Peer-vasive Computing: Leveraging Peers to Enhance the Accuracy of Self-Reports in Mobile Human Studies

Allan Berrocal
Quality of Life Technologies Lab
Center for Informatics
University of Geneva,
Switzerland
allan.berrocal@unige.ch

Katarzyna Wac
Quality of Life Technologies Lab
Center for Informatics
University of Geneva,
Switzerland
katarzyna.wac@unige.ch
University of Copenhagen,
Denmark
wac@di.ku.dk

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UbiComp/ISWC'18 Adjunct, October 8–12, 2018, Singapore, Singapore
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5966-5/18/10...$15.00.
https://doi.org/10.1145/3267305.3267542

Abstract
We discuss two methods designed to increase the accuracy of human-labeled data. First, *Peer-ceived Momentary Assessment* (*Peer-MA*), a novel data collection method inspired by the concept of Observer Reported Outcomes in clinical care. Second, *mQoL-Peer*, a platform aiming to equip researchers with tools to assess and maintain the accuracy of the data collected by participants and peers during mobile human studies. We describe the state of the research and specific contributions.

Author Keywords
Self-Assessment; Observer's Assessment; Ecological Momentary Assessment; Peer-ceived Momentary Assessment

ACM Classification Keywords
H.1.2 [User/Machine Systems]: Human Information Processing; H.5.3 [Group and Organization Interfaces]: Evaluation/Methodology; J.4 [Social and Behavioral Sciences]: Psychology

Introduction
The self-report, or self-assessment method is commonly used in human studies to collect both, qualitative (e.g. *Please describe your experience in the seminar?*), and quantitative data (e.g. *How many countries have you visited?*). Paulhus and Vazire [17] described the method, including its

EMA:
Ecological Momentary Assessment [12, 20], which implies putting forward a short survey to a study participant multiple times a day, collecting in-situ self-assessments about a condition being studied; as well as potentially context related data from smartphone sensors (if *EMA* is deployed in the participant's smartphone) to enrich the response.

Peers:
Peers are close, trusted friends or family members of a study participant, taking the role of observers to provide *Observer Reported Outcomes* (ObsRO) [15] about this participant during a study.

Peer-MA:
Is an *EMA* completed by an observer (peer) of a study participant, during the same time window when this participant completes an *EMA*.

main virtues, and strategies to counteract its known weaknesses (e.g. socially desirable, acquiescent and extreme responding). Self-assessments tend to be lengthy surveys, paper or computer-based, and typically applied only once during a human subjects study to capture the participants' overall state. With the advent of *Ecological Momentary Assessment* (*EMA*) [12, 20] and the widespread availability of smartphones, self-assessments can be collected more frequently than before (and usually in the form of short surveys). Nonetheless, more quantity does not directly imply better quality; self-assessments keep posing issues regarding the accuracy of the assessed variable. Consequently, obtaining accurate ground truth for the studied phenomena is still a challenge.

We describe ongoing work aiming to improve the accuracy of human-labeled data in mobile studies: (*a*) *Peer-MA*, a novel data collection method leveraging subjective peer-assessments inspired by clinical care practices [15]; and (*b*) *mQoL-Peer*, an extension of the *mQoL Lab* platform (established in 2010) [4] that equips researchers with better tools to assess and maintain the accuracy of the data collected during, and after mobile human studies.

(In)accuracy of Human Data Contributions

The following problems affect the accuracy of self-assessments [17]: (*a*) *self-presentation*, the most prevalent type is *social desirable responding*, when respondents answer untruthfully to attain a desirable outcome for a given situation (e.g. getting a job); (*b*) *acquiescent responding*, when respondents always agree or, (*c*) *reactant responding*, when respondents always disagree regardless of the content of statements; (*d*) *extreme responding*, when respondents always pick one extreme given a list of ranked choices; or (*e*) *random responding* which is even more difficult to detect.

