available at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/molonc

Review

Emerging uses of patient generated health data in clinical research

William A. Wood*, Antonia V. Bennett, Ethan Basch

Cancer Outcomes Research Program, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

ARTICLE INFO

Article history:
Received 12 June 2014
Received in revised form
12 August 2014
Accepted 18 August 2014
Available online 27 August 2014

Keywords:
Information technology
Patient reported outcomes
Quality of care
Clinical trials

ABSTRACT

Recent advancements in consumer directed personal computing technology have led to the generation of biomedically-relevant data streams with potential health applications. This has catalyzed international interest in Patient Generated Health Data (PGHD), defined as "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 (i.e. care partners or those who assist them) to help address a health concern."(Shapiro et al., 2012) PGHD offers several opportunities to improve the efficiency and output of clinical trials, particularly within oncology. These range from using PGHD to understand mechanisms of action of therapeutic strategies, to understanding and predicting treatment-related toxicity, to designing interventions to improve adherence and clinical outcomes. To facilitate the optimal use of PGHD, methodological research around considerations related to feasibility, validation, measure selection, and modeling of PGHD streams is needed. With successful integration, PGHD can catalyze the application of "big data" to cancer clinical research, creating both "n of 1" and population-level observations, and generating new insights into the nature of health and disease.

1. Background

In recent years, technological advancements have enabled consumers to interact with personal computing devices in ways that produce large amounts of consumer-specific data. As personal devices have grown more portable and powerful, consumer-directed applications have proliferated and have exponentially increased the breadth and depth of these data streams. Accelerometers, geolocators, and physiological sensors are now embedded in many personal computing devices. Some devices continue to exist in standalone, multipurpose computing form (e.g. smartphones, tablets, laptops, and desktops), others in uni- or oligo-purpose "wearable" form (e.g. wristbands, belt clips, skin patches), and still others that are

* Corresponding author.
E-mail address: wawood@med.unc.edu (W.A. Wood).

http://dx.doi.org/10.1016/j.molonc.2014.08.006

a hybrid of the two models (e.g. "smartwatches"). With varying amounts of active or passive consumer data entry, these devices can provide day to day or even hour to hour information about a person's location, diet, movement, symptoms, blood pressure, and heart rate.

Concurrently with these trends, the potential for "big data" to reveal insights about the external environment has gripped the public consciousness. Integrating multiple longitudinal data sources to predictively model complex events has long been a mainstay of activities as diverse as forecasting weather, choosing stocks, or assembling professional sports teams (Lewis, 2003). Entities in the for-profit, non-profit and academic spheres have recognized the ability of newer consumer-specific data streams to predict human behavior and outcomes. For example, large retailers like Target use data on consumer habits to identify and engage specific consumers for marketing purposes (Duhigg, 2012). The increasing amounts of data from personal devices promise to further improve these capabilities.

2. Patient-generated health data

In clinical care, we recognize that our patients' pathophysiological trends and events outside of clinic are at least as relevant to their health and disease as the brief snapshots of pathophysiology that are provided at the time of clinic visits. In the "big data" era, we can imagine using this information to predictively model disease states and to inform health-promoting interventions. Indeed, many of the newer consumer-specific data streams produce information that is biomedically relevant and which could inform research and clinical care. In this regard, an international dialog has emerged around health-related data that come specifically from patients, outside of the more general consumer context. These data are termed "Patient-generated health data" (PGHD) and defined as "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 (i.e. care partners or those who assist them) to help address a health concern" (Shapiro et al., 2012).

As interest in PGHD has increased, we are now seeing a convergence in consumer-directed personal technology and health-related applications. Samsung and Apple have recently announced major digital health initiatives, with Apple's features integrated into their new operating system (iOS8) as "HealthKit" and partnerships announced with the Mayo Clinic and the EPIC electronic health record, (Weise, 2014; Munro, 2014).

From a research standpoint, some of the device-generated PGHD of greatest interest include vital signs, stress levels, mood, physical activity, weight, diet, blood levels, medications, sleep patterns, tobacco and alcohol use, and environmental exposures (California Institute for Telecommunications and Information Technology, 2014). Under the more expansive PGHD definition, patient-curated histories, diaries, risk assessments, and reports of health and functional status are also likely to contribute valuable information within the research context. Additionally, other types of data that are not specifically health-related could be co-opted to generate health related insights, such as geolocation, social, and financial information. Examples of PGHD with potential relevance to clinical research are provided in Table 1. In general, key features of PGHD are that: patients, not providers, capture and record these data; PGHD is obtainable outside of clinical encounters; PGHD is longitudinal, with the potential for repeated measures over time; and PGHD can be collected at high frequency intervals, enabling nearly continuous data streams over extended periods of observation, depending on the metric of interest.

3. Improving clinical trials efficiency

As a separate issue, it is increasingly clear that there is a major need to improve the design and conduct of clinical trials in biomedical research. In the current era, clinical trials are expensive, inefficient, and time-consuming. While much has been written on these topics (Institute of Medicine 2010), these issues have had tangible consequences, including increasing political pressure on large clinical trial cooperative groups, and internal mandates among drug and device manufacturers

Table 1 – Patient generated health data with potential usefulness for clinical research.

Mode	Elements	Attributes/units
Sensor		
	Pedometry/accelerometry	Steps, activity intensity
	Sleep	Sleep duration, latency, interruption
	Weight	Pounds/kilograms
	Blood Pressure	mmHg
	Heart Rate	Beats per minute
	Temperature	Celsius/Fahrenheit
	Environmental exposure	Exposure-dependent
	Blood levels	Glucose, medication levels
	Falls	Times fallen
	Geolocation	Coordinates
Data entry		
	Exercise testing	Self-administered 6 min walk distance, others
	Diet	Calories, composition
	Mood/stress levels	Type, severity, frequency, interference
	Symptoms	Type, severity, frequency, interference
	Health-related quality of life	Scale, instrument-dependent
	Functional status	Scale, instrument-dependent
	Social support	Scale, instrument-dependent
	Medications (including opiates)	Type, frequency
	Tobacco use	Type, frequency
	Alcohol use	Type, frequency
Other		
	Social connectedness	Activity (e.g. Facebook, Twitter, others)
	Financial data	Medication and health care expenditure co-pays