Never Too Old, Cold or Dry to Watch the Sky: A Survival Analysis of Citizen Science Volunteerism

S. ANDREW SHEPPARD, University of Minnesota and Houston Engineering, Inc.
JULIAN TURNER, Colorado State University
JACOB THEBAULT-SPIEKER, University of Minnesota
HAIYI ZHU, University of Minnesota
LOREN TERVEEN, University of Minnesota

CoCoRaHS is a multinational citizen science project for observing precipitation. Like many citizen science projects, volunteer retention is a key measure of engagement and data quality. Through survival analysis, we found that participant age (self-reported at account creation) is a significant predictor of retention. Compared to all other age groups, participants aged 60-70 are much more likely to sign up for CoCoRaHS, *and* to remain active for several years. We also measured the influence of task difficulty and the relative frequency of rain, finding small but statistically significant and counterintuitive effects. Finally, we confirmed previous work showing that participation levels within the first month are highly predictive of eventual retention. We conclude with implications for observational citizen science projects and crowdsourcing research in general.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; • **Mathematics of computing** → *Survival analysis*; • **Social and professional topics** → *Seniors*; • **Applied computing** → Environmental sciences;

Additional Key Words and Phrases: citizen science; crowdsourcing; volunteer retention

ACM Reference format:
S. Andrew Sheppard, Julian Turner, Jacob Thebault-Spieker, Haiyi Zhu, and Loren Terveen. 2017. Never Too Old, Cold or Dry to Watch the Sky: A Survival Analysis of Citizen Science Volunteerism. *Proc. ACM Hum.-Comput. Interact.* 1, 2, Article 94 (November 2017), 21 pages.
https://doi.org/10.1145/3134729

1 INTRODUCTION

Citizen science provides researchers the opportunity to crowdsource large quantities of environmental observations or classifications in an era of limited resources but ubiquitous technology. For science-intensive projects, or those with specific policy goals, data quality is essential to project success [51, 57]. Other projects leverage citizen science as an opportunity to engage non-professionals in a rich shared appreciation of the natural environment and the scientific method. This leads to research focusing on participant motivations and learning outcomes [3, 25, 30, 45, 59]. For many projects, both data quality and engagement are essential – providing a unique opportunity for CSCW and HCI to explore the design space that addresses the tensions, tradeoffs and synergy between these goals [4, 15, 21, 26, 49, 50].

2573-0142/2017/11-ART94 $15.00
https://doi.org/10.1145/3134729

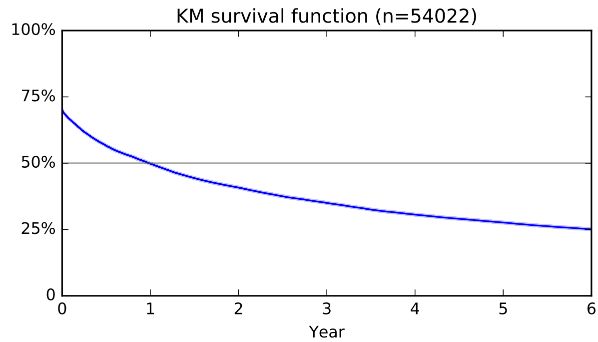

Fig. 1. Kaplan-Meier survival curve for CoCoRaHS participants. 50% of participants drop out by the end of their first year, including 28% of accounts that never submit an observation. However, those who make it past the first year often stay for many more (c.f. [42]). Confidence intervals are indistinguishable until 10 years.

While data quality and participant engagement can be measured in many ways, **volunteer retention** stands out as one key metric that addresses both goals. This is certainly the case for *CoCoRaHS*, or the Community Collaborative Rain, Hail, and Snow Network. CoCoRaHS volunteers submit observations that are used as nearly real-time input into the meteorological outputs of the U.S. National Weather Service and other organizations. Quality control is important - but so is inclusivity and ensuring that participants know their effort is appreciated. In fact, volunteer retention is itself a key component of quality control. While every CoCoRaHS contribution is useful, only participants with 100 or more daily contributions are included in analysis by the Global Historical Climatology Network. As Figure 1 shows, 50% of CoCoRaHS participants drop out by the end of their first year. As it turns out, 50% is also the ratio of CoCoRaHS participants who contribute 100 records or more, and the two metrics are highly correlated[1].

While user activity patterns are well-studied in fully online systems like Wikipedia [18] and Galaxy Zoo [29], little quantitative work to date has measured activity and retention in field-based citizen science. A better understanding of the predictors of retention would help inform recruitment strategies as well as potential interventions [44]. We are interested in the following research questions:

> *How well do **participant characteristics**, **task characteristics**, and **early activity** predict retention?*
>
> *How does retention relate to other measures of **data quality**?*

This paper provides three contributions toward understanding these questions.

1. We find that participant age at signup is a particularly good predictor of retention *and* (to a lesser extent) several other measures of data quality. Since very few studies have directly compared participant demographics to actual activity levels, this is the primary contribution of this paper.
2. We find that exposure to more below-freezing days *positively* correlates with retention. This is counterintuitive given that cold weather presumably increases CoCoRaHS task difficulty.
3. Finally, we show that activity within the first month after signup is one of the strongest predictors of long term retention. While this finding largely replicates previous work in online peer production communities, it leads to important implications for the design of citizen science and crowdsourcing projects.

[1] 92% of those who make it past the first year also pass the 100 contribution threshold, vs. 14% of those who drop out sooner.

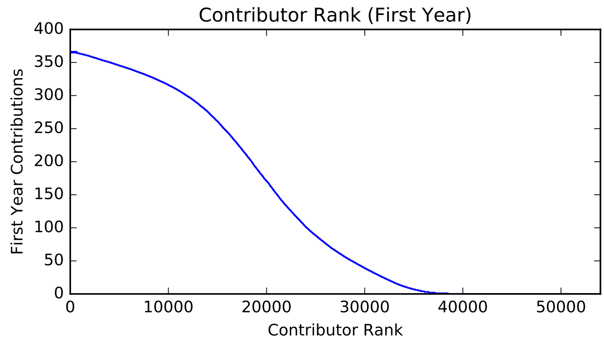

Fig. 2. While there is a long tail of CoCoRaHS contributors, the power law for any given year is truncated due to the maximum contribution rate of once per day.

In the remainder of this paper, we review related work in volunteerism and citizen science, before turning to discuss the design of our survival analysis. We then explore the results showing how our measured characteristics correspond to volunteer retention and data quality. We discuss potential explanations for why our results contrast with earlier work and intuition, before concluding with recommendations and directions for future work.

2 RELATED WORK

2.1 Volunteerism

While crowdsourcing and computer-aided citizen science are relatively new phenomena, they are related to much older fields of study such as volunteerism. It is known that retirees are generally less likely to volunteer than younger age groups, with volunteerism peaking at around 40-45 [37]. However, the difference in volunteering rates is much smaller in recent years than it was a few decades ago [8], and retirees who do volunteer tend to spend more actual time per month than younger volunteers [9]. There are also large differences in peak age depending on activity type, and some evidence that motivations for volunteering change as participants age [33, 34].

Several studies have explored factors that influence retention and other measures of commitment to volunteering. Participant motivations, attitudes, and beliefs are often the primary focus of research [5, 45, 46]. Cnaan et al. studied demographic, personality, and situational factors, noting that volunteerism is qualitatively different than paid work and that different theoretical models should be used [10]. They found that age was positively correlated with self-reported duration of volunteering. Komp et al. found that chronological age by itself is a poor predictor of time spent volunteering, at least for the oldest age groups [24]. By contrast, we found that numeric age was useful as a predictor in our model.

In addition to age, gender is also known to influence the types and duration of volunteer activity [37]. While not directly related to volunteerism, models for technology adoption also incorporate both age and gender. Most relevant to our work, age and gender are known to interact to influence habit forming, with the effect being strongest for older men [53]. Research on online social networks has shown that perceptions of existing gender imbalances can create a self-perpetuating skew toward one gender or the other [31].

Traditionally, research on volunteerism has relied on self-reported surveys to determine activity levels and duration [58]. Relatively recently, studies of online systems have been able to combine theory-backed survey results with actual activity logs [16, 32].

