**Article:**

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

**BMC Public Health**

# Broadening horizons: the case for capturing function and the role of health informatics in its use

Denis Newman-Griffis[1,2]*, Julia Porcino[1], Ayah Zirikly[1], Thanh Thieu[3], Jonathan Camacho Maldonado[1], Pei-Shu Ho[1], Min Ding[4], Leighton Chan[1] and Elizabeth Rasch[1]

## Abstract

**Background:** Human activity and the interaction between health conditions and activity is a critical part of understanding the overall function of individuals. The World Health Organization's International Classification of Functioning, Disability and Health (ICF) models function as all aspects of an individual's interaction with the world, including organismal concepts such as individual body structures, functions, and pathologies, as well as the outcomes of the individual's interaction with their environment, referred to as activity and participation. Function, particularly activity and participation outcomes, is an important indicator of health at both the level of an individual and the population level, as it is highly correlated with quality of life and a critical component of identifying resource needs. Since it reflects the cumulative impact of health conditions on individuals and is not disease specific, its use as a health indicator helps to address major barriers to holistic, patient-centered care that result from multiple, and often competing, disease specific interventions. While the need for better information on function has been widely endorsed, this has not translated into its routine incorporation into modern health systems.

**Purpose:** We present the importance of capturing information on activity as a core component of modern health systems and identify specific steps and analytic methods that can be used to make it more available to utilize in improving patient care. We identify challenges in the use of activity and participation information, such as a lack of consistent documentation and diversity of data specificity and representation across providers, health systems, and national surveys. We describe how activity and participation information can be more effectively captured, and how health informatics methodologies, including natural language processing (NLP), can enable automatically locating, extracting, and organizing this information on a large scale, supporting standardization and utilization with minimal additional provider burden. We examine the analytic requirements and potential challenges of capturing this information with informatics, and describe how data-driven techniques can combine with common standards and documentation practices to make activity and participation information standardized and accessible for improving patient care.

(Continued on next page)

* Correspondence: denis.griffis@nih.gov
[1]Rehabilitation Medicine Department, National Institutes of Health, Mark O. Hatfield Clinical Research Center, 6707 Democracy Boulevard, Suite 856, MSC 5493, Bethesda, MD 20892, USA
[2]Department of Computer Science and Engineering, The Ohio State University, 2015 Neil Avenue, DL 395, Columbus, OH 43210, USA
Full list of author information is available at the end of the article

(Continued from previous page)

**Recommendations:** We recommend four specific actions to improve the capture and analysis of activity and participation information throughout the continuum of care: (1) make activity and participation annotation standards and datasets available to the broader research community; (2) define common research problems in automatically processing activity and participation information; (3) develop robust, machine-readable ontologies for function that describe the components of activity and participation information and their relationships; and (4) establish standards for how and when to document activity and participation status during clinical encounters. We further provide specific short-term goals to make significant progress in each of these areas within a reasonable time frame.

**Keywords:** Disability evaluation, Electronic health records, Health informatics, Clinical informatics, Public health informatics, Natural language processing

# Background

The way in which we learn about our world as individuals and how we willfully act within it is fundamental to human existence. Human activity, and the impact of health conditions on it, is an important component of contemporary conceptualizations of health. This article is the product of a collaborative effort between an interdisciplinary group of health professionals and scientists to summarize the importance of capturing information on activity in health systems, and to identify analytic tools and techniques that can support effective utilization of this information for improving patient care. We draw on particularly relevant references in the salient fields to highlight the concepts, techniques, and evidence for a broader incorporation of activity information into healthcare and data analytics. We first describe conceptualizations of human activity and its role in models of health and disability, and existing methods and applications for measurements of activity. We then identify current information gaps regarding activity, along with methods for improving the rate and quality of capture of activity information and analyzing it to inform care decisions. Finally, we suggest four specific actions that can be taken towards more effective use of activity information in health systems, and identify practical short-term goals to make meaningful progress towards each.
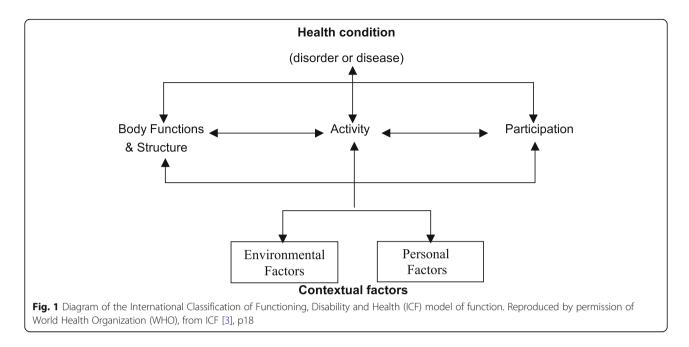
## Activity and disability

In sociology, action theory describes human activity, and its purposeful nature, in the context of environments and societies in which activities take place. Although first described in 1937 [1], the concept of human action has more recently been applied to the fields of medicine and health sciences to characterize the consequences of health conditions as an important and meaningful indicator of health. This concept is reflected in contemporary models of disability, for instance, where disability is conceptualized as the outcome of the interaction between the capabilities of individuals and the demands of environments with which individuals interact. The premise that disability reflects how people function given a particular context was articulated by Saad Nagi in the

early 1960's [2] and formed the basis for every contemporary model of disability that followed. Now codified in the World Health Organization's (WHO) International Classification of Functioning, Disability, and Health (ICF) [3] and adopted internationally, human action is embodied in the domain of <u>activity and participation</u>, where activity represents the execution of an action by an individual and participation represents actions through involvement in life situations. Actions, which take place at the level of the individual, are distinguished from organ or organ system function (ICF body structures/functions), or cellular/tissue function (ICF health conditions).

## What is function?

Human function can be broadly conceptualized as a continuum from body structures and functions to outcomes of interactions between individuals and their environments [4, 5], and has been argued to reflect "the lived experience of health" [6, 7]. The ICF defines <u>function</u> as an umbrella term encompassing all aspects of the interaction "between an individual (with a health condition) and that individual's contextual factors (environmental and personal factors)" [4]. Within the ICF model, function is broken down into several components, illustrated in Fig. 1. This model encompasses all aspects of an individual's interaction with the world, including organismal concepts such as individual body functions/structures and pathologies, as well as activity and participation, and all the environmental factors that affect these interactions. Importantly, activity and participation reflect volitional actions that take place at the level of the whole person, such as walking, communicating, applying knowledge, etc., which take place in, and are influenced by, a life situation or social context. For the purposes of this article, we operationalize the term "function" at this whole person level, and refer primarily to "activity and participation" in detailed discussion.

## Why are activity and participation important health indicators?

At both the individual and population levels, the ability of people to engage in activities and their participation

**Fig. 1** Diagram of the International Classification of Functioning, Disability and Health (ICF) model of function. Reproduced by permission of World Health Organization (WHO), from ICF [3], p18

in social roles shapes the need for resources and the associated response from national agencies, health systems, home and community-based organizations, and other support entities [8]. One timely example of the need for information about activities and participation on a global scale is a consequence of the dramatic shift in the world's demographic profile due to population aging. Among figures that the United Nations (UN) calculates in relationship to population ageing is the support ratio, which is the number of workers per retiree. By 2050, 36 countries, including the U.S., are expected to have support ratios below 2 [9], meaning that there will be fewer than 2 working persons to support each person over the age of 60. Ultimately, an individual's independence and ability to participate in meaningful life activities (i.e., quality of life) will heavily influence resource needs [10] and, at the population level, will have an overwhelming impact on national public health, pensions, and social programs serving the elderly. As noted in the WHO World Report on Ageing and Health, complex health states resulting from the coexistence of multiple chronic conditions (which can exist at any age) are not adequately represented by identifying or treating one disease at a time. As a result, there is a need for measures that are more meaningful to individuals [5].

