

Added Value from Secondary Use of Person Generated Health Data in Consumer Health Informatics

Contribution of the Consumer Health Informatics IMIA Working Group

P.-Y. Hsueh¹, Y.-K. Cheung², S. Dey³, K. K. Kim⁴, F. J. Martin-Sanchez⁵, S. K. Petersen⁶, T. Wetter⁷

¹ Center for Computational Health, IBM T.J. Watson Research Center, Yorktown Heights, NY, USA

² Mailman School of Public Health, Columbia University, New York, NY, USA

³ Center for Computational Health, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

⁴ Betty Irene Moore School of Nursing, University of California Davis, Sacramento, CA, USA

⁵ Department of Healthcare Policy and Research, Division of Health Informatics, Environmental and Participatory Health Informatics (ENaPHI) Research Group, Weill Cornell Medicine, New York, NY, USA

⁶ University of Texas MD Anderson Cancer Center, Houston, TX, USA

⁷ Institute of Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany and Dept. of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA

Summary

Introduction: Various health-related data, subsequently called Person Generated Health Data (PGHD), is being collected by patients or presumably healthy individuals as well as about them as much as they become available as measurable properties in their work, home, and other environments. Despite that such data was originally just collected and used for dedicated predefined purposes, more recently it is regarded as untapped resources that call for secondary use.

Method: Since the secondary use of PGHD is still at its early evolving stage, we have chosen, in this paper, to produce an outline of best practices, as opposed to a systematic review. To this end, we identified key directions of secondary use and invited protagonists of each of these directions to present their takes on the primary and secondary use of PGHD in their sub-fields. We then put secondary use in a wider perspective of overarching themes such as privacy, interpretability, interoperability, utility, and ethics.

Results: We present the primary and secondary use of PGHD in four focus areas: (1) making sense of PGHD in augmented Shared Care Plans for care coordination across multiple conditions; (2) making sense of PGHD from patient-held sensors to inform cancer care; (3) fitting situational use of PGHD to evaluate personal informatics tools in adaptive concurrent trials; (4) making sense of environment risk exposure data in an integrated context with clinical and omics-data for biomedical research.

Discussion: Fast technological progress in all the four focus areas calls for a societal debate and decision-making process on a multitude of challenges: how emerging or foreseeable results transform privacy; how new data modalities can be interpreted in light of clinical data and vice versa; how the sheer mass and partially abstract mathematical properties of the achieved insights can be interpreted to a broad public and can consequently facilitate the development of patient-centered services; and how the remaining risks and uncertainties can be evaluated against new benefits. This paper is an initial summary of the status quo of the challenges and proposals that address these issues. The opportunities and barriers identified can serve as action items individuals can bring to their organizations when facing challenges to add value from the secondary use of patient-generated health data.

Keywords

Secondary use of data; consumer health informatics, Person Generated Health Data; participatory health; adaptive trial.

Yearb Med Inform 2017;160-71

<http://dx.doi.org/10.15265/IY-2017-009>

Published online May 8, 2017

1 Introduction

Person Generated Health Data (PGHD), which is sometimes also used for Patient Generated Health Data, has received strongly increasing attention in numerous publications (cf. [1, 2] for a small selection), and at events such as the AMIA 2016 symposium. Example data sources include blood glucose and blood pressure monitoring data from devices, exercise, and nutrition logs from mobile applications, and questionnaires (e.g., screening, medication adherence, risk assessment, and intake). The growing availability of devices, sensors, smartphone apps, and Direct-to-Consumer Services (e.g. personal genome, microbiome) enables citizens to collect relevant health information (e.g. physical activity, sleep patterns, mood, genetic or exposure information, etc.) that can be shared among them, or with health providers and researchers.

In the last years, we have been witnessing a dizzying convergence between the digital revolution and the world of health [3]. In particular, two significant trends in the healthcare landscape are emerging. First, the movement of quantified self has led to the increasing prevalence of health

tracking capabilities on consumer devices and to the emerging citizen science [4, 5, 6], which both entail with rising health consumer awareness. The statistics shown in recent large-scale consumer surveys have clearly revealed this trend: a 2016 survey conducted in the US [7] showed that 78% of the citizens were willing to wear technology for health tracking. In fact, 33%, resp. 21%, of the US consumers were already using health-tracking mobile apps, resp. wearable devices. In another 2015 survey [8], 45%, resp. 32%, were intrigued to use wearable devices to track health in the US, resp. Europe. Gownder JP, et al. further showed that a majority of the respondents believed that PGHD would be beneficial for maintaining their own health [9].

In this respect, with presumably healthy *citizens* tracking health-related behaviors, “Person” certainly better fits the P in PGHD than “Patient”. Also the term “Generated” is misleading when presumably health-related data such as environmental exposure is collected *about* citizens, rather than *by* citizens. However, since PGHD has been widely introduced, we stick to the acronym also in somewhat atypical situations. It should also be noted that for the comprehensive management of health problems, PGHD couldn't be regarded in isolation. In many cases, PGHD only gain meaning in light of clinical data, or vice versa. Therefore, health-related data from sources other than the citizen or her attached devices (health history, treatment history, symptoms, biometric data from lab tests/devices) will be incorporated wherever medically mandated.

Second, the model of value-based care has led to healthcare ecosystem forming and payment reform. To achieve this, some countries have attempted a more coordinated approach to consolidate efforts from multiple providers to include patients in full cycles of care [10]. In many other countries, this trend has led to increasingly diverse channels of health service delivery such as nurse-led clinics and case management in UK, Italy, and Spain [11], and retail clinics in the US [12]. Also in the US, 85% of provider payments under Medicare were expected to tie into quality of care by the end of 2016 [13]. Under the

Medicare Access and CHIP Reauthorization Act of 2015 (MACRA)[14], which is the regulation currently being finalized to determine how Medicare providers would be paid, some of the proposed quality measures require patient contribution. To name a few: functional status, shared decisions, ambulatory assessments, and preferences [15]. In addition, the perceived need of shared decision-making [16] has also changed significantly: in a 2005 survey, only 48% of patients preferred not to leave final decisions to their physicians [17]; a decade later, in a 2012 survey [18], there are almost 76% requesting to be involved in the final decision-making.

Owing to these trends in quantified self [4, 5, 6] and value-based care [10, 11, 12], a plethora of patient-centered data generated by devices, self-reporting portals/apps, and care processes are produced. However, quality and value of this data, objectively and in patients' perceptions, are unknown. Health informaticians are challenged by the amount of the accumulated PGHD and by their heterogeneous and temporal nature, which call for innovation in data collection, storage, standardization, integration, analysis, and visualization [19].

While this challenges data analytics which has begun to deliver algorithms and heuristics to separate good from bad, and noise from meaningful data, we will hereby address the next challenge in “secondary use of data”, i.e., how to utilize cleansed and approved data for purposes beyond their primary context and motivation of collection. Examples of such usages beyond the primary purposes of PGHD include: (i) feeding information back into Medical Product development [20], (ii) use of data as trial outcome measures [21], (iii) identifying predictors of disease progression [22], and (iv) returning usefull information to the patient to assess self-efficacy and induce behavior changes [23,24]. To date, this is still an active research area that interdisciplinary researchers are exploring to collect evidence. The nascency of this field may not yet allow for a systematic review. However, the field still needs an account of the status quo. This overview therefore aims to provide an initial reading into exemplary application areas wherein added value in these areas is demonstrated.

