MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-333

Quantifying the Activities of Self-quantifiers: Management of Data, Time and Health

Manal Almalki[a], Fernando Martin Sanchez[a], Kathleen Gray[a]

[a] *Health and Biomedical Informatics Centre, University of Melbourne, Melbourne, 3010, Australia*

Abstract

Current self-quantification systems (SQS) are limited in their ability to support the acquisition of health-related information essential for individuals to make informed decisions based on their health status. They do not offer services such as data handling and data aggregation in a single place, and using multiple types of tools for this purpose complicates data and health self-management for self-quantifiers. An online survey was used to elicit information from self-quantifiers about the methods they used to undertake key activities related to health self-management. This paper provides empirical evidence about self-quantifiers' time spent using different data collection, data handling, data analysis, and data sharing tools and draws implications for health self-management activities.

Keywords:

Self-quantification; health self-management; participatory health; consumer health informatics; patient activation.

Introduction

A self-quantification system (SQS) is a tool for capturing personal data and an application for processing that data (e.g., analysing and visualising) according to an individual's objectives. SQS have the potential to track and measure various health aspects, such as sleep, weight, diet, and physical activity, and may offer opportunities for self-quantifiers to change behaviours and attitudes toward health self-care. In one study [1], 69% of US adults said that they tracked themselves, and 46% of those reported changing their behaviour based on the collected data.

There are two types of SQS: A primary SQS collects data directly from the users; health-related examples are MoodPanda for tracking mood and iBGStar for monitoring blood glucose. A secondary SQS can aggregate data delivered by multiple primary tools or apps, to facilitate data analysis; TicTrac, BodyTrack, and Argus are health-related examples [2].

Self-quantification for health consists of two main stages of activities in which a self-quantifier not only collects personal health measurements, but also manages and transform these data into actionable knowledge [2]. The first stage, data management, involves a chain of activites that are necessary to establish, use, and maintain a mapping between the individual's objectives and the information available. The second stage is health management; here the knowledge obtained from self-quantification is activley transformed into the desired health outcomes, a process that may be called health activation [3]. In these activities, the person attempts to establish self-awareness of one's health and functional status to actively engage with maintaining or improving it. Aggregating data from different devices and apps is a major facilitator in gaining a holistic view of one's health-related indicators [4].

To undertake this chain of activities for health data self-management, a self-quantifier may use multiple unconnected primary and secondary SQSs, and also ancillary tools – such as iPhone note apps, Microsoft Excel spreadsheet, or Dropbox. With their data and services scattered among different tools self-quantifiers may devote significant time for these activities. The aim of this paper is to characterise the experience of using combinations of such tools, particularly its impact on self-quantifiers' time and data, and to draw implications for health self-management and health activation.

Methods

Sample and demographic characteristics

An international online survey was conducted from December 2013 to March 2014 to elicit information from adult self-quantifiers, about the methods they used to perform key activities related to health data self-management - data collection, handling, analysis and aggregation, and sharing. Participants were recruited through Quantified Self meetup forums, Twitter, Facebook, and LinkedIn. People who used one or more self-quantification tools or apps as part of their health self-care were included. 103 individuals provided sufficient information for analysis.

Respondents from USA were the highest proportion (60%), followed by Australia (11%). Nearly two thirds (62%) were aged between 20 and 39 years. 75% were male. Almost one third (32.3%) had completed high school or equivalent; the rest had at least a university undergraduate degree. 68% reported good to very good health status. The top three motivations to use SQS were: to know if a certain health-related variable could affect another variable (64%); to find answers to specific questions related to health (62%); to proactively minimise possible future health problems (61%). The Eurostat's ICT (information and communication technology) model [5] was used to classify respondents into high (53%), medium (39%), and low (8%) level of ICT skills.

From the surveyed self-quantifiers, 89% used at least one tangible tracking device for self-quantification whereas 11% used only apps. The most popular primary tangible self-quantification devices were for tracking and quantifying physical activities, sleep, weight, calories burned, and diet: the Fitbit collection (e.g., Fitbit Ultra, Fitbit One, Fitbit Zip, etc.) with nearly 25% of the respondents; followed by Withings tools (e.g., Withings scale, and Withings activity tracker) with 19%; and Jawbone Up with about 12%. The most popular primary tracking apps were 80Bites (for tracking food consumption), 42Goals (for tracking personal life goals such as quit smoking, lose weight, and reduce coffee consumption), and Moves (records walking, cycling, and running activities through step counter), among approximately 13%, 11%, and 9% of respondents respectively.

