Published in final edited form as:

AIDS Educ Prev. 2018 October ; 30(5): 434–447. doi:10.1521/aeap.2018.30.5.434.

Paradata Analysis of an eHealth HIV Testing Intervention for Young Men who have Sex with Men

Stephen Bonett, BA, BSN[1], **Daniel Connochie, MPH**[1], **Jesse M Golinkoff, MPH**[1], **Keith J Horvath, PhD**[2], and **José A. Bauermeister, MPH PhD**[1]

[1]University of Pennsylvania, School of Nursing, Philadelphia, PA, United States

[2]University of Minnesota, School of Public Health, Minneapolis, MN, United States

Abstract

Background: Little is known about users' intervention engagement and use patterns within eHealth interventions. We describe these patterns among young men who have sex with men (YMSM) who participated in a brief eHealth intervention designed to increase HIV testing.

Methods: We merged pilot trial participants' survey data (N=86) with their paradata (e.g., system data recorded during interaction with the intervention). We created engagement (time spent on components) and use (interaction with features) metrics, and explored whether they differed by participant characteristics.

Results: Engagement (mean=322.67 seconds, SD= 385.40) and use (mean=10.28 clicks on features) varied between groups. Racial/ethnic minorities clicked on fewer features (mean=8.30) than their Non-Hispanic White men (mean=12.00). Use was associated with older age (r=.19), greater educational attainment (r=.25), and a greater number of methods to connect online (r=.38).

Conclusions: Paradata can help researchers understand how users interact with eHealth interventions, and inform which components to retain or redesign. Efforts to systematically collect, analyze, and report paradata in eHealth HIV prevention and care interventions are warranted.

Keywords

Engagement; HIV/AIDS; Prevention; technology

A substantial number of annual HIV infections are attributed to lack of HIV/STI status awareness (Marks, Crepaz, & Janssen, 2006). In light of the high incidence of HIV among young men who have sex with men (YMSM) in the United States, the Centers for Disease Control and Prevention (CDC) recommend that YMSM test for HIV/STI regularly. Successful engagement in HIV testing services, however, requires that YMSM overcome a series of barriers. Building on the efficacy of the CDC's Project Connect Health Systems Intervention to link heterosexual adolescents to competent comprehensive sexual health care services (Dittus et al., 2014), we developed *Get Connected!* as a brief online intervention that employs individual and systems-level tailoring technology to reduce barriers to linkage

Corresponding Author: José A. Bauermeister, MPH, PhD, bjose@upenn.edu, University of Pennsylvania, 418 Curie Blvd, Suite 222L, Philadelphia, PA 19109, Phone: (215) 898-9993.

to competent preventative care (e.g., HIV/STI testing) for YMSM. The deployment of *Get Connected!* through a mobile-friendly web application sought to optimize the online intervention's accessibility, availability, long-term affordability, and acceptability among youth (Allison et al., 2012; Chavez, Shearer, & Rosenthal, 2014; Muessig, Nekkanti, Bauermeister, Bull, & Hightow-Weidman, 2015). Data from the pilot randomized control trial (N=130 YMSM; ages 15–24) indicated high acceptability and feasibility for *Get Connected!*, and preliminary efficacy of the intervention to increase HIV/STI testing behavior at the 30-day follow-up (Bauermeister et al., 2015).

Increasingly, researchers have emphasized the need to describe how participants use technology-related devices and features in mHealth studies (see reviews by Alkhaldi et al., 2016 & Simblett et al., 2018). Paradata (i.e., process data automatically captured in a database that records actions performed on a website by its users) is one approach that can be employed to estimate how much time users spent on different sections of an online intervention and what content they viewed. These data can help researchers characterize participants' engagement with and use of an intervention, understand how users' interactions within the intervention are associated with changes in proposed theoretical mechanisms, and refine an intervention's design and features if a second iteration of is planned (Couper et al., 2010). Within HIV research, however, little is known about participants' interactions with the features embedded within eHealth interventions. As a result, researchers have called for greater attention to user engagement (i.e., amount of time a user spent on the intervention or on a specific feature) and use (i.e., the frequency with which users interacted with different features or types of content) within HIV interventions (Baltierra et al., 2016; Bauermeister et al., 2017).

The goal of this paper is to describe participants' engagement and use in *Get Connected!* based on paradata metrics. Our study has two primary objectives. First, we describe our process for developing paradata measures based on the intervention's theoretical framework and design. Second, we explore whether participants' sociodemographic characteristics are associated with their overall and feature-specific use and engagement with the intervention. We discuss how paradata can aid in understanding patterns of participants interaction with online interventions and inform development of future interventions.