2 Method

Two of the authors (SH, TW) identified perspectives on PGHD leading to a characteristic type of secondary use and respective value-added services. The value-added services range from enhancing PGHD services by *pre-existing evidence* through feeding personalized decision support with *evidence newly discovered* in PGHD. In particular, two key dimensions have emerged to differentiate the emerging PGHD value-added services: (a) person's role in data capture and (b) scope of value added to the PGHD. The first dimension differentiates *how* PGHD was collected for secondary use, i.e., whether a “person” was passively involved during data collection, or has been actively participating via some reporting tools (such as patient-reported outcome measure questionnaires and ecological momentary assessment apps). The second dimension differentiates *where* PGHD are adding values to consumers (via research, app surveillance, clinical decision support, or personalized care coordination support). Four application areas from the different person's roles and scopes of value-added services were selected to provide deep dives in this survey.

SH and TW then solicited concise descriptions from pioneers of four focus areas of secondary use identified by scholarly publications in the years 2015-2017. Figure 1 illustrates the chosen focus areas on the two key dimensions. Each of these will be treated as a focus area where we describe an original setting of collecting or processing data and contrast it with the secondary usage methods that address open problems or new opportunities and by that token add value to the original purposes of the efforts.

Therefore, this article serves as a survey of emerging practices that have demonstrated preliminary evidence of their values. The four focus areas were selected with the following rationale: (1) **PGHD for care coordination across multiple conditions**: we discuss the Shared Care Plan (SCP) to demonstrate that the value of dispersed data that patients maintain about their multiple diseases can be enhanced by applying existing evidence to aggregate them into a personalized comprehensive plan for care

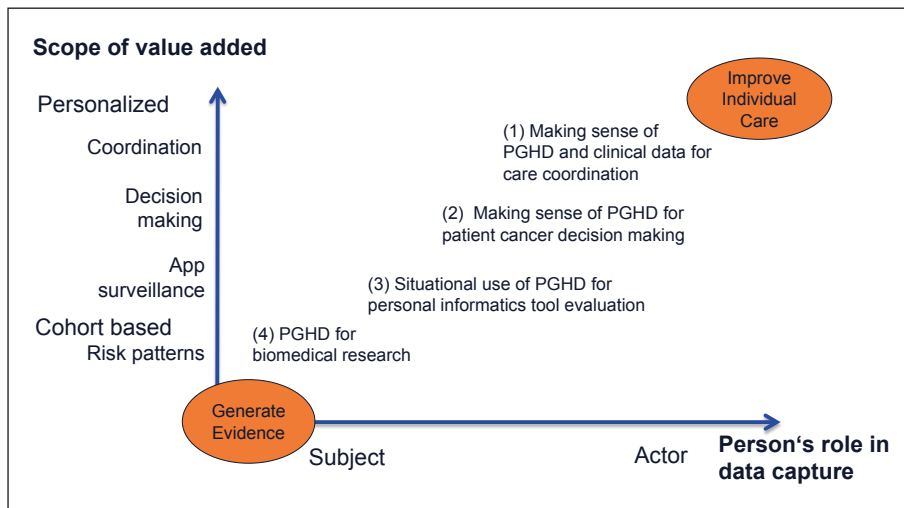


Fig. 1 Four focus areas on the key dimensions of secondary use of PGHD

coordination; (2) **PGHD for patient cancer decision-making**: we discuss primary data collected by cancer patients to demonstrate that new insight can emerge from identifying patterns in large collections of such data; (3) **PGHD for personal informatics tool evaluation**: we discuss the secondary use of mobile app data that were originally collected for behavioral coaching to demonstrate that methods from advanced biostatistics can be used to identify the most successful apps concurrently to their use in a realistic setting; (4) **PGHD for biomedical research**: we discuss environmental risk exposures to demonstrate that patterns and effects found in large data collections can flow back to the patient to infer cohort risk patterns for intervention allocation at a later stage.

3 Results

3.1 Focus Area 1: PGHD for Care Coordination across Multiple Conditions

Primary purpose. Managing chronic illness frequently involves managing multiple chronic conditions (MCC). A quarter of US adults have MCC, with 18% having two or three conditions and 7% having four or

more [25]. Common co-morbidities across all groups are hypertension, hyperlipidemia, and diabetes [26]. The proportion of MCC is higher for older age groups [25]. There is a substantial burden on family caregivers who assist individuals with MCC. They are unpaid and not included in cost estimates of the healthcare system. A recent national survey estimated that 16.6% or 39.8 million Americans provide care for an adult [27], and on average, caregivers spend 24.4 hours per week caring their parent.

Care coordination – the deliberate synchronization of care – is an evolving concept that attempts to address the challenges faced by those with complex care needs. However, care coordination is also more difficult for those with MCC due to the involvement of larger numbers of clinicians and differing advice received from them [28, 29], more medications to manage [30], and higher rates of adverse drug events [31, 32]. A multinational survey indicates that serious gaps in care coordination exist in the US: 44% of respondents said a doctor or a pharmacist sometimes/rarely/never reviewed medications, 24% that medical information was not available during scheduled visits, a rate which is 30% higher than in other countries, and 38% experienced a failure on at least one component of discharge care coordination

such as receiving a written plan of care or arrangements for follow up care [33]. Therefore, while fragmented data and information are maintained by the patient, her caregivers and her healthcare providers for her different conditions, integration of these disperse data is lacking. This lack of integration makes enrichment of cross-disciplinary care processes such as adverse drug events surveillance or application of evidence-based practices impossible.

Secondary use. Shared Care Plan (SCP) is a secondary use of the fragmented threads of health documentation. It is a comprehensive, evidence-based plan of care that is collaboratively developed with participation of the patient, family, and health care team [34]. The objective is to address all of the patient's health-related needs in the context of the patient's values, requirements, and preferences. A number of authors have suggested specific elements for inclusion in SCPs (c.f. Table 1) [35, 36, 37, 38, 39, 40]. The breadth of requirements for care coordination and SCPs clearly indicate the need for both person- and healthcare-generated data for comprehensive situation assessment and coordination.

While much of clinical data in Table 1 exists already dispersed across EHRs, it is likely not available to patients, caregivers, and healthcare teams across multiple settings. Similarly, PGHD may reside in siloed applications in use by patients but not integrated together, or shared with healthcare teams.

SCPs need to be collaboratively built and dynamic over time. In order to enable SCP requirements, newer collaborative technologies such as mobile applications, social-network-styled systems and wikis should be considered to complement the EHR-based care plans in place today. These technologies should be implemented with attention to technology standards that enable interoperability and preserve privacy and security of health information, concerns that previous studies have shown to be important for consumer acceptance of health technology and electronic data sharing [41].

One example of a shared platform that demonstrates owner-controlled coordinated collection and distribution of information is LinkedIn, a professional networking

Table 1 Informational Components of SCPs

Content Categories	Person Generated Health Data	Clinical Data
Contact Information	<ul style="list-style-type: none"> ▪ Patient preferred contacts 	<ul style="list-style-type: none"> ▪ Responsible clinician ▪ Number(s) to call for results
Health History	<ul style="list-style-type: none"> ▪ Detailed health concerns ▪ Allergies 	<ul style="list-style-type: none"> ▪ Conditions, diagnoses ▪ Health status evaluation populated with computable, standardized data.
Goals and Preferences	<ul style="list-style-type: none"> ▪ Patient's goals ▪ Expectations of care ▪ Challenges and concerns ▪ Self-management capabilities ▪ Family or caregiver resources ▪ Patient-reported health status ▪ Advanced directives ▪ Patient likes and dislikes 	<ul style="list-style-type: none"> ▪ Problem list ▪ Clinical goals ▪ Treatment plans
Actions	<ul style="list-style-type: none"> ▪ Self-tracking measures (e.g. blood glucose, weight) ▪ Tracking of observations of daily living ▪ Patient self-management plan/behavior change action plan ▪ Side effects and symptoms ▪ Tracking SCP items 	<ul style="list-style-type: none"> ▪ Appointments ▪ Interventions and treatments ▪ Test results ▪ Tests and orders pending at discharge/transfer ▪ Responsible individual for follow up ▪ Evidence-based guidelines ▪ Tracking SCP items
Health Education	<ul style="list-style-type: none"> ▪ Identified learner for education if patient is unable to receive it ▪ Information about health condition 	<ul style="list-style-type: none"> ▪ Clinical instructions given to patient
Medications	<ul style="list-style-type: none"> ▪ Medication concordance and adherence plan and tracking ▪ Over the counter medications ▪ Medications that are not being taken 	<ul style="list-style-type: none"> ▪ Prescribed medications ▪ Medications during hospitalization ▪ Pre-admission medication list ▪ New discharge medications with start date, duration, route, dose, frequency, date, indication