To manage this issue, researchers take advantage of advancements in sensing devices and ubiquitous technologies. Self-assessments can be complemented by objective data collected from smartphones and wearable devices. For instance, Harari et al. [11] reviewed studies using Smartphone Sensing Methods to identify physical movement, social interactions and other daily activities, which can be used as objective and automated measures of behaviour. More recently, Gresham et al. [10] leveraged objective data from activity monitors to predict the risk of adverse events, hospitalisations and hazard for death, in a single-cohort clinical trial of advanced cancer patients. Their results highlight an opportunity to increase the accuracy of a health assessment, reducing subjectivity and bias known to be present in traditional assessments. Finally, in the research of human stress assessment via non-invasive methods, the combination of subjective self-assessments with objective data from smartphone sensors and wearable devices has produced favourable results [2, 9, 18, 19].

However, despite its value, objective data does not always enable accurate modelling of highly subjective individual's perceptions and states (e.g., intimacy). Consequently, self-assessments are leveraged at some point during mobile human studies. Additionally, Lazar et al. [14] reported that abandonment rates are high for many smart devices. Some users perceive that the data collected from these devices is not useful to them, and they perceive maintenance (e.g. battery charging), as well as privacy risks to be high.

To better assess the accuracy of human-labeled data, particularly self-assessments, we extend *EMA* with *Peer-ceived Momentary Assessment* denoted as *Peer-MA*. *Peer-MA* is a type of *EMA* completed by an observer of an individual, during the same time window when such an individual completes an *EMA*. We will measure the value of *Peer-MA* as

Figure 1: (a) Example of an *EMA* about stress. (b) Example of corresponding *Peer-MA* about stress. (c) Example of the confidence assessment provided for this *Peer-MA*.

a novel, and timely data collection method for the goal of increasing data accuracy in mobile human studies.

Our approach is inspired by clinical care [15], which distinguishes between patient-reported outcomes (PROs) (e.g. self-assessed stress level) and observer-reported outcomes (ObsRO) (e.g. when a family member reports about the perceived stress level of a dementia patient or child). This method is not well explored to date in healthy populations. So, given the availability of smartphones and frequency of social interactions, *Peer-MA* has the potential to enrich the self-assessment datasets with peers as pervasive data providers, whose observations could help researchers identify and manage data accuracy issues in human studies.

Previous Research Involving Peers

Peers are involved in stroke rehabilitation [6, 21], to assess general health conditions [16], community integration after severe injuries [3], and chronic stress detection [13]; typically via paper-based methods and infrequent clinical care visits in vulnerable populations (e.g. elderly or children).

For many individuals, interactions with peers about their emotional or health states are frequent and meaningful. For instance, Eisenberg et al. [7, 8] reported that 80% of college students in the US (N=14175, Age>18) with a mental health problem received counselling from a nonprofessional (70.5% friends, 52.5% family members). Drum et al. [5] reported that two-thirds of college students (N>26000) with suicidal intentions disclosed it to a peer first (a romantic partner, roommate or friend). Additionally, Vazire [22] promoted 'informant' reports in psychology with evidence to refute some misconceptions (e.g. high cost, less accuracy).

To our knowledge, no other research is currently exploring the role of peers/observers as pervasive data contributors in human studies, including healthy populations.

Contributions

First: Peer-ceived Momentary Assessment Method

We research *Peer-MA* as a new data collection method which makes it possible to complement an individual's self-assessment of a given condition (e.g. perceived physical or mental state), with assessments of said condition provided by friends or family members chosen by the individual.

During a study, participants and peers contribute quantitative and qualitative datasets via *EMAs* and *Peer-MAs* using *mQoL-Peer*. Peers report what they perceive the participant is projecting with regards to the condition/state being assessed (e.g. stress, depression, anxiety), and they indicate their degree of confidence for each assessment (because sometimes peers may be legitimately unable to answer). Figure 1 shows an example of an ongoing study about human stress assessment using peer-assessments. Besides subjective self and peer-assessments during the study, *mQoL-Peer* automatically collects objective smartphone sensed data (e.g. category of applications used, screen touches, physical activity) in the background from participants (not from peers).

Second: mQoL-Peer Researcher Support Platform

mQoL-Peer is designed to help researchers to procure an accurate dataset during a human subjects study.

