Patient-Generated Health Data: Dimensions, Challenges, and Open Questions
Other titles in Foundations and Trends® in Human–Computer Interaction

*Human-Food Interaction*
Rohit Ashok Khot and Florian Mueller

*10 Lenses to Design Sports-HCI*
Florian Mueller and Damon Young
ISBN: 978-1-68083-528-1

*Values and Ethics in Human-Computer Interaction*
Katie Shilton
ISBN: 978-1-68083-466-6

*Research Fiction and Thought Experiments in Design*
Mark Blythe and Enrique Encinas
Patient-Generated Health Data: Dimensions, Challenges, and Open Questions

Mayara Costa Figueiredo
University of California
Irvine, USA

Yunan Chen
University of California
Irvine, USA
Editorial Scope

Topics

Foundations and Trends® in Human–Computer Interaction publishes survey and tutorial articles in the following topics:

- History of the research community
- Theory
- Technology
- Computer Supported Cooperative Work
- Interdisciplinary influence
- Advanced topics and trends

Information for Librarians

Foundations and Trends® in Human–Computer Interaction, 2020, Volume 13, 4 issues. ISSN paper version 1551-3955. ISSN online version 1551-3963. Also available as a combined paper and online subscription.
# Contents

1 Introduction 2
  1.1 Why Patient-Generated Health Data – The Benefits . . . . 4
  1.2 Definitions: Data and Practices . . . . . . . . . . . . . . . . . 6
  1.3 Overview of This Review . . . . . . . . . . . . . . . . . . . . . . . 19

2 Dimensions of PGHD 22
  2.1 Health Focus . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
  2.2 Type of Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
  2.3 Who Initiates Data Collection . . . . . . . . . . . . . . . . . 36
  2.4 Whose Data Are Collected . . . . . . . . . . . . . . . . . . . 40
  2.5 Who Uses the Data . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
  2.6 Mechanisms of Data Collection and Use . . . . . . . . . . . . 50
  2.7 Duration of PGHD Use . . . . . . . . . . . . . . . . . . . . . . . 53

3 Challenges 58
  3.1 Patient-Provider Collaboration . . . . . . . . . . . . . . . . . 58
  3.2 Complex Conditions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73
  3.3 Goals . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
  3.4 Reflection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
  3.5 Access to PGHD Technologies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89
  3.6 Expected and Unexpected Users . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92

Full text available at: http://dx.doi.org/10.1561/1100000080
4 Open Questions for PGHD Research

4.1 The Complex Social Context

4.2 The Evolving Scope of Health

4.3 The Design Culture

4.4 Final Remarks

References
Patient-Generated Health Data: Dimensions, Challenges, and Open Questions

Mayara Costa Figueiredo\textsuperscript{1} and Yunan Chen\textsuperscript{2}

\textsuperscript{1}University of California, Irvine; mcostafi@uci.edu
\textsuperscript{2}University of California, Irvine; yunanc@ics.uci.edu

ABSTRACT

In this review, we present an overview of patient-generated health data (PGHD) research, focusing on important aspects that inform and define studies in the area. We start by exploring a fundamental question: what is patient-generated health data? We list the main terms and definitions identified from previous research and generate a set of seven key dimensions for understanding PGHD: (1) the health focus of the study, (2) the type of data, (3) who proposes the use of PGHD, (4) whose data are collected, (5) who are the intended users, (6) how PGHD is collected and used, and (7) the duration of PGHD use. We describe these dimensions and discuss their importance to research PGHD. We then present a discussion of the impact of PGHD and related practices in people’s lives and the debates concerning the consequences, both positive and negative, that may arise.
The practices of collecting and using individuals’ or patients’ personal health data are not new (Cortez et al., 2018) and people have been recording data about themselves for hundreds of years (Rettberg, 2014), often with specific self-management goals (Neff and Nafus, 2016). It is well known that medical practices are essentially information-centric, and the decisions are largely based on patients’ data: it is through patients’ descriptions of symptoms that healthcare providers can order further tests and provide treatment plans (Loos and Davidson, 2016; Schroeder et al., 2017). To better access and utilize patients’ data, numerous studies have advocated patients’ active participation as a key factor for enhancing the quality of their healthcare. These studies suggest that patients should have access and contribute to the generation of their health data, and be directly involved in their own healthcare decisions (Shapiro et al., 2012).

Over the past years, individuals’ and patients’ health data have been increasingly present in public debate: from stories about the benefits of accessing and using these data to reports of privacy breaches and potential negative consequences (e.g., Harwell, 2019; Rowl, 2019; Siegel, 2019). Recently there has been a proliferation of new technologies,
particularly sensor and mobile apps, produced to measure and track different aspects of a person’s health and behavior: industry reports state that as of 2018 there were more than 325,000 health related apps available for consumers (Dabbs, 2018). These data are often termed patient-generated health data (PGHD), a research topic that has been increasingly examined by multiple fields of studies, such as Human-Computer Interaction, Computer Science, Health Informatics, Medicine, Psychology, Science and Technology Studies, and Social Sciences, to cite a few.