application (mobile and social) that supports the construction of a dynamic, longitudinal career record with multimedia capabilities. The record owner controls how information is ordered and prioritized and how it is communicated to the member's connections. Another example is the wiki (a type of social network platform). The best-known wiki, Wikipedia, leverages the crowd wisdom to aggregate information and constantly update and verify the accuracy of that information. But many wikis are used as tools for knowledge management and collaboration [42, 43], and health professions education [44, 45]. By that token, this presents a blueprint for the coordinated aggregation of data and knowledge related to a patient's conditions. Few wikis, though, have focused

on patient information needs. These types of technologies, which have rarely been applied to healthcare, could be helpful to the requirements for a SCP.

New solutions that foster the inclusion of PGHD must be flexible to allow for changes in healthcare delivery models and integration of new technologies, as demonstrated by the authors of [46]. EHR-based care plans and clinical decision-support systems are built on enterprise software platforms that require that communication pathways, workflows, and evidence guidelines be set in advance, making any change cumbersome and time-consuming. Technology-enabled SCPs could offer solutions for personalized care processes and be supportive of collaboration of individuals with MCC and their healthcare teams.

3.2 Focus Area 2: PGHD for Patient Cancer Decision-making

Primary purpose. The growing use of mobile and sensor technology offers opportunities for data collection and intervention delivery to prevent cancer and improve cancer-related outcomes. In 2016, 64% of US residents own smartphones (up from 58% in 2014), which enable collection of health data and delivery of intervention content through mobile apps, text messaging, and video [47]. In addition, 45% report owning consumer-grade wearable sensors (e.g., Fitbit®, smart watch). This number has doubled since 2014 [48]. Having the capability to deliver mobile and sensor technology and achieve desired and sustained behavior changes depends on the user acceptance, engagement, and perceived clinical utility of the device. This focus area demonstrates that, when coordinated with relevant clinical data and encounters, these enabling technologies can facilitate the collection and integration of PGHD into clinical and patient decision-making, particularly during times of acute cancer care.

Most oncology care is provided on an outpatient basis and current standard of care largely relies on in-clinic interactions between providers and patients to assess and evaluate outcomes and provide feedback [49]. Information gained in relatively brief encounters can be biased due to underreporting and inaccurate recall [50]. Ecological validity of cancer interventions can be improved by tailoring health assessment “in-the-moment” when and where it is most needed [51]. Integrating mobile and sensor technology may overcome limitations and constraints in their ability to objectively and accurately evaluate important outcomes that impact cancer prevention and treatment regimens, and to provide meaningful health interventions at the most appropriate time [52, 53].

Secondary use. Navigating and managing cancer care is a complex task, with inherent challenges including information deficits, poor care coordination between primary care and specialty providers, and psychosocial support needs. Harnessing technological innovations for PGHD can potentially enhance individual management of cancer

care across the cancer continuum, from primary prevention through treatment and into survivorship [54]. For example, PGHD analysis focused on identifying emotional distress has facilitated the identification of persons in need for mental health care, assistance, and support [55]. Mobile technology applications that enable patients to identify and report symptoms or adverse effects from treatment between routine clinic visits may improve adherence to their care plans [53]. Systems that enable the collection of PGHD during cancer care can improve clinical decision-making and enhance patient engagement in their care [52, 53].

Despite increasing prevalence of sensors and mobile applications directed at improving health and psychosocial well being, there are relatively limited data on their optimal use for improving PGHD data collection and intervention delivery in cancer [56]. Many basic questions need to be addressed: what type of data is most useful, under what type of conditions, and for what patients or populations? What are the best practices to distill, analyze, and intervene using data generated by patients from mobile or continuous monitoring? Addressing these questions may help overcome acceptance and adoption barriers, understand and implement processes to integrate enabling technologies, and result in information that is clinically relevant and useful as well as enriching to the target audience's experience.

3.3 Focus Area 3: PGHD for Personal Informatics Tool Evaluation

Primary purpose. As shown in Focus Area 2 on cancer care, mobile health apps are an important source of PGHD. In this section, we describe a primary use case of behavioral intervention technology (BIT) [57], which uses mobile technologies (e.g., smartphones, tablets, sensors) to deliver behavioral intervention to support physical and mental health. Examples of BIT include passive monitoring such as activity tracking [58], and delivering psychological therapies via smartphone apps [59, 60]. The main advantage of BIT is that it can be deployed at low costs to a large number of patients who will otherwise not have access to the

traditional one-to-one exercise programs or psychological therapies, thus addressing important public health problems.

Secondary use. Data generated via BIT can be leveraged for app surveillance as a secondary use case of the data. There are currently websites that exist to facilitate the evaluation of health apps (e.g. psyberguide.org for mental health apps). While evaluation can be based on expert opinions, the opportunities of using PGHD for app monitoring will be enormous not only because of the large number of users and high volume of data, but also because data are collected in an environment that reflects behaviors as they naturally occur in real time. The large number of health apps available poses a challenge in how to evaluate them. There are over 97,000 health apps in the wild [61], while many care providers are introducing multiple mobile tools into the regulated processes of medical care in their own systems. In addition, due to the rapid development cycle of apps and constant update, each app will have a relatively short “shelf life”, thus limiting the user horizon that may benefit from the apps [62]. Therefore, knowing soon the health value of an app would benefit all stakeholders. This motivates the secondary use of PGHD via BIT for app surveillance using adaptive designs.

For the sake of argument we assume that patients look for an app for their health condition and, rather than choosing an app, accept an app assigned by a central system that utilizes a randomized clinical trial (RCT) to provide evidence for superiority among the tested apps. While generally in accordance with principles of therapy research, the RCT (or A/B testing) is impractical in light of the large number of health apps and short shelf-life problem described. To illustrate, consider a system comparing the utility of 10 apps by patients. Utility is measured by a pre-specified clinically meaningful use metric (e.g., total number of app sessions patients engaged in). In order to compare app A with a 30% utility rate and app B with a 40% utility rate, the RCT will require $n = 477$ subjects assigned to each app in order to have 90% power to declare app B has a superior utility based on a chi-squared test at 5% significance. Thus, to evaluate all 10

apps on this system simultaneously, it will roughly be a 5,000-subject study.

Adaptive designs are a collection of statistical tools for clinical trials that aim to increase the efficiency of evaluation and to streamline the drug development process [63]. The basic idea is to use interim data obtained in a study to inform the treatment decisions of the prospective enrolled subjects. One such adaptive design is the sequential elimination [64, 65], by which small batches of subjects will be sequentially enrolled, randomized evenly to different apps, and observed for their app utility. Apps that trail on the empirical use rate by a pre-specified margin (denoted d) will be eliminated and no longer be allocated to patients by the system.