2.2 User Generated Content and Peer Production

The near ubiquity of internet access has enabled a particular kind of volunteerism: the creation and curation of repositories of shared knowledge. Two of the largest platforms in this space are Wikipedia and OpenStreetMap. In these and most peer production systems, a small number of participants perform the bulk of the work. Ma et. all characterized this "power law" distribution in OpenStreetMap, finding significant skew in user contribution rates as well as in the size and number of edits for each geographic element [28]. As Figure 2 shows, CoCoRaHS has a similar distribution, though it is truncated due to a maximum contribution rate. Panciera et al. found that the most active Wikipedia contributors distinguish themselves almost immediately after signup [36]. Yasseri et al. evaluated temporal activity patterns in OSM, noting important differences with Wikipedia [61].

Survival analysis has been used to study retention outcomes in both Wikipedia and OSM. Ortega et al. describe the use of survival analysis to characterize median dropout times in several Wikipedia languages as well as open source projects [35]. Zhang et al. further characterize the distinct survival functions for two main categories of Wikipedia editors [62]. Zhu et al. showed how survival analysis could be used to predict the survival of wiki communities themselves [63]. Dittus et al. evaluated participation in three Humanitarian OSM Team initiatives, comparing early activity with long-term retention [12]. Like our study, they leveraged actual contribution history with an inactivity buffer to measure survival, and they incorporated task difficulty and early activity as predictor variables. However, they did not explore how participant demographics affect outcomes.

2.3 Virtual Citizen Science and Crowdsourcing

Citizen science is a relatively broad category encompassing a wide variety of scientific activities. Wiggins et al. identify five broad types of citizen science projects: *Action*, *Conservation*, *Investigation*, *Virtual*, and *Education* [55]. Common to all project types is a focus on scientific inquiry (often toward environmental questions), and usually a more hierarchical approach than peer production models. Rotman et al. note that the unique relationship between professional scientists and volunteers in citizen science means that domain-specific research is needed to understand how participant motivations change over time [44].

Crowdsourcing is also a vague term, used to describe both paid microtasking work (e.g. Mechanical Turk [6, 43]), as well as large-scale unpaid image classification projects like Galaxy Zoo [29]. This second definition is often used interchangeably with citizen science [55]. Throughout this paper, we use crowdsourcing to refer primarily to large-scale citizen science projects, and not paid microtask work.

Perhaps due to its technical focus and scalability, *Virtual* Citizen Science has received the bulk of attention in the literature, often under the banner of crowdsourcing. Perhaps the most well-known (and well studied) virtual citizen science platform is Galaxy Zoo, or more broadly, the Zooniverse family of artifact classification projects [41]. Raddick et al. found that older participants are slightly over-represented in surveys (versus what would be expected from the general internet population) [40]. However, Cox et al. found no significant effect of age on retention or activity [11].

Some work has explored the characteristics and experience of volunteer crowdsourcing in niche platforms. Kobayashi et al. found that seniors remained more active in contributing to an experimental OCR proofreading system [23]. Similarly, Baruch et al. found that the most active participants in the Tomnod platform tend to be over 50 [2], based on the results of several surveys. However, they do not report quantitative differences in the actual number of contributions or in participant retention.

2.4 Observational (Field-Based) Citizen Science

Relatively little work to date has explored participant contribution and retention patterns in citizen science projects that require real-world activities. Sullivan et al. briefly characterize participation patterns in eBird, noting that activity peaks in May and drops during the summer [52]. Wood et al. note that 90% of contributions come from the most active 10% of eBird contributors [60]. Wiggins and He explore the use of technology and its affect on data validation practices in iNaturalist [56].

We are unaware of published research that characterizes the demographics of eBird and iNaturalist contributors, or quantifies factors that lead to increased participation and retention. Together with CoCoRaHS, we believe that these are among the largest observational citizen science projects, making the relative lack of quantitative research particularly striking. We fill a gap in this research by applying techniques from peer production and crowdsourcing to study participation patterns in observational citizen science.

2.5 About CoCoRaHS

The Community Collaborative Rain, Hail, and Snow Network, or *CoCoRaHS*, is a multi-national citizen science project engaging volunteers in daily precipitation monitoring. The network started in Colorado, USA in 1998 after an underpredicted flash flood, and has since expanded to all 50 US states, as well as Canada and the Bahamas [42]. As of March 2017, over 50,000 participants have contributed over 33 million daily observations.

CoCoRaHS participants are asked to empty a rain gauge daily and report the total precipitation during each 24-hour period. During winter months, participants can either melt snow to report the equivalent rain amount, or take a break and rejoin in the spring. Like many citizen science projects, CoCoRaHS tries to balance data collection with the equally important goal of public outreach toward scientific awareness.

Reges et al. describe the demographics and participation patterns in CoCoRaHS, noting a skew toward older participants in surveys and providing an overview of seasonal activity patterns [42]. However, they do not measure whether the skew is due to differences in recruitment or in retention rates, or directly measure the relationship between age and actual contribution activity. We build on their work by measuring actual activity levels, and controlling for initial signup skew when computing retention rates. We demonstrate that *both* recruitment and retention are skewed toward older participants in CoCoRaHS, and that the relationship between age and retention is effectively monotonic.

In addition to a skew toward retirement age, the majority of CoCoRaHS participants are white and male. Given the project's dual purpose of data collection and scientific awareness, CoCoRaHS has set an explicit goal to expand the diversity of project participants [42]. This is important for expanding climate literacy among a broader segment of the population, but also for expanding geographic coverage in the dataset (since different regions have different demographic makeup). From a practical standpoint, applications for additional project funding often require a discussion of how the resources will be used to serve underrepresented populations.

Almost since its inception, CoCoRaHS has had a website[2] to support participant registration and online data entry. Since 2014, CoCoRaHS has also provided data entry apps for Android and iOS. The provenance and editing history is stored in the database with each observation, providing a unique opportunity to apply analysis techniques from UGC and crowdsourcing platforms to observational citizen science.

[2] https://cocorahs.org

3 HYPOTHESES

While parts of this research were exploratory in nature, we drew heavily from prior work and theoretical foundations, which we present as hypotheses below.

The most active participants in Wikipedia and other large peer production systems have tended to be younger [17]. However, domain-specific citizen science projects have attracted middle-age and even older participants [2, 40], and Reges et al. [42] describe a skew toward middle and retirement age participants in CoCoRaHS survey respondents. We hypothesize that this skew will pan out in activity levels and retention as well.

H1: Age positively correlates with retention in observational citizen science.

A critical component of participant retention in citizen science relates to the structure of the actual task being performed [51, 55]. In theory, participants in a citizen science project all follow the same protocol for collecting and reporting data. In practice, the experience of task completion may vary significantly for different participants depending on how often they experience the phenomena of interest. For example, in Galaxy Zoo, participants who see fewer interesting photos of galaxies are more likely to end their session early [29]. Thus, our second hypothesis is as follows.

H2: Frequent encounters with the monitored phenomena improve retention.

With field-based citizen science projects, the design of the data collection protocol is particularly important [50]. In some cases, participants may need to use a more complex protocol to adapt to changing external conditions. This may lead to increased task difficulty. We hypothesize that as task difficulty increases, so does the likelihood of participants dropping out. This could be due to exhaustion, or due to reduced participation levels. In particular, when the task difficulty changes based on geographic and seasonal factors, we would expect this to be reflected in retention patterns as well.

H3: Increased task difficulty leads to lower retention.

In many open collaboration systems, long-term participants differentiate themselves within a short period after signing up [36]. However, in Galaxy Zoo and OpenStreetMap, high initial activity is sometimes associated with early burnout [12, 38, 47]. Given the upper limit on CoCoRaHS participation of one record per day, we expect a relatively straightforward relationship between early activity and retention.

H4: Participants who are more active during their first month stay longer.

Finally, we expect that highly motivated participants will not only participate longer, but also contribute higher quality data than less active participants [1].

H5: Participants who are more active during their first month contribute higher quality data throughout their first year.