The most popular secondary self-quantification tool was BodyTrack (20% of respondents). In addition, two types of ancillary tools were used for collecting data and for handling the generated datasets. The most popular ancillary tools for data collection were diary apps such as Evernote, Microsoft One Note, and iPhone note apps (30% of respondents). Most popular tools for data handling were cloud storage services such as Dropbox, SkyDrive, and Google Drive, collectively used by nearly one quarter (24%) of respondents.

Statistical analysis

Data analysis to address the specific aim of this paper focused on responses to survey questions about the amount of time spent to undertake the key activities of health data self-management (i.e., data collection, handling, analysis and aggregation, and sharing). The survey devoted separate sections to each of the key activities where the participants were asked questions about the type of tools they used (i.e., primary, secondary, or ancillary); their experience in using these tools (i.e., less than 6 months, or 6 months or more); the frequency of tool usage (i.e., daily, weekly, or monthly) for these activities; the number of collected data types (such as blood pressure, weight, mood, coffee consumption, etc.); the methods used to collect these data (i.e., manual or automatic); and their ICT skills level. These questions represent a set of explanatory variables that were fitted in testing the statistical models.

General linear model for continuous outcomes, and logistic regression for binary outcomes, were conducted by using SPSS 22 Mac version. To construct these models, a manual backwards selection approach was used. We started with set of candidate explanatory variables and eliminated variables with large p-values one at a time. We stopped once all the remaining variables had relatively small p-values. It is worth noting that building these statistical models did not use only a simple analysis of concepts; rather, it involved exploring a more complex concept through testing for a two-way interaction. This two-way interaction is often thought of as a relationship between an explanatory variable and outcome variable that is moderated by a third explanatory variable. Once a two-way interaction between two varables was found, the main effects of the explanatory variables are not discussed given the presence of these interactions. Bootstrapping results were examined to check the robustness of our models. Also, SPSS 22 was used to run descriptive statistics, frequencies, and cross-tabulations to determine related demographics.

Due to the sample size limitation , we combined the two categories of medium and low ICT skills into one group (low to medium level). Also, the ICT skills level variable could not be fitted with other explanatory variables in a single logistic regression model. Adding this variable increases the number of categorical variables in the model which can result in empty cells and problems in model fitting. Therefore, two logistic regression models were built for these cases, one based on the ICT skills variable and the other including all other variables.

Results

Combined use of primary, secondary, and ancillary tools

No survey respondents were using only primary SQS for health data self-management. Nearly a third used all types of tools, i.e. primary, secondary, and ancillary. Nearly 80% used three primary tools to collect health related data; however, 51% stated that their primary tools did not offer the kind of data they needed and therefore had to use ancillary tools to collect the desired data.

75% of respondents used another type of ancillary tool for handling the collected data from these primary SQS, for example, to organise the collected data into folders, import or export data, update data or files. 70% of respondents were performing most of the related tasks manually. Thus, over two thirds (68%) had lost files permanently, and over three quarters (77%) had failed to locate old files.

75% of respondents used secondary SQS to aggregate the data generated from the primary SQS. In over a third (34%) of these cases the secondary tools did not connect automatically with the primary tools, requiring users to upload the datasets manually into the secondary SQS.

Time devoted to data collection

Survey respondents who used multiple primary and ancillary tools to collect the required data spent more time on average than those who used fewer tools. The data collection time was calculated by counting the cases that took over 10 minutes to acquire the needed data using these tools. For representation simplicity, we refer to this count with the symbol CTM (Count of Taking more than 10 Minutes). To measure the impact of using these tools on CTM, we identified a set of factors (explanatory variables) that had a statistically significant effect on the mean of CTM. These factors were: the number of primary and ancillary tools used for data collection; the number of collected data types; the method used to collect these data; and the ICT skills level. Three two-way interactions between these variables existed in this model (see Table 1 model number 1). The first two-way interaction was between the number of tools used and the number of data types collected (by the primary and ancillary tools used). The second interaction was between the number of data types collected and the method used to collect these data(manual or automatic). The last interaction was the number of collected data types and the ICT skills level. For reporting the results of these two-way interactions, the main effects of the explanatory variables that involevd in these interaction are not presented in the Table 1.