The use of PGHD is not a new phenomenon; its popularity is associated with the recently increasing interest in patients’ data and the rapid development of technologies that can facilitate data collection and use (Consolvo et al., 2008; Cortez et al., 2018). Specifically, two developments have influenced the popularization of PGHD (Neff and Nafus, 2016): the first one is the technology itself. Mobile phones, sensors, and connectivity are pervasive, expanding the presence of technology in our lives and providing the basis for the development of systems that can track a greater number of aspects in a greater frequency and detail. The second development is related to a culture of biomedicalization, or the expansion of “medical jurisdiction, authority, and practices” “through the new social forms of highly technoscientific biomedicine” (Clarke et al., 2003). This culture is related to the increased interest in measuring and medicalizing aspects of life that were not previously medicalized, particularly through new technologies such as sensors and smartphone applications. For example, exercising is not only a habit anymore, it needs to be quantified and measured too (Brown, 2019). Together, these aspects promoted the growth of the social phenomenon of people tracking their own data (Neff and Nafus, 2016), including collecting health data outside of traditional clinical settings from multiple sources and with a rapidly increasing volume (Cortez et al., 2018; Shapiro et al., 2012). Recent movements such as the Quantified Self (Quantified Self, n.d.; Wolf, 2009, 2010) add up to this increased popularity of systems, habits, and research concerning health data generated by patients (and non-patients) as they go about their daily lives.

Many researchers point to a paradigm shift in healthcare from a clinical-centered to a more patient-centric practice, in which patients
Introduction

have a more active role in their care (Demiris et al., 2008; Grönvall and Verdezoto, 2013a; Hong et al., 2016; Loos and Davidson, 2016; Mamykina et al., 2008; O’Kane and Mentis, 2012; O’Kane et al., 2016; Paton et al., 2012; Zhu et al., 2016). The new view puts patients and their own health data at the center of the healthcare practices. Considering the wide impact and potential benefits of PGHD in healthcare, it is necessary to understand the current landscape and scope of PGHD research, so we can support good practices, work to improve areas that need more attention, promote PGHD benefits, and avoid negative consequences, such as reinforcing negative social stereotypes or increasing health disparities.

In this review we present an overview of the extensive literature related to PGHD, ranging from an attempt to characterize the research to a discussion of the impact of these practices on people’s lives and the debates concerning the consequences, both positive and negative, that may arise. Based on the literature, we identified important dimensions to define the research and design scope, and pinpoint several challenges in researching and developing technologies for PGHD. We defined these dimensions through examining several streams of literature related to PGHD and also through our own previous research in the area.

The remainder of the review is organized as follows. In the first section we summarize the potential benefits of using PGHD and explore vocabularies, definitions, and scopes used in a diverse set of studies on health and health-related data generated and used by patients and non-patients. In the second section we translate this discussion into seven dimensions that can be used to categorize and define the scope of studies related to PGHD. The third section focuses on the main challenges of researching and developing for PGHD. Finally, the fourth section explores important open questions for PGHD research.

1.1 Why Patient-Generated Health Data – The Benefits

PGHD have several benefits to healthcare, including changing healthcare practices to provide more information concerning patients’ health and quality of life (Cortez et al., 2018; Jacobs et al., 2015; Raj et al., 2019). These benefits can impact multiple stakeholders, especially healthcare
providers and patients. In this subsection, we briefly describe the multiple benefits from using PGHD as reported by the literature.

Providers can benefit from PGHD in numerous ways. These data can support personalized care (Cortez et al., 2018; Loos and Davidson, 2016; Zhu et al., 2016), potentially leading to new insights about patients’ health status, conditions, or treatment results (Zhu et al., 2016) and improving or facilitating diagnosis and treatment plans (Chen, 2011; Chung et al., 2016, 2019; Jacobs et al., 2015; Loos and Davidson, 2016; Schroeder et al., 2017; West et al., 2016) by providing important measures of lifestyle and personal behavior that may be missed during consultations (Schroeder et al., 2017; Zhu et al., 2016). PGHD can also provide further context about patients’ health and health behaviors, reveal unexpected side effects, enable timely and cost-effective interventions (Cheng et al., 2015; Cortez et al., 2018; Frost et al., 2011; Loos and Davidson, 2016; Nundy et al., 2014; West et al., 2016), and provide crucial support to continuity of care or patient adherence (Chung et al., 2019; Demiris et al., 2008; Murnane et al., 2018; Nundy et al., 2014). Some studies also describe increased benefits for specific “sub-areas” of healthcare. For example, Grönvall and Verdezoto (2013a) highlight the potential benefits of PGHD for elderly care, e.g., supporting a more independent life outside of clinical settings. Other benefits are related to patient-provider interaction. Sanger et al. (2016) mention that PGHD can improve “clinical outcomes and patient satisfaction” (Sanger et al., 2016) by making providers more accountable and improving patients’ engagement and self-management. Many studies also argue that these data can improve patient-provider communication and foster shared decision-making (Cheng et al., 2015; Chung et al., 2016, 2019; Cortez et al., 2018; Jacobs et al., 2015; Loos and Davidson, 2016; O’Kane and Mentis, 2012; Schroeder et al., 2017; Zhu et al., 2016).

