

# Achieving Inclusivity by Design: Social and Contextual Information in Medical Knowledge

Janna Hastings<sup>1,2</sup>

<sup>1</sup> Department of Clinical, Educational and Health Psychology, University College London, UK

<sup>2</sup> Institute for Intelligent Interacting Systems, Otto-von-Guericke University Magdeburg, Germany

## Summary

**Objectives:** To select, present, and summarize the most relevant papers published in 2020 and 2021 in the field of Knowledge Representation and Knowledge Management, Medical Vocabularies and Ontologies, with a particular focus on health inclusivity and bias.

**Methods:** A broad search of the medical literature indexed in PubMed was conducted. The search terms ‘ontology’/‘ontologies’ or ‘medical knowledge management’ for the dates 2020-2021 (search conducted November 26, 2021) returned 9,608 records. These were pre-screened based on a review of the titles for relevance to health inclusivity, bias, social and contextual factors, and health behaviours. Among these, 109 papers were selected for in-depth reviewing based on full text, from which 22 were selected for inclusion in this survey.

**Results:** Selected papers were grouped into three themes, each addressing one aspect of the overall challenge for medical knowledge management. The first theme addressed the development of ontologies for social and contextual factors broadening the scope of health information. The second theme addressed the need for synthesis and translation of knowledge across historical disciplinary boundaries to address inequities and bias. The third theme encompassed a growing interest in the semantics of datasets used to train medical artificial intelligence systems and on how to ensure they are free of bias.

**Conclusions:** Medical knowledge management and semantic resources have much to offer efforts to tackle bias and enhance health inclusivity. Tackling inequities and biases requires relevant, semantically rich data, which needs to be captured and exchanged.

## Keywords

Medical knowledge management, ontologies, vocabularies, health inequities, bias

Yearb Med Inform 2022;228-35

<http://dx.doi.org/10.1055/s-0042-1742509>

## 1 Introduction

After the second year of the ongoing global COVID-19 pandemic, the global medical research community has had ample opportunity to witness, evaluate, track, compare, and learn in real-time from the unfolding of a global medical crisis. One fact that has been apparent throughout is that while all may have been facing the same storm, we have not been in the same boat: the impact of the pandemic has differed, from country to country, community to community, individual to individual, in ways that often tracked the pre-existing fault-lines that already delineated health inequalities before COVID-19 struck [1].

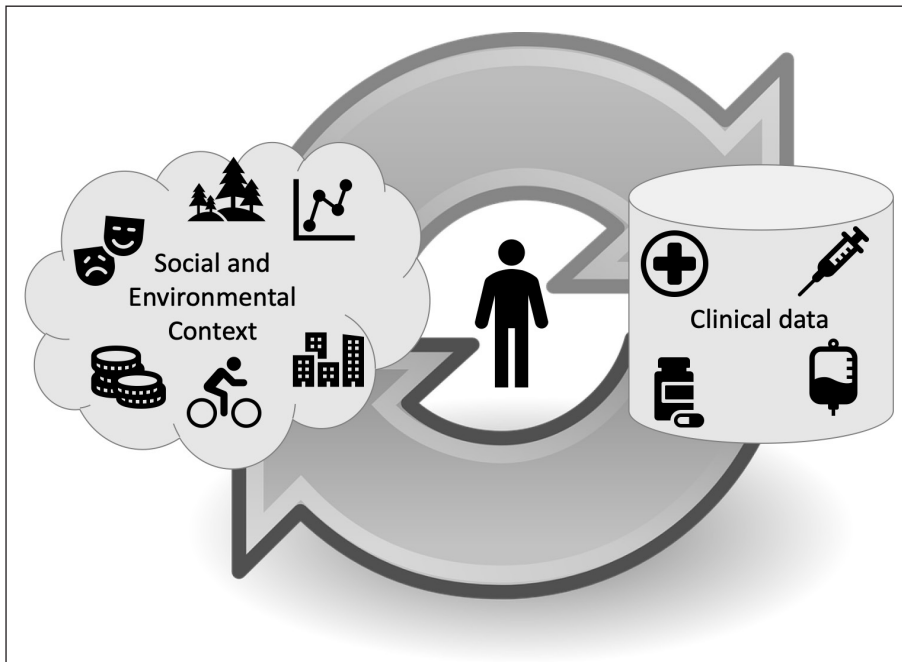
The recent edition of the Global Burden of Disease study [2] emphasises the strong correlation of health outcomes with socioeconomic factors – including educational achievements, gender, housing quality and safety, nutrition, and access to opportunities – and the accompanying need to broaden the scope of medical concern beyond the clinic. What the pandemic has further taught us is that the health of communities, societies, nations and the world as a whole are inextricably intertwined [3, 4].

The previous year’s survey article already discussed the role of knowledge synthesis in tackling the pandemic by enabling a global learning health system [5]. This year, we ask what medical knowledge representation and management can contribute to addressing global health disparities, reducing bias and ensuring health inclusivity – the answer, as we will see in what follows, is ‘a lot’ (Figure 1). Among the contributions identified in the survey, we see calls for an urgent need to transcend historical boundaries between countries, disciplines, and perspectives, for

finding a common vocabulary for public health and for clinical medicine, for paying more attention to data about contextual and social factors within clinics, for working towards semantic interoperability between clinical, research and public health data, and for collecting sufficient data and the right sorts of data that are needed to counteract bias in health systems and address or mitigate the blockers to inclusivity.