The need for better information on activity and participation at the individual level has also been widely endorsed [11, 12]. Activity and participation reflect the cumulative outcome of disease burden, i.e. multimorbidity. In the U.S., it has been reported that over half of working age adults experience one or more chronic conditions [13]. It is well established that there is a strong and consistent association between a greater

number of chronic conditions and the existence and severity of limitations in activities and participation [14, 15]. Thus, the effect of multiple chronic conditions on the lives of individuals is realized in their overall function [6, 7]. Since function reflects, among other factors, the cumulative impact of health conditions on the person, and is not disease specific [16], its use as a health indicator helps to address major barriers to holistic, patient-centered care, such as fragmentation in care resulting from multiple and often competing disease-specific interventions [17].

In clinical settings, the inclusion of information on activity and participation in case mix calculations has been shown to improve the prediction of patient needs and resource use [8]. Evidence suggests that in cases of multi-morbidity, reducing the complexity of an individual's overall health state to approaches focusing on each disease individually fails to provide adequate care for this growing segment of the global population [18]. Viewing the outcome of these complexities in the form of whole person function, i.e., activity and participation, is therefore likely to clarify approaches to intervention [8, 10]. Function reflects a health continuum and thus is more comprehensive in its characterization of health than other endpoints like morbidity or mortality [17]. Indicators of function are strongly predictive of mortality [19] but have the additional advantage of being more proximal health indicators, permitting earlier and potentially more effective interventions [10, 20]. Simple and objective tests of physical performance have been included as biomarkers in studies of ageing, outperforming more traditional impairment measures in models predicting mortality [20]. Markers of frailty that include physical

function have been associated with employment difficulties in late middle age [21]. In addition to predicting mortality, indicators of physical function have been shown to predict other important and more immediate outcomes such as subsequent disability [22] and dementia [23] among older adults. In the context of population ageing, the prevalence of multi-morbidity within populations and within individuals will have associated consequences in function. Thus, information about function at both the individual and population level is critical for the design of healthcare systems, home and community-based supports, and for resource allocation.

### How have activity and participation been measured?

Models of function have historically been developed in the context of discussing disability, which is often described in terms of limitations in function [2, 24, 25]. However, these are *conceptual* models, describing the broad components that contribute to function, and have proven difficult to translate to *data* models that can capture specific aspects of function in context and how they relate to one another. Even the ICF, the most detailed framework developed for function, does not formally describe the relationships between different structures, activities, and environmental factors. Thus, how best to measure function, and particularly activity and participation, remains an open question despite international efforts [26, 27]. Many of the existing measurements are at the population level, in the form of national survey questions (see [26] for a detailed review of many such survey instruments). While these are relatively easy to administer with high coverage, they are necessarily limited in detail, in order to minimize respondent burden, and are unable to capture the individual perspective. Some efforts have been made to systematically capture information on activities of daily living (ADLs) in individual healthcare encounters; however, these have been captured relatively rarely and only present one small piece of the overall picture of activity and participation [27, 28]. Notably, information about the *environment* in which an individual functions is rarely captured under either approach, despite being central to concepts of function and disability. This continuing debate and development of instruments to measure function means that even where measurements of activity and/or participation are captured, they cannot easily be recognized as such or mapped to standardized vocabularies and data models for analysis.

### Definitions and examples of terms

One effect of the malleable definitions of function and its measurement is that language used for these concepts varies widely, particularly between different scientific fields. For clarity, we define our key terms here, and provide examples of each.

– **Function** – "a dynamic interaction between a person's health condition, environmental factors, and personal factors" [3]. This is an umbrella term including cellular and tissue function, organ and body structure function, and whole person function.
– **Activity and participation** – the outcome of the interaction between an individual (with some health condition) and their environment, including specific activities and participation, as well as personal contextual factors; also referred to as whole person function. This encompasses basic willful actions, specific tasks, organized activities, and role participation [26, 29]. Examples include walking (including the environment being walked on, anything used to assist in performing the activity, etc), taking public transportation (which combines walking with other activities such as identifying a destination, sitting, etc), or participating in work.
– **Activity report** – a recorded observation of activity and/or participation, which identifies relevant components of a specific activity or participation outcome and records them in structured or unstructured data. Examples include, "Patient walked one lap in the hallway," or "Sue reports to work every day at 9 and works with no limitations until 5pm." Prior work has referred to information samples of this type variously as "functioning information" [30], "functional status terms" [31], "functional status information" [32], "functional health status" [33], and other terms. However, prior studies have not specifically distinguished information about activity and participation from information about other elements of function; thus, we adopt the term "activity report" to clearly distinguish activity and participation information from other types of health information.

### The information gap: What's missing?

While information on pathology, and even impairments of individual body functions, has been captured at a high rate for use in many modern health systems [34], information on activity and participation is captured relatively rarely and remains difficult to use effectively [7, 35]. In order to utilize data on activity and participation, i.e., activity reports, the healthcare field has two primary needs: (1) standardized procedures and tools for capturing activity reports routinely and quickly (both in and out of the clinic), and (2) methods for analyzing activity reports to support evidence-based decision making. We suggest approaches towards meeting both of these needs, and provide four concrete calls to action, with example

short term goals for each, to improve both the availability and the utility of activity and participation information for modern health systems.

### How can information on activity and participation be captured?

At the population level, most countries collect basic information on function via national censuses and surveys [36], but this information is rarely captured in sufficient detail or frequency to have an impact on healthcare systems [7]. Thus, national surveys cannot be responsive to information needs in real time. At the individual level, some self-administered surveys for measuring specific aspects of functional status have been developed [37], and social media technologies have been shown to passively capture some information about individual function [31]; wearable devices are also an emerging technology for capturing individuals' activity-related information. However, these tools are, at least currently, difficult to standardize and apply to reliably capture information on activity and participation at scale. Health systems, which many individuals encounter fairly regularly, offer another logical source for capturing information about activity and participation, which can be combined with other sources for a fuller picture of individual function. While some information about activity and participation is already collected during healthcare encounters, there remains significant variability in terms of how often and on whom it is collected, as well as what information is captured [7, 17, 20, 35]. In addition to objective observations of activity and participation, expanded documentation of activity reports in health records can also capture self-reported data, which complements clinical assessments [28, 38].

The current scarcity of activity reports at the individual level, recorded via diverse modalities, instruments, and language, presents challenges for their use in decision making. Firstly, to support evidence-based decision making in health systems, health information must be standardized and interoperable to optimize its potential usefulness [17]. Usefulness, in turn, can only be achieved when raw data are translated into knowledge that can change practice, requiring analytics. An extraordinary volume of data is generated in health systems [39], and many of these data may include errors that impact analytics [40, 41]. Coordination with data from surveys, self-reported tools, and other media can improve accuracy, but increases the volume of data that must be processed. Thus, concerted efforts are needed to tap into the potential of these sources of information on activity and participation. A data-driven approach leveraging current techniques in health informatics to extract information about function, in particular activity and participation, is needed and represents an effort that requires the involvement and coordination of many entities [5].