The *mQoL-Peer* platform, which extends [4] allows to design and launch human studies meeting strict study protocol requirements. For instance, the specification of an informed consent (see Figure 2), which is stored and always available to the user via *mQoL Lab*, entry surveys (e.g. demographics, standardised scales) (see Figure 3), *EMA*-based self-assessments, exit survey, study duration, among others. Some aspects of *EMA* can be adjusted: (1) number of *EMAs* per day; (2) desired hours range; (3) trig-

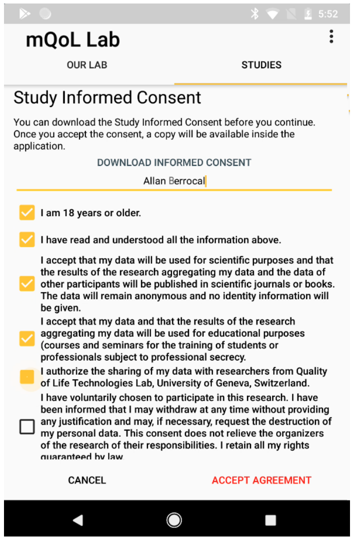

Figure 2: Signing an informed consent in the *mQoL Lab*.

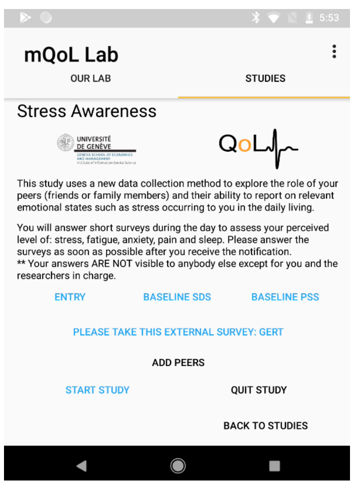

Figure 3: Study home-screen with buttons to launch pre-study surveys.

ger (either random or schedule based); (4) *Peer-MA*-survey corresponding to an *EMA*-survey. Similarly, some aspects of the *Peer-MA* method deployment can be adjusted: (1) number of peers (NoP) for each individual; (2) frequency of the *Peer-MA* (FoP), being either *paired* (in which case a *Peer-MA* is only triggered when a participant is completing an *EMA*) or *unpaired* (in which case a *Peer-MA* can be completed voluntarily at any moment); (3) trigger of *Peer-MA* (ToP), *EMA*-based (happening as a result of an *EMA*), or proximity-based (occurring after the detection of physical or virtual proximity between participants and peers). Furthermore, the order of questions in *EMA* or *Peer-MA* can be randomised to prevent issues such as pattern answering.

To join a study, users download our Android application *mQoL Lab*, and follow a few steps guided by videos. Participants invite peers by clicking a button. Peers join the study via an email link which downloads the *mQoL Lab* application and guides them through the informed consent, after which they start contributing data.

During the study, the *mQoL Lab* application running on the participant's smartphone collects objective data in the background and synchronises it with the dedicated server every time the phone is plugged into a charger and connected to WiFi. *EMA* and *Peer-MA*, however, are synchronised with the server immediately after they are completed by the participants and the peers respectively. This real-time availability of data allows researchers to quickly identify if someone stopped contributing data. Researchers can easily send notifications to participants (preserving anonymity) to encourage them to keep active in the study.

Additionally, as shown in Figure 4 for a case of two participants each having two peers, researchers can analyze quantitative aspects about the collected data, e.g. (1) *recall*: (% of *EMAs* and *Peer-MAs* completed from all triggered ones); (2) *gaps*: (% of *EMAs* with no corresponding *Peer-MA*); (3) *voids*: (% of *Peer-MAs* reported as zero confidence); (4) *halts*: (hours/days with no *EMAs*/*Peer-MAs*) and temporal distribution of such *halts*; (5) *validity*: of answers (e.g. acquiescent, reactant, extreme or random responding); (6) *pace*: mean duration to complete *EMAs*/*Peer-MAs*, and mean duration for each question; (7) *lapse*: mean time interval between the answer of an *EMA* and its corresponding *Peer-MA*.