A broad search of the medical literature indexed in PubMed returned 9,608 records for the terms ‘ontology’/‘ontologies’ or ‘medical knowledge management’ for the dates 2020-2021 (search conducted November 26, 2021). These were pre-screened based on their titles only for relevance to health inclusivity, bias, social and contextual factors, and health behaviours. Among these, 109 papers were selected for in-depth reviewing based on full text, of which 22 were selected for inclusion. Articles contributing primary results (not reviews) addressing ontologies, semantic resources, or knowledge-based information systems relevant for health inclusivity and bias were included. Selected papers were grouped into three themes, each addressing one aspect of the overall challenge of tackling bias and health inequalities within medical knowledge management.

First, the largest theme includes ontologies, vocabularies and semantic resources that have been developed to enhance the description and reporting of the social and contextual factors that drive health inequalities and bias. These include ontologies and vocabularies encompassing the social determinants of health, behavioural interventions to improve health, and for enhancing clinician-patient communication and dialogue. Capturing social and contextual information



**Fig. 1** Achieving health inclusivity by design: social and contextual information in medical knowledge.

in a structured and reproducible way is an essential first step to understanding and developing systems that mitigate against biases and inequities. However, it is only a first step: such information needs to inform downstream research and practice if it is to lead to a mitigation of bias and inequity, and for that it may need to be integrated, translated and transformed. Thus, a second theme encompasses medical knowledge translation across boundaries, closing the medical and evidential knowledge and translation gaps between different perspectives and domains, and exploring the barriers to such translation, including data privacy requirements. Finally, we consider the role of a different type of knowledge in the form of medical artificial intelligence systems, which are trained on existing data in order to make predictions for unseen data and thereby support the work of clinicians. Trained predictive models carry the biases of the data they are trained on, but due to their many parameters they may obscure those biases in ways that are difficult to detect. There is a growing emphasis on training data semantics in mitigating against bias, and on enhancing the transparency and interpretabil-

ity of those systems. Thus, the third and final theme encompasses semantic resource-based efforts to reduce and mitigate against bias in medical artificial intelligence systems.

## 2 Results

### 2.1 Ontologies for Social and Contextual Information

#### The Social Determinants of Health

The COVID-19 pandemic has highlighted health inequalities, and one aspect of this is the role of data inequalities in exacerbating those health inequalities. Outside of clinical care, the factors that affect health outcomes include the social and economic influences on the life course of an individual, which have been called the ‘social determinants of health’ [6] or ‘non-clinical determinants of health’. We will include in this scope both non-modifiable factors of individuals’ situations and environments, and modifiable risk factors such as smoking and other health-related behaviours. Addressing socioeconomic dis-

parities often improves health – for example, antipoverty programs were found to reduce psychopathologies in US-based youth [7]. Tackling social and economic determinants of health, and predicting differential responses to treatments, involves annotation of social and contextual information that is generally under-reported even in public health contexts [8] and particularly in clinical data.

There is a wide gap between the variables that are relevant for research on the social determinants of health and the associated datatypes that are available in the clinic. There have been explicit calls to collect more data on social determinants of health in order to make health care more equitable [9, 10]. A recent rapid review [11] on combining data on social determinants of health with clinical data found 744 different social determinants of health across the literature, including socioeconomic status, material conditions such as housing and income, social circumstances, demographic measures such as household income, marital status, education (including health literacy), race and ethnicity, health-related behaviours such as smoking, diet, substance abuse, and environmental conditions such as transportation, buildings, and air quality. However, in clinical studies, the overwhelming majority of the studies only used aggregate, area-level information and combined this with individual-level clinical data by matching geographic location.

In order to enable clinical studies to incorporate important contextual information relevant for the social determinants of health, it would be ideal if existing clinical standard vocabularies could encompass these attributes. Although clinical standards such as Logical Observation Identifiers Names & Codes (LOINC) [12], the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) [13, 14] and the International Classification of Diseases (ICD) [15] do incorporate some relevant terminology [16], gaps remain that need to be filled, and concerns about interoperability have led some to call for a dedicated effort to create a harmonised clinical ontology for this domain [17]. For example, ICD includes codes for ‘unemployment’, ‘change of job’ and ‘threat of job loss’, SNOMED-CT includes codes for ‘unemployed’, ‘chronic unemployment’, ‘changed job’, ‘dismissed from job’ and

'loss of job', and LOINC includes codes for 'homemaker' vs. 'employed' vs. 'retired'. Semantic alignment between these different coding schemes will not be trivial at all, and similar complexity plagues many aspects of social and contextual information.

One aspect of broadening the scope of clinical concern to address health inequalities is the use of social prescribing: clinical prescriptions of social and community-based interventions that aim to improve health by addressing the social determinants of health. Examples of interventions that might be available for social prescribing include referrals to sporting or social activities, training programmes, or social support services such as income or housing support. The UK's National Health Service uses primary care social prescribing, and in order to enable better reporting of and analysis of the use of social prescribing, an ontology was created to formally describe types of social prescribing, which was implemented in primary care electronic health systems [18]. The ontology encompasses 668 codes for social prescriptions. Prior to the introduction of this new ontology within primary care systems, it was possible to record only 'social prescribing offered', 'social prescribing declined' and 'referral to social prescribing service' as classified beneath 'finding', 'situation' and 'procedure' respectively in SNOMED-CT. The new ontology encompasses detail about the specific types of social prescriptions, including addiction support services, housing support services, parenting support services, physical activity management services, and stress reduction support services among others. The ontology was evaluated in use by recording how often the codes were used in primary care. They found over 5 million instances of use compared to just 29,000 for the older less semantically specified codes.