On the patient side, PGHD are seen as useful for patient empowerment (Ayobi et al., 2017; Demiris et al., 2008; Grönvall and Verdezoto, 2013a; Tang et al., 2012). These data can serve as important memory aids for patients during time-constrained medical consultations (i.e., recording important facts that happen in the sometimes long period between appointments) (Cheng et al., 2015; Jacobs et al., 2015; Loos and Davidson, 2016; Mishra et al., 2019; Nundy et al., 2014;
Introduction

Tang et al., 2012). They can also provide support in monitoring and mitigating symptoms and delaying or preventing progression of chronic diseases (Chung et al., 2016; Demiris et al., 2008). Additionally, PGHD are believed to enhance patients’ knowledge about their health condition, self-awareness, and understanding of their own health, behavior, and lifestyle—aspects that are fundamental for individuals’ general wellness and illness management (Choe et al., 2015; Grönvall and Verdezoto, 2013a; Li et al., 2011; Mamykina et al., 2008; O’Kane et al., 2016; Pina et al., 2017). PGHD can also be used to identify possible associations in health events, e.g., identifying the trigger of an allergic reaction (Chung et al., 2019; Karkar et al., 2015a,b, 2017; Pina et al., 2017). In this sense, these data can be used to support patients’ reasoning regarding, e.g., their current health status and future trends for their conditions, the relationship between their health status and daily health behaviors, and the important and effective ways to manage illness and health (Barbarin et al., 2016; Johansen and Kanstrup, 2016). Therefore, PGHD can be used to explore alternative approaches of self-management beyond clinical interventions.

1.2 Definitions: Data and Practices

Although PGHD have been extensively studied and frequently mentioned in prior literature, to date there isn’t a unified definition for PGHD, largely because of their interdisciplinary nature and multiple fields of inquiry. The following definition was proposed by the Office of the National Coordinator (ONC) for Health Information Technology of the U.S. Department of Health and Human Services in its attempts to explore PGHD opportunities and challenges (Shapiro et al., 2012):

PGHD are health-related data—including health history, symptoms, biometric data, treatment history, lifestyle choices, and other information—created, recorded, gathered, or inferred by or from patients or their designees [...] to help address a health concern. PGHD are distinct from data generated in clinical settings and through encounters with providers in two important ways. First, patients, not
providers, are primarily responsible for capturing or recording these data. Second, patients direct the sharing or distributing of these data to health care providers and other stakeholders. In these ways, PGHD complement provider-directed capture and flow of health-related data across the health care system.

As this definition shows, PGHD is a broad and loosely defined term that encompasses health-related data generated by individuals outside of traditional care settings. The data can be in different types: physiological indicators measured by patients (e.g., temperature, weight), lifestyle data (e.g., exercise, diet), quality of life data (e.g., mood, sleep quality), symptoms of medical conditions, or any other information that helps in personalizing patients’ situations (Shapiro et al., 2012).

This definition of PGHD overlaps with many other related terms used in the literature, such as “personal health” (Sherman, 2016), “data relevant for healthcare” (Estrin et al., 2016), “personal health experience” (Chen, 2010), “patient-logged data” or “self-logged data” (West et al., 2016). These terms significantly overlap with PGHD but offer slightly different emphases and foci based on the field of study. In understanding and deciding the scope of this review, we first reviewed relevant terms and definitions, particularly the ones commonly used in medical and technology-oriented domains. In this review, we will briefly explore the following terms:

Definitions commonly used in health practices:

(a) Patient Health Outcomes (e.g., Street et al., 2009) and Patient Reported Outcomes (e.g., Black, 2013; Zhang et al., 2019)

(b) Journaling (e.g., Zhu et al., 2016)

(c) Self-management and variants, e.g., personal health information management, home care (e.g., Civan et al., 2006; Davies et al., 2019; Havas et al., 2016; Moen and Brennan, 2005)