### How can information on activity and participation be analyzed?

The field of health informatics involves the use of health-related data for scientific inquiry and discovery and for decision making in healthcare and government [42]. This definition encompasses a wide variety of analytic methods, which can be broadly separated into analyses of structured data (i.e., data fields such as vital signs, demographics, lab readings, etc) and unstructured data (e.g., free-text health records or medical images). Analysis of structured data has proven invaluable in advances in medical informatics and public health, such as monitoring cancer incidence and treatment at a population level [34], predicting the need for specific interventions in individual breast cancer treatment [43], cohort identification in Nordic countries [44], and many others [45, 46]. In the area of functional status measures and its correlation to mortality risk, factors such as age, gender, and some ADL information have been used to predict 2-year mortality [47]. However, a lack of standardized data models means that activity reports are difficult to capture in structured form. Even where some simpler aspects such as ADLs are captured in health records, they are difficult to correlate across samples [35]; existing structured judgments also often lack the granularity to capture functional limitations informatively [48]. Ongoing development of standards for recording information relevant to activity, such as physical therapy outcomes, offers one way to improve capture of structured data for analysis [49]. Further, imaging techniques are growing as an area of assessing impairments and associated functional limitations [50, 51], although such techniques impose high provider burden. Thus, we focus our discussion on unstructured text—particularly in health data—where activity reports have historically been captured [28, 48], and which offers flexibility to capture relevant details such as environmental or personal factors. While this flexibility can contribute both to provider burden in writing documentation and analytic burden in extracting useful information from it [52, 53], technologies such as speech recognition and natural language processing (NLP) can be used to reduce this burden while enabling automatic extraction, organization, and summarization of relevant information [53–55].

### How has NLP been used in clinical care and research?

Natural Language Processing (NLP) is a broad field of research that has been used for a variety of purposes in processing health-related text data. The most common application of NLP for health has been automatically extracting and recognizing health-related information in text [56–58], such as symptoms, procedures, and diseases [59–61], medications [62, 63] health events [64, 65], and patient characteristics [66], among other examples. Many

advances in NLP for health have been enabled through shared tasks [67], which engage a wider research community to solve a specific research problem such as detecting smoking status [68] or heart risk factors [69]. NLP has a long history of research and operational use in clinical informatics [70], and has proven especially helpful for several tasks that are difficult or expensive for humans to complete, such as detecting rates of patient readmission to different facilities [71]. NLP methods have also been incorporated operationally in diverse decision support systems including modeling disease progression, identifying cancer-related information in pathology reports, and risk assessment tools [72, 73].

While NLP for healthcare applications has historically focused on diagnostic information such as diseases, symptoms, medications, and procedures, more recent research is expanding both within and outside the clinic to consider contextual factors and other data sources. For example, homelessness is an important social indicator of health that can be extracted from the text of clinical encounters [74, 75]. NLP techniques have also been instrumental in leveraging pervasive social media data for diverse applications, from detecting adverse drug reactions to epidemiological surveillance [72]. Social media data have been particularly transformative for monitoring and analyzing mental health, a critical component of function. For instance, NLP techniques have been used to assist moderators of online forums by automatically flagging posts suggesting a mental health crisis—such as suicide risk—for immediate human intervention [76]. Current efforts are also being put into creating datasets that would further application of NLP techniques in this domain [77, 78].

### How has unstructured activity and participation information been analyzed?

Structured data about activities, participation, and associated limitations are central to disability research, assistive technology development, and many other fields. These data can be gathered from national surveys [79, 80], obtained via specialized research instruments [81], or modeled from available clinical information [82], although use of this information in healthcare delivery remains relatively limited [83]. Analyzing unstructured text information about activity and participation, however, along with associated environmental and personal factors, is an emerging area of interest in health informatics research. Recent work has included collecting self-reported function terms by manually reviewing clinical documents and online forums [31], and identifying groups of phrases describing various aspects of function via clinical chart review [33]; notably, the majority of these terms were not found in established terminological resources like the Unified Medical Language System (UMLS) [84]. To address this issue

of coverage, some researchers interested in activity and participation have utilized application-specific vocabularies compiled by clinical staff. Such handcrafted approaches have been successful in various applications, including automatically assigning some ICF codes in discharge summaries [85], using ICF codes for information retrieval [86], and predicting patients' rehospitalization risk [87]. Other work has avoided the coverage issue by using vocabulary-agnostic methods that are targeted to specific types of activity reports [88]. Additionally, activity and participation information has been used in the extraction and modeling of other functional outcomes, such as frailty or grave illnesses, from clinical text [89–91]. These studies represent significant initial efforts in analyzing activity and participation information with NLP, but the lack of systematic alignment with an overall conceptual framework for activity and participation and lack of shared definitions of the analytic tasks pose challenges for synthesizing and building on these efforts.

### What is needed to improve analysis of activity and participation information?

While activity reports may not yet be commonplace or a robust part of medical records, important information on activity and participation is currently being recorded, and is most often located in the free text portions of clinical notes. Thus, we focus on NLP as a critical tool for capturing this information for use and analysis. NLP, like other techniques used in health informatics, is a complex field that relies on a multitude of resources to achieve optimal performance. In the following sections, we walk through several factors in effective informatics, what is needed to support them, and the particular challenges of supporting these needs in the context of activity and participation information analysis. These points are also summarized in Table 1.

### What data are needed for successful informatics?

Much of the potential of health informatics is predicated on the availability of data. To develop and evaluate informatics methods for activity and participation, it is necessary to have data that have been annotated, or marked by experts as to what relevant information it contains and where that information can be found. Annotation serves two primary roles in informatics: to tell analysts and machine learning systems what specific information to focus on; and to serve as a gold standard for evaluating proposed automated methods and supporting benchmarking and comparison within a broader research community.

Examples of annotations for activity and participation information might include highlighting descriptions of specific actions (e.g., walking, climbing, shopping, cleaning) or life situations in free text, or even what type of

**Table 1** Four approaches to addressing the information gap on activity and participation

| Approach: | Common datasets for research | Shared understanding of analytic tasks | Expert knowledge of activity and participation | Records of activity and participation |
|---|---|---|---|---|
| Analytic Needs: | • Volume: sufficient data to support modern methods of analysis.<br>• Representation: data must be widely representative.<br>• Annotation: gold standard descriptions of activity reports for benchmarking and comparison. | • Problem definitions: common definitions of analytic tasks and evaluation.<br>• Problem sharing: information exchange in the community.<br>• Interdisciplinary collaboration: input from clinical and analytic stakeholders. | • Standardized information structure: clear standards of information components and their relationships.<br>• Robust sources of information: capture variation and common usage of language and data. | • Recorded observations: activity reports explicitly recorded during patient encounters. |
| Challenges for Activity and Participation: | • Records from general encounters often have few activity reports.<br>• Activity reports are expressed in diverse language and in varying levels of detail.<br>• No common datasets with activity and participation information available for community research. | • Prior NLP work on activity and participation information has been highly specific and does not generalize easily.<br>• Requires both data science and clinical expertise to effectively adapt existing methods to data that contain activity reports. | • Existing resources lack sufficient structure to accurately represent activity and participation information in practice.<br>• Current vocabularies have poor coverage of activity and participation concepts and terms. | • Multiple competing standards exist for documenting information in rehabilitation medicine.<br>• Standards are not widely adopted outside of rehab for standard clinical care. |
| Action: | • Develop and publish standards for annotating activity reports.<br>• Develop resources for research that can be shared through regulatory frameworks. | • Identify and define common research problems and applications for processing activity reports. | • Develop a clinically-informed ontology for activity and participation information, along with representative terminologies from multiple sources. | • Establish common standards for observing and documenting activity reports in patient encounters. |
| Short-term Goals: | • Develop and publish annotation schema for 1–2 specific aspects of activity and participation. Make small sets of annotated data available through existing data sharing mechanisms. | • Establish shared tasks for extracting particular activity reports from an annotated dataset. | • Develop mappings across existing conceptual frameworks, such as ICF and SNOMED. | • Identify minimal interventions that can capture high-impact activity and participation status. |

clinical evaluation is being described. Annotating such information requires both identifying and standardizing the components of activity reports in clinical records. Function is defined within the ICF as the outcome of the interaction of individuals with various contextual factors, which means that descriptions of activity and participation tend to be complex and rely on multiple pieces of evidence. For example, a therapist might observe that a patient is able to walk with a rolling walker for 300 ft. While the activity report that needs to be captured is focused on the action ("walk"), this information is contextualized by other factors such as the assistive device ("rolling walker"), and these relationships must be captured in annotation as well.