Many electronic health records and clinical study questionnaires allow only a range of predefined codes for important demographic variables, thus there is an equity burden to ensure that such fields allow for comprehensive data capture to ensure data representativeness. Data that cannot be captured may result in minority personal attributes and experiences being erased. Within this scope, it is increasingly being recognised that there is a need for a more de-

tailed and sophisticated ontology for gender, sex and sexuality. A new ontology that aims to address this gap is the Gender, Sex and Sexual Orientation ontology (GSSO) [19] which has been added to the Open Biological and Biomedical Ontology (OBO) Library repository [20]. GSSO includes a wider range of gender and sexual identities than other clinical ontologies, such as SNOMED-CT, and is mapped to clinical reference vocabularies to enable interoperability with existing datasets. Although it does not report yet having been implemented in an electronic health record (EHR) setting, it was evaluated for comprehensiveness by means of a study of the terminology in the published literature. A particularly novel feature of the GSSO is its coverage of culturally specific gender identities within global geographic regions.

### Health Behaviours and Health Behaviour Change

The pandemic has highlighted and underscored the enduring and inescapable relevance of human behaviour for health, and therefore also the necessity of developing effective interventions that aim to change human behaviour. A large-scale effort to develop ontologies for the formal description and annotation of behaviour change intervention data, and thereby improve the evidence base in this domain, is the Behaviour Change Intervention Ontology [21–23]. This ontology elaborates a model for behavioural interventions and studies thereof, and provides detailed and systematic vocabularies for different aspects of such interventions including their setting (where they happen) [24], how they are delivered [25], and by whom they are delivered [26]. Several additional modules of this ontology are still being developed [23], including a module for *how* behaviour changes ('mechanisms of action') and one for the methods and techniques that aim to facilitate behaviour change.

Dramatic and sudden changes in working and learning environments accompanied the pandemic, with many workplaces and classrooms becoming fully or partially remote during lockdowns, with consequently reduced quantities of social encounters and potentially reduced access to social support. In one study, an ontology-driven representation of

individual psychological states and traits was used to characterise learner well-being during the pandemic [27]. An online psychological survey used a mix of measures including big five personality traits, depression screening, state and trait anxiety measures, feelings towards the pandemic, worries about mental and physical health, sentiments about social distancing, perceived difficulties identifying and representing emotions, changes in behavioural patterns, and general emotions. These diverse psychological measures were then classified into an ontology, and the ontology was used to reveal additional patterns in the survey data. For example, the ontology contains rules that map from personality traits to sleep behaviours, corresponding to known associations mined from the sleep hygiene literature. The ontology-based analysis highlighted entities around emotional and physical well-being, health-related behaviours such as sleep, and the individual's social situation. Patterns of habitual behaviour were contrasted with once-off behaviours, and behaviour change was defined as a change in a pattern of behaviour rather than as a once-off behaviour.

Continuing the theme of ontologies for behavioural medicine, a lifestyle ontology for lifestyle medicine promotion to prevent prostate cancer was developed [28]. This ontology was built with 397 lifestyles and lifestyle-related factors associated with prostate cancer, including diet, habitual behaviours, social and environmental factors, sleep, sun exposure and others. The ontology draws on broader classes from the National Cancer Institute thesaurus (NCIt) [29] and SNOMED-CT for the development of its hierarchy. The objective of the ontology is to be used to clarify the evidence base in the face of contradictory findings about the relationships between various health behaviours and prostate cancer, e.g. what amount of alcohol consumption and of what type is safe? The need to clarify contradictory findings within a large and growing evidence base and the role of ontologies for that purpose was highlighted more than once: with smoking cessation still being one of the most effective behavioural changes to improve health, yet the evidence around new smoking cessation tools such as e-cigarettes being disputed, ontologies have been suggested as a path towards enhanced clarity for this evidence base [30, 31].

In an alternative application scenario, ontologies have been proposed to support eHealth coaching interventions which target behaviour change in real-time [32]. In this report, an eHealth ontology is developed to support an eHealth coach which is proposed to be able to annotate personal, physiological, behavioural, and contextual data from heterogeneous sources including sensor measurements, questionnaires, and interviews, and thereafter allow meaningful, practical, personalized, and contextual lifestyle recommendations to be generated automatically from ontology-structured data based on defined rules. The effort targets obesity as the initial development use case health challenge. It re-uses content from the Semantic Sensor Network Ontology [33] and SNOMED-CT. However, the proposed framework is evaluated based on simulated data only, so it remains to be seen how it would perform with real sources of unstructured data as inputs.

### Clinicians and Patients in Dialogue

The pandemic has shined a spotlight on the health and well-being of clinicians and all health workers [34], and on the ways that the interpersonal domain, including clinician-patient relationships and conversations, in turn affects health outcomes [35].