Methods

Study Information

Data from this study come from the *Get Connected!* pilot trial (Bauermeister et al., 2015; Horvath & Bauermeister, 2017). *Get Connected!* was an eHealth intervention developed to increase HIV/STI testing among gay, bisexual, and other young men who have sex with men (MSM) between 15 and 24 years of age (referred to hereafter as young MSM or YMSM) (N=130). Briefly, *Get Connected!* employed tailoring algorithms based on key characteristics of participants (e.g., age, race/ethnicity, relationship status, sexual identity) to tailor imagery (e.g., Black YMSM were shown pictures of similar other Black young men) and intervention content (e.g., participants who reported never having tested for HIV received messages promoting testing, while those who had been tested received messages reinforcing their testing behaviors). At 30-day follow-up, YMSM in the *Get Connected!*

tailored intervention condition perceived the intervention to be more accurate than men in the non-tailored control condition, and more reported getting tested than those in the non-tailored control condition (see Bauermeister et al., 2015 for additional information on the intervention's development process).

Of the 382 individuals screened for inclusion, 180 were eligible and invited to the study. Fifty eligible participants did not complete the intervention (36 were given an access code but never accessed the intervention; 14 stopped participation at the consent form). The remaining 130 participants completed the intervention for a 72.20% participation rate; 104 participants were retained at the 30-day follow-up (80% retention rate). Within the intervention arm, 77 of 86 participants completed all measures included in the current analyses (see Table 1).

Measures

Participant Characteristics and Technology Use

Participants self-reported their age (in years). Participants self-identified their race(s) and Latino ethnicity from the following categories: Black/African American, Hispanic, White, Middle Eastern, Asian or Pacific Islander, and Other. Due to the small size of the sample, we collapsed participants based on whether they identified as Non-Hispanic White or as a Racial/Ethnic minority. Participants were asked which sexual identity most closely matched their own from the following categories: Gay/Homosexual, Bisexual, Straight/Heterosexual, Same Gender Loving, Queer, or Other. The majority of participants identified as gay, so we dichotomized sexual orientation to indicate whether a participant was gay-identified or not. Participants' relationship status was measured by asking if they were currently in one or more relationships.

Participants indicated their highest level of educational attainment (1 – 8^{th} grade or less; 2 – Some high school; 3 – Graduated high school/GED; 4 – Technical school; 5 – Associate degree; 6 – Some college; 7 – College; 8 – Some graduate school; 9 – Graduate school). Participants reported whether they were currently a full-time student, part-time student, or not a student, and these categories were collapsed into student and non-student for analysis.

Participants were asked to indicate if they had spent at least one night in the past 30 days in one of the following locations: a shelter; a public place not intended for sleeping; on the street or outside; temporarily doubled up with a friend or family member; in a temporary housing program; in a welfare or voucher motel/hotel; in jail, prison or a halfway house; in a drug treatment facility, a detox unit, or a drug program house; or in a hospital, nursing home or hospice. We dichotomized this variable to indicate whether participants had experienced recent residential instability.

Participants reported at baseline whether they had ever had an HIV or STI test in their lifetime. We assessed health insurance coverage by asking if the participant was covered by any health insurance plan, including government assistance.

We also asked participants to indicate which methods they regularly used to access the internet (family/household computer, personal computer, workplace computer, school computer, mobile phone, computer at a public library, computer at a community agency, and other). We created a summed score based on participants' total number of methods used. We also asked participants to report the number of hours they spent each day on the internet outside of their school and work responsibilities. The responses for time spent on the internet per day were collapsed into four categories: less than one hour; 1–3 hours; 4–6 hours; and 7 or more hours. Finally, we measured participants' eHealth literacy using an adapted, shortened 6-item version of the eHealth Literacy Scale (Norman & Skinner, 2006) (eHEALS). eHEALS asks participants about their knowledge of online health resources including where to locate them, how to use information they obtain, and comfort with assessing the quality of online health information on a 5-point scale (1=Strongly Disagree; 5=Strongly Agree). We computed a mean score based on participants' answers; higher scores indicate greater eHealth literacy (α=.94).

Participant Paradata

Paradata were collected as participants navigated the *Get Connected!* site. The site's database timestamped every action performed by each participant. We restructured the paradata output, aggregating participants' engagement and use within each of the four theoretically-informed sections of the intervention: (1) knowledge-related content regarding HIV/STI infection and testing; (2) motivations and decisional balance regarding HIV/STI testing; (3) addressing barriers to HIV/STI testing and recognizing participants' personal strengths, and; (4) an HIV/STI testing site locator in the participant's community.

Use.—We measured intervention use by counting the number of theoretically informed content features that participants clicked on while navigating each section of the intervention (Figure 1). Use of knowledge-related content was measured by the number of clicks on information about specific STIs (11 features), facts about sexual health (13 features), and information about STI testing methods (5 features). Use of motivations and decisional balance content was measured by the number of clicks on reasons to get tested (3 features) and pros and cons of getting tested (1 feature). Use of barriers to HIV/STI testing and personal strengths content was measured by the number of clicks on the barriers to getting tested (3–5 features) and personal strengths and supports (3–4 features). Participants were shown different numbers of features about barriers and strengths based on the algorithm used tailor the content to that user. Use of the HIV/STI testing site locator was measured by the number of filters (e.g. zip code, has evening hours, does not require insurance, distance from bus stop) selected while using the locator tool (9 options).