Table 2 gives an example of how a sequential elimination process may run on a 10-app platform, with an elimination margin $d=10$. In the simulated trial, after 18 subjects were randomized to each app, one out of 18 subjects had a meaningful use of app E, whereas 11 out of 18 subjects used app C; the former (app E) was eliminated and discontinued from the platform. Similarly, after 25 subjects in each remaining app, app C (15 uses out of 25) eliminated app F (5 out of 25). The process continued in a similar fashion until only one app remained: This simulated trial reached the conclusion with recommending app C after a total of 1,664 subjects. On average, under a scenario where one app has a 40% use rate and the other nine have 30% use rate, this design will reach a conclusion with 941 subjects and will be able to recommend the correct app with 92% probability (“power”). This represents a five-fold reduction of the sample size required with the conventional RCT.

Sequential elimination can be viewed as a special case of adaptive randomization (AR) that alters the allocation ratios favorably to interventions with superior empirical performance during a study [66]. Table 2 gives the simplest form of AR for illustration purposes. More sophisticated statistical models and reinforcement learning techniques can be used to account for user characteristics and accommodate delayed observations [67], and incorporate the dynamic nature of app use [68]. All these methods consider a “closed system” in that the app pool is static and the

Table 2 A hypothetical trial using sequential elimination: an app is eliminated at an interim analysis N if the number of meaningful uses trails another app by at least $d = 10$. Shaded areas indicate enrollment has stopped for a given app.

N*	Number meaningful use (“success”)									
	App A	App B	App C	App D	App E	App F	App G	App H	App I	App J
18	8	4	11	5	1	4	4	4	5	5
25	8	8	15	6		5	6	6	7	7
31	12	12	16	6			8	8	11	9
56	25	24	22				15	20	22	18
100	34	35	37					40	34	30
111	37	40	40					45	35	
115	37	42	41					47		
240		72	75					82		
484			150					140		

*N = Number of users assigned to each app that has not be eliminated at the interim analysis.

number of apps evaluated is fixed. However, in the context of an open app platform (e.g. Google Play), the number of apps is constantly changing. Even for app platforms within a care organization, the app ecosystem will also change with new apps or updates occurring in a staggered fashion. Extensions of these AR procedures are currently under investigation for these important realistic settings [59, 69]. Importantly, in all these designs, assigning health apps using AR will enhance the compliance of patients, as they will receive the better apps with higher likelihoods. Thus, this is an example where the secondary use will indeed enhance the primary purpose of the intervention.

3.4 Focus Area 4: PGHD for Biomedical Research

Primary purpose. In recent years, and in parallel to the evolution of digital health, biomedical research has recognized the central role played by environmental factors in disease progression, and is beginning to

study these risk factors at the individual level. The concept of the exposome, coined by Wild [70], proposes a formal and systematic framework for the study of all environmental exposures to which an individual is subject, from conception to the end of her life, and demands a systematic research effort equivalent to what has been done to characterize the human genome (and also the human phenome).

Although the elaboration of the complete exposome of an individual is still far from the reach of research laboratories, because of its enormous complexity [71] and relatively recent definition, it is now possible to carry out studies of partial exposomes, for example, focused on a disease [72], health condition [73], organ [74], geographical location [75], or employment status [76]. Exposure information in the broad sense comprises all non-genetic data of an individual (including behavioral factors, social determinants of health, physico-chemical exposures), and these data can be obtained from multiple sources, including biomarkers (molecules reporting exposure to particular environmental agents such as smoking

[77]), geographic information systems, environmental questionnaires (e.g. NHANES) [78], EHRs [79], but also through PGHD (surveys, sensors, self-quantification systems, mobile apps, and Direct to Consumer Services) [80]. These partial exposures should be analyzed in conjunction with genetic and phenotypic data in order to better understand gene-environment interactions and the underlying causes of diseases.

Secondary use. Scientific endeavors such as the US **Precision Medicine Initiative (PMI)** (*All of Us* Research Program) explicitly recognize that to understand diseases better, it will not be sufficient to sequence a large number of personal genomes [81]. On the contrary, the initiative raises the need to collect primary PGHD (or participant-provided information, as it is called in this initiative), making use of personal digital health (participatory) technologies such as web-based surveys, wearable devices, and smartphone apps. This research program, which aims at generating a data repository of one million

participants or more, joins other projects that also use digital health technologies to collect primary PGHD data for biomedical research, such as the Health eHeart Study [82], or the Health Data Exploration Network [83]. In all of them, the investigators aim at collecting PGHD that reports environmental risk factors making use of digital technology in large populations of healthy individuals and patients.

There are still major challenges to making PGHD routinely used in medicine (incomplete data, reliability, noise, standardization, etc.), but PGHD offers a unique opportunity for biomedical research [84]. The use of digital health technologies, coupled with advances in the characterization of individual exposomes and the development of participatory medicine, converges in projects (e.g., PMI) of high potential to support truly integrative research approaches (gene, environment, phenotype) for secondary use of PGHD. This can further enable the generation of new hypotheses about health maintenance and disease development, encourage advances in prevention, and ultimately bring new solutions for predicting individual risk and generating new truly individualized diagnostic and therapeutic solutions.

So, data associated with environmental characteristics to which individuals are exposed is being primarily collected, complementing omics-data, for research purposes, and to explain under which circumstances dormant risks are likely to materialize. This data will serve the secondary purpose of informing the development of new preventive, diagnostic, and therapeutic solutions applicable in patient care.

4 Discussion

In the four focus areas, the secondary use cases of PGHD fill the following gaps: (1) disease process management, (2) continuous data flow over time for decision-making, (3) personal informatics tooling evaluation, and (4) environmental cohort-based risk pattern mining for precision disease understanding. These focus areas push the envelope of varying innovations, ranging

from integrating and interpreting highly diverse data (Focus Areas 1 & 4), achieving easy access to various data for patients who already expend significant efforts to manage their conditions (Focus Areas 2 & 3), and efficiently selecting quality of services from a plethora of candidates (Focus Area 3). All presented innovations also face the challenges of quality/reliability, interpretability, integration, and ethical judgment. In the remaining of the paper, we will outline the challenges and discuss the possible ways to make possible the synergy among initiatives and platforms that is necessary for the success of the secondary use of PGHD.

4.1 Quality and Reliability Issues of PGHD

Despite all the opportunities identified in the four focus areas in Section 3, the quality and reliability of the different PGHD data sources vary. Take the patient health mobile/wearable/sensor data used in Focus Areas 2 and 3 as example: evidence is strong for the reliability of physical activity measures [85], but weaker for sleep patterns and heart rate measures [86]. Another dimension of quality and reliability in view of the patients' right to data integrity is about information security and trust management. With the increasingly complex healthcare ecosystem of PGHD, the attack surface for malevolent intruders gets larger as we speak [87]. The issues are especially paramount as "Bring-Your-Own-Device" (BYOD) is considered as a norm to enhance patients' convenience in a healthcare environment. This has further expanded the risk to all the focus areas that use connected devices (e.g., environment monitoring IOT devices in Focus Area 3) and Patient Reported Outcome Measures (e.g., Patient Health Questionnaire-2 PHQ2) for tracking emotional distress for care coordination purposes in Focus Area 1).

To further understand the implications of PGHD with varying quality and reliability, clinical trials have been conducted to examine the clinical value of PGHD in care flows. Yet the results vary widely, with some showing no behavioral improvement over standard care [88, 89]. Therefore, al-

most across all focus areas, the secondary analysis of PGHD is expected to integrate some machine learning and statistical tools to assess quality and reliability, filter noise, and derive useful information from the vast and raw PGHD.

4.2 Interpretability Issues of PGHD

Another common challenge across the four focus areas is related to the interpretability of the secondary analysis of PGHD. The interpretability issues that impede the progress of using PGHD in practice come from two major sources: (1) the interpretability of the clinical content that is needed for putting PGHD in context, and (2) the interpretability of the computational models.