In addition to annotating data, it is important to devote research and administrative efforts to collecting and sharing large volumes of data that represent activity and participation information. Many recent advances in statistical methods for NLP, particularly deep learning technologies, have relied on the availability of thousands or millions of documents [92], but virtually no documents with activity and participation information are available to the broader research community at present. Semantic approaches leveraging expert knowledge have been used to great effect in low-data settings in the past [93]; however, such methods have typically relied on robust standardized resources that are lacking for activity and

participation, emphasizing the value of statistical learning from large datasets.

In medical data, which often contains protected health information (PHI), there are two main strategies for collecting such datasets. First, research groups within a single institution or collaboration may collect private data under an IRB-approved protocol. These data may be re-used or shared after the initial study via mechanisms such as protocol amendments, designing new protocols, and developing business or data use agreements. While these tend to be limited to specific named parties included in the protocol or legal agreements, and may involve lengthy approval processes, such mechanisms have been effectively used for a large variety of data sharing scenarios in health research [94]. A second strategy is to curate de-identified datasets that remove PHI and are then made more widely available while taking appropriate precautions for data stewardship. This is not a simple task: though de-identification can be performed without significantly reducing relevant clinical information [95], it is by no means a perfect process [96, 97], and defining what qualifies as de-identified requires agreement between all relevant stakeholders, such as IRBs, privacy offices, government entities, and most certainly patients. De-identified datasets are thus rare, but have an outsize impact in supporting rapid and effective research within a whole community. Under any chosen

mechanism, sharable datasets of activity reports will contribute significantly to informatics research and applications using activity and participation information.

## How do we make use of these data?

Applying informatic methods to use activity and participation information in clinical and administrative practice requires addressing a wide variety of analytic challenges. One challenge is that many specific analytic tasks do not clearly correspond to existing informatics research problems. For example, activity reports, such as "walks without gait aid 50 feet in hallway", involve the interaction of several concepts. Recognizing and extracting such reports from text requires both identifying the component concepts (e.g., the action "walks", environmental factors "in hallway" and "without gait aid", and the specific distance "50 ft") and linking them together. Walking in an indoor hallway is significantly different from walking across rough terrain outside; connecting these elements is necessary to extract the atomic outcome being recorded. This task is further complicated when multiple outcomes are described in a single report; for example, "ambulate in the hallway and stairs" refers both to walking and to climbing (two distinct activities in the ICF). Thus, modeling the complex semantics of activity reports may involve combining multiple existing research problems, such as named entity recognition, syntactic dependency parsing, and even conceptual inference.

Even well-studied problems such as information retrieval or relation extraction can face new challenges for activity and participation information. For example, some patient records, such as History and Physical Examinations, often contain only a few sentences describing physical and mental function among a much larger concentration of diagnostic history, past procedures, etc. For a healthcare provider or administrator attempting to locate activity and participation information about a patient, such as a physical therapist tracking activity history or an analyst surveying inpatient functional outcomes, it is therefore necessary to pinpoint which sections or paragraphs of a long document include important information to review. Furthermore, such users must be able to quickly access and intuitively organize patient records from a variety of disciplines. These applications encompass diverse NLP tasks, including information extraction and retrieval, for identifying and organizing activity and participation information in the medical record; knowledge representation, for capturing clinically-informed relationships between activity and participation concepts; and determining the relevance of documents with respect to particular criteria, such as potential limitations in function. As with all complex tasks and modern problem solving approaches, addressing these issues for practical care will require interdisciplinary collaboration between clinical or domain experts, knowledge representation specialists, and informaticians at all stages of the analytic process, from defining goals to practical implementation in healthcare systems.

## What resources do we need?

Beyond the quantity and quality of available data, many successful clinical applications of NLP have been enabled by robust medical knowledge sources. These sources are referred to by various names, including (but not limited to) taxonomies, terminologies, and ontologies. These terms are used inconsistently in the literature, so we define each of them for this article as follows. Terminologies capture the diverse names used to refer to biomedical concepts, such as diseases, substances, measurements, etc., and are intended to both catalogue distinct concepts and provide a more or less comprehensive reference for the ways these concepts can be referred to. Biomedical terminologies often include elements of domain-specific ontology in their structure, which describe invariant classes of concept, such as diseases, symptoms, biological processes, functions, etc. Ontology also describes relations that hold universally between these classes: for example, that convulsions are a symptom of seizure [98]. Many terminologies have been developed as formalized coding systems, and can be referred to as classifications or taxonomies; the International Classification of Diseases (ICD), another WHO reference classification, being a salient example. As a result, the organization of many terminologies distinguishes not only between ontologically different classes (e.g., febrile vs afebrile seizure), but also epistemologically distinct observations (e.g., tuberculosis identified via microscopy or bacterial culture) [98]. Both types have been critical components of many successes in health informatics [45, 99].

However, comparable knowledge sources are few and far between for non-medical aspects of function. The ICF, originally developed in 1980 as the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) and revised in 2001 to better model environmental aspects of function [100], is a conceptual terminology that was designed to provide a common language for a wide variety of administrative and policy needs such as reporting, service coordination, and policy development [4]. Though the ICF has been integrated into the UMLS, and some efforts have been made to map it to other ontological resources [101], comprehensive coverage of practical vocabulary has never been its intent, and mappings to other well-developed terminologies such as SNOMED CT or LOINC are minimal. As a result, its coverage and granularity for coding practical information on activity and participation has been shown to lag behind higher-coverage medical terminologies [102]. Additionally, the distinctions it draws do not necessarily reflect a clinically-

based organization of knowledge. As a practical example, the mobility-related action of *walking* is not linked within the ICF to terms commonly used in practice, such as *ambulation*. A recent review found several other criticisms of the organization of the ICF, such as its emphasis of the health condition component, the ambiguity of concepts, and its "lack of a clear ontological structure" [103]. Some of these criticisms may be related to the lack of revisions to the ICF over the years. While the WHO publishes updates to the language of the ICF each year, it has never been revised, unlike the ICD, which is currently under its 11th revision. Thus, while the ICF has been hailed as the "best prospect for an internationally recognized, sufficiently complete and powerful information reference for the documentation of functioning information" [17], and it has the potential to be effectively combined with other vocabularies for coding purposes [104], a number of practical shortcomings make it difficult to utilize for successful NLP methods relying on dictionary definitions or common patterns in order to extract activity and participation information.

## A call to action

Incorporating information on activity and participation into the operation of health systems is not a simple task, and fully utilizing activity and participation status to improve the quality of life of populations and individuals will require concerted long-term efforts. In the following sections, we describe four major components of this overall goal. These approaches are highly inter-related, but reflect distinct steps to be taken by the medical and research communities to enable greater capture and utilization of activity reports. While these steps are complex and may require coordination between international entities, we have identified short-term goals that can achieve significant initial progress within a reasonable time frame.