In support of better data capture about this aspect of health care, an ontology was developed to capture and characterise the notion of ‘presence’, encompassing relationality and interpersonal connection in clinical encounters [36]. Presence has been shown to be an important factor in the quality of the healthcare encounter both in terms of health outcomes and in terms of reducing stress and improving the well-being of clinicians [37]. The presence ontology includes factors related to presence, such as communication, emotions, tools, environments, empathy, and trust. The ontology is being developed with the aim of supporting investigative methods to improve healthcare processes for both patients and healthcare providers, including annotation of videotaped encounters, development of clinical instruments to measure presence, and even the implementation of EHR reminders for providers to support improved presence.

Clinicians are often on the frontline of the effort to counteract the effects of health misinformation on vaccine hesitancy [38]. As such, there is a need to develop supporting tools that are able to assist clinicians in managing these dialogues. One proposed tool [39] uses an ontology-based automated conversation agent to manage dialogues with patients about vaccines. This tool is based on a Patient Health Information Dialogue Ontology (PHIDO) [40] which provides terminology to support questions and answers around vaccines. The overall system that is developed includes a speech to text module, followed by a natural language processing module for the resulting text, which matches the free text to questions in the ontology using salient noun and verb phrases. The ontology and accompanying ontology-based reasoning is used to match questions with answers, after which the best-matched answer is returned. For evaluation, simulated questions are used rather than real questions from the target population, and the scoring of the answers is done manually. It would be interesting to see how this approach performed in real-world conversations.

An ontology-based automated dialogue system has also been proposed for use in determining the existence of adverse childhood experiences [41] and making accompanying recommendations designed to reduce the incidence of mental health conditions. This system is based on an ontology for adverse childhood experiences [42] which encompasses entities such as witnessing crimes, lack of housing, bullying, emotional neglect, food insecurity, poverty, placement in social care, involvement of justice department among others. Such adverse childhood experiences can have a profound impact on health outcomes across the life course of an individual.

## 2.2 Closing Knowledge and Translation Gaps

### Interoperability and Translation

Interoperability of data and vocabularies is a perennial concern within the field of medical knowledge, and the accompanying concerns can hinder the adoption of new ontologies and vocabularies such as those

surveyed above, if the new resources are not interoperable with the existing technologies and vocabularies. However, there is widespread agreement on shared underlying standards and best practices, such as the use of the Web Ontology Language for ontologies. In support of the exchange and sharing of annotated and structured data, there is a need for clinical data infrastructures that support the flexible and interoperable application of exchangeable - “pluggable” - semantic resources. One example of such an infrastructure is that developed by a national health data exchange programme in Switzerland [43].

There are well-known barriers to data exchange and interpretation between different health perspectives and teams, such as population health and the clinic, or between different health systems [5]. Supporting data exchange and semantic interoperability across a wider range of different contexts has the potential to bring benefits both for individual patients and healthcare workers, and for addressing broader objectives in health systems research. Knowledge work, including semantic harmonisation, is therefore needed at the boundaries between different disciplines, practices and healthcare systems, while at present, such work is largely carried out within and for specific focus areas rather than between them.

For example, it is necessary to enable data to be shared between healthcare and social care organizations [44] – currently very challenging in most health care ecosystems – and thereby promote the integration of social care with health care to create a more equitable health provision. Taking a social care perspective on clinical data may also reveal additional challenges in data quality: a recent study [45] found evidence of bias in the patterns of missing or present data in a clinical EHR, such that for certain groups, data were systematically under-reported (i.e., missing not at random). Evidence-based data quality improvement techniques for clinical EHR data on social determinants of health may reduce bias and improve quality overall.

Data interoperability across different health systems has also been proposed to be supported by an ontology [46] for continuity of care and a continuous care

relationship. This effort, building on a Hospital at Home model, aims to support the management of individual care at home as far as possible and thereby enable the intelligent management of healthcare resources, using remote patient monitoring as far as possible together with a continuous flow of information.

There are other knowledge boundaries that do not primarily involve different data-types or disciplines or practices. For example, there are boundaries based on culture and language, which may operate at the interface between countries or communities. One paper calls for increased awareness of the fact that digital interventions may need to be culturally adapted [47]. This point may have particular relevance in mental health contexts, which form a large and growing application area for digital health interventions [48]. Translation of interoperable health information standards into different languages perennially requires nuance and sophisticated attention to intended semantics. In this light, a new rule-based method [49] is proposed to choose the best translation of Human Phenotype Ontology [50] labels into Japanese from among four candidate translations.

### Personalised Medical Knowledge

A high profile recent article [51] calls for personalised profiles for disease risk prediction to capture a wider scope of health determinants, including environmental factors e.g., income, exposures, culture, lifestyle and other risk factors. The article suggests that these factors need to be combined with clinical measurements in order for each to contextualise the other and their interactions with genetic risk scores to achieve personalised medicine objectives. Several shortcomings of the current data available for personalised medicine are noted. Firstly, genetic information is largely biased towards white, middle class populations, and needs a broader basis among more-diverse populations. Secondly, there is a need to include non-genetic risk factors – data about social, cultural and economic factors – and also to include more real-time measurements of clinical state, to go beyond static one-time data. In this respect,

it is interesting to note a recent ontology for the representation of time in clinical events, and of clinical temporal relations [52]. Finally, the article suggests a need to move away from tendency to collapse rich individual-level data into rigid (thresholded, binarised) clinical categories. Related to the non-representativeness of publicly available omics data for personalised medicine, another study [53] explored sex biases in publicly available gene expression data and found that while sex bias was low overall, it was high in some specific sub-fields dealing with particular conditions.