For improving the interpretability of clinical content, it would be important to start investigating how to further incorporate data-driven, cohort-based risk patterns to foster personalized treatment [90]. For example, as mentioned in Focus Areas 2 and 3, mobile/wearable data-driven insights can help identify emotional distress and any adverse reactions of cancer treatment in an early stage. These insights, when interpreted properly, can help locate appropriate mental health care providers for assistance and support [55]. The same is applicable to many other exogenous health determinants, which are often difficult to detect in traditional EHRs using prospective study designs. In Focus Areas 1, 2 and 4, the development of clinical content interpretation for secondary analysis of PGHD, with appropriate tools to select the right patient, right data, right problem, at the right time, has a huge potential to discover new risk factors for decision support and program design of disease management.

For improving the interpretability of computational models, a few studies in the vein of Focus Area 4 have started emerging and building informatics tools to discover novel risk factors associated with a variety of disease phenotypes. Similarly in Focus Area 2, once the risk factors for a particular disease phenotype have been identified, digital devices, sensors, and smartphones can be used for designing new interventions to

mitigate the effect of those risk factors [91, 92]. One method to do this is to analyze the parameters obtained from the computational model and then to map them back to the original features space. Afterwards, those features that are selected by the models can act as the potential risk factors of the disease. For example, model coefficients obtained from logistic regression can easily be transformed into log odds to provide clinical interpretations [93].

Another popular method of interpreting complex models is given with the rule-based or decision tree based model, which both can provide nice intuitive interpretations about risk factors and their interactions [94]. Other popular techniques include visualization techniques for depicting complex clinical models into an easily digestible form [95]. For example, visualizations can leverage the sequences of events occurred in a patient's healthcare utilization during a particular time period to build a patient's health trajectory [96] (Focus Areas 1 and 2). Likewise, stepwise elimination of inferior therapies like in Focus Area 3 can be easily communicated by comparing with the inferiority margin.

Despite initial evidence on the feasibility of developing tools to address the interpretability issues for PGHD secondary use, the validity of such tools is still an area calling for investigation.

4.3 Integration of PGHD with other Types of Healthcare Data

As human diseases are often complex being affected by multiple health determinants, the value from the secondary use of PGHD often hinges on the ability to uncover PGHD-driven insights in perspective with other data sources. For example, in Focus Area 4, multiple data sources (ranging from EHRs, genetics, environmental data, and PGHD) are combined to infer patient status and disease progression. However, there still exist challenges that require special caution while building integrated solutions for clinical decision-making.

First, the heterogeneous nature of the diverse data sources incurs pre-integration barriers that are not trivial. As it is most

distinctly shown in Focus Areas 1 and 4, each of the data sources provides a snapshot of the patient's health trajectories collected by different stakeholders, at different time points, and even using different technologies. Table 1 has given a detailed account on how the diverse types of EHR and PGHD are not even accessible uniformly to patients and caregivers. Moreover, differences in the nature of data such as differences in data formats, types, dimensionalities, volumes, and properties also create problems for building analytical methods integrating those datasets [97].

Second, the technologies for collecting both PGHD and EHR data can change over time. As exemplified in Focus Area 2 and 3, mobile- and sensor-based apps change rapidly over time. Therefore, the overall integration framework needs to be flexible and dynamic to allow better interoperability and electronic data sharing [75].

Third, there exist intricate relationships among health determinants across multiple data sources. For example, in Focus Area 4 interaction relationships can exist between a genetic determinant and an environment exposure, and the combined effect of the two determinants can incur a greater risk than the marginal effects of each determinant alone [98]. In addition, some might even inherit causal relationships with certain clinical markers. For example, a particular intervention can have some unwanted downstream effect on patient's behavioral aspects, which may ultimately lead to the disease under consideration. Accurately identifying such relationships, present among different health determinants, is among the future opportunities of Focus Areas 1, 2, and 4, is of great importance for proper clinical decision-making. Computational techniques should aim to find such type of subtle relationships present among different types of markers.

Fourth, while prior knowledge that can illustrate the relationships among multiple data sources is available, it is not always well captured. For example, to construct SCP (Focus Area 1), it is important to integrate prior knowledge about drug categories, hierarchical relationships among diagnostic codes, and clinical guidelines among interventions [99,46].

4.4 Societal and Authenticity Issues

One last recurring challenge throughout the focus areas is the privacy issue incurred by the collected and analyzed PGHD and authentication mechanisms. For example, monitoring data from fitness trackers, smartphone apps and posts in social networks (as used in Focus Area 2 and 3) are presumably authentic, but they flow outside the health care and public health arena. By that token, they are not meant for research and as a rule they have not been authorized for research use. Using traces of human behavior belongs to human subjects research for which either individual consent must be solicited or very good reason must be provided to Institutional Review Boards (IRBs) to exempt from subject consent. The act of consenting, however, jeopardizes authenticity. The Hawthorne effect [100] is an early manifestation of what can happen if human beings are informed that they will be subjects of an experiment. A comprehensive treatment of such effects and how they may invalidate results can be found in Chapter 15, especially Table 15.1 of [101].

While consent mechanisms have been established to authenticate data volunteered by patients (such as those in Focus Areas 1 through 3), we face wider ranging challenges with data collected *about* the patient (such as those in Focus Area 4). The latter characterize citizens through their physico-chemical and social exposure and hence immensely intrude on their privacy. Even if such data are collected anonymously, their combination with other data modalities as may occur in Focus Areas 1-3 increases the risk of re-identification.

4.5 Addressing Challenges with Multi-level Initiatives and Platforms for Behavior Modeling

The consumer and pervasive health informatics community are increasingly handicapped by the problem of not being able to leverage the secondary analysis of PGHD directly in care flows. As observed in Sections 4.1- 4.4, these challenges are multi-level and will require the whole community to start enabling technologies and standards so as to decrease the burden of collecting and lever-

aging evidence from PGHD. A multitude of proposals have been brought up to address each of the different challenges.

First, the quality and reliability issues call for attention to assess user behavior and, in any PGHD-integrated solutions and platforms (e.g., the Shared Care Plan in Focus Area 1, the *All of Us* Research Program in Focus Area 4), to further standardize report metrics as parts of the finer-grained PGHD management strategy [39, 40, 84].

Secondly, to further address the interpretability issues, it has been conjectured that the secondary use of PGHD needs to be coupled with user behavior-based nudging mechanism (e.g., incentives, incremental feedback, other strategic initiatives) and careful program designs to make it work. In any case, improving interpretability of clinical content and computational models in the context of user behavior is a crucial step to infer potential knowledge that is complementary to the already known clinical practices and guidelines.

Last but not least, to remove the barriers for integrating across multiple data sources, innovations are needed to build common data models (including standard communication pathways, reporting guidelines and PROM templates) that can capture intricate relationships among health determinants and prior knowledge in a flexible and dynamic framework.

Going forward, it would be essential to continue collecting evidence and examining the synergistic areas where an end-to-end solution of secondary use of PGHD can realize its value. In fact, across the various proposals for addressing the challenges, there has been one missing piece, that is, a behavior learning mechanism that can scan through PGHD to identify outcome-differential patient behavioral patterns. We expect more future research in this field will bridge the gap between the PGHD-driven insights and care practices.

4.6 Addressing Challenges with Ethics as a Tool

As observed in Section 4.4, societal and authenticity issues will need solutions that go beyond enabling technology. Among

all the possible solutions, an ethics-based utilitarian approach has shown potential to serve as a tool to address the challenge as well as other dilemmas that follow. Starting points that are widely recognized in biomedical ethics are four maxims [102], which include: (i) (respect for patient) autonomy, (ii) benevolence, (iii) non-maleficence, and (iv) distributive justice. Since these maxims can get into conflict, the utilitarian approach can be applied to evaluate desired and undesired outcomes of a decision. A decision is made in favor of that action that maximizes the positive minus the negative utilities. As an example, privacy violations may enter the “calculation” as negative utilities, while new general insights or individual treatment options are typical positive utilities.