In a general linear model with the set of these factors and their interactions (Table 1 model 1), the effect of the number of tools on the mean of CTM was greater for users who collected relatively fewer types of data. For example, if the number of data types was five, and three tools (lower quartile) were used to collect these data, then the average CTM was 2.3; in comparison, adding one more tool (upper quartile) to collect these five data types increased the mean of CTM to 2.8 (Figure 1-A). The mean of CTM increased with the rise in the number of data collected manually. For example, if the number of data types was eleven and all were collected automatically (lower quartile), then the mean of CTM to collect these data was 1.7; however, if two (upper quartile) out of these eleven data types were collected manually, then the mean of CTM activity was 2.9 (see Figure 1-B). Similarly, the mean difference in time taken for data collection between users with low to medium ICT skills and users with high ICT skills was statistically significant (mean difference=1.65). On average, for collecting five types of data, people with high level ICT skills took less time (0.4) relative to those with low to medium ICT skills (2.4) (Figure 1-C).

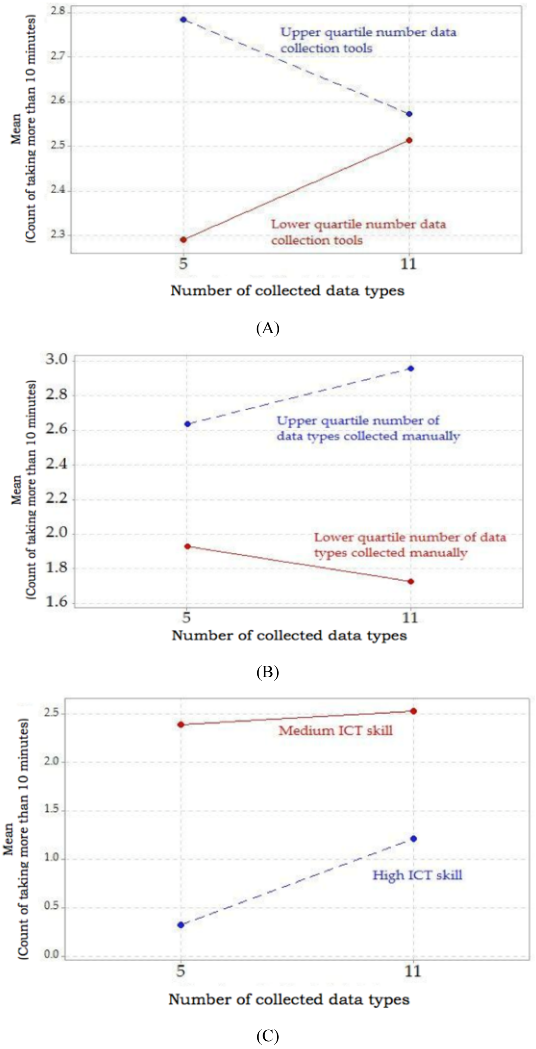

Figure 1 - Effects of three different two-way interactions on the mean of count of taking more than 10 minutes

Time devoted to data handling

Using ancillary tools for data handling increased the time needed to complete the activity. To measure the tools' impact, we identified three factors that had a statistically significant effect on the odds ratio of the duration of data handling. These factors were: using ancillary tools for data handling; the frequency the tool usage; and ICT skills level. Two models were built to examine this, due to model fitting problems as explained previously. One model included the first two factors and the other included the ICT skills level.

In the first logistic regression model, the odds of taking more than 10 minutes to complete the data handling activity were statistically significant - six times higher among users of ancillary tools relative to non-users (Table 1 model 2). With growing data volume, among people who used the tool less frequently (on a weekly or monthly basis), the odds of a longer duration of data handling activity were four times higher than among those who used the tool on a daily basis.

In the second logistic regression model (Table 1 model 3), the effect of having low to medium ICT skills on the odds of the data handling duration was significant - 18 times higher compared to users with high ICT skills.

Time devoted to data aggregation and analysis

Self-quantifiers who used secondary SQS took longer times to perform this activity. To measure the impact of using these tools, we identified four factors that had statistically significant effects on the odds ratio of the longer duration of data analysis. These factors were: using secondary SQS; the number of collected data types; the method used to collect these data manually; and ICT skills level. There was a two-way interaction between the number of collected data types and the method used to collect the data. These variables were fitted in two models (Table 1 model 4). One model included the first three factors and the two-way interaction, and the other included the ICT skills level.

In the first logistic regression model, the odds of taking more than 10 minutes to complete this activity were five times higher among secondary SQS users relative to the non-users. People who manually collected multiple types of data took a longer time to analyse these data. The longer duration odds increased with each additional type of data collected manually (see Figure 2). For example, if the number of data types was eleven (third quartile) and all were collected automatically (Manual data collection=0), the odds ratio of data analysis duration (taking more than 10 minutes) was 2.04; in comparison, if two out of these eleven data types were collected manually (Manual data collection=2), then the odds ratio was higher (Odds ratio=6.79) (Figure 2).