4 METHODOLOGY

Our primary dataset was the full archive of 35,581,914 CoCoRaHS daily observations from June 1998 to February 2017[3]. We excluded multi-day observations, which make up a relatively small fraction of the dataset and are considered less useful by meteorological analysts. Since the CoCoRaHS database incorporates data from other monitoring networks, we limited our analysis to participants marked as being members of CoCoRaHS. We also incorporated information entered in the account registration form on the CoCoRaHS website, which includes information about each participant and the geographic location of their rain gauge.

[3] As determined by observation date, which is not always the same as data entry date.

Our dependent variables were volunteer retention and data quality, as defined below. We operationalized 10 characteristics of the participant, the task, and early activity as independent variables. All independent variables were defined using only information that would be known within a month after a participant signed up.

4.1 Dependent Variables

4.1.1 Volunteer Retention. To test hypotheses 1-4, we conducted a survival analysis using the Cox proportional hazards model. Survival analysis has an advantage over other statistical methods in that it can handle "censored" dropout events (i.e. for participants who remain active past the end of the study window). In addition, the use of survival analysis allows us to separate the *outcomes* of various predictor variables from any biases in their initial distribution.

Since participants do not announce when they are leaving, we determined dropout based on the period of inactivity after the last contribution. While Dittus et al. use 180 days as a cutoff [12], Karumar et al. use 365 days [22]. Given the large seasonal variability in CoCoRaHS participation, we limited our choice of time periods to multiples of a year. We settled on a one year cutoff, which corresponds to CoCoRaHS' internal inactivity determination rule [42]. As it turns out, about 8% of participants who we counted as having dropped out rejoined again after a break of over one year. For these participants, we excluded the second activity period from our analysis.

We used the account creation date as the start date for our analysis, unless the participant submitted an observation for an earlier date (about 8.5% of accounts). We necessarily excluded any participants who signed up after February 2016. We counted participants who signed up and never contributed anything (about 28% of accounts) as having dropped out on their first day.

We conducted our exploratory analysis with Kaplan Meier curves using the `lifelines` survival analysis library for Python [14]. We used `pandas` and `matplotlib` to create the graphics in this paper. We used R's `survival` package to run the final Cox analysis due to its support for frailty terms in the model (see the subsection on control variables).

4.1.2 Data Quality. Data quality is inherently context-specific, but is most commonly thought of in terms such as *accuracy*, *reliability*, *timeliness*, and *consistency* [54]. We can view participant retention as a measure of reliability. To test H5, we operationalized three additional definitions of data quality as follows:

- **timeliness**: the proportion of observations entered in the system on the same day they were observed
- **consistency**: the proportion of observations that immediately followed a previous day observation (i.e. without multi-day gaps, which are less useful for analysis)
- **accuracy**: the proportion of observations that were never edited[4]

Since we are already measuring retention (reliability) separately, we designed the three other metrics as percentages to minimize the effect of the total number of contributions. However, even with this factoring, the metric for consistency is necessarily correlated with higher contribution rates. We computed these metrics for each participant's first year of contributions, excluding participants who never contributed anything. We then conducted a separate linear regression for each metric.

[4]While an edited record is likely higher quality than the unedited version, we take the act of editing as a proxy for an accuracy issue in the original data. This is based on prior work on Wikipedia which treats the persistence (i.e. non-editing) of each word as an indicator of article quality (c.f. [19]). However, note that in CoCoRaHS, editing is relatively rare and usually done by the contributors themselves. More broadly, it is challenging to define a universally meaningful notion of quality and accuracy [50]. We consider this particular operationalization of accuracy to be good enough for this particular use as one of several proxies for participant data quality.

4.2 Participant Characteristics

4.2.1 Age. To test Hypothesis 1, we included participant age in the model. We were able to obtain this without a survey, as participants can optionally provide their age in years to CoCoRaHS when creating an account. However, only around 30% of participants actually provide their age. Since the Cox PH model does not support null values in predictor variables, we used multiple imputation to fill in plausible random values for age [7]. We validated our results by running a second model containing only the participants who entered age, and by comparing the actual median survival for participants who entered ages versus those who did not. As we discuss in the next section, the results for each method were compatible.

4.2.2 Gender. As noted previously, gender has also been shown to influence volunteerism and technology adoption. Participants are not asked for their gender when signing up for CoCoRaHS, but they do enter their full name[5]. We were able to estimate gender using the relative frequencies of participants' first names in the United States, using the equation $\frac{M(x)-F(x)}{M(x)+F(x)}$ as proposed by Liu et al. [27]. While they use U.S. Census data for their name corpus, we used birth names as registered with the U.S. Social Security service between 1950 and 2016.

We used 0.5 as the association threshold for our analysis. 86.6% of accounts met this threshold and were assigned a gender, with an average absolute score of 0.97. This means that, on average, fewer than 1 / 75 people with each given name had the opposite gender than our assignment, at least based on the U.S. Social Security dataset. While this does not guarantee a correct assignment in every case (particularly for group accounts), we believe the accuracy is high enough that our results will not be affected. The 13.4% of accounts with an absolute score less than 0.5 were assigned to the Unknown / Other group.

Table 1. First Name Gender Association

score	assigned gender	n	%
< -0.9	Female	12,618	23.4%
-0.9 to -0.5	Female	1,167	2.2%
-0.5 to 0.5	Unknown / Other	7,260	13.4%
0.5 to 0.9	Male	1,692	3.1%
> 0.9	Male	31,285	57.9%

4.3 Task Characteristics

The CoCoRaHS experience may be different for participants depending on how often they experience rain or snow. We can take the number of rain days as a measurement of **phenomena frequency**, and the number of below-freezing days as a proxy for **task difficulty**.

4.3.1 Rain Days (year before signup). Participants are regularly reminded that reporting a zero is highly preferred to not reporting at all. Nevertheless, we expected to find that participants from drier climates would have more trouble with consistent data entry, per Hypothesis 2.

We spatially joined each CoCoRaHS participant's reported latitude and longitude to the PRISM climate grid of historical daily rainfall and temperature data [39]. Due to the nature of the PRISM dataset, we limited our analysis to participants with locations that were within the 48 continental

[5]While most CoCoRaHS data is open to the public, full name and address are protected by privacy policy. We used only first names for this analysis.

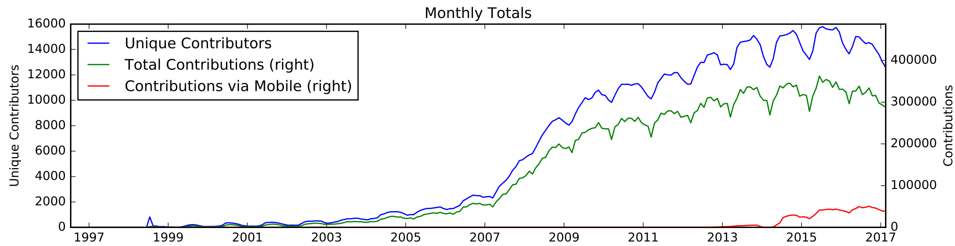

Fig. 3. CoCoRaHS monthly contribution rates; note the pronounced seasonal effect. The right axis is scaled to ~16000×31 contributions per month

U.S. states. We counted any daily grid cell with more than 0mm of precipitation as a rainy day. We used the 365 days before signup in order to avoid measuring the effect of rainy days that occurred after participants may have already dropped out. Ideally we would have included the rain they experienced after signing up, but this is highly correlated across years.

4.3.2 Freezing Days (year before signup). Per Hypothesis 3, we expected to find that participants who experienced more below-freezing days would drop out sooner, either due to the increased difficulty of the snow protocol, or to forgetting about the project during the break. As Figure 3 shows, there are substantial drops in activity levels during the winter months.

We calculated freezing days using the same technique as for rain days, but instead counted days where the average temperature was less than zero degrees Celsius. Given that precipitation itself changes when temperatures are below zero, we included an interaction term combining rain and freezing days.

4.4 Early Activity

4.4.1 First Month Observations. To test Hypothesis 4, we counted the number of observations each participant submitted during their first 30 days of activity. While previous work has used sessions measured in minutes or hours as a metric for activity, a timespan of a month made more sense for CoCoRaHS given the maximum contribution rate of once per day. Due to differences in signup time, some participants were able to submit 31 records during their first 30 days; we counted these participants as having submitted 30. We included an interaction term combining age and first month observations, since habit-forming is known to be stronger for older adults [53].