Many insights can be gained from re-use of interoperable, patient history-oriented clinical data together with other layers such as -omics data [54] for the better understanding of diseases, including viral diseases such as COVID-19. For example, better linking of patient history descriptions to viral genomic features has been proposed [55] and to share ‘patient status’ alongside viral sequencing information. However, in practice, such information is largely being supplied in free text, with semantically vacuous expressions such as ‘unknown’ and ‘not provided’ making up the largest fraction.

While much has been written in recent years about the potential for the use of clinical data for data-driven, personalised medicine, less has been written about the need to have data available for clinical decision-making based on the actual characteristics of patients in the clinic, which may not correspond to those of participants in randomised controlled trials after all applicable exclusions have been applied to arrive at a clean study group, and challenges remain in delineating to whom findings from randomised controlled trials can be taken to apply [56]. This points to the need for enhanced metadata systematically describing the population groups associated with such trials. An evaluation of the metadata [57] associated with clinical trial descriptions in the ClinicalTrials.gov repository concluded that trials are poorly annotated with structured terminologies overall – even with respect to the main condition targeted by the respective studies – and, importantly, the eligibility criteria for study participants lacked standardisation, which is a crucial element for interpreting their applicability.

### Data Privacy

While efforts to enhance the reporting of social and contextual factors affecting health outcomes are welcome, and can be expected to lead to improved inclusivity and reduced bias in clinical datasets and the resulting research, better reporting of this type of information nevertheless poses an additional privacy challenge. The more information that is collected about an individual patient or research participant, the more difficult it is to ensure that person’s privacy even after overtly identifying information such as their name or date of birth have been removed. Different data-sharing strategies have been proposed to deal with the challenge while still enabling re-use of data for research purposes [58]. Data managers have a responsibility and legal obligation to ensure that information is securely stored, and only made accessible for those purposes for which the individual has given their consent. In some cases, only aggregate or encrypted data may be shared. Another strategy is to execute analysis algorithms in a distributed fashion on secure servers and return only their results without gaining direct access to secured datasets. However, even with distributed and secure processes, care must be taken as some approaches - such as ‘deep’ neural networks - can accidentally ‘leak’ private information from secured sources into their results.

## 2.3 Achieving Inclusivity by Design in Medical AI

Medical artificial intelligence (AI) encompasses the use of advanced technologies for machine learning and reasoning to solve challenges in healthcare, including image analysis and decision support. In many cases, such systems are sophisticated models with many parameters that have been *trained*, and their behaviour is dependent on the dataset on which they have been trained, on the structures and relationships that they have learned therefrom. In this sense, they can be said to be knowledge-based systems, even though their internal representation of knowledge is not explicit but rather implicit.

Interest in, and expectations of, medical artificial intelligence have been growing throughout the last decade, although recently concerns have also been growing about their risk of bias and exacerbating pre-existing inequalities [59, 60]. The World Health Organisation has issued a report [61] on AI in health describing six guiding principles for its design and use: ‘protecting human autonomy’, ‘promoting human well-being and safety and the public interest’, ‘ensuring transparency, explainability and intelligibility’, ‘fostering responsibility and accountability’, ‘ensuring inclusiveness and equity’ and ‘promoting AI that is responsive and sustainable’. It is increasingly being recognised that achieving many of these objectives will require a deeper inspection of the underlying data on which predictive models are based: inequalities and biases in medical artificial intelligence systems are driven by biases in the underlying training data.

Some biases that machine learning systems have been found to show reflect socioeconomic disparities with negative effects on health. For example, one study [62] found that a widely used artificial intelligence system that aims to identify health risks in the US systematically disadvantages Black persons. For a given risk prediction score, Black patients were found to be considerably sicker than White patients at the same risk prediction score. This disparity reduced the help that was made available to Black patients on the basis of the prediction system. The study determined that the bias arose because the predictive model used data relating to health care costs as a proxy for health care need, but pre-existing systematic economic disadvantages meant that the health care expenditures for Black persons in the dataset had been lower without the health care needs being concomitantly lower.

This finding is a reminder that data semantics cannot be ignored in the development of medical artificial intelligence systems, and also that clinicians and other healthcare workers must have the training and resources that they need to inspect and critically interpret the results of algorithms that are deployed in their healthcare settings. Effective quantification and communication of uncertainty in a model’s

predictions and its sources could help to increase trust in the model by healthcare workers, and provide safeguards against known failure modes of current learning approaches [63]. Uncertainty can stem from different sources – noise, model parameters and model selection, and dataset shift. The sorts of high-dimensional data that typically form the input to train medical artificial intelligence systems has many more features than samples (the ‘curse’ of dimensionality) [64]. Thus, it is common that the resulting learned models are underspecified and prone to overfitting, and their performance may be overestimated even when using evaluation techniques such as cross-validation that hide a part of the input data in order to use it as a test of model performance. Model complexity and feature space should be limited during development and care should be taken to formally validate that the contextual features do not differ in unexpected ways between the training dataset and the data expected in the final deployment.