### Action 1: Develop annotation standards and data

In order to understand how to process activity and participation information as it is currently documented, it is necessary to **develop and publish standards for annotating activity reports in structured and unstructured data, and develop data resources for research that can be shared through regulatory frameworks.** Preliminary investigations into the variety of ways in which activity reports are documented in various text sources can lay the groundwork for this effort, but published annotation standards establish a common base for communication and comparison within the research community. Development of sharable datasets regarding individuals' health data faces significant challenges in data privacy and interoperability, as well as a lack of robust legal frameworks or incentives for development

[105]. However, there are well-developed risk-tolerant mechanisms for data exchange, including IRB procedures, data use agreements, and business agreements [106], and when such mechanisms are used, sharable datasets contribute significantly to rapid advancement of research. For example, the MIMIC Critical Care database is a de-identified dataset made available through a signed data use agreement that, through active maintenance, has expanded to include over 2 million text documents in addition to lab readings, vital signs, etc. [107]. MIMIC has been invaluable for clinical informatics and NLP research into extracting diagnoses, symptoms, medications, modeling patients' course of care, and many other purposes. While more datasets of the scale of MIMIC are needed, they are achievable only through long-term effort. In the short term, significant first steps could be made for activity and participation information by developing and publishing an annotation schema for one or two specific aspects of activity, and by making a small set of annotated data available to the research community through existing data sharing mechanisms. This will enable rapid, effective communication in research via common reference points and shared benchmarking for evaluation.

### Action 2: Define analytic tasks

As a companion effort to developing these data resources and standards, we must also **identify and clearly define common research problems and applications for processing activity reports.** In computational research communities such as NLP, shared definitions of analytic tasks are the bones of effective research and evaluation. Identifying the characteristics of activity reports in structured and unstructured data and evaluating how these problems fit existing frameworks in NLP and other fields will enable development and adaptation of methods within the research community. Together with identifying downstream analytic tasks where information on activity and participation can be leveraged, such as cohort selection or rehospitalization risk prediction, this process will also help identify relevant data needs in collecting and storing activity reports. This task is thus interdependent with documentation and annotation standards; the challenge for analysis is to define how the information is to be automatically extracted and used. These problems and tasks must be defined with input from clinicians and data scientists alike. A major first step in this direction could be to develop a shared task for extracting one particular type of activity report from an annotated dataset. Such efforts promote broader research by laying the groundwork for the collaborative effort in developing and evaluating analytic methods.

### Action 3: Develop machine-readable ontologies

For both capturing and analyzing activity reports, it is critical to **develop a robust ontology that describes**

Newman-Griffis *et al. BMC Public Health* (2019) 19:1288

Page 10 of 13

the components of activity and participation infor-mation and their relationships to one another and to other biomedical, psychological, and social concepts. Such an effort has two major components: formalizing the conceptual framework and developing machine-readable resources. The first component involves defin-ing the concepts necessary to represent activity and participation and activity reports, and capturing the ne-cessary relationships between these concepts to describe their interaction. Many such resources and conceptual models—such as the ICF—already exist in rehabilitation medicine, mental health research, etc., and drawing on and connecting these proven resources should be the starting point for any analytically-focused effort. In addition, some important elements of activity and par-ticipation have coverage in other biomedical vocabular-ies such as SNOMED CT and LOINC; by mapping to these resources, well-developed analytic methods for clinical information can inform work on analyzing activ-ity reports. As such models and mappings are developed, machine-readable implementations, similar to the UMLS, will enable analytic methods to build directly on the conceptual structure. One initial step towards this goal could be leveraging previous findings of activity and participation information in SNOMED [102] to develop mappings from SNOMED concepts to the ICF framework, providing a powerful tool for identifying and analyzing components of activity information. Development of onto-logical models needs to be a clinically-motivated process that is verified empirically, and thus must be developed in concert with engaged practitioners and researchers. Such standardized resources will support training in document-ing activity and participation, as well as methods for analyzing it.

### Action 4: Establish documentation standards

A key step in improving the availability of information on activity and participation within healthcare delivery is to establish standards for how and when to document activity and participation status during clinical en-counters. While this is a much larger task than a single paper can accomplish, potentially involving the coordi-nated efforts of international entities, multiple such stan-dards have already been developed within rehabilitation medicine, as mentioned in the previous section; addition-ally, the Institute of Medicine has made some specific recommendations for documenting social and behavioral information in EHRs, including some activity and partici-pation information [108]. However, awareness and adop-tion of these standards by the broader medical community are limited, and different standards compete even within the rehabilitation community. Establishing a single stand-ard for the medical field at large to use is a long-term effort, but in the short term, small, focused efforts can be

made within local institutions or health systems to in-crease the availability of activity reports. In some cases, such as team settings involving an occupational or phys-ical therapist, activity reports are likely already being captured, and need only be intentionally analyzed. In other settings, relatively minimal interventions can cap-ture high-impact activity and participation status. For example, a clinician could regularly note a patient's ability to move through the clinic independently, and ask the patient if they are currently experiencing any limitations in their regular activities. Developing small sets of such practices can significantly improve the availability of activ-ity reports within health records while broader standards are established.

### Conclusion

Function is an important indicator of health from both population and individual perspectives. However, infor-mation on function, and particularly on activity and participation, has not been used in a routine and stan-dardized way when evaluating and monitoring the health of individuals from a holistic viewpoint. We believe rapid advances in data management and analytic tools have the potential to address barriers facing the effective use of activity and participation information, by locating, extracting, organizing, and summarizing activity reports from massive quantities of medical records. We find health informatics, and natural language processing in particular, to be a promising avenue for accelerating these efforts. Informatics can enable identification, ex-traction, and organization of activity and participation information for applications such as disability assess-ment and health monitoring [90, 91], and can also be used in software or devices to assist people with disabil-ities to engage in daily activities effectively [109, 110]. While existing applications of informatics methodologies to activity and participation information have shown promise, they face several challenges, including reliance on manual collection of non-standardized terminologies in text by domain experts, a lack of a shared systematic framework for activity and participation analysis, and a lack of relevant data. To drive informatics forward as a tool for capturing and utilizing activity and participation information, we recommend four important steps: (1) make activity and participation annotation standards and datasets available to the broader research community; (2) define common research problems in automatically processing activity and participation information; (3) de-velop robust, machine-readable ontologies for function that describe the components of activity and participation information and their relationships; and (4) establish standards for how and when to document activity and participation status during clinical encounters. These are challenging steps, requiring international coordination,

but we provide short-term goals for each that can be accomplished in a reasonable timeframe and measurably improve ability to capture and use activity and participation data.

Whole-person function, as embodied by activity and participation, is a strong predictor of mortality, disability, employment, and resource utilization. Moreover, it outperforms comorbidities in predicting acute care readmissions in medically complex patients. We envision that standardized and accessible activity and participation information yielded from these efforts will provide valuable evidence-based knowledge that can be translated into practice by helping provide holistic and patient-centered care and ultimately improving the efficiency and effectiveness of health care delivery, management, and planning.