4.5 Control Variables

4.5.1 U.S. State (Fixed/Frailty). CoCoRaHS is organized into separate leadership structures for each U.S. state. Each state joined at different times, with Colorado starting in 1998 and Minnesota joining in 2009. In addition, each local organization is given considerable flexibility in recruiting and participation methods, particularly during the annual March Madness recruitment drives. Some state coordinators offer free rain gauges, others create YouTube videos, while others put out press releases or letters to the editor. Some coordinators are paid to work with CoCoRaHS as part of their official responsibilities, while others are purely volunteer based. Finally, the level of enthusiasm for project recruitment differs between coordinators.

To reduce the risk of misattributing differences to climate that were instead due to differences in the local CoCoRaHS leadership structure, we included participants' U.S. state of residence as a *frailty* (fixed effect) term in the model. This may have overcorrected for effects that were in fact due to climate differences.

Table 2. Volunteer Retention - Survival Analysis (n=52154)

Predictor	H.R.		95% CI	effect size	mean	stdev
First Month Observations	0.488	****	0.482 - 0.495	+1334 days	10.6 obs.	11.28
% Submitted via Mobile App	0.936	***	0.920 - 0.953	+52 days	1.6%	11.6%
Age (imputed, see discussion)	0.867	****	0.844 - 0.892	+117 days	48.3 yrs	10.44
Rain Days (year before signup)	1.027	**	1.007 - 1.047	-24 days	71.6 days	26.4
Freezing Days (“ ”)	0.948	***	0.922 - 0.975	+43 days	99.1 days	57.1
Signup rank within State (log)	1.082	****	1.065 - 1.099	-66 days	6.37	1.31
Had gauge at signup (or later)	0.662	****	0.636 - 0.690	+441 days	8.2%	
Daily internet access	0.961	**	0.938 - 0.986	+34 days	75.3%	
Signup during March Madness	1.043	**	1.012 - 1.073	-36 days	14.5%	
Female (guess from name)	1.009		0.975 - 1.043		25.5%	
Male (guess from name)	0.928	***	0.900 - 0.956	+61 days	61.3%	
First Month Obs. × % via Mobile	1.020	*	1.002 - 1.040	-18 days		
Age × Female	0.964	*	0.932 - 0.998	+27 days		
Age × Male	0.973		0.944 - 1.002			
Age × First Month Obs	0.972	***	0.960 - 0.984	+21 days		
Rain Days × Freezing Days	0.992		0.977 - 1.007			
U.S. State	(frailty)	****				
Concordance: *0.772*						
Median Retention: *295 days*						

p-values: *<0.05 **<0.01 ***<0.001 ****<2e-16

4.5.2 Timing & Other Factors. We also included a few variables to control for various factors related to the timing and nature of participant signup and early activity. We controlled for an early adopter effect by including participant signup rank within their state. We also flagged March signups since they are likely to be part of an annual recruitment drive (March Madness). We accounted for participants who had a rain gauge when signing up[6], since we expected them to be more intrinsically motivated to start reporting right away. We also included a variable for daily internet access, since participants without it presumably need to call to submit their observations. Finally, we measured relative use of the CoCoRaHS observer mobile app for first month data entry. The apps were only available for two years of our study period, and the effects were nuanced. We hope to investigate the effects of mobile usage more fully in future work.

4.6 Correlation

To ensure accurate model specification, we verified that none of our independent variables were strongly correlated by evaluating Spearman’s ρ for each pair of coefficients. The only correlation stronger than 0.15 was between *Signup rank within state (log)* and *Rain days (year before signup)*, which are slightly negatively correlated (-0.21).

[6]Technically, this flag can also be set after signup, if the contributor interacts with a coordinator who has access to update their account.

5 RESULTS

The results of the Cox proportional hazards model are shown in Table 2. For the purpose of evaluating effect size, we compare the median survival day, i.e., the day by which 50% of participants have dropped out. For continuous variables, we compare the mean value (shown at right) and the median baseline survival (295 days) with the effect of an increase of one standard deviation in the predictor. For logistic variables, the effect size shows the difference between the false and true conditions.

We also generated figures 4-8 to further explore the effect of certain predictors on measured retention rates. We computed these by generating small bins for each continuous variable and then extracting the median dropout date and 95% confidence intervals from the Kaplan-Meier curve for each bin. While this approach loses information about the actual shape of each curve, it makes it feasible to interpret differential outcomes for continuous variables.

To provide insight into the differences between initial and long-term skews in the data, we also plot the initial distribution (as measured at or shortly after account creation) as a histogram under the retention chart for each figure. Since this is essentially the n for each measurement, it is inversely correlated with the confidence interval for the median outcome shown on the upper chart. There is no necessary correlation between the initial n and the outcome, other than potential shared underlying mechanisms.

5.1 Participant Characteristics

CoCoRaHS is heavily skewed toward older participants: the average age at signup is 48, while the median is 52 and the mode is 60. Compared to all other age groups, participants aged 60-70 are more likely to sign up - and even more likely to stay for several years. As Table 2 shows, the model effect size is quite large: an additional 10 years of age corresponds to 117 additional days of participation in the program. Since the age used in Table 2 contains a large number of imputed values, we also ran a second model with only the participants who entered an age. The results are shown in Table 3. In the smaller model, the hazard ratio for older participants is even smaller and the effect is larger (+247 days).

This striking result can be explored further by examining the actual median survival for different ages, as shown in Figure 4. Between the ages of 20 and 70, and in particular between 45 and 65, there is an almost perfectly monotonic relationship between age and retention. The median dropout for participants aged 19-20 is 13 days, while the median dropout for ages 69-70 is 3.12 years. Thus,

Table 3. Volunteer Retention for known age (n=16196)

Predictor	hazard ratio		effect
First Month Observations	0.4710	****	+1164d
Age	0.7578	****	+247d
Female (guess from name)	0.9384	*	+41d
Male (guess from name)	0.8720	***	+105d
Age × Female	0.9561		
Age × Male	0.9614		
Age × First Month Obs	0.9559	***	+40d
Concordance: 0.781			
Median Retention: 278 days			

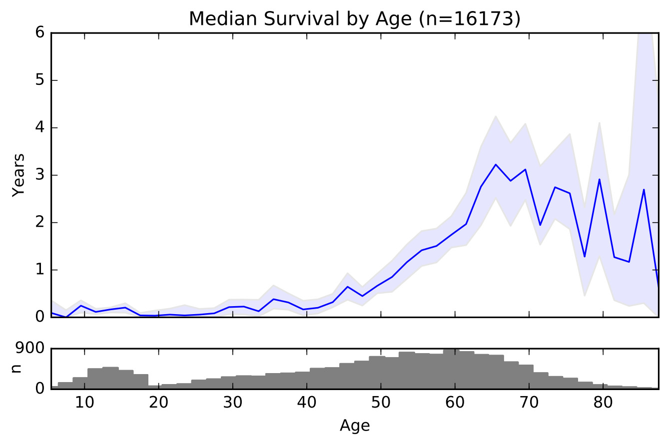

Fig. 4. Retirement age participants are the largest group and participate for the longest. (Bin size=2 years)

Hypothesis 1 is strongly confirmed. Interestingly, the age with the highest retention (65-66) appears to be older than the peak signup age (59-60).

We also computed the median survival for participants who did not enter an age, to verify that the results could be generalized between the groups. If age was truly missing at random, we would expect the median survival for unknown age to be roughly equal to the survival for all known ages (253 days). Instead, we found that the median survival for participants who did not provide an age is 405 days, which is close to the survival for participants aged 53-54. A plausible explanation is that older participants are less comfortable providing their age, and are thus are even more over-represented in CoCoRaHS participation than these results show.

In Table 2, each identified gender is separately contrasted with accounts for whom a gender could not be automatically determined (e.g. group accounts). Compared to women and other accounts, men are much more likely to sign up for CoCoRaHS, and participate for 61 days longer (according to the model). The actual median survival for men is 444 days, while the median survival for women is 207 days.