Healthcare artificial intelligence has also been found to cause concern to patients [65], who are concerned about the safety of artificially ‘intelligent’ devices, and who would like to see regulatory oversight and ultimate clinician responsibility over the use of such devices, with the discretion for treatment decisions remaining in the realm of clinicians and shared clinician-patient decision-making processes. Patients are concerned as to whether artificial intelligence could lead to new forms of discrimination, e.g., higher insurance premiums. These concerns relate at least in part to the contextual and semantic attributes of the underlying data on which such models are trained: is it accurate, representative, and unbiased? In order to evaluate whether this is the case, a necessary precondition is to collect and be able to explore information about the social and contextual variables that drive biases and inequalities in health-related datasets. And in case the available data is indeed determined to be biased, solutions must be found which enable training unbiased models nevertheless: for example, synthetic minority group cohort data may be generated [66] that explicitly enhances the representativeness of training datasets.

### 3 Conclusions

Each of the articles included in this survey contribute in some way to the overall goal of tackling bias and ensuring health inclusivity, by addressing different problems and barriers, many of which have been exacerbated by the ongoing pandemic crisis.

Differential distributions of health outcomes track pre-existing socioeconomic disparities, and patients are not always able to access treatments that they need. This has been shown to be particularly true in the context of *global* health through the pandemic, during which we have seen avoidable mortality and morbidity in part due to differential social determinants of health as globally distributed. The need for better understanding this problem necessitates describing and keeping track of the associated variables better, and for the ability to exchange such data globally, even while remaining mindful of the challenges associated with privacy preservation for individual patients and research participants.

Some aspects of the wider body of medical knowledge have been brought into the focus more as a result of the pandemic and are becoming more central, including for example behavioural science and its role in health promotion such as by reducing viral transmission. Better evidence from behavioural science needs to be combined with evidence arising from other aspects of clinical science, and in general the need for integration of the whole medical evidence base transcending disciplinary sub-divisions has been highlighted.

Finally, there are concerns about ways in which health information systems might be themselves exacerbating pre-existing biases and inequalities, for example due to poorer evidence bases for some sub-populations, or due to biases in algorithms. Health informaticians should be vigilant in canvassing our evidence to identify gaps and where representativeness is failing, as well as ensuring the interpretability and transparency of algorithms and predictive models linked to representativeness of underlying data. In other words, inclusivity will only be achieved by deliberate design – which, for health information systems, means *semantic* design.