## Abbreviations
ADLs: Activities of daily living; CMS: Centers for Medicare and Medicaid Services; DALYs: Disability adjusted life years; EHR: Electronic health record; ICD: International Classification of Diseases; ICF: International Classification of Functioning, Disability, and Health; ICIDH: International Classification of Impairments, Disabilities, and Handicaps; IRB: Institutional Review Board; LOINC: Logical Observation Identifiers Names and Codes; NLP: Natural language processing; PHI: Protected health information; SNOMED CT: SNOMED Clinical Terms; SNOMED: Systematized Nomenclature of Medicine; UMLS: Unified Medical Language System; UN: United Nations; WHO: World Health Organization

## Authors' contributions
ER conceived of the article and designed initial organization. ER, DNG, and JP led the writing and organization process. DNG, JP, AZ, TT, JCM, PSH, MD, LC, and ER all contributed to gathering references, framing discussions, and the writing and revisions of the manuscript; DNG led the editing and revision processes. All authors read and approved the final manuscript.

## Authors' information
LC is the chief of the Rehabilitation Medicine Department (RMD) of the NIH Clinical Center.
ER is chief of the Epidemiology and Biostatistics section in RMD.

## Availability of data and materials
Not applicable

## Ethics approval and consent to participate
Not applicable

## Consent for publication
Not applicable

## Competing interests
The authors declare that they have no competing interests.

## Author details
[1]Rehabilitation Medicine Department, National Institutes of Health, Mark O. Hatfield Clinical Research Center, 6707 Democracy Boulevard, Suite 856, MSC 5493, Bethesda, MD 20892, USA. [2]Department of Computer Science and Engineering, The Ohio State University, 2015 Neil Avenue, DL 395, Columbus, OH 43210, USA. [3]Department of Computer Science, Oklahoma State University, 116-A MSCS, Stillwater, OK 74078, USA. [4]Information Technology Laboratory, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA.

## References
1. Parsons T, Durkheim É, Marshall A, Pareto V: The structure of social action. A study in social theory with special reference to a group of recent European writers (Alfred Marshall, Vilfredo Pareto, Émile Durkheim, Max Weber); 1937.
2. Nagi SZ. In: Sussman MB, editor. Some conceptual issues in disability and rehabilitation. In: Sociology and Rehabilitation. Washington, DC: American Sociological Association; 1965. p. 100–13.
3. World Health Organization: International Classification of Functioning, Disability and Health: ICF. 2001.
4. World Health Organization: How to use the ICF: a practical manual for using the International Classification of Functioning, Disability, and Health (ICF). Exposure draft for comment edn. Geneva: WHO; 2013.
5. Beard JR, Officer A, de Carvalho IA, Sadana R, Pot AM, Michel J-P, Lloyd-Sherlock P, Epping-Jordan JE, Peeters GMEE, Mahanani WR, et al. The world report on ageing and health: a policy framework for healthy ageing. Lancet. 2016;387(10033):2145–54.
6. Stucki G, Bickenbach J, Melvin J. Strengthening rehabilitation in health systems worldwide by integrating information on functioning in national health information systems. Am J Phys Med Rehabil. 2017;96(9):677–81.
7. Stucki G, Bickenbach J. Functioning information in the learning health system. Eur J Phys Rehabil Med. 2017;53(1):139–43.
8. Hopfe M, Stucki G, Marshall R, Twomey CD, Ustun TB, Prodinger B. Capturing patients' needs in casemix: a systematic literature review on the value of adding functioning information in reimbursement systems. BMC Health Serv Res. 2016;16:40.
9. United Nations, Department of Economic and Social Affairs, population division: world population prospects: the 2017 revision, key findings and advance tables; 2017.
10. Taniguchi Y, Kitamura A, Nofuji Y, Ishizaki T, Seino S, Yokoyama Y, Shinozaki T, Murayama H, Mitsutake S, Amano H et al: Association of trajectories of higher-level functional capacity with mortality and medical and long-term care costs among community-dwelling older japanese. J Gerontol Ser A 2018:gly024-gly024.
11. Seals DR, Justice JN, LaRocca TJ. Physiological geroscience: targeting function to increase healthspan and achieve optimal longevity. J Physiol. 2016;594(8):2001–24.
12. Stucki G, Bickenbach J, Gutenbrunner C, Melvin J. Rehabilitation: the health strategy of the 21st century. J Rehabil Med. 2018;50(4):309–16.
13. Gulley SP, Rasch EK, Chan L. If we build it, who will come?: working-age adults with chronic health care needs and the medical home. Med Care. 2011;49(2):149–55.
14. Verbrugge LM, Lepkowski JM, Imanaka Y. Comorbidity and its impact on disability. Milbank Q. 1989;67(3–4):450–84.
15. Jones GC, Bell K. Adverse health behaviors and chronic conditions in working-age women with disabilities. Fam Community Health. 2004; 27(1):22–36.
16. Cooper R, Kuh D, Hardy R. Mortality review G, on behalf of the F, teams HAs: objectively measured physical capability levels and mortality: systematic review and meta-analysis. BMJ. 2010;341:c4467.
17. Hopfe M, Prodinger B, Bickenbach JE, Stucki G. Optimizing health system response to patient's needs: an argument for the importance of functioning information. Disabil Rehabil. 2017:1–6.
18. Banerjee S. Multimorbidity--older adults need health care that can count past one. Lancet. 2015;385(9968):587–9.
19. Keevil VL, Luben R, Hayat S, Sayer AA, Wareham NJ, Khaw K-T. Physical capability predicts mortality in late mid-life as well as in old age: findings from a large British cohort study. Arch Gerontol Geriatr. 2018;74:77–82.

20. Cooper R, Strand BH, Hardy R, Patel KV, Kuh D. Physical capability in mid-life and survival over 13 years of follow-up: British birth cohort study. BMJ. 2014; 348:g2219.

21. Palmer KT, D'Angelo S, Harris EC, Linaker C, Gale CR, Evandrou M, Syddall H, van Staa T, Cooper C, Sayer AA, et al. Frailty, pre-frailty and employment outcomes in the health and employment after fifty (HEAF) study. Occup Environ Med. 2017;74(7):476–82.

22. Perera S, Patel KV, Rosano C, Rubin SM, Satterfield S, Harris T, Ensrud K, Orwoll E, Lee CG, Chandler JM, et al. Gait speed predicts incident disability: a pooled analysis. J Gerontol Ser A Biol Sci Med Sci. 2016;71(1):63–71.

23. Beauchet O, Annweiler C, Callisaya ML, De Cock A-M, Helbostad JL, Kressig RW, Srikanth V, Steinmetz J-P, Blumen HM, Verghese J, et al. Poor gait performance and prediction of dementia: results from a meta-analysis. J Am Med Directors Assoc. 2016;17(6):482–90.

24. Institute of Medicine: Disability in America: toward a national agenda for prevention. Committee on a National Agenda for the Prevention of Disabilities. A.M. Pope and A.R. Tarlov, Eds. Washington, DC: National Academies Press; 1991.

25. Institute of Medicine: Enabling America: assessing the role of rehabilitation science and engineering. In: Pope AM, Brandt EN, editors. Committee on a National Agenda for the Prevention of Disabilities. Washington, DC: National Academies Press; 1997.

26. Altman BM. In: Wunderlich GSNRC, editor. Population survey measures of functioning: strengths and weaknesses. In: Improving the Measurement of Late-Life Disability in Population Surveys: Beyond ADLs and IADLs: Summary of a Workshop. Washington, D.C.: The National Academies Press; 2009. p. 99–156.

27. Verbrugge LM. Disability experience and measurement. J Aging Health. 2016;28(7):1124–58.

28. Bogardus ST, Towle V, Williams CS, Desai MM, Inouye S. What does the medical record reveal about functional status? J Gen Intern Med. 2004;16:728–36.

29. Madans J, Altman B, Rasch E, Synneborn M, Banda J, Mbogoni M, Me A, DePalma E. Proposed purpose of an internationally comparable general disability measure. In: Washington Group Meeting, Brussels, Belgium. 2004;2004.