For three of the four theoretically informed intervention sections (excluding the HIV/STI testing site locator section), we classified participants as users of a given section if they clicked at least one feature in that section, and non-users if they clicked no features in that section. Participants who clicked five or more filter options in the HIV/STI testing site locator section were classified as high users, and those who clicked four or fewer were classified as low users; this division constituted a median split of the data. Total use was

measured by summing clicks from all four sections (51 potential total features), and then transformed using a square root function to approximate normality.

Engagement.—Using paradata, we measured intervention engagement by extracting the time participants spent in each of the four sections of the intervention. We also measured overall engagement with the intervention (total time spent on intervention).

Time spent in each section ranged from a few seconds to nearly 30 minutes. The database was not designed to accommodate participants who did not "complete" the intervention (move sequentially through all four sections) in one sitting. Therefore, for the two participants who completed the intervention in more than one sitting (i.e., logged off and logged on later to finish the intervention), the database recorded erroneously lengthy total time spent on the intervention. In order to minimize the influence of these outliers, we winsorized their times to a value of 35 minutes, which is approximately five minutes longer than the next highest time value for total time. We used a log-plus-one transformation on all time variables to approximate normality.

Statistical Procedures

Our analysis explored the relationships between participants' characteristics and measures of engagement and use extracted from the intervention's paradata. Nine of the 86 participants (10.5%) in the intervention arm had missing data on one or more of the sociodemographic variables included in the current analysis. There were no differences between the participants with missing data and those without on any of our measured sociodemographic factors.

We estimated the association between participants' demographic characteristics and their engagement and use metrics using bivariate statistics (e.g., correlations for continuous associations; t-test and Fisher's exact tests for dichotomous comparisons). We used a $p<.10$ cutoff to determine significant results of statistical tests given the small sample size and exploratory nature of this study. All analyses were done using R 3.3.1 (R Core Team, 2013).

Results

Overall Intervention Use and Engagement

Use.—On average, participants clicked a total of 10.28 features (SD=6.94, range=0–34) across all four sections of the intervention. When we examined differences by sociodemographic characteristics (see Table 2), we found that racial/ethnic minority participants clicked fewer features than Non-Hispanic White counterparts (8.30 clicks vs. 12.0 clicks, $p=.03$). Participants who were currently enrolled in school clicked on a greater number of features than peers not in school (11.23 clicks vs. 7.96 clicks, $p=.08$). We observed a positive association between higher intervention use and older age ($r=.19$, $p=.08$), greater educational attainment (Spearman's rho=.25, $p=.02$), greater number of methods to access the internet (Spearman's rho=.38, $p<.01$), and greater number of hours spent online outside of school and work responsibilities (Spearman's rho=.18, $p=.09$).

Engagement.—The average time spent on the intervention was 322.67 seconds (SD=385.40, median=216.00, IQR= 233–349, range=54–2100). We observed no associations between overall engagement and participants' sociodemographic characteristics (see Table 3).

Use and Engagement Differences by Theoretically Driven Content

In subsequent analyses, we examined whether participants' use of and engagement with any of the four theoretically informed sections of the intervention varied based on their demographic characteristics.

Knowledge-Related Content Regarding HIV/STI Infection and Testing

Use.: We examined whether interaction with knowledge-related content regarding HIV/STI infection and testing differed based on participants' characteristics. This section was broken into three topic areas: information about specific STIs and HIV; sexual health facts; and information about HIV/STI testing methods. Overall, 48% of participants were users of features containing information about specific STIs, 44% were users of features containing sexual health facts, and 49% were users of features containing information about STI testing methods.

Men were more likely to use information about specific STIs and HIV if they were in school (OR=2.50, 90% CI=1.00–6.67, p=.10) or if they were residentially unstable (OR= 4.17, 90% CI=1.33–14.29, p=.02). Participants were more likely to use information about specific STIs and HIV if they had been previously tested for STIs (OR=2.98, 90% CI=1.30–7.03, p=.02).

Participants who reported being single were more likely to use sexual health facts than their counterparts who reported being in relationships (OR=2.75, 90% CI=1.20–6.47, p=.03). We observed no other sociodemographic differences regarding content related to sexual health facts.

Racial/ethnic minorities were less likely than Non-Hispanic Whites to use information about HIV/STI testing methods (OR=.43, 90%CI=.18–.96, p=.06). We observed no other sociodemographic differences regarding use of information about HIV/STI testing methods.

Engagement.: On average, participants spent 86.55 seconds (SD=89.60) in the knowledge-related content regarding HIV/STI infection and testing section (median = 57.00, IQR = 26–120, range=0–537). Participants with a greater number of methods of access to the internet spent more time on knowledge-related content regarding HIV/STI infection and testing (Spearman's rho=.21, p=.06). We observed no other associations between engagement with knowledge-related content regarding HIV/STI infection and testing and participants' sociodemographic characteristics.

Motivations and Decisional Balance Regarding HIV/STI Testing

Use.: We examined whether interaction with motivational content (e.g., reasons for getting tested; decisional balance to get tested) differed based on participants' characteristics. In this

section, 52% of participants were users of at least one feature with content on reasons to test, and 26% were users of the feature focused on decisional balance.