(d) Remote Patient Monitoring (e.g., Cheng et al., 2015; Raj et al., 2019)
Introduction

(e) Self-monitoring (e.g., Choe et al., 2014, 2015; Grönvall and Verdezoto, 2013b; Paay et al., 2015; Snyder, 1974)

Concepts originated from Human Computer Interaction (HCI), Computer Science (CS), and Health Informatics:

(a) Self-tracking and Personal Informatics (e.g., Li et al., 2010, 2011; MacLeod et al., 2013; McKillop et al., 2018; Pina et al., 2017)

(b) Self-Experimentation (e.g., Karkar et al., 2015a,b, 2017)

(c) Observations of Daily Living – ODL (e.g., Brennan and Casper, 2015)

(d) Quantified Self (e.g., Choe et al., 2014; Gregory and Bowker, 2016; Johansen and Kanstrup, 2016; Neff and Nafus, 2016; Quantified Self, n.d.; West et al., 2016; Wolf, 2009, 2010).

1.2.1 Definitions Commonly Used in Health Practices

In medical research, a concept frequently used and close to PGHD is “Patient Health Outcomes.” This term concerns direct outcomes of treatment, such as disease markers (e.g., blood pressure, glucose levels), survival rates, and quality of life measures, such as “functioning and well-being in physical, psychological and social domains” (Street et al., 2009). Street et al. (2009) summarize health outcomes in terms of survival rates, cure or remission, decreased suffering, emotional well-being, pain control, functional ability, and vitality. But they also describe proximal outcomes, (understanding, satisfaction, clinician-patient agreement, trust, feeling ‘known,’ feeling involved, rapport, motivation) and intermediate outcomes (access to care, quality medical decisions, commitment to treatment, trust in the system, social support, self-care skills, emotional management) that can lead to the previously mentioned health outcomes (Street et al., 2009). Measuring these outcomes is a common goal and also a challenge. Some of these outcomes can be evaluated through clinical measures, e.g., recovery rate or remission. However, many others require considerable patient input, e.g., emotional well-being, pain levels, and vitality. These are often measured
through data that can be requested by healthcare providers, but that are generated, collected, and provided by patients as part of PGHD.

Similarly, Patient Reported Outcomes (PRO) focus on collecting and measuring the outcomes of healthcare, aiming to increase patient involvement. PROs are a “key measurement of the effectiveness of patient-centered care. PROs include patients’ self-reported symptoms, functional status, and health-related quality of life” (Zhang et al., 2019). They represent patients’ own views about their health status and care and can be used to compare providers’ performances (Black, 2013). Different measurements aim to determine PROs. Many of them are disease specific measures, which “are tailored to the symptoms and impact on function of a specific condition” (Black, 2013). Others are generic PRO measures, aiming to consider general aspects (e.g., self-care) common to multiple medical conditions (Black, 2013). These measures, although self-reported by patients (thus, related to PGHD), are often collected or requested by healthcare providers through questionnaires or questions during clinical appointments.

Journaling is another frequently used term in healthcare. Historically health providers ask patients to write their symptoms and other related information as they go in their daily lives in a journal, so that they can discuss them during appointments (Zhu et al., 2016). Studies in the medical field on chronic diseases also commonly use terms like self-management, or an individuals’ abilities to “manage the symptoms, treatment, physical and psychosocial consequences and lifestyle changes inherent to living with a chronic condition” (Davies et al., 2019). “Personal health information management” and “home care” are other terms similar to self-management, both referring to the activities patients perform in their daily lives to manage their health conditions, involving monitoring the status and progress of the condition, treatment adherence and outcomes, and quality of life (Civan et al., 2006; Havas et al., 2016; Moen and Brennan, 2005).

A more provider-oriented term is “remote patient monitoring,” which focuses on near real-time patients’ monitoring, through personal data collection at home and direct transmission of data to providers’ databases (Cheng et al., 2015). Remote patient monitoring often means that
patients have no or low access to, or influence over, the collection and use of their own data.

Finally, a common term rooted in healthcare practices is self-monitoring. Self-monitoring is traditionally used in behavioral psychology as the practice of recording one’s own thoughts, feelings, and behavior as part of assessments or treatments in behavior therapy (Choe et al., 2014, 2015). It originally focused on expressive behaviors, i.e., the practice of observing and controlling one’s self-presentation (Snyder, 1974). Although commonly applied in health practices, self-monitoring has also been used in research in technology-oriented areas, such as HCI, CS, and health informatics. In this case, it has been used with the connotation of monitoring health parameters, focusing on prevention or early detection of medical conditions (Grönnvall and Verdezoto, 2013b). Some of these studies also see self-monitoring as a technique for persuading people to improve health behavior (Paay et al., 2015). Additionally, self-monitoring is often used as a synonym of self-tracking or personal informatics (PI) (Choe et al., 2014; Li et al., 2010; Sanger et al., 2016), one of the currently most used terms for the practice of collecting and using PGHD, especially in HCI and other technology-related areas.

