national health, aging, and sexuality/gender study (NHAS). *Gerontologist* 2017; 57(Suppl 1): S1–14.

48. Hammoud MA, Jin F, Degenhardt L, *et al*. Following lives undergoing change (Flux) study: implementation and baseline prevalence of drug use in an online cohort study of gay and bisexual men in Australia. *Int J Drug Policy* 2017; 41: 41–50.

49. Davis RE, Couper MP, Janz NK, Caldwell CH, Resnicow K. Interviewer effects in public health surveys. *Health Educ Res* 2010; 25(1): 14–26.