Racial/ethnic minorities were less likely to interact with content focused on reasons to test than Non-Hispanic Whites (OR=.26, 90% CI=.11–.61, p<.01). Users without insurance were less likely to interact with content focused on reasons to test than men with insurance (OR=.35, 90% CI=.12–.91, p=.05). Users were more likely to interact with content focused on reasons to get tested if they reported greater educational attainment (Spearman's rho=.18, p=.09), having more methods of access to the internet (Spearman's rho=.34, p<.01), and spending more time online (Spearman's rho=.18, p=.10). We observed no other associations between reasons to get tested content and participants' characteristics.

Participants who had recently experienced residential instability were four times more likely to interact with the decisional balance content than residentially stable peers (OR=4.00, 90% CI=1.30–12.50, p=.02). We observed no other associations between use of the decisional balance content and participants' sociodemographic characteristics.

Engagement.: On average, participants spent 33.30 seconds (SD=45.10) in the motivations and decisional balance regarding HIV/STI testing section (median = 18.00, IQR = 9–38.5, range= 3–321). Time spent in this section was greater among participants with higher educational attainment (Spearman's rho=.20, p=.07), and those who spent a greater number of hours outside school and work online per day (Spearman's rho=.28, p=.01). No other associations were observed.

Addressing Barriers to HIV/STI Testing and Recognizing Participants' Personal Strengths

Use.: We examined whether interaction with content around barriers and strengths (e.g., barriers to getting tested; personal strengths and supports) differed based on users' characteristics. Overall, 47% of participants were users of at least one feature focused on barriers to HIV/STI testing, and 10% of participants were users of at least one feature focused on personal strengths.

Participants were more likely to be users of with content addressing barriers to HIV/STI testing if they were Non-Hispanic Whites (OR=.28, 90% CI=.12–.64, p=.01), were currently students (OR=.43, 90%CI=.16–1.07, p=.10), had health insurance (OR=.36, 90% CI=.12–.96, p=.08), and had a greater number of methods to access the internet (Spearman's rho=.23, p=.03). Participants who had never tested for STIs were more likely to be users of barriers content than peers who had tested (OR=.33, 90%CI= .14–.75, p=.02). No associations were observed between the use of personal strength content and sociodemographic characteristics.

Engagement.: On average, participants spent 26.70 seconds in the section on addressing barriers to HIV/STI testing and recognizing participants' personal strengths (SD=34.00, median = 19.50, IQR= 8–31, range= 3–275). Time spent in this section was greater among participants who had never tested for HIV (M=40.87, SD=57.80, n=23) compared to those who had tested for HIV (M=21.68, SD=17.57, n=60; p=.076). Participants who had never tested for STIs (M=34.81, SD=47.0, n=36) spent more time in this section than those who

had tested for STIs in the past (M=21.02, SD=19, n=47; p=.026). No other associations between engagement and sociodemographic characteristics were observed.

HIV/STI Testing Site Locator in the Participant's Community

Use.: We examined whether interaction with the HIV/STI testing site locator tool differed based on participants' characteristics. Participants clicked an average of 4.20 filters (SD=2.50, range=0–9) when using the HIV/STI testing site locator. Fifty-two percent of participants clicked five or more filters and were classified as high users of the filter feature. Those who had previously been tested for an STI were more likely to be high users of the filter feature than those who had never been tested (OR=2.29, 90% CI=1.02–5.26, p=.08). Participants with a greater number of methods to access the internet were more likely to be high users of filters (Spearman's rho=.25, p=.02). No other associations between use of the HIV/STI testing site locator and sociodemographic characteristics were observed.

Engagement.: On average, participants spent 128.20 seconds on the HIV/STI testing site locator (SD=198.60, median= 84.50, IQR = 60–130, range=27–1604). Participants who reported being single spent more time on the HIV/STI testing site locator (M=147.82, SD=235.86, n=45) than counterparts who reported being in a relationship (M=105.46, SD=144, n=39; p=.10). No other associations between engagement with the HIV/STI testing site locator and sociodemographic characteristics were observed.

Discussion

Online and other eHealth interventions have become more prevalent over the past decade, yet paradata reporting and analysis in these eHealth trials remains an underexplored area of research. Paradata analysis allows us to better understand how participants are exposed to specific content in an intervention (Baltierra et al., 2016; Bauermeister et al., 2017; Couper et al., 2010). By measuring participant exposure to particular features that that correspond to theoretical constructs and outcomes, future iterations of eHealth interventions may be refined in a systematic way that advances those features that are most impactful. In the current study, we illustrate how paradata can be used to develop use and engagement metrics in an eHealth intervention (*Get Connected!*), and to test whether users' characteristics were differentially associated with their use of and engagement with the intervention.