1.2.2 Concepts Rooted in HCI, CS, and Health Informatics

Li et al. (2010) introduced the term personal informatics (PI; also called self-tracking) to refer to systems that help people collect and reflect on personal information to gain and improve self-knowledge (Li et al., 2010). The definition includes the collection and use of health data as well as personal finances, emails, and other types of data (Ayobi et al., 2017; Rooksby et al., 2014). Later, Li et al. (2011) extended the definition to broadly encompass activities of self-tracking, not only technologies and systems (Li et al., 2011). Self-tracking and PI usually have a characteristic of repetition or periodicity: of repeatedly collecting and reflecting on one’s personal data to acquire self-knowledge or achieve a goal (Li et al., 2010). Concerning the data, Li et al. (2011) describe that PI data may include data about behavior and physiology, qualitative and quantitative current and historical data, and external data considered to be personally relevant (e.g., weather, if it impacts mood or exercise).
Pina et al. (2017) similarly use the term PI, but explicitly focus on health, defining it as the process of tracking behaviors, outcomes, and context to observe and adapt behavior. Their work calls attention to the “personal” part of the term, and they state that many aspects of one’s health impact and are influenced by other people, especially family members. Considering this aspect of who is affected or involved in self-tracking activities for health, Nissenbaum and Patterson’s (2016) taxonomy of health self-tracking lists three different types of actors “involved in the circuits of information flow:” initiators (who initiate the data tracking), data subjects (whose data is tracked), and data recipients (who use the data). In another study using the term PI and explicitly focusing on health, MacLeod et al. (2013) highlight the particular goals that patients with chronic or other serious conditions would have: they would be more interested in questions related to episodes (trends, how to prevent and deal with episodes, and consequences), medication (how to change dosage, efficacy, and side effects), and triggers (trends, and how to deal with triggers) than the general population analyzed by Li et al. (2010, 2011).

Karkar et al. (2015a,b, 2017) rigorously examine the focus on health triggers. They use the term “diagnostic self-tracking,” proposed by Rooksby et al. (2014) as a type of self-tracking that refers to “the recording of personal information to diagnose or manage a health condition” (Karkar et al., 2017). Karkar et al. (2015b) propose a “self-experimentation” framework, described as a subset of self-tracking based on single case designs or n-of-1 trials. They focus on providing some level of scientific rigor to people interested in associations of health-related events, indicators, and symptoms (e.g., if certain food triggers headaches), because often these people perform such analyses without the support of health providers.

Another related term is Observations of Daily Living (ODL), which directly concerns patient’s or individual’s experiences. Brennan and Casper (2015) define ODL as a type of PGHD. ODL consist of patient-defined and patient-generated data that reflect “concepts uniquely defined and uniquely important to the patient,” which are especially useful to indicate idiosyncratically if the person is well or if they should seek healthcare support (Brennan and Casper, 2015). These data come
from the person or her/his environment, and can be as different as indicators of health status (e.g., the presence of pain), indicators of behavior (e.g., eating more when feeling anxious), and “exposures” such as environmental measures (e.g., pollution). These indicators often are not used as symptoms of pathology, but to indicate a “need for action” (Brennan and Casper, 2015). ODL is a term for the data, and not for the practices that generate them.

Finally, a term often conflated with self-tracking or PI, is quantified self (QS) (Neff and Nafus, 2016; Quantified Self, n.d.; Wolf, 2009, 2010). QS originated as a movement of technology enthusiasts who monitor themselves and build technologies to support these activities, based on an interest in self-experimentation and self-knowledge (Choe et al., 2014; Quantified Self, n.d.; West et al., 2016). However, the term evolved to be also used as the general practice of self-tracking (Choe et al., 2014), a “pervasive social trend” (Gregory and Bowker, 2016), and to refer to technologies that support bodily and emotional quantification (Johansen and Kanstrup, 2016) or the cultural movement of self-optimization (Neff and Nafus, 2016). These other meanings may even contradict the original goals of the QS community. Members of the QS community define the movement as a subset of self-tracking that emphasizes self-experimentation or n-of-1 studies; they also often gather in meetups to discuss their idiosyncratic experiences. These experiences often include but are not limited to health concerns. Members of the QS movement also often self-define and self-initiate their experiments of data collection and analysis, also highlighting the initiation of PGHD practices (Neff and Nafus, 2016).