While a full analysis of motivations for participation in CoCoRaHS is beyond the scope of this paper, one key question is whether the skews in age and gender are due to inherent interest in the task, or due to biases in the recruitment process (c.f. [31]). Upon account creation, CoCoRaHS participants are provided an opportunity to enter the method by which they were referred to the project. We analyzed this free-form text to determine which words were used most often.

Out of the 37,438 participants who entered referral information, the five most common words used were "NWS" (12.0%), "weather" (8.6%), "newspaper" (8.6%), "friend" (5.5%), and "NOAA" (4.9%). NOAA and NWS both refer to the U.S. National Weather Service, as do most instances of "weather". While not conclusive, this corresponds with CoCoRaHS' internal estimation that a large subset of active CoCoRaHS participants find the project through interaction with National Weather Service programs, such as in-person severe weather training or online weather analysis tools. Thus, it is likely that the largest driver behind CoCoRaHS participation is inherent interest in the domain and/or task, rather than relationships with existing contributors[7].

[7]Certainly, there may be demographic biases in NWS program participation that then transfer to CoCoRaHS. For what it's worth, CoCoRaHS regional coordinators (many of whom are NWS staff) are younger, but more likely to be male, than the average CoCoRaHS contributor.

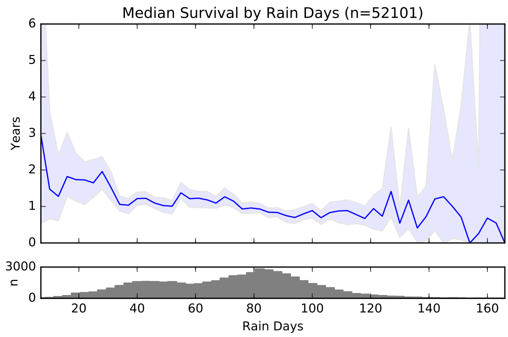

Fig. 5. Participants who experience more rain drop out a bit sooner. (Bin size=3 rainy days)

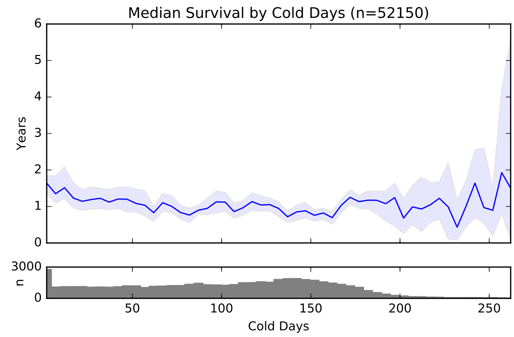

Fig. 6. The upward trend for cold days is less apparent in this chart, which does not account for state organizational differences. (Bin size=5 freezing days)

5.2 Task Characteristics

Contrary to our expectations, increased precipitation appears to be correlated with a slight decrease in retention. According to the model results, an additional 26 days of rain during the year before signup corresponds to a median dropout date 24 days earlier. This relationship bears out in Figure 5, which shows that the actual effect is even stronger when not adjusting for state differences. The median dropout for participants who experienced 60-62 rainy days is 448 days, while the median dropout for participants who experienced 90-92 rainy days is 274 days. This contradicts Hypothesis 2, which proposed that more exposure to rain would improve retention.

Also contrary to our expectations, more cold weather does not appear to negatively correlate with retention. According to the model, 57 additional freezing days corresponds to a moderate *increase* in retention of 43 days. Interestingly, when running the model without controlling for U.S. state, the effect for freezing days is to *decrease* retention by 17 days, which is more in line with Hypothesis 3.

To further understand this discrepancy, we wanted to explore if the state-normalized result was being skewed by one or two geographically unique states that just happened to have more registered CoCoRaHS accounts. To check this, we analyzed the effect of freezing days on retention outcomes in each of the 48 continental states. Within each state, we split participants into those who had more or less than the average freezing days for all participants in that state. For the sake of completeness, we performed the same calculation for rain. We used a logrank test to compare the statistical significance between each pair of populations.

Table 4. Between-State Differences (n=51969)

Statistical Effect	States	% Volunteers[8]
Cold $\Uparrow$ Retention	CO,TX,NC,KS,NY,OH,NJ,NH,VT	37%
Cold $\Downarrow$ Retention	MN,WY	4%
Cold Not Signif.	37 states	59%
Rain $\Downarrow$ Retention	NC,FL,TN,MO,WA,GA,AL,IA	22%
Rain $\Uparrow$ Retention	NY,MN,AZ,ME,NH,VT	9%
Rain Not Signif.	34 states	69%

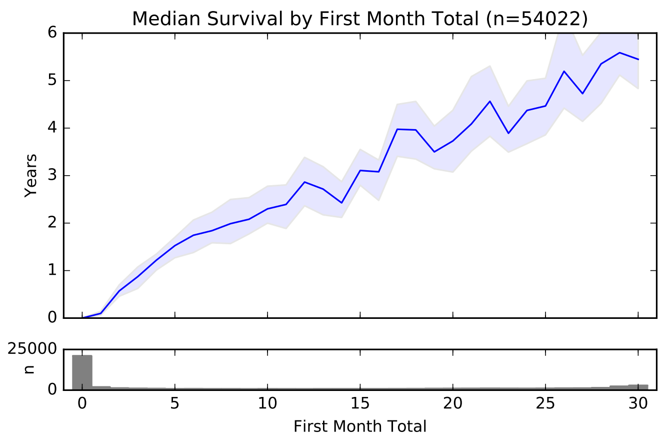

Fig. 7. Initial activity is highly predictive of retention. (Bin size=1 daily contribution)

The results are shown in Table 4. We found that in 9 states (representing 37% of CoCoRaHS participants[8]), participants in colder areas remain active longer. These include Colorado and Texas, the two most active CoCoRaHS states with over 5,000 registered accounts each. Only 2 states see an opposite trend: Minnesota and Wyoming. As it turns out, these two are among the coldest states in the U.S - though New Hampshire and Vermont are almost as cold. While more work is needed, it would appear that the positive effect of cold on retention only holds to a point. Still, the plurality of evidence points to freezing days being an indicator of improved retention, disconfirming Hypothesis 3.

5.3 Early Activity

CoCoRaHS participants submit an average of 10.6 records during their first month. As Figure 7 shows, the actual distribution is bimodal: the two largest groups are those who submit nothing during their first month (38.9%), followed by those who submit every day (5.2%). As is common with many collaborative projects, this early activity is a very strong predictor of eventual longevity. According to the model, an additional 11 contributions during the first month corresponds to an additional 4 years of participation in the program. In actuality, participants who do not contribute during their first month often never do, while the median survival for those who submitted reports for all 30 days is 5.4 years. Thus, Hypothesis 4 is confirmed.

5.4 Data Quality

We also measured the effect of each predictor on the three supplemental data quality metrics, to contrast with the primary metric of reliability measured as retention. The full results are listed in Table 5. Note that the goal of this research is to test the effect of the predictors on several alternative formulations of data quality, not to find the best fitting model per se[9]. Also, note that the coefficients

[8]The percentage of volunteers that live in one of the listed states. Note that this does not necessarily correspond to the actual population of each state.