In the second logistic regression model (Table 1 model 5), the effect of having low to medium ICT skills on the data analysis duration odds was significant 17.6 times higher compared to users with high ICT skills.

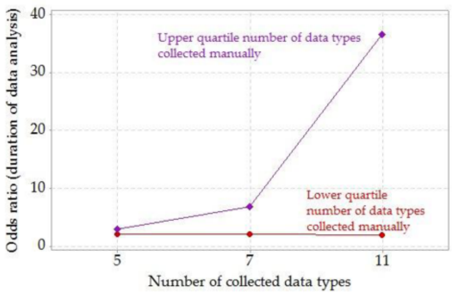

Figure 2 – Effect of the number of manually collected data types on the odds ratio of data analysis duration

Table 1 – Results from statistical models

Factors	Description of Estimate	Estimate Value	*P*	95% CI
Model (1): General linear model for the duration of data collection activity				
First interaction	Difference in regression slope between 'number of data types' and 'number of tools used to collect these data'	-0.073	0.012	-0.129, -0.017
Second interaction	Difference in regression slope between 'number of data types' and 'method used to collect these data (Manual)'	0.044	0.041	0.002, 0.086
Third interaction	Difference in regression slope between group with 'high' ICT skills and group with 'low to medium' ICT skills	0.126	0.024	0.017, 0.236
Model (2): Multiple binary logistic regression model for the duration of data handling activity				
Use ancillary tool	Odds ratio of taking more than 10 minutes in users compared to non-users groups	6.368	0.001	2.268, 17.876
Frequency of use	Odds ratio of taking more than 10 minutes in users who used this tool less frequently (weekly or monthly) compared to frequent users (daily)	4.061	0.004	1.556, 10.600
Model (3): Simple binary logistic regression model for the duration of data handling activity				
ICT skills	Odds ratio of taking more than 10 minutes in users with low to medium ICT skills compared to high level users	18.000	< 0.001	4.987, 64.971
Model (4): Multiple binary logistic regression model for the duration of data analysis activity				
Use secondary SQS	Odds ratio of taking more than 10 minutes in users compared to non-users	5.362	0.001	1.905, 15.091
Number of manually collected data (interaction)	Odds ratio of taking more than 10 minutes with increase in the number of data types collected manually	1.244	0.016	1.042, 1.486
Model (5): Simple binary logistic regression model for the duration of data analysis activity				
ICT skills	Odds ratio of taking more than 10 minutes for users with low to medium ICT skills compared to high level users	17.600	< 0.001	4.872, 63.575
Model (6): Multiple binary logistic regression model for the duration of data sharing activity				
Use secondary SQS	Odds ratio of taking more than 10 minutes in users compared to non-users	31.829	0.003	3.282, 308.7
Number of data collection tools (a)	Odds ratio of taking more than 10 minutes when (a) increased by $\frac{a+1}{a}$	0.283	0.012	0.106, 0.758
Method used for data collection (Manual) (b)	Odds ratio of taking more than 10 minutes when (b) increased by $\frac{b+1}{b}$	3.281	0.003	1.512, 7.122
Experience in using the tool (c)	Odds ratio of taking more than 10 minutes when (c) increased by $\frac{c+1}{c}$	2.667	0.023	1.146, 6.207
Model (7): Simple binary logistic regression model for the duration of data sharing activity				
ICT skills	Odds ratio of taking more than 10 minutes in users with low to medium ICT skills compared to high level users	50.667	< 0.001	11.567, 221.936

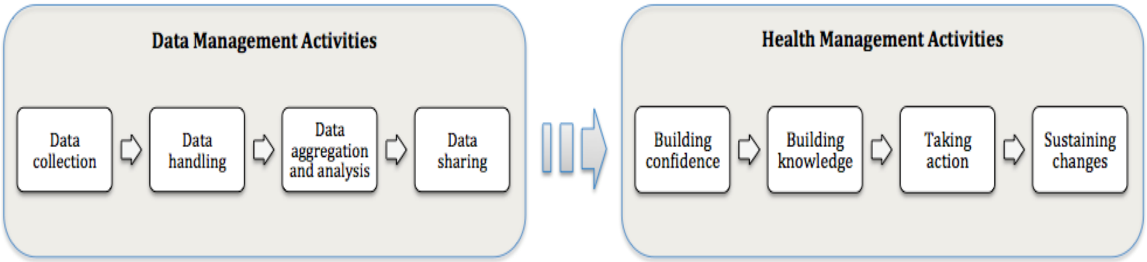

Figure 3 – The gap between data management activities and health management activities

Time devoted to data sharing

Most self-quantifiers in our study (73%) shared their data and findings with other people using primary and secondary SQS. Multiple primary tools usage was found to reduce the time needed to perform this activity; however, using secondary SQS increased the duration of this activity. To measure the impact of using these tools on the odds of the increased duration of data sharing we built two models. In the first logistic regression model, the variables were: using a secondary tool; the number of tools used for data collection; the method used for data collection (e.g., manual or automatica); the degree of experience using these tools; and ICT skills level.