Although health is its usual focus, QS is not only about health. Similar to self-tracking, it can also refer to the practices of quantifying other aspects of lives, such as energy consumption or finance. Regarding health, QS includes a wide range of repeated measures such as self-reported mood status, glucose readings from automatic pumps, cognitive performance, etc. (Choe et al., 2014), as well as single time measures such as genetic tests (Gregory and Bowker, 2016). Katz et al. (2018) argue that although QS (and they use the term more to refer to self-tracking practices than to a specific community) has many overlaps with chronic disease management, QS does not necessarily involve important
and common concerns of chronic disease management, such as: “the non-elective nature of disease; frequency of treatment decisions; need for continuous monitoring, greater unpredictability of measurements; affective impact of unwanted results due to justifiable fears of health complications; and the critical nature of situated decision-making based on personal data” (Katz et al., 2018).

1.2.3 Defining the Scope of PGHD in This Review

No single term or definition is universally used for PGHD in the literature, and current definitions mostly point to several broad uses. These are only a few terms offered by the literature that are related or overlap with PGHD. Although we searched for literature in several related fields of study, as HCI researchers we primarily approached the PGHD-related literature and its scope from this perspective. While there may be other related terms that were not included in this review, there is no standardized term that works in multiple research areas. As our review of the terminology shows, many related terms cover PGHD, but they either include other types of data, such as non-health data, or data collected during medical consultations, or cover only a subset of PGHD, such as data collected solely by technology. Each term and definition, with differences and similarities, focus on specific aspects in detriment of others. This multiplicity makes it complicated to define the general scope of PGHD, especially considering the varied research areas interested in the theme (e.g., medical vs. technology fields).

We cannot precisely demonstrate the relationship among these terms because their boundaries are fuzzy. From an HCI perspective, we consider PGHD as data collected and used by patients (or their caregivers). Therefore, we do not consider data automatically sent to healthcare providers’ databases, although we analyzed a few papers that use this approach. We also focus on data collected by patients themselves in their everyday lives, not by professionals in clinical settings. For example, we do not include in our review data generated through health assessments requested and conducted by healthcare providers during medical appointments. Since we focus on practices, we include both technologically assisted and traditional manual collection. Although
Table 1.1: Our PGHD space in relation to other terms.

<table>
<thead>
<tr>
<th>Term</th>
<th>Scope of this review</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Patient health outcomes and patient reported outcomes</em></td>
<td>Patient health outcomes and patient reported outcomes have many overlaps with PGHD. However, in the scope of this review we focus on data primarily generated by patients (or their caregivers). We do not include data generated through health assessments requested and conducted by healthcare providers during medical appointments.</td>
</tr>
<tr>
<td><em>Journaling</em></td>
<td>We consider journaling and self-management as practices that generate PGHD. However, PGHD can encompass more data than these practices traditionally generate, such as data individuals collect for their own health or general wellness management that are not necessarily related to a medical condition, nor to the intent to share this data with healthcare providers.</td>
</tr>
<tr>
<td><em>Self-management</em></td>
<td>The scope of this review focuses only on cases in which patients can at least see some of their data. Remote patient monitoring can generate PGHD, but often it generates data that do not fit the scope we approach in this review: data used or accessed by patients.</td>
</tr>
<tr>
<td><em>Remote patient monitoring</em></td>
<td></td>
</tr>
</tbody>
</table>

Continued.
1.2. Definitions: Data and Practices

<table>
<thead>
<tr>
<th>Term</th>
<th>Scope of this review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-monitoring</strong> – traditionally used in behavioral psychology as the practices of recording a person’s own thoughts, feelings, and behavior as part of assessments or treatments in behavior therapy (Choe et al., 2014, 2015). It originally focused on expressive behaviors, i.e., the practice of observing and controlling one’s own self-presentation (Snyder, 1974). Also used in research in technology-oriented areas, with the connotation of monitoring health parameters focusing on prevention or early detection of medical conditions (Grönvall and Verdezoto, 2013b).</td>
<td>Similar to self-management and journaling (with which it has many overlaps), we consider self-monitoring as a practice that generates a subset of what we examine as PGHD. However, it can also generate data out of our scope, for example data used in behavioral psychology and behavior therapy.</td>
</tr>
<tr>
<td><strong>Self-tracking and Personal Informatics</strong> – the practices of collecting personal data on which to reflect (Li et al., 2010, 2011). Self-tracking or PI are not only about health, encompassing multiple aspects of people’s lives such as finances, social interactions, and productivity (Li et al., 2010, 2011).</td>
<td>Following Li et al. (2010, 2011), this review considers self-tracking and personal informatics as synonyms. However, we focus only on self-tracking for health. We also consider “secondary tracking” to characterize health-related self-tracking that is not performed by “the self,” as in cases involving families (Pina et al., 2017).</td>
</tr>
<tr>
<td><strong>Self-experimentation</strong> – described as a subset of self-tracking based on single case designs or n-of-1 trials. It focuses on finding associations between health-related events, indicators, and symptoms (e.g., if certain food triggers headaches).</td>
<td>As a subset of self-tracking, we consider self-experimentation another term for practices that generate some types of PGHD. However, not every PGHD are collected with the intent of testing associations or triggers.</td>
</tr>
</tbody>
</table>