[9]As discussed previously, the total number of contributions is a particularly strong signal for data quality in this domain. Since the number of contributions is directly tied to retention, the survival analysis itself is likely already the best model for measuring data quality. We defined the other three metrics as ratios to factor out the total number of contributions, to see if the primary results would hold up under alternative operationalizations of quality. Even with this factoring, participants who contribute more records overall will still necessarily score higher on the consistency metric, since there will be fewer

Table 5. Data Quality Linear Models (n=37491)

Predictor	Timeliness		Consistency		Accuracy	
First Month Observations	-2.39	****	16.16	****	0.93	****
% Submitted via Mobile App	1.41	****	-0.62	***	0.30	***
Age	1.48	***	3.08	****	0.78	***
Rain Days (year before signup)	1.58	***	0.45	*	0.13	
Freezing Days (year before signup)	-0.31		-0.09		0.28	**
Signup rank within State (log)	2.29	****	-2.11	****	-0.45	***
Had gauge at signup (or later)	-1.65	***	3.54	****	-0.16	
Daily internet access	2.62	***	0.80	**	-0.09	
Signup during March Madness	0.83	*	2.39	***	0.44	***
Female (guess from name)	0.55		2.45	***	0.47	**
Male (guess from name)	2.37	***	0.82	*	0.17	
First Month Obs. × % via Mobile	-0.40	**	0.46	***	-0.08	
Age × Female	0.16		-0.31		-0.40	**
Age × Male	0.07		-0.32		-0.25	
Age × First Month Obs.	1.50	***	-1.71	***	-0.21	***
Rain Days × Freezing Days	-0.10		0.47	**	0.07	
U.S. State	(fixed)	****	(fixed)	****	(fixed)	****
Linear model fit (Adjusted R^2)	*0.02*		*0.32*		*0.02*	
Mean score for metric	*69.9%*		*69.3%*		*96.9%*	

p-values: *<0.05 **<0.01 ***<0.001 ****<2e-16

for the three models are not directly comparable. For example, the effective range of the accuracy metric is about one fifth that of consistency and timeliness, since fewer than 4% of records are ever edited.

Interestingly, age positively correlates with retention / reliability *and* all three of the other quality metrics, though the effects for the latter were relatively small. Older participants are not only more reliable, but are also more likely to enter their data in a timely, consistent, and accurate manner. These effects are further explored in Figure 8. We calculated effect size by computing Cohen's d as $\frac{m2-m1}{S}$ where $m1$ and $m2$ the smoothed average scores for participants aged 39 and 63 (representing the 25th and 75th percentile ages, respectively) and S is the standard deviation of all scores. The effect sizes for consistency (d=0.32) and timeliness (d=0.21) are small while the effect for accuracy (d=0.13) is very small (c.f [48]).

The effects for other predictors are not as consistent as those for age. While men are somewhat more reliable and timely, women are more consistent and accurate in their reports. This illuminates the importance of understanding the multifaceted nature of data quality. There were generally limited interaction effects between age and gender, other than minimal evidence that older women

gaps between submissions. Thus, the consistency model has the best fit, since the total contribution signal shows up both in the predictors and (indirectly) in the dependent variable.

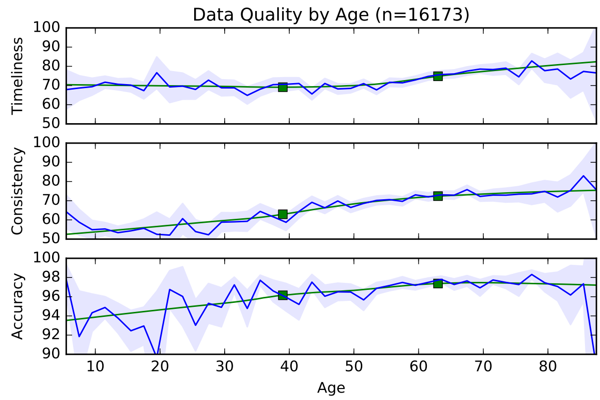

Fig. 8. Older participants are more timely, consistent, and accurate in their data entry, though the effect is small. Blue is the binned average scores, while green shows a Lowess curve with zero iterations. Green squares represent the 25th and 75th percentiles used to compute Cohen's d.

are slightly more reliable and slightly less accurate than would be expected from their age and gender separately.

While participants who experience more rain drop out slightly sooner, they are more timely in their reports (d=0.31 for participants with 89 versus 50 days of rain). It is likely that rain prompts a more immediate response, while those who get no rain may consistently report the 0 but not in a timely manner.

As expected, there is a high correlation between the number of records entered during the first month and overall consistency for the first year. The effect size is large (d=1.13 for participants with 25 versus 3 submissions). However, timeliness is somewhat decreased for the highest contributors (d=-0.12). Thus, Hypothesis 5 is not fully confirmed.

6 DISCUSSION

6.1 Participant Characteristics

There is a definite skew toward older participants in CoCoRaHS, not only in terms of initial sign-up rates, but also in terms of retention and (to a lesser extent) other measures of data quality. This is in line with our first hypothesis. While this finding may not be surprising to those familiar with CoCoRaHS and similar observational citizen science programs, it is relatively unheard of in the broader domains of user-generated content and peer production. For example, Wikipedia is heavily skewed toward younger contributors, with half of survey respondents being younger than 22 [17]. Within volunteerism research, the general consensus is that that volunteer activity peaks at age 45 [37].

With this in mind, what makes CoCoRaHS different than Wikipedia and many other forms of volunteerism? We propose three key mechanisms that may be responsible for this finding:

- First, CoCoRaHS is designed to be incorporated as a daily routine, and the strength of habit-forming is known to increase with age [53]. Several elderly participants (and their spouses) have sent messages to CoCoRaHS thanking the project for providing a stable routine and a reason to get up early every morning (c.f. [5]). Many elderly CoCoRaHS participants remain active until they are physically unable to check their gauge.

- Second, CoCoRaHS is structured around repeated monitoring at a single location, usually in the participant's own backyard. From a purely practical standpoint, this means it is more accessible to individuals with a backyard, and those who spend more time at home.
- Third, the task itself is relatively intuitive and structured around a topic of immediate general interest and applicability - local rainfall.

We suggest that further quantitative research is needed to measure the effect of different task designs on demographic interest in participation in peer production systems and citizen science. With regard to age distribution, Wikipedia and CoCoRaHS appear to be opposite ends of a spectrum, while Galaxy Zoo appears to be near the middle (c.f. [11]). We predict that the peak retention age in eBird is older than Galaxy Zoo, but younger than CoCoRaHS, given that birding often requires travel.

While increased diversity is a goal, CoCoRaHS' skew toward older participants could also be seen as an opportunity - e.g., to promote broader scientific literacy among an influential demographic.

6.2 Task Structure

More rainy days correspond with slightly decreased retention, while more freezing days correspond with a moderate increase, meaning Hypotheses 2 and 3 were not supported. The result for freezing days is particularly counterintuitive, as we intended it to be a measure of task complexity. As Dittus et al. and others have found, task complexity is generally known to negatively affect retention [12]. One plausible explanation is that the complexity of the CoCoRaHS task is well known and relatively stable. Participants from moderately cold locations who decide to join CoCoRaHS are already likely highly motivated and self-selecting, which would correspond to our findings related to early activity. It would be helpful for future work to quantify the geographic disparity in CoCoRaHS signups - both in terms of population and in terms of climatology.

6.3 Early Activity

First month contributions are this strongest predictor of long-term retention, which confirms Hypothesis 4 and replicates prior work on Wikipedia [36]. A key follow-up question, then, is whether early interactions with a project extrinsically influence retention, or whether both early activity and retention merely reflect existing intrinsic participant motivation. In Wikipedia, at least, the latter, "intrinsic" view seems to be more applicable [36].

Indeed, the majority of CoCoRaHS participants appear to join with an existing interest in the project, and early intervention efforts to date have shown mixed results. In addition, CoCoRaHS staff report encountering certain "personality types" that simply enjoy the daily routine and data management aspects of the project - sometimes even without a particular interest in the weather. With this in mind, one potential application of this research might be to recruit a wide range of participants and then focus follow up efforts on the highest-contributing participants.

However, this approach is complicated by the importance of diversity and inclusivity to the goals of CoCoRaHS and other citizen science projects. If project resources are devoted only to engaging the most active contributors, existing biases may become more entrenched and opportunities for growth may be missed (c.f. [31]). Broader recruitment activity can help expand the project's demographic base, but it may be just as important to understand why certain participants leave, as why they are less likely to sign up. In addition, research on Wikipedia has shown that there is value in using targeted interventions to promote the retention of newcomers [18].

The requirement of a rain gauge is a potential barrier to entry in CoCoRaHS, which is why many citizen science projects do not require any equipment at all [51]. On the other hand, CoCoRaHS staff note that asking participants to spend at least the cost of shipping on a gauge can help participants gain a sense of commitment that they might not have if the gauge is given for free.