Based on the ethics-based utilitarian approach, when trying to resolve the dilemma between scientific insights at the price of privacy violations, a utility function can then be derived to assign utility scores to a designated achievable set of insights and their consequences. For example, to determine whether a research proposal in Focus Area 4 should be approved, the utilitarian approach can be applied to evaluate both the negative utility of privacy breaches and the positive utility of a found exposome-genome pattern that might enable the early detection of certain cancers. Only if the former outperforms the latter could respective research achieve approval.

While the above example in Focus Area 4 was a dilemma between the maxims of benevolence and maleficence, in the Focus Area 3 we face a dilemma between autonomy and benevolence. Total autonomy would mean that patients freely choose among all treatment options and eventually pick one and cling to one with a mediocre outcome because they so will. Adaptive trials with sequential elimination dramatically cut autonomy: instead of myriads to choose from, the patients gets what he is assigned to. His remaining autonomy is to request an app or not. However, the likelihood to get an effective app is higher than in the free choice situation. So the approach deserves the attribute benevolent. Here as well, the negative utility of reducing autonomy must be weighed against two positive utilities: for the patient to have an effective app with a higher likelihood and for the community to find out which are the effective apps.

The sequential elimination has other positive utilities that do not come to mind easily. It protects the prospective patient from unwarranted opinion-related effects. It has been observed inside [101, section 3.3.1.1] and outside [103] the internet that clever rhetoric in the absence of true knowledge, maybe better characterized as bullying, can influence the selection and decision-making process in communities. It has also been observed that patients are not truly capable of judging physicians’ competences [104]. Therefore patients may actually be better off by declining their autonomy and trusting an insight formation process like in Focus Area 3.

5 Conclusion

Secondary use of PGHD brings Medical Informatics into a role of creating new clinical insight, be it about processes (Focus Area 1), risk assessments (Focus Areas 2 and 4), or therapeutic effectiveness (Focus Area 3). Medical Informatics has to find and define its role in how to deliver such insights. The new 2016 IMIA Code of Ethics for Health Informatics Professionals (HIPs) [105] has added as Part II-A various duties that HIPs have directly towards the patient. It is commented in [106] that through the advent of eHealth, HIPs become so central to the delivery of health care that they can no longer regard themselves as “supportive technical players”, but “acquired a fiduciary role *sui generis* that can no longer be ignored”. This may be seen as an invitation to market direct to the consumer. However, whether insights achieved through data gathering and computation alone and bypassing the medical profession truly advances the field is an open question.

References

1. Staccini P, Fernandez-Luque L. Health Social Media and Patient-Centered Care: Buzz or Evidence? *Yearb Med Inform* 2015;10(1):160–3.
2. Kutun T, Foller-Nord M, Wetter T. Characterizing Consumer Health Informatics in Low and Middle Income Countries. *Stud Health Technol Inform* 2015;216:973.
3. Miller EA, West DM. Where’s the revolution? Digital technology and health care in the Internet