Continued.
Table 1.1: Continued

<table>
<thead>
<tr>
<th>Term</th>
<th>Scope of this review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observations of Daily Living</strong> – Defined as a type of PGHD, ODL consist of patient-defined and patient-generated data that reflect “concepts uniquely defined and uniquely important to the patient” (Brennan and Casper, 2015).</td>
<td>This review considers ODL as a subset of PGHD, because we also include data that is clinically-defined but patient-generated (Brennan and Casper, 2015).</td>
</tr>
<tr>
<td><strong>Quantified Self</strong> – QS is originally a movement of technology enthusiasts who independently monitor themselves and build technologies to support these activities, focusing specifically on self-experimentation, and often gathering in meet-ups to discuss their idiosyncratic experiences (Choe et al., 2014; Quantified Self, n.d.; West et al., 2016). The term evolved to be also used as the general practice of self-tracking (Choe et al., 2014). QS is not only about health, it can also refer to the practices of quantifying other aspects of individuals’ lives (e.g., finances).</td>
<td>In this review, we consider QS as another practice that can generate PGHD. However, QS is not restricted to health, while we focus only on health-related data. Also, the QS community often highlights they perform these activities independent of healthcare providers. In this review, we approach both cases in which individuals collect and use PGHD by themselves as well as cases in which they share the data with their healthcare providers.</td>
</tr>
</tbody>
</table>

Technology provides many benefits, patients have been collecting their data for decades and studies that do not focus on technology use may provide valuable insights for future technology development.

We also focus on measures directly related to health, excluding, for example, data related to payments, insurance, or other data patients may generate. Also, although this review describes PGHD use in the context of both medical conditions and general wellbeing, we are especially interested in the former. We focus on data collected and used by patients or individuals and related to a medical condition or health concern they have, even if their healthcare providers do not value these data. We do not include or discuss in this review data automatically and
implicitly tracked while individuals are engaged in activities that may
influence but are not directly connected to their health (e.g., online
shopping or government data), unless they are explicitly used for health-
related issues by the person or caregivers, possibly together with their
healthcare providers.

We also consider data that are collected and used by caregivers, as in
cases involving families (Pina et al., 2017). In these cases, the data can
be collected by both patients and caregivers or only by the latter. For
example, PGHD can be collected by older adults and adolescents, but
also by caregivers or family members (Hong et al., 2016), collaboratively
or not. Table 1.1 presents a comparison between the terms described in
the previous subsection and the scope used in this review.

In summary, in this review we considered the following aspects when
defining our PGHD scope:

(1) We include data related to a medical condition or to general
wellness, excluding management of other life aspects, such as
finances, unless this data is used explicitly for health-related
interests;

(2) Data can be directly related to a medical condition, general
health, or wellbeing; or provide context for patients’ health-related
interests;

(3) Data collection can be patient- or provider-initiated, but patients
(or their caretakers) are responsible for collecting the data outside
of traditional clinical settings;

(4) Data can be collected by the patient, caregivers, or both;

(5) Data can be primarily used by patients (and/or their caregivers),
healthcare providers, or both patients and providers. Although we
analyzed a few papers focusing on data intended for exclusive use
by healthcare providers, we focused on studies in which patients
(or their caregivers) could access at least part of the data;

(6) Data can be both technology and manually generated;

(7) And data can be intended to be used in the short and long term.
These aspects helped us delimit the space of our review. However, even within this roughly delimited space, defining the PGHD scope is not straightforward. The ONC definition (Shapiro et al., 2012) is a good summary, but we propose further analysis. Based on the analyzed literature, particularly on the definitions explored in the previous subsection, we derived the following questions to define the space of PGHD research and technology around key aspects:

(1) What is the focus of the study? Is it strictly medical-related, or does it relate to general health and lifestyle?

(2) What types of data are used? How are they used? What data can be considered PGHD?

(3) Who initiates or proposes (or is intended to initiate or propose) the use of PGHD? Patients? Healthcare providers? Other actors?

(4) Whose data are collected? Is the person a patient? Is the person tracking her/himself?

(5) Who is intended to use the data? Who uses the data? Patients? Healthcare providers? Others?