It is still possible to promote early data entry while waiting for equipment to arrive. CoCoRaHS staff occasionally encourage participants to submit their initial record on a dry day (since the value is sure to be zero). However, there is no place in the CoCoRaHS database for rain measurements made without an official gauge. As an alternative, CoCoRaHS has an informal relationship with mPING, an independent mobile app for reporting precipitation that does not require an established site or any custom equipment [13]. Participants unable to commit to the full CoCoRaHS protocol can start by contributing to mPING instead.

We suggest that these types of partnerships provide a rich opportunity to engage a broad range of participants in the projects that best suit their interests and abilities. In addition, centralized volunteer recruitment platforms like SciStarter [20] can facilitate these relationships while also tracking participant demographics and activity across multiple citizen science projects.

6.4 Data Quality

While age consistently correlated with higher scores on all quality metrics, first month activity did not, so Hypothesis 5 was not fully supported. This may in part be due to the challenges in operationalizing data quality; our efforts to factor out the strongest signal (number of contributions) caused the remaining metrics to be quite noisy. Nevertheless, some interesting patterns arose from the data. Participants who contributed more data were more consistent but less timely than other participants. There is likely a moderate subset of users who consistently make observations every day but only occasionally enter them in the website *en masse*. This would confirm previous work noting that data entry is not considered a desirable task by many citizen science participants [51].

7 LIMITATIONS & FUTURE WORK

This project focused only on predictors that could be determined within one month of signup. This simplified the model at the expense of some practical applicability. In future work, it would be valuable to use time-dependent covariates: e.g. how likely are you to drop out this month, given the number of rainy days last month? It would also be informative to examine the 10% or so of participants who did not contribute anything during their first month, but later became active (perhaps after obtaining a rain gauge). Further, while median retention was useful as a comparative metric, it also masked a large variability in individual outcomes.

This project focused only on a descriptive analysis of existing data, and our hypotheses were not fully formed prior to starting the exploratory analysis. A next step to continue this work would be to experimentally evaluate the effect of one or more interventions on retention and data quality. Another next step would be to interview a number of CoCoRaHS participants to shed more light on the differences we found in activity levels.

To simplify our analysis, we assumed that each CoCoRaHS reporting station was run by a single observer, accounting for group accounts only in our treatment of gender. In fact, a small but sizeable subset of accounts belong to school teams and other groups. Future work could examine other indicators of group accounts and evaluate differences between groups and individuals. In addition, we did not account for the possibility of participants remaining active after moving to another house, which would require establishing a new monitoring site and account.

Finally, we focused on factors that predict individual outcomes, without fully measuring the effect of organizational structure on retention. We included a term for U.S. state, but only as a fixed effect. In future work, it would be valuable to quantitatively measure the effect of specific organizational factors and recruitment strategies.

8 CONCLUSION

As is often the case for citizen science, the implications of this study depend on the goals of each particular project. Some predictors of retention are amenable to intervention, while others are not. But more fundamentally, there is a potential tension between different intervention strategies. If robust data collection is prioritized, a project might focus on recruiting and retaining participants who demonstrate an ability to remain active long term. On the other hand, if educational and inclusion goals are prioritized, projects might focus recruitment and intervention strategies toward students and under-represented groups. Projects that focus on both goals (like CoCoRaHS) will need to carefully weigh the benefits of each approach.

This project calls into question the assumption that crowdsourcing output is always dominated by younger contributors. It is clear that more comparative work is needed to determine what types of projects are more likely to attract older volunteers. Together with the rest of the CSCW community, we look forward to mapping out the full spectrum of task types and participant interests in observational citizen science and beyond.

REFERENCES

[1] Denise Anthony, Sean W Smith, and Tim Williamson. 2005. Explaining quality in internet collective goods: Zealots and good samaritans in the case of Wikipedia. *Hanover: Dartmouth College* (2005).

[2] Avinoam Baruch, Andrew May, and Dapeng Yu. 2016. The motivations, enablers and barriers for voluntary participation in an online crowdsourcing platform. *Computers in Human Behavior* 64 (2016), 923–931.

[3] Christopher Beirne and Xavier Lambin. 2013. Understanding the Determinants of Volunteer Retention Through Capture-Recapture Analysis: Answering Social Science Questions Using a Wildlife Ecology Toolkit. *Conservation Letters* 6, 6 (2013), 391–401.

[4] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, and Jennifer Preece. 2013. Using gamification to inspire new citizen science volunteers. In *Proceedings of the first international conference on gameful design, research, and applications*. ACM, 18–25.

[5] Dana Burr Bradley. 1999. A reason to rise each morning: The meaning of volunteering in the lives of older adults. *Generations* 23, 4 (1999), 45.

[6] Robin Brewer, Meredith Ringel Morris, and Anne Marie Piper. 2016. Why would anybody do this?: Understanding Older Adults' Motivations and Challenges in Crowd Work. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2246–2257.

[7] S van Buuren and CGM Oudshoorn. 2000. *Multivariate imputation by chained equations: MICE V1. 0 user's manual.* Technical Report. TNO.

[8] Susan M Chambré. 1993. Volunteerism by elders: Past trends and future prospects. *The Gerontologist* 33, 2 (1993), 221–229.

[9] Alice Ming Lin Chong, Tina Louisa Rochelle, and Susu Liu. 2013. Volunteerism and positive aging in Hong Kong: A cultural perspective. *The International Journal of Aging and Human Development* 77, 3 (2013), 211–231.

[10] Ram A Cnaan and Toni A Cascio. 1998. Performance and commitment: Issues in management of volunteers in human service organizations. *Journal of Social Service Research* 24, 3-4 (1998), 1–37.

[11] Joe Cox, Eun Young Oh, Brooke Simmons, Gary Graham, Anita Greenhill, Chris Lintott, Karen Masters, et al. 2016. Doing Good Online: An Investigation into the Characteristics and Motivations of Digital Volunteers. In *12th International Conference of the International Society for Third Sector Research.*

[12] Martin Dittus, Giovanni Quattrone, and Licia Capra. 2016. Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 108–118.

[13] Kimberly L Elmore, ZL Flamig, V Lakshmanan, BT Kaney, V Farmer, Heather D Reeves, and Lans P Rothfusz. 2014. mPING: Crowd-sourcing weather reports for research. *Bulletin of the American Meteorological Society* 95, 9 (2014), 1335–1342.

[14] Cameron Davidson-Pilon et al. 2017. CamDavidsonPilon/lifelines. (June 2017). https://doi.org/10.5281/zenodo.815943

[15] Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L Cox. 2014. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2985–2994.

[16] Paul Fugelstad, Patrick Dwyer, Jennifer Filson Moses, John Kim, Cleila Anna Mannino, Loren Terveen, and Mark Snyder. 2012. What makes users rate (share, tag, edit...)?: predicting patterns of participation in online communities. In

Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM, 969–978.

[17] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. 2010. Wikipedia survey–overview of results. *United Nations University: Collaborative Creativity Group* (2010).

[18] Aaron Halfaker, R Stuart Geiger, Jonathan T Morgan, and John Riedl. 2013. The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist* 57, 5 (2013), 664–688.

[19] Aaron Halfaker, Aniket Kittur, Robert Kraut, and John Riedl. 2009. A jury of your peers: quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. ACM, 15.

[20] Catherine Hoffman, Caren B Cooper, Eric B Kennedy, Mahmud Farooque, and Darlene Cavalier. 2017. SciStarter 2.0: A Digital Platform to Foster and Study Sustained Engagement in Citizen Science. In *Analyzing the Role of Citizen Science in Modern Research*. IGI Global, 50–61.

[21] Caroline Jay, Robert Dunne, David Gelsthorpe, and Markel Vigo. 2016. To Sign Up, or not to Sign Up?: Maximizing Citizen Science Contribution Rates through Optional Registration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1827–1832.

[22] Raghav Pavan Karumur, Tien T Nguyen, and Joseph A Konstan. 2016. Early Activity Diversity: Assessing Newcomer Retention from First-Session Activity. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 595–608.