30. Stucki G, Bickenbach J. Functioning: the third health indicator in the health system and the key indicator for rehabilitation. Eur J Phys Rehabil Med. 2017;53:134–8.

31. Kuang J, Mohanty AF, Rashmi VH, Weir CR, Bray BE, Zeng-Treitler Q. Representation of functional status concepts from clinical documents and social media sources by standard terminologies: AMIA Annu Symp. American Medical Informatics Association; 2015. p. 795–803.

32. Thieu T, Camacho J, Ho P-S, Brandt D, Porcino J, Newman-Griffis D, Yuan A, Ding M, Nelson L, Rasch E, et al. Inductive identification of functional status information and establishing a gold standard corpus A case study on the Mobility domain. In: 2017 IEEE Int Conf Bioinform Biomed (BIBM): 2017; 2017. p. 2300–2.

33. Skube SJ, Lindemann EA, Arsoniadis EG, Wick EC, Melton GB. Characterizing functional health status of surgical patients in clinical notes. In: 2018 AMIA summit Clin res inform. American Medical Informatics Association, vol. 2018.

34. White MC, Babcock F, Hayes NS, Mariotto AB, Wong FL, Kohler BA, Weir HK. The history and use of cancer registry data by public health cancer control programs in the United States. Cancer. 2017;123(Suppl 24):4969–76.

35. Brown RT, Komaiko KD, Shi Y, Fung KZ, Boscardin WJ, Au-Yeung A, Tarasovsky G, Jacob R, Steinman MA. Bringing functional status into a big data world: validation of national veterans affairs functional status data. PLoS One. 2017;12(6):e0178726.

36. McPherson A, Durham J, Richards N, Gouda H, Rampatige R, Whittaker M. Strengthening health information systems for disability-related rehabilitation in LMICs. Health Policy Plan. 2017;32(3):384–94.

37. Bowie CR, Twamley EW, Anderson H, Halpern B, Patterson TL, Harvey PD. Self-assessment of functional status in schizophrenia. J Psychiatr Res. 2007;41(12):1012–8.

38. Burns RB, Moskowitz MA, Ash A, Kane RL, Finch MD, Bak SM. Self-report versus medical record functional status. Med Care. 1992;30(5):MS85–95.

39. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst. 2014;2(1):3.

40. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc. 2004;11(2):104–12.

41. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013;20(1):144–51.

42. Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, Kalet IJ, Lenert LA, Musen MA, Ozbolt JG, et al. AMIA board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. J Am Med Inform Assoc. 2012;19(6):931–8.

43. Specht MC, Kattan MW, Gonen M, Fey J, Van Zee KJ. Predicting nonsentinel node status after positive sentinel lymph biopsy for breast cancer: clinicians versus nomogram. Ann Surg Oncol. 2005;12(8):654–9.

44. Maret-Ouda J, Tao W, Wahlin K, Lagergren J. Nordic registry-based cohort studies: Possibilities and pitfalls when combining Nordic registry data. Scand J Public Health. 2017;45(17_suppl):14–9.

45. Oellrich A, Collier N, Groza T, Rebholz-Schuhmann D, Shah N, Bodenreider O, Boland MR, Georgiev I, Liu H, Livingston K, et al. The digital revolution in phenotyping. Brief Bioinform. 2015:bbv083.

46. Shortreed SM, Cook AJ, Coley RY, Bobb JF, Nelson JC. Challenges and opportunities for using big health care data to advance medical science and public health. Am J Epidemiol. 2019.

47. Carey EC, Walter LC, Lindquist K, Covinsky KE. Development and validation of a functional morbidity index to predict mortality in community-dwelling elders. J Gen Intern Med. 2004;19(10):1027–33.

48. Nicosia FM, Spar MJ, Steinman MA, Lee SJ, Brown RT. Making function part of the conversation: clinician perspectives on measuring functional status in primary care. J Am Geriatr Soc. 2019;67(3):493–502.

49. Physical Therapy Outcomes Registry Scientific Advisory P, Chesbrough K, Elrod M, Irrgang JJ. Systems science in rehabilitation practice realized. Phys Ther. 2018;98(11):909–10.

50. Steinheimer S, Dorn JF, Morrison C, Sarkar A, D'Souza M, Boisvert J, Bedi R, Burggraaff J, Kontschieder P, Dahlke F, et al. Setwise comparison: efficient fine-grained rating of movement videos using algorithmic support - a proof of concept study. Disabil Rehabil. 2019:1–7.

51. Crawford RJ, Fortin M, Weber KA 2nd, Smith A, Elliott JM. Are magnetic resonance imaging technologies crucial to our understanding of spinal conditions? J Orthop Sports Phys Ther. 2019:1–32.

52. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc. 2011;18(2):181–6.

53. Payne TH, Tang PC, Tierney WM, Weaver C, Weir CR, Zaroukian MH, Corley S, Cullen TA, Gandhi TK, Harrington L, et al. Report of the AMIA EHR-2020 task force on the status and future direction of EHRs. J Am Med Inform Assoc. 2015;22(5):1102–10.

54. Hoyt R, Yoshihashi A: Lessons learned from implementation of voice recognition for documentation in the military electronic health record system. Perspect Health Information Manag 2010, 7(Winter):1e-1e.

55. Blackley SV, Huynh J, Zhou L, Wang L, Korach Z. Speech recognition for clinical documentation from 1990 to 2018: a systematic review. J Am Med Inform Assoc. 2019;26(4):324–38.

56. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform. 2008;35(8):128–44.

57. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, Forshee R, Walderhaug M, Botsis T. Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. J Biomed Inform. 2017;73:14–29.

58. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, Liu S, Zeng Y, Mehrabi S, Sohn S, et al. Clinical information extraction applications: a literature review. J Biomed Inform. 2018;77:34–49.

59. Doğan RI, Leaman R, Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. J Biomed Inform. 2014;47:1–10.

60. Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H, Xu H. CLAMP – a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc. 2018;25:331–6.

61. Uzuner Ö, DuVall SL, South BR, Shen S. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011;18(5):552–6.

62. Grouin C, Zweigenbaum P, Deléger L. Extracting medical information from narrative patient records: the case of medication-related information. J Am Med Inform Assoc. 2010;17(5):555–8.

63. Uzuner Ö, Cadag E, Solti I. Extracting medication information from clinical text. J Am Med Inform Assoc. 2010;17(5):514–8.

64. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform. 2015;54:202–12.

65. Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. Clin Pharmacol Ther. 2012;92(2):228–34.

66. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, Lai AM. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc. 2013;21:221–30.

67. Huang C-C, Lu Z. Community challenges in biomedical text mining over 10 years: success, failure and the future. Brief Bioinform. 2015;17(1):132–44.

68. Uzuner Ö, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. J Am Med Inform Assoc. 2008;15:14–24.

69. Stubbs A, Kotfila C, Xu H, Uzuner Ö. Identifying risk factors for heart disease over time: Overview of 2014 i2b2/UTHealth shared task track 2. J Biomed Inform. 2015(58):S67–77.

70. Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Nat Lang Eng. 1995;1:83–108.

71. Rastegar-Mojarad M, Lovely JK, Pankratz J, Sohn S, Ihrke DM, Merchea A, Larson DW, Liu H. Using unstructured data to identify readmitted patients. In: 2017 IEEE Int Conf Healthc Inform (ICHI); 2017. p. 1–4.

72. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. Yearb Med Inform. 2017;26:214–27.

73. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform. 2009;42(5):760–72.