The odds of taking more than 10 minutes to complete the data sharing activity increased 32 times in the secondary tools users compared to the non-users. On the other hand, the odds for longer duration reduced somewhat (0.3 times) for each additional primary tool the self-quantifiers used to share the data (Table 1 model 6). The range of the predicted estimates (confidence interval (CI)) of the odds of longer duration of data sharing when using secondary tools was wider due to limitation in the sample size. The odds of data sharing activity taking longer increased three times with additional types of data collected manually. The odds of longer duration also increased when self-quantifiers' experience in using the tools increased (Odds ratio=2.6).

In the second logistic regression model, the effect of having low to medium ICT skills on the odds of data sharing activity taking longer was 50 times higher compared to users with high ICT skills (Table 1 model 7).

Discussion

This study offers a detailed understanding of the SQS usage experience for health data self-management activities. The results show that managing day-to-day health requires individuals to interact directly with the chosen tools to collect, handle, analyse, or share the generated data. The time needed to go through each of these activities towards achieving one's desired health outcomes depends on the tools' ability to provide the required information and services. Having high ICT skills can slightly reduce the amount of time self-quantifiers have to devote to these activities. These findings are consistent with previous research studies [4, 6, 7].

Currently, few single primary self-quantification systems can support all information needs of a user. This limitation could generate a gap between the two critical stages of health self-quantification (Figure 3). We argue that as the gap gets wider, a self-quantifier has to focus more on managing data instead of managing health; thereby, facing information and time limitations on taking an active role to turn the collected data into knowledge, and subsequently take informed actions. Additionly, using multiple types of tools could diminish the data quality and increase the risk of losing data (files). The senario gets worse when the individuals do not have high ICT skills.

Conclusion

Using multiple types of tools for the purpose of health self-quantification complicates the health data management activities. In particular, it has implications for data and time management that pose challenges for achieving good health outcomes. We suggest that a framework representing and describing all key activities in health self-quantification could be beneficial. Application developers and health researchers could use this framework to improve the design and evaluation of health self-quantification solutions. Healthcare providers could use this framework to determine the applicability of these solutions and plan their implementation, based on the expected outcomes. Thus our current research focuses on the development of a framework, based on the theories of behaviour changes and human interactions with SQS tools, allowing a comprehensive view of all the essential activities and components of health self-quantification practice. We are also conducting further empirical research on the relationship between self-quantifiers' data management activities and health activation, to better understand the factors involved in their ability to achieve desired health objectives.

Acknowledgments

Funding for this project was provided by the University of Melbourne's Institute for a Broadband-Enabled Society (IBES). Almalki's PhD research is supported by a scholarship from Jazan University and a top-up scholarship from IBES.

References

[1] Fox S, Duggan M. Tracking for Health. *Pew Research Center's Internet & American Life Project. 2013*. Available from: http://www.pewinternet.org/2013/01/28/tracking-for-health

[2] Almalki M, Gray K, Martin-Sanchez F. The Use of Self-Quantification Systems for Personal Health Information: Big Data Management Activities and Prospects. *Health Information Science and Systems HISS*; Australia, Melbourne. 2014.

[3] Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers. *Health Services Research*. 2004;39(4p1):1005-26.

[4] Li I. Personal Informatics and Context: Using Context to Reveal Factors that Affect Behavior. *Journal of Ambient Intelligence and Smart Environments*. 2012;4(1):71-2.

[5] Eurostat. 2011. Available from: https://circabc.europa.eu/sd/a/2ea7b600-242c-46db-9e4f-afb25e03ba87/Questionnaire%20HH%202011%20v.3.1.pdf

[6] Jones W. Personal Information Management. *Annual Review of Information Science and Technology*. 2007;41(1):453-504.

[7] Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Canada, Toronto. 2014;1143-1152.

Address for correspondence

Manal Almalki. University of Melbourne, 3010, Australia. E-mail: malmalki@student.unimelb.edu.au