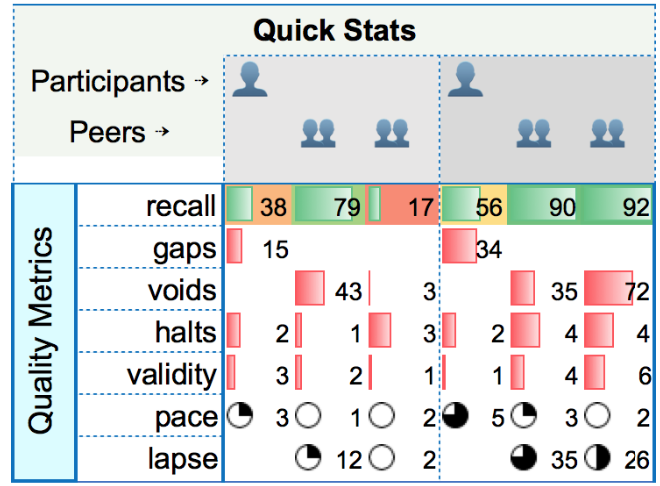

Figure 4: *EMA* and *Peer-MA* Quality metrics readily available during a study.

Finally, *mQoL-Peer* provides the means for the researcher to download complete datasets from all participants and peers in the study. These datasets can be ported to data analysis tools such as R or Python.

Achievements and Ongoing Work

We demoed the *mQoL-Peer* application and platform at the 2018 ACM Digital Health Conference [1]. We received feedback from health researchers, and we are improving the platform since then. For example, we are testing a simplified method to deploy *Peer-MA* as a web link via SMS, so that peers with non-Android smartphones can participate.

We have an ethically approved, ongoing study about human stress assessment with support from peers in Geneva, Switzerland. Also, we are preparing two collaborations to apply our method in clinical contexts: (1) Geneva University Hospital, Department of Psychiatry (depression patients). (2) Department of Surgery at Stanford University Medical Center (US) with patients undergoing organ transplant whose peers have an instrumental role in their quality of life.

Summary of Scientific Contributions

We aim to scientifically contribute to the research with: (1) Exploration of a new computational method to improve the accuracy of data collection in mobile human studies, capitalising on the collective assessments by trusted peers. (2) A set of empirical data obtained from real-life experiments with individuals and their peers, in various experimental conditions. Also, (3) a set of design principles that can enhance the field of human-computer interaction when it comes to creating digital solutions to enable more accurate assessment of human well-being in healthy populations by leveraging the peers.

Acknowledgements

Swiss Federal Commission for Foreign Students Scholarships (2016-2019), H2020 WellCo (no.769765, 2018-2021), AAL CoME (2014-7-127) and the University of Costa Rica.

REFERENCES

1. A. Berrocal, A. De Masi, M. Gustarini, and K.B Wac. 2018. mQoL-Peer: Assessing the Individual's State via the Just-in-Context Individual's Peers'. *ACM Digital Health Conference* (2018).

2. M. Ciman and K. Wac. 2016. Individuals' Stress Assessment using Human-Smartphone Interaction Analysis. *IEEE Transactions on Affective Computing* 9, 1 (2016), 51–65.

3. C. P. Cusick, C. A. Brooks, and G. G. Whiteneck. 2001. The use of Proxies in Community Integration Research. *Archives of Physical Medicine and Rehabilitation* 82, 8 (2001), 1018–1024.

4. A. De Masi, M. Ciman, M. Gustarini, and K. Wac. 2016. mQoL Smart Lab: Quality of Life Living Lab for Interdisciplinary Experiments. *UbiMI Workshop, ACM Int. Conf. on Pervasive and Ubiquitous Computing* (2016), 635–640.

5. D. J. Drum, C. Brownson, A. B. Denmark, and S. E. Smith. 2009. New Data on the Nature of Suicidal Crises in College Students: Shifting the Paradigm. *Professional Psychology: Research and Practice* 40, 3 (2009), 213–222.