(6) How are the data generated? What means are used to collect and use PGHD? Are the data collected with or without the use of technology?

(7) For how long are the data expected to be used? How long are the data actually used?

Based on these questions, we generated a set of seven dimensions important for understanding the space of PGHD: (1) the health focus of the study, (2) the type of data, (3) who proposes the use of PGHD, (4) whose data are collected, (5) who is intended to use the data, (6) what are the mechanisms of PGHD collection and use, and (7) what is the duration of PGHD use. The next subsection describes each of these dimensions and how different studies approach them. We argue that it is important to consider these seven dimensions when researching and developing technologies and solutions focusing on PGHD.
1.3 Overview of This Review

We primarily draw on the research literature published in the ACM digital library and PubMed repositories. We chose the ACM Digital Library because it contains papers from most of the relevant conferences and journals related to HCI and technology-oriented research, e.g., CHI, CSCW, Pervasive Health. As this study is primarily pursued from the HCI perspective, it is necessary for us to include papers published in ACM. We chose PubMed because it covers a wide range of medical and health studies that are particularly relevant for PGHD. Many studies in technology-oriented areas focus on technology design and the use of such technologies by healthcare consumers, while medical and health studies focus on supporting clinical practices. Analyzing these two repositories allows us to include both the patients’ and providers’ views and address their individual challenges in using PGHD. We did not intend to provide a comprehensive review of technologies and design approaches focused on PGHD collection and use, such as pervasive or mobile computing. Other reviews have explored the common features and design challenges of health and wellness applications (e.g., Consolvo et al., 2014; Tentori et al., 2012). Instead of examining aspects of the engineering design and features of technologies, since our focus in this review is from an HCI perspective we examine the data, the practices of collecting and using these data, and the consequences of this use. However, readers should be aware that there is much more relevant literature scattered among other medical, nursing, public health, mental health, science and technology studies, social sciences, media studies, and other general conferences, journals, and repositories. Due to the broad applications and diverse relevant concepts of PGHD, it is not feasible for us to comprehensively review all existing literature in this highly multidisciplinary area. In this review, we aimed to describe the current PGHD space, articulate a set of important dimensions to consider when researching and developing within this space, and discuss the main challenges identified in prior literature regarding PGHD. Although as HCI researchers we have a special interest in technology support, our focus in this review is broader than studies proposing new technologies, since people use health-related data in varied ways, and often without technology support.
To identify representative papers to include in this review, we performed multiple searches in the two databases, using combinations of different keywords, such as quantified self, patient-generated data, self-tracking, personal informatics, self-experimentation, personal data, self-monitoring, self-management, log, journal, diary, daily living, patient-reported, combined with health-related words such as health, healthcare, patient, illness, chronic, disease, conditions, symptoms, and outcomes. However, due to the broadness of the research space, each individual research query resulted in a large number of irrelevant studies, and many important studies did not appear in the results. As mentioned in the earlier subsection, PGHD literature is broad and interdisciplinary, with different terms used in different areas to refer to the same concept. This complex space makes it extremely difficult to conduct a comprehensive and systematic review. Instead of going through all the research results and attempting to review a complete list of papers, we opted to conduct a narrative review in which we selected relevant papers identified through our initial search, incorporated papers we are aware of in this area, and further searched for papers citing the key literature in the area. Besides these searches, we also draw on our own research experiences and those of our colleagues and fellow researchers, as well as from the list of references of several key publications.

Our general inclusion criteria for the papers analyzed in this review also included studies, pilot studies, and case studies focused on:

(a) Data collection and use to support medical conditions or health-related concerns;

(b) Health-related data used by patients or individuals, caregivers, and/or healthcare providers;

(c) New technology solutions, such as wearable devices or self-tracking systems;

(d) Individuals (both patients and caregivers) outside of traditional clinical settings; to understand their healthcare needs and their use of PGHD (e.g., interviews with patients or users, social media analysis, surveys);
Overview of This Review

(e) Healthcare providers; to understand their views and experiences concerning PGHD use (e.g., interviews with healthcare providers, observation studies).

Based on this literature review, we identified seven important dimensions to define the research and design scope in researching and developing technologies for PGHD. We defined these dimensions through examining representative key literature in the area, emphasizing challenges in defining the term and the scope around PGHD. We articulate these dimensions to create a general guideline for researchers and developers to better study and understand the opportunities and challenges in studying PGHD.

The remainder of this review is organized as follows: Section 2 describes each dimension and their classifications, highlighting the challenges in classifying the papers; and Section 3 presents different open challenges related to PGHD, including the consequences of using these data for patients, caregivers, and healthcare providers.


References


References


References


Full text available at: http://dx.doi.org/10.1561/1100000080


References


References