It is important to note that these paradata metrics were developed within the specific context of the *Get Connected!* intervention. *Get Connected!* was a brief intervention that aimed to increase rates of HIV and STI testing among participants. Paradata collection and analysis must be tailored to the aims of each unique intervention. For instance, an eHealth intervention that aims to increase medication adherence with daily mobile notifications may benefit more from collecting paradata on participants' login behaviors (e.g. the time of day and number of logins per day/week) than from participants' exposure to specific content. Researchers and other program developers should reserve time before deploying eHealth interventions to determine which features will be most useful to track and how best to capture and report that data.

In this study, we used paradata to characterize two dimensions of participants' interactions with the intervention. The first dimension, termed "use," counted the number of different features a participant clicked within each section of the intervention and served as a measure of what content they interacted with. The second dimension, termed "engagement," measured the depth of interaction participants had with each content area. Engagement was measured by the duration of a participant's interaction with each section, and gave us a sense of their exposure to that content. By measuring these two dimensions of interaction, we get a picture of *how* a participant is interacting with the intervention (e.g., on which features did they click) and *how much* a participant is interacting with the intervention (e.g., how much time did they spend in a given section).

Paradata analyses from *Get Connected!* suggest group differences in overall intervention use. Use differed based on participants' sociodemographic characteristics, with younger participants and racial/ethnic minorities exhibiting lower overall use of the intervention. Survey data from a previous analysis of this intervention suggest that this finding may be related to the user experience of the intervention. Horvath and Bauermeister (2017) found that racial/ethnic minority YMSM in this study were more likely than Non-Hispanic White peers to report higher overall satisfaction with the intervention and higher ratings on system quality (i.e., how easy the intervention was to navigate and its technical responsiveness), after adjusting for age, education and technology use (Horvath & Bauermeister, 2017). Less overall use, then, could indicate the intervention had a user-friendly design whereby participants, specifically racial/ethnic minorities and younger MSM, could cognitively process desired content quickly and efficiently without having to click on features irrelevant to them and read content they already know. Alternatively, these results may suggest that racially/ethnically diverse YMSM need additional incentives to use intervention content compared to white and somewhat older counterparts. Taken together, these findings underscore the importance of supplementing paradata analysis with participants' subjective experiences of the intervention to better contextualize paradata results.

Overall use of the intervention was positively correlated with measures of internet use, suggesting that access to the internet both at home and school may be associated with how participants interact with eHealth interventions. Future research in this area should measure both participant access to technology (e.g., ownership/access to smartphones, laptops, and desktops) and participant access to high-quality internet connection (e.g., mobile phone data plans, Wi-Fi access). User device type and web browser could also be collected to better understand if hardware or software affects intervention use and engagement, and if there are associations between demographic groups and those paradata metrics (Bauermeister et al., 2017; Kreuter, Farrell, Olevitch, & Brennan, 2013).

Our examination of how use and engagement varied by theoretically-informed content revealed important differences by participant characteristics. Knowledge-related content was used more often by participants who reported being single and those who reported a prior STI test. From a stages of change perspective (Prochaska & Velicer, 1997), it is possible that YMSM who were single and/or had tested for STIs previously interacted with more knowledge-related features to explore whether repeat testing for HIV/STIs was relevant to them (i.e., pre-contemplators and contemplators). Use of motivational content, on the other

hand, varied based on participants' overall ability to navigate structural barriers to testing. Both use and engagement with features highlighting specific motivations for HIV/STI testing, for instance, was higher in YMSM with more education; use of motivational content was greater amongst those who had access to health insurance. Conversely, YMSM were more likely to use the decisional balance feature if they had experienced recent residential instability. YMSM who had never tested for HIV/STIs were more likely to use and had higher engagement with content related to overcoming barriers to HIV/STI testing. Taken together, these findings also align with stages of change model perspectives, as prior research has suggested that individuals must perceive that the benefits of doing a behavior outweigh its barriers to feel motivated to engage in that behavior (Prochaska, Redding, & Evers, 2008; Prochaska & Velicer, 1997; Prochaska et al., 1994). Although we were unable to measure participants' location within the stages of change, the fact that participants at high risk (single YMSM, those who have tested for STIs, YMSM with housing instability) were engaging with these types of motivational content suggests that the content is engaging its intended audience. Future research examining whether participants' stage of change is associated with intervention use and engagement may be warranted. These data may help further refine online interventions, and offer content specific to a participant's current stage of change.

HIV/STI testing site locators are potentially powerful tools to connect YMSM to HIV/STI prevention resources (Bauermeister et al., 2015; Habel et al., 2011), yet little is known about participants' use of filters to refine their options for HIV testing sites in their communities. Within *Get Connected!*, YMSM who had previously tested for STIs were more likely to use the filtering feature than their peers who had never been tested for STIs and who were in relationships. These users may click on greater number of filters because they are more selective about where they get tested, and/or because their previous experience with testing has made them more cognizant of their particular needs. Participants who reported being single spent more time engaging with the testing site locator, a finding that suggests testing site locators are particularly relevant to those without partners. Future research, both qualitative and quantitative, may be warranted to examine how YMSM decide where to get tested, as it may inform how to create and optimize filtering features for testing site locators.