74. Bejan CA, Angiolillo J, Conway D, Nash R, Shirey-Rice JK, Lipworth L, Cronin RM, Pulley J, Kripalani S, Barkin S, et al. Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. J Am Med Inform Assoc. 2017:ocx059.

75. Gundlapalli AV, Carter ME, Palmer M, Ginter T, Redd A, Pickard S, Shen S, South B, Divita G, Duvall S, et al. Using natural language processing on the free text of clinical documents to screen for evidence of homelessness among US veterans. In: AMIA Annu Symp. American Medical Informatics Association; 2013. p. 537–46.

76. Zirikly A, Kumar V, Resnik P. The GW/UMD CLPsych 2016 shared task system. In: Third Workshop Comp Ling Clin Psychol, vol. 2016. p. 166–70.

77. Shing H-C, Nair S, Zirikly A, Friedenberg M, Daumé Iii H, Resnik P. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Fifth workshop comp Ling Clin Psychol. New Orleans, LA: Association for Computational Linguistics; 2018. p. 25–36.

78. Zirikly A, Resnik P, Uzuner O, Hollingshead K. CLPsych 2019 Shared task: predicting the degree of suicide risk in Reddit posts. In: sixth workshop comp Ling Clin Psychol. Association for Computational Linguistics. 2019:24–33.

79. Frochen S, Mehdizadeh S. Functional status and adaptation: measuring activities of daily living and device use in the National Health and aging trends study. J Aging Health. 2017;30(7):1136–55.

80. Lin IF, Wu H-S. Activity limitations, use of assistive devices or personal help, and well-being: variation by education. J Gerontol Ser B. 2014;69(Suppl_1):S16–25.

81. Zahuranec DB, Skolarus LE, Feng C, Freedman VA, Burke JF. Activity limitations and subjective well-being after stroke. Neurology. 2017;89(9):944.

82. Hart DL, Werneke MW, Deutscher D, George SZ, Stratford PW, Mioduski JE. Using intake and change in multiple psychosocial measures to predict functional status outcomes in people with lumbar spine syndromes: a preliminary analysis. Phys Ther. 2011;91(12):1812–25.

83. Garçon L, Lapitan J, Ross A, Nakatani Y, Velazquez Berumen A, Khasnabis C, Walker L, Borg J. Medical and assistive health technology: meeting the needs of aging populations. Gerontologist 2016. 56(Suppl_2):S293–302.

84. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32:D267–70.

85. Kukafka R, Bales ME, Burkhardt A, Friedman C. Human and automated coding of rehabilitation discharge summaries according to the international classification of functioning, disability, and health. J Am Med Inform Assoc. 2006;13(5):508–15.

86. Sundar V, Daumen ME, Conley DJ, Stone JH. The use of ICF codes for information retrieval in rehabilitation research: an empirical study. Disabil Rehabil. 2008;30(12–13):955–62.

87. Greenwald JL, Cronin PR, Carballo V, Danaei G, Choy G. A novel model for predicting rehospitalization risk incorporating physical function, cognitive status, and psychosocial support using natural language processing. Med Care. 2017;55(3):261–6.

88. Newman-Griffis D, Zirikly A. Embedding transfer for low-resource medical named entity recognition: a case study on patient mobility. In: BioNLP. Association for Computational Linguistics: Melbourne, Australia; 2018.

89. Shao Y, Mohanty AF, Ahmed A, Weir CR, Bray BE, Shah RU, Redd D, Zeng-Treitler Q: Identification and use of frailty indicators from text to examine associations with clinical outcomes among patients with heart failure. In: AMIA Annu Symp. vol. 2017, 2017/03/09 edn; 2017: 1110–1118.

90. Abbott K, Ho Y-Y, Erickson J. Automatic health record review to help prioritize gravely ill social security disability applicants. J Am Med Inform Assoc. 2017;24(4):709–16.

91. Davis MF, Sriram S, Bush WS, Denny JC, Haines JL. Automated extraction of clinical traits of multiple sclerosis in electronic medical records. J Am Med Inform Assoc. 2013;20(e2):e334–40.

92. Hirschberg J, Manning CD. Advances in natural language processing. Science. 2015;349:261–6.

93. Jovanovic J, Bagheri E. Semantic annotation in biomedicine: the current landscape. J Biomed Semant. 2017;8(1):44.

94. Pisani E, Aaby P, Breugelmans JG, Carr D, Groves T, Helinski M, Kamuya D, Kern S, Littler K, Marsh V, et al. Beyond open data: realising the health benefits of sharing data. BMJ. 2016;355:i5295.

95. Meystre SM, Ferrández Ó, Friedlin FJ, South BR, Shen S, Samore MH. Text de-identification for privacy protection: a study of its impact on clinical text information content. J Biomed Inform. 2014;50:142–50.

96. Cimino JJ. The false security of blind dates. Appl Clin Inform. 2012;03(04):392–403.

97. Hripcsak G, Mirhaji P, Low AFH, Malin BA. Preserving temporal relations in clinical data while maintaining privacy. J Am Med Inform Assoc. 2016;23(6):1040–5.

98. Bodenreider O, Smith B, Burgun A: The ontology-epistemology divide: a case study in medical terminology. In: Third Int Conf Form Ontol Inf Syst. Edited by Varzi AC, Vieu L: IOS Press; 2004: 185–195.

99. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. N Engl J Med. 2018;379(15):1452–62.

100. Simeonsson RJ, Lollar D, Hollowell J, Adams M. Revision of the international classification of impairments, disabilities, and handicaps: developmental issues. J Clin Epidemiol. 2000;53(2):113–24.

101. Della Mea V, Simoncello A. An ontology-based exploration of the concepts and relationships in the activities and participation component of the international classification of functioning, disability and health. J Biomed Semant. 2012;3(1):1.

102. Tu SW, Nyulas CI, Tudorache T, Musen MA. A method to compare ICF and SNOMED CT for coverage of US Social Security Administration's disability listing criteria. In: AMIA Annu Symp. vol. 2015: American Medical Informatics Association; 2015. p. 1224–33.

103. Heerkens YF, de Weerd M, Huber M, de Brouwer CPM, van der Veen S, Perenboom RJM, van Gool CH, Huib TN, Marja v B-M, Stallinga Hillegonda A, van Meeteren NLU. Reconsideration of the scheme of the international classification of functioning, disability and health: incentives from the Netherlands for a global debate. Disabil Rehabil. 2018;40(5):603–11.

104. Vreeman DJ, Richoz C. Possibilities and implications of using the ICF and other vocabulary standards in electronic health records. Physiother Res Int. 2015;20:210–9.

105. van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, Heymann D, Burke DS. A systematic review of barriers to data sharing in public health. BMC Public Health. 2014;14(1):1144.

106. A robust health data infrastructure. Prepared by JASON at the MITRE Corporation under Contract No. 13-717F-13. Rockville, MD: Agency for Health Research and Quality. April 2014. AHRQ Publication No. 14–0041-EF.

107. Johnson AEW, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035.

108. Institute of Medicine. Capturing social and behavioral domains and measures in electronic health records: phase 2. Washington, DC: The National Academies Press; 2014.

109. Sorna C, Steele R, Inoue A. Word prediction in assistive technologies for aphasia rehabilitation using Systemic Functional Grammar. In: 2009 Annu Meet N Am Fuzzy Inf Proc Soc (NAFIPS); 2009. p. 1–6.

110. Newell A, Langer S, Hickey M. The role of natural language processing in alternative and augmentative communication. Nat Lang Eng. 1998;4(1):1–16.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.