- age. *J Health Polit Policy Law* 2009;34(2):261-84.
4. Fawcett T. Mining the Quantified Self: Personal Knowledge Discovery as a Challenge for Data Science. *Big Data* 2015;3(4):249-66.
 5. Naci H, Ioannidis JPA. Evaluation of Wellness Determinants and Interventions by Citizen Scientists. *JAMA* 2015;314(2):121-2.
 6. Societize Project. Citizen science for Europe: towards a better society of empowered citizens and enhanced research. Societize Consortium; 2013. Available from: https://www.researchgate.net/publication/259230549_Green_Paper_on_Citizen_Science
 7. Safavi K, Ratliff R, Webb K, MacCracken L. Accenture Consumer Survey of Patient Engagement. Accenture Consulting; 2016 [updated June 2016; cited Sep 2016]. Available from: <https://www.accenture.com/us-en/insight-research-shows-patients-united-states-want-heavy>
 8. PricewaterhouseCoopers. Personal Health Management: The Rise of the Empowered Health Consumers. PwC Strategy Report; 2015 [updated in May 2015; cited in Sep 2016]. Available from: https://www.pwc.com/us/en/advisory/customer/publications/assets/pwc-personal-health-management_final.pdf
 9. Gownder JP, McQuivey JL, Reitsma R, Gillett FE, Husson T, Ask JA, et al. Five Key Truths About Wearables That Every Leader Should Know: Wearables Are Poised to Change The Marketing Landscape. Forrester Research Report; 2015 [updated in Dec 2015; cited in Sep 2016]. <https://www.forrester.com/report/Five+Key+Truths+About+Wearables+That+Every+Leader+Should+Know/-/E-RES119783>
 10. Kummervold PE, Chronaki CE, Lausen B, Prokosch HU, Rasmussen J, Santana S, et al. eHealth Trends in Europe 2005-2007: A Population-Based Survey. *J Med Internet Res* 2008;10(4):e42.
 11. Charlesworth A, Davies A, Dixon J. Reforming payment for health care in Europe to achieve better value. London: The Nuffield Trust; 2012.
 12. Spetz J, Parente ST, Town RJ, Bazarko D. Scope-Of-Practice Laws For Nurse Practitioners Limit Cost Savings That Can Be Achieved In Retail Clinics. *Health Aff (Millwood)* 2013;32(11):1977-84.
 13. Catalyst for Payment Reform. First of Its Kind Scorecard on Medicare Payment Shows Widespread Payment Reform. Berkeley, California: CPR; 2015.
 14. Medicare Access and CHIP Reauthorization Act of 2015. Pub. L. No. 114-10; 2015 [updated in April 16; cited in Sep 2016]. Available from: <https://www.congress.gov/bill/114th-congress/house-bill/2/text/pl>
 15. American Medical Association. Medicare Physician Payment Reform: Understanding How MACRA Will Impact Physicians. Chicago, IL: AMA; 2016.
 16. Lee EO, Emanuel EJ. Shared Decision Making to Improve Care and Reduce Costs. *N Engl J Med* 2013;368:6-8.
 17. Levinson W, Kao A, Kuby A, Thisted, RA. Not all patients want to participate in decision making. A national study of public preferences. *J Gen Intern Med* 2005;20(6):531-5.
 18. Lynch W, Smith B, Slover M. Altarum Institute Survey of Consumer Health Care Opinions 2012 [updated fall 2013; cited in Sep 2016]. Available from: <http://altarum.org/sites/default/files/uploaded-related-files/Altarum%20Fall%202013%20Survey%20of%20Consumer%20Opinions.pdf>
 19. Martin-Sanchez FJ, Lopez-Campos GH. The New Role of Biomedical Informatics in the Age of Digital Medicine. *Methods Inf Med* 2016 Oct 17;55(5):392-402.
 20. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health and Quality of Life Outcomes. 2006;4:79. Available from: <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>
 21. Litt MD, Cooney NL, Morse P. Ecological momentary assessment (EMA) with treated alcoholics: Methodological problems and potential solutions. *Health Psychology* 1998;17(1):48-52.
 22. Sung M, Marci C, Pentland AS. Objective Physiological and Behavioral Measures for Identifying and Tracking Depression State in Clinically Depressed Patients. MIT Technical Report. 2005;595:1-20.
 23. Hsueh PS, Zhu X, Deng V, Ramarishnan S, Ball M. Dynamic and accretive composition of patient engagement instruments for personalized plan generation. *Stud Health Technol Inform* 2014;201:447-51.
 24. Teixeira PJ, Carraça EV, Marques MM, Rutter H, Oppert J-M, Bourdeaudhuij ID, et al. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. *BMC Med* 2015;13:84.
 25. Machlin SR, Soni A. Health Care Expenditures for Adults With Multiple Treated Chronic Conditions: Estimates From the Medical Expenditure Panel Survey 2009. *Prev Chronic Dis* 2013;10:120172.
 26. Ward BW, Schiller JS. Prevalence of Multiple Chronic Conditions Among US Adults: Estimates From the National Health Interview Survey, 2010. *Prev Chronic Dis* 2013;10:E65.
 27. Gibson HG, Reinhard S. Caregiving in the US. Washington, DC: AARP Public Policy Institute; 2015 [updated June 2015]. Available from: <http://www.caregiving.org/caregiving2015/>
 28. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple Chronic Conditions: Prevalence, Health Consequences, and Implications for Quality, Care Management, and Costs. *J Gen Intern Med* 2007;22(3):391-5.
 29. Eichner J, Blumenthal D. Medicare in the 21st century: building a better chronic care system. SSRN [updated June 2003]. Available from: <https://ssrn.com/abstract=555632>
 30. Gerteis J, Izrael D, Deitz D, LeRoy L, Ricciardi R, Miller T, et al. Multiple chronic conditions chartbook. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ) Publications. 2014 (AHRQ Pub. No. 14-0038).
 31. Boyd CM, Darer J, Boulton C, Fried LP, Boulton L, Wu AW. Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: Implications for pay for performance. *JAMA* 2005;294(6):716-24.
 32. Tinetti ME, Bogardus ST, Jr., Agostini JV. Potential pitfalls of disease-specific guidelines for patients with multiple conditions. *N Engl J Med* 2004 Dec 30;351(27):2870-4.
 33. Schoen C, Osborn R, How SKH, Doty MM, Peugh J. In Chronic Condition: Experiences Of Patients With Complex Health Care Needs, In Eight Countries, 2008. *Health Aff (Millwood)* 2009;28(1):w1-w16.
 34. Osborn R, Squires D, Doty MM, Sarnak DO, Schneider EC. In New Survey Of Eleven Countries, US Adults Still Struggle With Access To And Affordability Of Health Care. *Health Aff (Millwood)* 2016; 35:2327- 36.
 35. Whatcom County Pursuing Perfection Project. My Shared Care Plan. Institute for Healthcare Improvement [last visited March 22 2017] Available from: <http://www.ihl.org/resources/Pages/Tools/MySharedCarePlan.aspx>
 36. Samal L, Hasan O, Venkatesh AK, Volk LA, Bates DW. Health Information Technology to Support Care Coordination and Care Transitions: Data Needs, Capabilities, Technical and Organizational Barriers, and Approaches to Improvement. Washington, DC: National Quality Forum; 2012.
 37. Cheng C, Stokes TH, Wang MD, editors. caRE-MOTE: the design of a cancer reporting and monitoring telemedicine system for domestic care. Engineering in Medicine and Biology Society (EMBC) 2011 Annual International Conference of the IEEE; 2011.
 38. Hesse BW, Hanna C, Massett HA, Hesse NK. Outside the Box: Will Information Technology Be a Viable Intervention to Improve the Quality of Cancer Care? *J Natl Cancer Inst Monog* 2010;2010(40):81-9.
 39. Dykes PC, Samal L, Donahue M, Greenberg JO, Hurley AC, Hasan O, et al. A patient-centered longitudinal care plan: vision versus reality. *J Am Med Inform Assoc* 2014 Nov-Dec;21(6):1082-90.
 40. Standards & Interoperability Framework. Longitudinal Coordination of Care: Interoperable Care Plan Exchange Use Case v2.0. Washington, DC: Office of the National Coordinator for Health Information; 2014.
 41. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. *J Am Med Inform Assoc* 2015 Jul;22(4):821-30.
 42. Grace T. Wikis as a knowledge management tool. *Journal of Knowledge Management* 2009; 13(4):64-74.
 43. Molyneux T, Brumley J, editors. The use of wikis as a management tool to facilitate group project work. Proc AAEE Conference; 2007.
 44. Varga-Atkins T, Dangerfield P, Brigden D. Developing professionalism through the use of wikis: A study with first-year undergraduate medical students. *Med Teach* 2010;32(10):824-9.
 45. Morley DA. Enhancing networking and proactive learning skills in the first year university experience through the use of wikis. *Nurse Educ Today* 2012 4//;32(3):261-6.
 46. Mandl KD, Kohane IS. No Small Change for the Health Information Economy. *N Engl J Med* 2009;360(13):1278-81.