The current study has several limitations deserving mention. Given the small scale of the pilot trial, we had limited statistical power to extend our comparisons beyond a bivariate analysis. Future research seeking to examine differences in participants' use and engagement with eHealth interventions may benefit from larger sample sizes. Second, our estimate of overall use relies on counts (i.e., sum of all clicks on features) rather than proportion (i.e., what percentage of features were clicked) given that the intervention tailored content derived from participants' baseline surveys to display personalized content in some sections of the intervention. Therefore, direct comparisons in intervention use is complicated by the potential that different participant groups had more or fewer pieces of content available to them. Finally, our study relied on a convenience sample of YMSM living in Southeast Michigan. Given our inability to estimate a random sample of YMSM due to the absence of a population frame, our findings are not necessarily representative of the larger YMSM community. Replication and expansion of our methods with diverse participant groups and outcomes are warranted and encouraged. Despite these limitations, the current study is

among the first to use paradata metrics to characterize and evaluate participant engagement and use in an HIV eHealth intervention.

The collection and analysis of paradata remains an understudied area in eHealth intervention research. In the context of a pilot study, paradata analyses—knowing what content areas participants engaged with and for how long—can prove useful for intervention refinement and future study scale-up. In preparation for our efficacy trial with 480 YMSM living in three cities in the United States (Bauermeister et al., 2018), we used paradata analyses to understand what content received little use in the pilot, and replaced it with new content emerging from HIV prevention advances (e.g., information about PrEP). Furthermore, paradata metrics discussed in this paper may assist in the evaluation of future theoretically-informed eHealth interventions, especially as interventions are developed for mobile platforms (mHealth). In efficacy trials, these metrics may assist researchers to deconstruct their intervention's effects. For example, the *Get Connected!* team will be examining how paradata metrics are associated with the on-going efficacy trial's outcomes (see Bauermeister et al., 2018). Paradata may provide opportunities to test the proposed mechanisms of change through mediation analyses, or to examine whether the intervention's effects are subject to dose-response relationships. Future work in this area is warranted.

Acknowledgments:

This work was supported by the NIH Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN 139; MPI: Bauermeister & Stephenson) as part of and the UNC/Emory Center for Innovative Technology (iTech; PIs: Drs. Hightow-Weidman/Sullivan, 1U19HD089881). The content is solely the responsibility of the authors and does not represent the official views of the funding agencies.

References

Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, & Murray E (2016). The effectiveness of prompts to promote engagement with digital interventions: A systematic review. Journal of Medical Internet Research, 18(1), e6 10.2196/jmir.4790 [PubMed: 26747176]

Allison S, Bauermeister JA, Bull S, Lightfoot M, Mustanski B, Shegog R, & Levine D (2012). The intersection of youth, technology, and new media with sexual health: moving the research agenda forward. Journal of Adolescent Health, 51, 207–212. 10.1016/j.jadohealth.2012.06.012 [PubMed: 22921129]

Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, & Hightow-Weidman LB (2016). More than just tracking time: Complex measures of user engagement with an internet-based health promotion intervention. Journal of Biomedical Informatics, 59, 299–307. 10.1016/j.jbi.2015.12.015 [PubMed: 26732997]

Bauermeister JA, Golinkoff JM, Muessig KE, Horvath KJ, & Hightow-Weidman LB (2017). Addressing engagement in technology-based behavioural HIV interventions through paradata metrics. Current opinion in HIV and AIDS, 12(5), 442–446. 10.1097/COH.0000000000000396 [PubMed: 28617711]

Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, & Stephenson R (2018). A Multilevel Tailored Web App–Based Intervention for Linking Young Men Who Have Sex With Men to Quality Care (Get Connected): Protocol for a Randomized Controlled Trial. Journal of Internet Medical Research – Research Protocols, 7(8), e10444 10.2196/10444

Bauermeister JA, Pingel ES, Jadwin-Cakmak L, Harper GW, Horvath K, Weiss G, & Dittus P (2015). Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS & Behavior, 19(10), 1860–1874. 10.1007/s10461-015-1009-y [PubMed: 25638038]

Chavez NR, Shearer LS, & Rosenthal SL (2014). Use of digital media technology for primary prevention of STIs/HIV in youth. Journal of Pediatric and Adolescent Gynecology, 27(5), 244–257. 10.1016/j.jpag.2013.07.008 [PubMed: 24332613]

Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, … Cole Johnson C (2010). Engagement and retention: measuring breadth and depth of participant use of an online intervention. Journal of Medical Internet Research, 12(4), e52 10.2196/jmir.1430 [PubMed: 21087922]

Dittus PJ, De Rosa CJ, Jeffries R. a., Afifi A. a., Cumberland WG, Chung EQ, … Ethier K. a. (2014). The Project Connect Health Systems Intervention: Linking Sexually Experienced Youth to Sexual and Reproductive Health Care. Journal of Adolescent Health 10.1016/j.jadohealth.2014.04.005

Habel MA, Hood J, Desai S, Kachur R, Buhi ER, & Liddon N (2011). Google it: obtaining information about local STD/HIV testing services online. Sexually Transmitted Diseases, 38(4), 334–338. 10.1097/OLQ.0b013e3181fe64f2 [PubMed: 21099732]

Horvath KJ, & Bauermeister JA (2017). eHealth Literacy and Intervention Tailoring Impacts the Acceptability of a HIV/STI Testing Intervention and Sexual Decision Making Among Young Gay and Bisexual Men. AIDS Education & Prevention, 29(1), 14–23. 10.1521/aeap.2017.29.1.14 [PubMed: 28195779]

Kreuter MW, Farrell DW, Olevitch LR, & Brennan LK (2013). Tailoring Health Messages: Customizing Communication With Computer Technology Routledge: New York, New York.