[23] Masatomo Kobayashi, Shoma Arita, Toshinari Itoko, Shin Saito, and Hironobu Takagi. 2015. Motivating multi-generational crowd workers in social-purpose work. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1813–1824.

[24] Kathrin Komp, Theo Van Tilburg, and Marjolein Broese Van Groenou. 2012. Age, retirement, and health as factors in volunteering in later life. *Nonprofit and Voluntary Sector Quarterly* 41, 2 (2012), 280–299.

[25] Anne M Land-Zandstra, Jeroen LA Devilee, Frans Snik, Franka Buurmeijer, and Jos M van den Broek. 2016. Citizen science on a smartphone: Participants' motivations and learning. *Public Understanding of Science* 25, 1 (2016), 45–60.

[26] Jeffrey Laut, Francesco Cappa, Oded Nov, and Maurizio Porfiri. 2017. Increasing citizen science contribution using a virtual peer. *Journal of the Association for Information Science and Technology* 68, 3 (2017), 583–593.

[27] Wendy Liu and Derek Ruths. 2013. What's in a Name? Using First Names as Features for Gender Inference in Twitter.. In *AAAI spring symposium: Analyzing microtext*, Vol. 13. 01.

[28] Ding Ma, Mats Sandberg, and Bin Jiang. 2015. Characterizing the heterogeneity of the OpenStreetMap data and community. *ISPRS International Journal of Geo-Information* 4, 2 (2015), 535–550.

[29] Andrew Mao, Ece Kamar, and Eric Horvitz. 2013. Why stop now? predicting worker engagement in online crowdsourcing. In *First AAAI Conference on Human Computation and Crowdsourcing*.

[30] Adina M Merenlender, Alycia W Crall, Sabrina Drill, Michelle Prysby, and Heidi Ballard. 2016. Evaluating environmental education, citizen science, and stewardship through naturalist programs. *Conservation Biology* 30, 6 (2016), 1255–1265.

[31] Hannah J Miller, Shuo Chang, and Loren G Terveen. 2015. I LOVE THIS SITE! vs. It's a little girly: Perceptions of and Initial User Experience with Pinterest. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1728–1740.

[32] Oded Nov, Ofer Arazy, and David Anderson. 2011. Technology-Mediated Citizen Science Participation: A Motivational Model.. In *Proceedings of the International AAAI Conference on Web and Social Media*.

[33] Morris A Okun and Amy Schultz. 2003. Age and motives for volunteering: testing hypotheses derived from socioemotional selectivity theory. *Psychology and Aging* 18, 2 (2003), 231–239.

[34] Allen M Omoto, Mark Snyder, and Steven C Martino. 2000. Volunteerism and the life course: Investigating age-related agendas for action. *Basic and applied social psychology* 22, 3 (2000), 181–197.

[35] Felipe Ortega and Daniel Izquierdo-Cortazar. 2009. Survival analysis in open development projects. In *ICSE Workshop on Emerging Trends in Free/Libre/Open Source Software Research and Development*. IEEE, 7–12.

[36] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, 51–60.

[37] Jone L Pearce. 1993. *Volunteers: The organizational behavior of unpaid workers*. Routledge.

[38] Lesandro Ponciano, Francisco Brasileiro, Robert Simpson, and Arfon Smith. 2014. Volunteers' Engagement in Human Computation for Astronomy Projects. *Computing in Science & Engineering* 16, 6 (2014), 52–59.

[39] PRISM Climate Group, Oregon State University. 2017. (2017). http://prism.oregonstate.edu created 11 Mar 2017.

[40] M Jordan Raddick, Georgia Bracey, Pamela L Gay, Chris J Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. 2013. Galaxy Zoo: Motivations of citizen scientists. *Astronomy Education Review* 12, 1 (2013).

[41] Jason Reed, M Jordan Raddick, Andrea Lardner, and Karen Carney. 2013. An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. In *46th Hawaii International Conference on System Sciences*. IEEE, 610–619.

[42] Henry W Reges, Nolan Doesken, Julian Turner, Noah Newman, Antony Bergantino, and Zach Schwalbe. 2016. COCORAHS: The evolution and accomplishments of a volunteer rain gauge network. *Bulletin of the American Meteorological Society* 97, 10 (2016), 1831–1846.

[43] Joel Ross, Lilly Irani, M Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI'10 extended abstracts on Human factors in computing systems*. ACM, 2863–2872.

[44] Dana Rotman, Jen Hammock, Jenny Preece, Derek Hansen, Carol Boston, Anne Bowser, and Yurong He. 2014. Motivations affecting initial and long-term participation in citizen science projects in three countries. *iConference 2014 Proceedings* (2014).

[45] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. 2012. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 217–226.

[46] Robert L Ryan, Rachel Kaplan, and Robert E Grese. 2001. Predicting volunteer commitment in environmental stewardship programmes. *Journal of Environmental Planning and Management* 44, 5 (2001), 629–648.

[47] Henry Sauermann and Chiara Franzoni. 2015. Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences* 112, 3 (2015), 679–684.

[48] Shlomo S Sawilowsky. 2009. New effect size rules of thumb. *Journal of Modern Applied Statistical Methods* 8, 2 (2009), 597–599.

[49] Avi Segal, Ya'akov Gal, Ece Kamar, Eric Horvitz, Alex Bowyer, and Grant Miller. 2016. Intervention Strategies for Increasing Engagement in Crowdsourcing: Platform, Predictions, and Experiments. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*. AAAI Press, 3861–3867.

[50] S Andrew Sheppard and Loren Terveen. 2011. Quality is a verb: the operationalization of data quality in a citizen science community. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. ACM, 29–38.

[51] S Andrew Sheppard, Andrea Wiggins, and Loren Terveen. 2014. Capturing quality: retaining provenance for curated volunteer monitoring data. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1234–1245.

[52] Brian L Sullivan, Jocelyn L Aycrigg, Jessie H Barry, Rick E Bonney, Nicholas Bruns, Caren B Cooper, Theo Damoulas, André A Dhondt, Tom Dietterich, Andrew Farnsworth, et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. *Biological Conservation* 169 (2014), 31–40.

[53] Viswanath Venkatesh, James Y. L. Thong, and Xin Xu. 2012. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Q.* 36, 1 (March 2012), 157–178.

[54] Yair Wand and Richard Y Wang. 1996. Anchoring data quality dimensions in ontological foundations. *Commun. ACM* 39, 11 (1996), 86–95.

[55] Andrea Wiggins and Kevin Crowston. 2011. From conservation to crowdsourcing: A typology of citizen science. In *44th Hawaii international conference on System Sciences*. IEEE, 1–10.

[56] Andrea Wiggins and Yurong He. 2016. Community-based data validation practices in citizen science. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1548–1559.

[57] Andrea Wiggins, Greg Newman, Robert D Stevenson, and Kevin Crowston. 2011. Mechanisms for data quality and validation in citizen science. In *e-Science Workshops (eScienceW), 2011 IEEE Seventh International Conference on*. IEEE, 14–19.

[58] John Wilson. 2012. Volunteerism research: A review essay. *Nonprofit and Voluntary Sector Quarterly* 41, 2 (2012), 176–212.

[59] Ilene Wolcott, Dean Ingwersen, Michael A Weston, Chris Tzaros, et al. 2008. Sustainability of a long-term volunteer-based bird monitoring program: recruitment, retention and attrition. *Australian Journal on Volunteering* 13, 1 (2008), 48.

[60] Chris Wood, Brian Sullivan, Marshall Iliff, Daniel Fink, and Steve Kelling. 2011. eBird: engaging birders in science and conservation. *PLoS Biol* 9, 12 (2011), e1001220.

[61] Taha Yasseri, Giovanni Quattrone, and Afra Mashhadi. 2013. Temporal analysis of activity patterns of editors in collaborative mapping project of OpenStreetMap. In *Proceedings of the 9th International Symposium on Open Collaboration*. ACM, 13.

[62] Dell Zhang, Karl Prior, and Mark Levene. 2012. How long do Wikipedia editors keep active?. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*. ACM, 4.

[63] Haiyi Zhu, Robert E Kraut, and Aniket Kittur. 2014. The impact of membership overlap on the survival of online communities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 281–290.