47. The Smartphone Difference. Pew Internet Survey 2016 [last visit 2016-12-20]. Available from: <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/>
48. PwC. The wearable life 2.0: Connected living in a wearable world. Consumer Intelligence Series; 2016. <http://pwc.com/CISwearables> last visit 2016-12-20
49. American Society of Clinical Oncology. The state of cancer care in America, 2016: A report by the American Society of Clinical Oncology. *J Oncol Pract* 2016;12(4):339-83.
50. Nkolaichuk CL, Bruera E, Spachynski K, MacEachern T, Hanson J, Maguire TO. A comparison of patient and proxy symptom assessments in advanced cancer patients. *Palliat Med* 1999;13:311-23.
51. Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behavior treatments. *Br J Health Psychol* 2010;15(Pt.1):1-39.
52. Patrick K, Wolszon L, Basen-Engquist K, Demark-Wahnefried W, Prokhorov A, Barrera S, et al. CYberinfrastructure for COmparative effectiveness REsearch (CYCORE): improving data from cancer clinical trials. *Transl Behav Med* 2011;1:83-8.
53. Peterson SK, Shinn EH, Basen-Engquist K, Demark-Wahnefried W, Prokhorov AV, Baru C, et al. Identifying early dehydration risk with home-based sensors during radiation treatment: A feasibility study with head and neck cancer patients. *J Natl Cancer Inst Monogr* 2013;47:162-8.
54. Clauser SB, Wagner EH, Aiello Bowles EJ, Tuzzio L, Greene SH. Improving modern cancer care through information technology. *Am J Prev Med* 2011;40(5 Suppl 2):S198-S207.
55. Place S, Blanch-Hartigan D, Rubin C, Gorrostieta C, Mead C, Kane J, et al. Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. *J Med Internet Res* 2017;19(3):e75.
56. Wood WA, Bennett AV, Basch E. Emerging uses of Person Generated Health Data Person Generated Health Data in clinical research. *Mol Oncol* 2015;9(5):1018-24.
57. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen Hosp Psychiatry* 2013;35(4):332-8.
58. Hu X, Hsueh P-Y, Chen C, Qian M, Diaz K, Cheung Y-K. Estimation of adaptive and interpretable policy for fitness app users. *IBM Journal of Research and Development* [abstract accepted Oct 2016; to appear in 2017].
59. Mohr DC, Cheung K, Schueller SM, Brown CH, Duan N. Continuous evaluation of evolving behavioral intervention technologies. *Am J Prev Med* 2013;45:517-23.
60. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well being apps. *Internet Interv* 2016;4:152-8.
61. research2guidance. Mobile health market report 2013—2017: the commercialization of mHealth applications (Vol. 3). Berlin: research2guidance; 2013 [last visited 2017-04-03]. Available from: <https://research2guidance.com/?s=mobile+health+market+report+2013-2017>
62. Cheung K, Duan N. Design of implementation studies for quality improvement programs: an effectiveness-cost-effectiveness framework. *Am J Public Health* 2014;104(1):e23-e30.
63. Cheung K, Kaufmann P. Efficiency perspectives on adaptive designs in stroke clinical trials. *Stroke* 2011;42(10):2990-4.
64. Levin B, Robbins H. Selecting the highest probability in binomial or multinomial trials. *Proc Natl Acad Sci U S A* 1981;78(8):4663-6.
65. Cheung YK. Simple sequential boundaries for treatment selection in a multi-armed randomized clinical trials with a control. *Biometrics* 2008;64:940-9.
66. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY-2: An adaptive breast cancer trial design in the setting of neo-adjuvant chemotherapy. *Clin Pharmacol Ther* 2009;86:97-100.
67. Cheung YK, Inoue LYT, Wathen JK, Thall PF. Continuous Bayesian adaptive randomization based on event times with covariates. *StatMed* 2006;25:55-70.
68. Cheung YK, Chakraborty B, Davidson KW. Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics* 2015;71:450-9.
69. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. *J Med Internet Res* 2015;17(7):e166.
70. Wild CP. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;14(8):1847-50.
71. Cui Y, Balshaw DM, Kwok RK, Thompson CL, Collman GW, Birnbaum LS. The Exposome: Embracing the Complexity for Discovery in Environmental Health. *Environ Health Perspect* 2016;124(8):A137-40.
72. Smith MT, Zhang L, McHale CM, Skibola CF, Rappaport SM. Benzene: the exposome and future investigations of leukemia etiology. *Chem Biol Interact* 2011;192(1-2):155-9.
73. Goldfarb DS. The exposome for kidney stones. *Urolithiasis* 2016;44(1):3-7.
74. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The blood exposome and its role in discovering causes of disease. *Environ Health Perspect* 2014;122(8):769-74.
75. Donald CE, Scott RP, Blaustein KL, Halbleib ML, Sarr M, Jepson PC, et al. Silicone wristbands detect individuals' pesticide exposures in West Africa. *R Soc Open Sci* 2016;3(8):160433.
76. Faisandier L, Bonnetterre V, De Gaudemaris R, Bicout DJ. Occupational exposome: a network-based approach for characterizing Occupational Health Problems. *J Biomed Inform* 2011 Aug;44(4):545-52.
77. Ding YP, Ladeiro Y, Morilla I, Bouhnik Y, Marah A, Zaag H, et al. Integrative Network-based Analysis of Colonic Detoxification Gene Expression in Ulcerative Colitis According to Smoking Status. *J Crohns Colitis* 2016 Oct 4. pii: jcw179.
78. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2016 [last visit 2017-0403]. Available from: <https://www.cdc.gov/nchs/nhanes/>
79. Gottlieb L, Tobey R, Cantor J, Hessler D, Adler NE. Integrating Social And Medical Data To Improve Population Health: Opportunities And Barriers. *Health Aff (Millwood)* 2016 Nov 1;35(11):2116-23.
80. Martin Sanchez F, Gray K, Bellazzi R, Lopez-Campos G. Exposome informatics: considerations for the design of future biomedical research information systems. *J Am Med Inform Assoc* 2014;21(3):386-90.
81. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372(9):793-5.
82. Health eHeart. [last visited 2017-04-03]. Available from: <https://www.health-ehartstudy.org/>
83. Health Data Exploration Project. [last visited 2017-04-03; cited 19/07/2015]; Available from: <http://hdxplore.calit2.net/wp/>
84. Swan M. Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *J Pers Med* 2012;2(3):93-118..
85. Scuffham PA, Nikles J, Michell GK, Yelland MJ, Vine N, Pillans PI et al.. Using N-of-1 trials to improve patient management and save costs. *J Gen Intern Med* 2010; 25:906–13.
86. Case MA, Burwick HA, Volpp KG, and Patel. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *JAMA* 2015;313(6):625-6.
87. 7 largest data breaches of 2015: <http://www.healthcareitnews.com/news/7-largest-data-breaches-2015> [last visited 2017-04-03]
88. Bloss CS, Wineinger NE, Peters M, Boeldt DL, Ariniello L, Kim JY et al., A prospective randomized trial examining health care utilization in individuals using multiple smartphone-enabled biosensors. *PeerJ* 2016;4:1554.
89. SANS Institute Survey: New Threats Drive Improved Practices: State of Cybersecurity in Health Care Organizations. 2014 [last visited 2017-0403]. [Avahttps://www.sans.org/reading-room/whitepapers/analyst/threats-drive-improved-practices-state-cybersecurity-health-care-organizations-35652](https://www.sans.org/reading-room/whitepapers/analyst/threats-drive-improved-practices-state-cybersecurity-health-care-organizations-35652)
90. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. *J Am Med Inform Assoc* 2008;15(4):473-83.
91. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med* 2013;10(1):e1001362.
92. Davenport A, Gura V, Ronco C, Beizai M, Ezon C, Rambod E. A wearable haemodialysis device for patients with end-stage renal failure: a pilot study. *Lancet* 2007;370(9604):2005-10.

93. Hailpern SM, Visintainer PF. Odds ratios and logistic regression: further examples of their use and interpretation. *Stata J* 2003;3(3): 213-25.
94. Westra BL, Dey S, Fang G, Steinbach M, Kumar V Oancea C, et al. Interpretable Predictive Models for Knowledge Discovery from Home-Care Electronic Health Records. *J Healthc Eng* 2016;2(1):55-74.
95. Krause J, Perer A, Ng K, editors. *Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models*. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM;2016.
96. Kwon BC, Verma J, Perer A. Peekquence: Visual analytics for event sequence data. In: *ACM SIGKDD 2016 Workshop on Interactive Data Exploration and Analytics*; 2016.
97. Dey S, Gupta R, Steinbach M, Kumar V. Predictive Models for Integrating Clinical and Genomic Data. In: Reddy CK, Aggarwal CC. *Healthcare Data Analytics* Boca Raton (CRC); 2015. p. 433-66.
98. Dey S, Lim K, Atluri G, MacDonald III A, Steinbach M, Kumar V. A pattern mining based integrative framework for biomarker discovery. In: *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*. ACM; 2012. p. 498-505.
99. Dey S, Simon G, Westra B, Steinbach M, Kumar V. Mining interpretable and predictive diagnosis codes from multi-source electronic health records. In: *Proceedings of the 2014 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics*; 2014. p. 1055-63.
100. Roethlisberger FJ, Dickinson WJ. *Management and the worker*. Cambridge: Harvard University Press; 1939.
101. Wetter T. *Consumer Health Informatics: New Services, Roles and Responsibilities*; Cham (Springer); 2016.
102. Beauchamp TL, Childress JF. *Principles of biomedical ethics*. 7th ed. Oxford/New York: Oxford University Press; 2013
103. Muchnik L, Aral S, Taylor SJ. Social influence bias: a randomized experiment. *Science* 2013;341:647-50.
104. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association Between Physician Online Rating and Quality of Care. *J Med Internet Res* 2016;18(12):e324
105. Kluge HE. The IMIA Code of Ethics for Health Information Professionals. IMIA endorsed document 2016. [last visited 2017-04-03]. Available from: <http://imia-medinfo.org/wp/wp-content/uploads/2015/07/IMIA-Code-of-Ethics-2016.pdf>
106. Kluge HE. The IMIA Code of Ethics: Revision 2016. IMIA endorsed document 2016-08-28. [last visited 2017-04-03]. Available from: <http://imia-medinfo.org/wp/wp-content/uploads/2015/07/IMIA-CODE-of-Ethics-Short.pdf>

Correspondence to:

Dr. Thomas Wetter
 Institute of Medical Biometry and Informatics
 Heidelberg University
 Im Neuenheimer Feld 130.3
 D-69120 Heidelberg
 Germany
 Tel +49 6221 56 7490
 Fax +49 6221 56 4997
 E-Mail: thomas.wetter@urz.uni-hd.de