Marks G, Crepaz N, & Janssen RS (2006). Estimating sexual transmission of HIV from persons aware and unaware that they are infected with the virus in the USA. AIDS, 20(10), 1447–1450. 10.1097/01.aids.0000233579.79714.8d [PubMed: 16791020]

Muessig KE, Nekkanti M, Bauermeister JA, Bull S, & Hightow-Weidman LB (2015). A Systematic Review of Recent Smartphone, Internet and Web 2.0 Interventions to Address the HIV Continuum of Care. Current HIV/AIDS reports 10.1007/s11904-014-0239-3

Norman CD, & Skinner HA (2006). eHEALS: The eHealth Literacy Scale. Journal of Medical Internet Research, 8(4), e27. [PubMed: 17213046]

Prochaska JO, & Velicer WF (1997). The transtheoretical model of health behavior change. American journal of health promotion, 12, 38–48. 10.4278/0890-1171-12.1.38 [PubMed: 10170434]

Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, … Rosenbloom D (1994). Stages of change and decisional balance for 12 problem behaviors. Health psychology, 13, 39–46. 10.1037/0278-6133.13.1.39 [PubMed: 8168470]

R Core Team (2013). R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing Retrieved from http://www.R-project.org/

Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrao J, Gamble P, & Wykes T (2018). Barriers to and facilitators of engagement with remote measurement technology for managing health: Systematic review and content analysis of findings. Journal of Medical Internet Research, 20(7), e10480 10.2196/10480 [PubMed: 30001997]