## References

- Mishra V, Seyedenouzi G, Almohtadi A, Chowdhury T, Khashkhusa A, Axiaq A, et al. Health Inequalities During COVID-19 and Their Effects on Morbidity and Mortality. *J Health Leadersh* 2021 Jan 19;13:19-26.
- The Lancet. Global health: time for radical change? *Lancet* 2020 Oct 17;396(10258):1129.
- Fontanet A, Autran B, Lina B, Kieny MP, Karim SSA, Sridhar D. SARS-CoV-2 variants and ending the COVID-19 pandemic. *Lancet* 2021 Mar 13;397(10278):952-4.
- Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*. 2022 Feb;602(7898):657-63.
- Verspoor K. The Evolution of Clinical Knowledge During COVID-19: Towards a Global Learning Health System. *Yearb Med Inform* 2021 Aug;30(1):176-84.
- Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep* 2014 Jan-Feb;129 Suppl 2(Suppl 2):19-31.
- Weissman DG, Hatzenbuehler ML, Cikara M, Barch D, McLaughlin KA. Antipoverty programs mitigate socioeconomic disparities in brain structure and psychopathology among U.S. youths [Internet]. *PsyArXiv*; 2021 [cited 2021 Dec 1]. Available from: <https://psyarxiv.com/8nhej/>
- Buckeridge DL. Precision, Equity, and Public Health and Epidemiology Informatics - A Scoping Review. *Yearb Med Inform* 2020 Aug;29(1):226-30.
- Weir RC, Proser M, Jester M, Li V, Hood-Ronick CM, Gurewich D. Collecting Social Determinants of Health Data in the Clinical Setting: Findings from National PRAPARE Implementation. *J Health Care Poor Underserved* 2020;31(2):1018-35.
- Breen N, Berrigan D, Jackson JS, Wong DWS, Wood FB, Denny JC, et al. Translational Health Disparities Research in a Data-Rich World. *Health Equity* 2019 Nov 8;3(1):588-600.
- Golembiewski E, Allen KS, Blackmon AM, Hinrichs RJ, Vest JR. Combining Nonclinical Determinants of Health and Clinical Data for Research and Evaluation: Rapid Review. *JMIR Public Health Surveill* 2019 Oct 7;5(4):e12846.
- McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem* 2003 Apr;49(4):624-33.
- Millar J. The Need for a Global Language - SNOMED CT Introduction. *Stud Health Technol Inform* 2016;225:683-5.
- International Health Terminology Standards Development Organization. Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT); 2022.
- World Health Organization. International Classification of Diseases and Related Health Problems [Internet]. Twelfth. World Health Organization; 2020. Available from: <https://icd.who.int/>
- Arons A, DeSilvey S, Fichtenberg C, Gottlieb L. Documenting social determinants of health-related clinical activities using standardized medical vocabularies. *JAMIA Open* 2018 Dec 24;2(1):81-8.
- Watkins M, Viernes B, Nguyen V, Rojas Mezarina L, Silva Valencia J, Borbolla D. Translating Social Determinants of Health into Standardized Clinical Entities. *Stud Health Technol Inform* 2020 Jun 16;270:474-8.
- Jani A, Liyanage H, Okusi C, Sherlock J, Hoang U, Ferreira F, et al. Using an Ontology to Facilitate More Accurate Coding of Social Prescriptions Addressing Social Determinants of Health: Feasibility Study. *J Med Internet Res* 2020 Dec 11;22(12):e23721.
- Kronk CA, Dexheimer JW. Development of the Gender, Sex, and Sexual Orientation ontology: Evaluation and workflow. *J Am Med Inform Assoc* 2020 Jul 1;27(7):1110-5.
- Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 2007 Nov;25(11):1251-5.
- Michie S, Thomas J, Mac Aonghusa P, West R, Johnston M, Kelly MP, et al. The Human Behaviour-Change Project: An artificial intelligence system to answer questions about changing behaviour. *Wellcome Open Res* 2020 Jun 10;5:122.
- Wright AJ, Norris E, Finnerty AN, Marques MM, Johnston M, Kelly MP, et al. Ontologies relevant to behaviour change interventions: a method for their development. *Wellcome Open Res* 2020 Dec 18;5:126.
- Michie S, West R, Finnerty AN, Norris E, Wright AJ, Marques MM, et al. Representation of behaviour change interventions and their evaluation: Development of the Upper Level of the Behaviour Change Intervention Ontology. *Wellcome Open Res* 2021 Jan 6;5:123.
- Norris E, Marques MM, Finnerty AN, Wright AJ, West R, Hastings J, et al. Development of an Intervention Setting Ontology for behaviour change: Specifying where interventions take place. *Wellcome Open Res* 2020 Jun 10;5:124.
- Marques MM, Carey RN, Norris E, Evans F, Finnerty AN, Hastings J, et al. Delivering Behaviour Change Interventions: Development of a Mode of Delivery Ontology. *Wellcome Open Res* 2021 Feb 26;5:125.
- Norris E, Wright AJ, Hastings J, West R, Boyd N, Michie S. Specifying who delivers behaviour change interventions: development of an Intervention Source Ontology. *Wellcome Open Res* 2021 Apr 8;6:77.
- Bolock AE, Abdennadher S, Herbert C. An Ontology-Based Framework for Psychological Monitoring in Education During the COVID-19 Pandemic. *Front Psychol* 2021 Jul 22;12:673586.
- Chen Y, Yu C, Liu X, Xi T, Xu G, Sun Y, et al. PCLION: An Ontology for Data Standardization and Sharing of Prostate Cancer Associated Lifestyles. *Int J Med Inform* 2021 Jan;145:104332.
- NClT Developers. NCI Thesaurus OBO Edition: Disease or Disorder [Internet]. Ontology Lookup Service. 2020 [cited 2020 Sep 24]. Available from: [https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT\\_C2991](https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C2991)
- Cox S, Hastings J, West R, Notley C. The case for development of an E-cigarette Ontology (E-CigO) to improve quality, efficiency and clarity in the conduct and interpretation of research. *Qeios* [Internet]. 2020 Apr 3 [cited 2020 Nov 17]. Available from: <https://www.queios.com/read/5YYRJP>
- Cox S, Notley C. Cleaning up the science: the need for an ontology of consensus scientific terms in e-cigarette research. *Addiction* 2021 May;116(5):997-8.
- Chatterjee A, Prinz A, Gerdes M, Martinez S. An Automatic Ontology-Based Approach to Support Logical Representation of Observable and Measurable Data for Healthy Lifestyle Management: Proof-of-Concept Study. *J Med Internet Res* 2021 Apr 9;23(4):e24656.
- Compton M, Barnaghi P, Bermudez L, Garcia-Castro R, Corcho O, Cox S, et al. The SSN ontology of the W3C semantic sensor network incubator group. *J Web Semant* 2012 Dec 1;17:25-32.
- Søvdal LE, Naslund JA, Kousoulis AA, Saxena S, Qoronfle MW, Grobler C, et al. Prioritizing the Mental Health and Well-Being of Healthcare Workers: An Urgent Global Public Health Priority. *Front Public Health* 2021 May 7;9:679397.
- Zhou Y, Yang WZF, Ma Y, Wu Q, Yang D, Liu T, Wu X. Doctor-Patient Relationship in the Eyes of Medical Professionals in China During the COVID-19 Pandemic: A Cross-Sectional Study. *Front Psychiatry* 2021 Oct 28;12:768089.
- Maitra A, Kamdar MR, Zulman DM, Haverfield MC, Brown-Johnson C, Schwartz R, et al. Using ethnographic methods to classify the human experience in medicine: a case study of the presence ontology. *J Am Med Inform Assoc*. 2021 Aug 13;28(9):1900-9.
- Brown-Johnson C, Schwartz R, Maitra A, Haverfield MC, Tierney A, Shaw JG, et al. What is clinician presence? A qualitative interview study comparing physician and non-physician insights about practices of human connection. *BMJ Open* 2019 Nov 3;9(11):e030831.
- Katzman JG, Katzman JW. Primary Care Clinicians as COVID-19 Vaccine Ambassadors. *J Prim Care Community Health* 2021 Jan-Dec;12:21501327211007026.
- Amith M, Lin RZ, Cui L, Wang D, Zhu A, Xiong G, et al. Conversational ontology operator: patient-centric vaccine dialogue management engine for spoken conversational agents. *BMC Med Inform Decis Mak* 2020 Dec 14;20(Suppl 4):259.
- Amith M, Roberts K, Tao C. Conceiving an application ontology to model patient human papillomavirus vaccine counseling for dialogue management. *BMC Bioinformatics* 2019 Dec 23;20(Suppl 21):706.
- Ammar N, Shaban-Nejad A. Explainable Artificial Intelligence Recommendation System by Leveraging the Semantics of Adverse Childhood Experiences: Proof-of-Concept Prototype Development. *JMIR Med Inform* 2020 Nov 4;8(11):e18752.
- Brenas JH, Shin EK, Shaban-Nejad A. Adverse Childhood Experiences Ontology for Mental Health Surveillance, Research, and Evaluation: Advanced Knowledge Representation and Semantic Web Techniques. *JMIR Ment Health* 2019 May 21;6(5):e13498.
- Gaudet-Blavignac C, Raisaro JL, Touré V, Österle