6. P. W. Duncan, S. M. Lai, D. Tyler, S. Perera, D. M. Reker, and S. Studenski. 2002. Evaluation of proxy responses to the Stroke Impact Scale. *Stroke* 33, 11 (2002), 2593–2599.

7. D. Eisenberg, J. Hunt, and N. Speer. 2012. Help Seeking for Mental Health on College Campuses: Review of Evidence and Next Steps for Research and Practice. *Harvard Review of Psychiatry* 20, 4 (2012), 222–232.

8. D. Eisenberg, J. Hunt, N. Speer, and K. Zivin. 2011. Mental Health Service Utilization Among College Students in the United States. *Journal of Nervous and Mental Disease* 199, 5 (2011), 301–308.

9. M. Gjoreski, H. Gjoreski, M. Lutrek, and M. Gams. 2015. Automatic Detection of Perceived Stress in Campus Students Using Smartphones. In *11th Int. Conf. on Intelligent Environments*. IEEE, 132–135.

10. G. Gresham, A. E. Hendifar, B. Spiegel, E. Neeman, R. Tuli, B. J. Rimel, R. A. Figlin, C. L. Meinert, S. Piantadosi, and A. M. Shinde. 2018. Wearable Activity Monitors to Assess Performance Status and Predict Clinical Outcomes in Advanced Cancer Patients. *npj Digital Medicine* 1, 1 (2018), 27.

11. G. M. Harari, S. R. Müller, M. S. Aung, and P. J. Rentfrow. 2017. Smartphone Sensing Methods for Studying Behavior in Everyday Life. *Current Opinion in Behavioral Sciences* 18 (2017), 83–90.

12. J. M. Hektner, J. A. Schmidt, and M.B Csikszentmihalyi. 2007. Experience Sampling Method: Measuring the Quality of Everyday Life. (2007).

13. W. Kromm, M. C Gadinger, and S. Schneider. 2010. Peer Ratings of Chronic Stress : Can Spouses and Friends Provide Reliable and Valid Assessments of a Target Person ' s Level of Chronic Stress ? *Stress and Health* 26, 4 (2010), 292–303.

14. A. Lazar, C. Koehler, J. Tanenbaum, and D. H. Nguyen. 2015. Why we Use and Abandon Smart Devices. *ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing - UbiComp '15* (2015), 635–646.

15. N. E. Mayo, S. Figueiredo, S. Ahmed, and S. J. Bartlett. 2017. Montreal Accord on Patient-Reported Outcomes Use Series-Paper 2: Terminology Proposed to Measure What Matters in Health. *Journal of Clinical Epidemiology* 89 (2017), 119–124.

16. P. J. Neumann, S. S. Araki, and E. M. Gutterman. 2000. The Use of Proxy Respondents in Studies of Older Adults: Lessons, Challenges, and Opportunities. *Journal of American Geriatrics Society* 48, 12 (2000), 1646–1654.

17. D.L.. Paulhus and S. Vazire. 2005. The Self-Report Method. *Handbook of Research Methods in Personality Psychology* (2005), 224–239.

18. A. Sano and R. W. Picard. 2013. Stress Recognition Using Wearable Sensors and Mobile Phones. In *Humaine Association Conference on Affective Computing and Intelligent Interaction Stress*. IEEE Computing Society, 671–676.

19. A. Sano, S. Taylor, A.W. McHill, A.J. Phillips, L.K. Barger, B. Klerman, and R. Picard. 2018. Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones. *Journal of Medical Internet Research* 20 (2018).

20. S. Shiffman, A.A. Stone, and M.R. Hufford. 2008. Ecological Momentary Assessment. *Annu Rev Clin Psychol.* 4 (2008), 1–32.

21. K.C. Sneeuw, N. K. AAaronson, R.J. de Haan, and M. Limburg. 1997. Assessing Quality of Life After Stroke. *Stroke, American Heart Association* 28, 8 (1997), 1541–1549.

22. S. Vazire. 2006. Informant Reports: A Cheap, Fast, and Easy Method for Personality Assessment. *Journal of Research in Personality* 40, 5 (2006), 472–481.