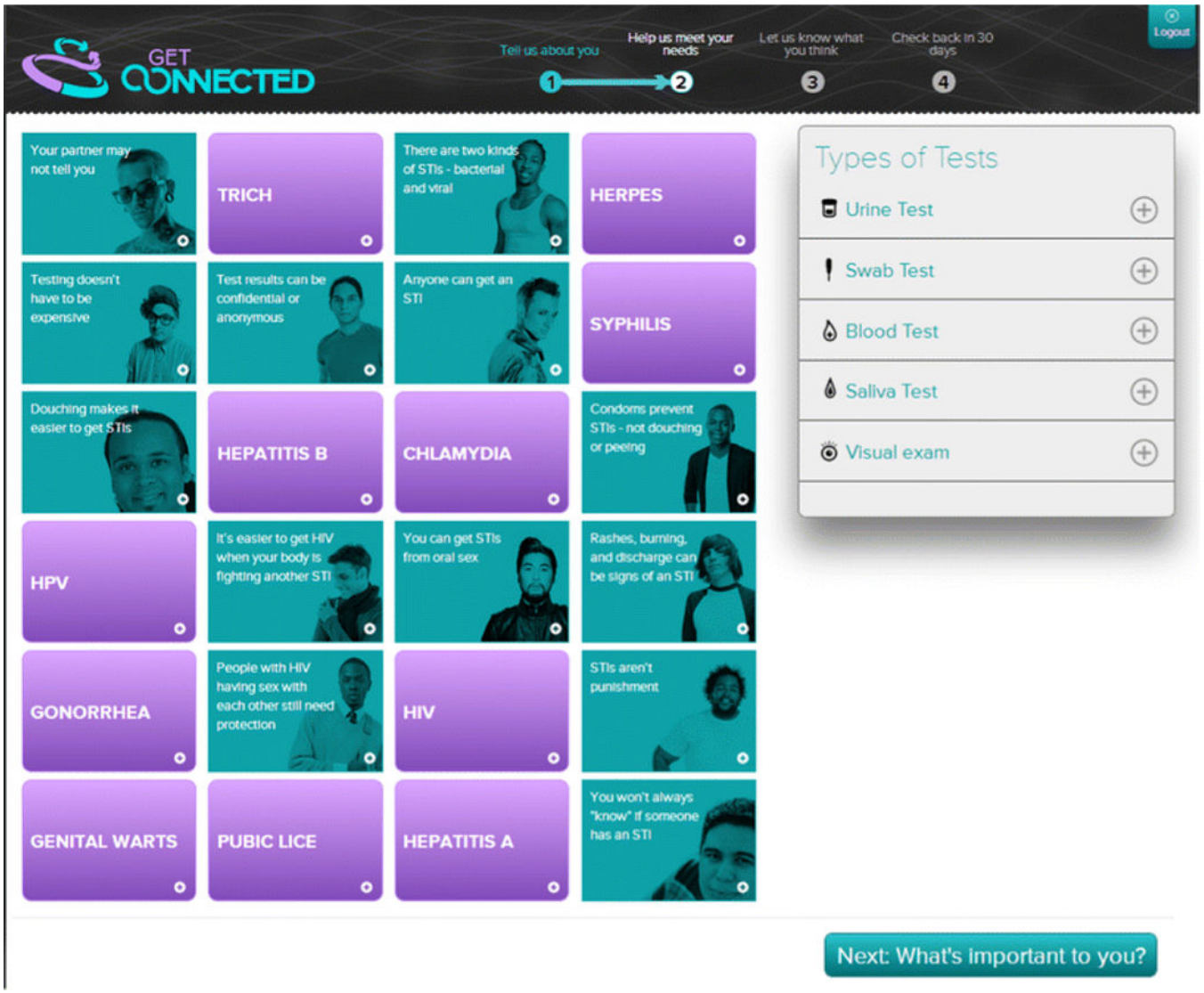

Figure 1.
Screenshot of the knowledge-related content section of the *Get Connected!* intervention

Table 1.

Descriptive Statistics and Paradata Metrics of Young Men who Have Sex with Men Assigned to the *Get Connected!* Intervention Arm (N=86)

	n	%
Racial/Ethnic Minority	40	47%
Gay-Identified	69	80%
In a Relationship	39	45%
Has Insurance	64	75%
Prior HIV Test	63	73%
Prior STI Test	49	57%
Residential Instability	16	19%
Currently a Student	61	71%
Age	mean=21.38	SD=2.13
Educational Attainment	-	-
8th grade or less	0	0%
Some high school	7	8%
Graduated high school/GED	13	15%
Technical school	2	2%
Associate degree	3	3%
Some college	36	42%
College	14	16%
Some graduate school	9	10%
Graduate school	2	2%
Internet Methods	mean=2.55	SD=1.22
Online Hours	-	-
Less than 1 hour	11	13%
1–3 hours	46	53%
4–6 hours	22	26%
7 or more hours	7	8%
eHEALS	mean=3.24	SD=.63
Paradata Metrics		
Overall Engagement	mean=322.67	SD= 385.40
Knowledge Engagement	mean= 86.55	SD=89.60
Motivation Engagement	mean= 33.3	SD=45.10
Barriers Engagement	mean= 26.70	SD=34.00
Locator Engagement	mean= 128.20	SD=198.60
Overall Use	mean=10.28	SD = 6.94
STI Facts Use	41	48%
Testing Facts Use	42	49%
Information about STI Tests	38	44%
Motivations Use	45	52%
Decisional Balance	22	26%

	n	%
Barriers Use	40	47%
Strengths Use	9	10%
Testing Site Locator Use	45	52%

Table 2.

Bivariate Comparisons of Overall Intervention Use in *Get Connected!* (N=86)

	Yes			No				
Overall Use	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**t/r statistic**	***p***
Racial/Ethnic Minority	8.30	4.96	40	12.00	7.95	46	−2.26	0.03
Gay-Identified	10.43	6.25	69	9.65	9.45	17	1.08	0.30
In a Relationship	9.56	6.59	39	10.87	7.23	47	−0.77	0.45
Has Insurance	10.95	7.52	64	8.52	4.48	21	1.06	0.29
Prior HIV Test	10.54	6.62	63	9.57	7.86	23	1.06	0.29
Prior STI Test	11.20	7.60	49	9.05	5.83	37	1.41	0.16
Residential Instability	9.81	4.52	16	10.39	7.41	70	−0.05	0.96
Currently a Student	11.23	7.52	61	7.96	4.64	25	1.77	0.08
Age							0.19	0.08
Educational Attainment							0.25	0.02
Internet Methods Sum							0.38	0.001
Online Hours Per Day							0.18	0.09
eHEALS							−0.03	0.78

Table 3.

Bivariate Comparisons of Overall Intervention Engagement in *Get Connected!* (N=83)

	Yes			No				
Overall Engagement	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**t/*r* statistic**	***p***
Racial/Ethnic Minority	333.82	432.47	38	313.27	345.47	45	−0.12	0.91
Gay-Identified	307.75	347.19	67	385.19	525.80	16	−0.12	0.90
In a Relationship	290.16	363.07	38	350.13	405.33	45	−0.78	0.44
Has Insurance	311.38	364.94	64	370.83	467.73	18	−0.65	0.52
Prior HIV Test	275.38	271.28	60	446.04	578.17	23	0.90	0.38
Prior STI Test	288.38	305.40	47	367.44	470.93	36	0.98	0.33
Residential Instability	349.00	450.16	15	316.87	373.18	68	0.50	0.62
Currently A Student	358.58	440.4	59	234.42	170.25	24	1.13	0.26
Age							0.05	0.64
Educational Attainment							0.07	0.51
Internet Methods Sum							0.07	0.52
Online Hours Per Day							0.11	0.33
eHEALS							−0.04	0.70