- S, Cramer K, Lovis C. A National, Semantic-Driven, Three-Pillar Strategy to Enable Health Data Secondary Usage Interoperability for Research Within the Swiss Personalized Health Network: Methodological Study. *JMIR Med Inform* 2021 Jun 24;9(6):e27591.
44. Bibbins-Domingo K. Integrating Social Care Into the Delivery of Health Care. *JAMA* 2019 Nov 12;322(18):1763-4.
  45. Cook LA, Sachs J, Weiskopf NG. The quality of social determinants data in the electronic health record: a systematic review. *J Am Med Inform Assoc* 2021 Dec 28;29(1):187-96.
  46. Das S, Hussey P. ContSonto: A Formal Ontology for Continuity of Care. *Stud Health Technol Inform* 2021 Oct 27;285:82-87.
  47. Marwaha JS, Kvedar JC. Cultural adaptation: a framework for addressing an often-overlooked dimension of digital health accessibility. *NPJ Digit Med* 2021 Oct 1;4(1):143.
  48. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of Engagement With Digital Interventions on Mental Health Outcomes: A Systematic Review and Meta-Analysis. *Front Digit Health* 2021 Nov 4;3:764079.
  49. Ninomiya K, Takatsuki T, Kushida T, Yamamoto Y, Ogishima S. Choosing preferable labels for the Japanese translation of the Human Phenotype Ontology. *Genomics Inform* 2020 Jun;18(2):e23.
  50. Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res* 2021 Jan 8;49(D1):D1207-D1217.
  51. McCarthy M, Birney E. Personalized profiles for disease risk must capture all facets of health. *Nature* 2021 Sep;597(7875):175-7.
  52. Li F, Du J, He Y, Song HY, Madkour M, Rao G, et al. Time event ontology (TEO): to support semantic representation and reasoning of complex temporal relations of clinical events. *J Am Med Inform Assoc* 2020 Jul 1;27(7):1046-56.
  53. Flynn E, Chang A, Altman RB. Large-scale labeling and assessment of sex bias in publicly available expression data. *BMC Bioinformatics* 2021 Mar 30;22(1):168.
  54. Robinson PN, Haendel MA. Ontologies, Knowledge Representation, and Machine Learning for Translational Research: Recent Contributions. *Yearb Med Inform* 2020 Aug;29(1):159-62.
  55. Bauer DC, Metke-Jimenez A, Maurer-Stroh S, Tiruvayipati S, Wilson LOW, Jain Yet al. Interoperable medical data: The missing link for understanding COVID-19. *Transbound Emerg Dis* 2021 Jul;68(4):1753-60.
  56. Averitt AJ, Weng C, Ryan P, Perotte A. Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations. *NPJ Digit Med* 2020 May 11;3:67.
  57. Miron L, Gonçalves RS, Musen MA. Obstacles to the reuse of study metadata in ClinicalTrials.gov. *Sci Data* 2020 Dec 18;7(1):443.
  58. Wirth FN, Meurers T, Johns M, Prasser F. Privacy-preserving data sharing infrastructures for medical research: systematization and comparison. *BMC Med Inform Decis Mak* 2021 Aug 12;21(1):242.
  59. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *Br Med Bull* 2021 Sep 10;139(1):4-15.
  60. Madai VI, Higgins DC. Artificial Intelligence in Healthcare: Lost In Translation? *ArXiv210713454 Cs* [Internet] 2021 Jul 28 [cited 2021 Sep 20]. Available from: <http://arxiv.org/abs/2107.13454>
  61. WHO issues first global report on Artificial Intelligence (AI) in health and six guiding principles for its design and use [Internet]. [cited 2021 Aug 26]. Available from: <https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use>
  62. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019 Oct 25;366(6464):447-53.
  63. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digit Med* 2021 Jan 5;4(1):4.
  64. Berisha V, Krantsevich C, Hahn PR, Hahn S, Dasarathy G, Turaga P, Liss J. Digital medicine and the curse of dimensionality. *NPJ Digit Med* 2021 Oct 28;4(1):153.
  65. Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digit Med* 2021 Sep 21;4(1):140.
  66. Bing S, Dittadi A, Bauer S, Schwab P. Conditional Generation of Medical Time Series for Extrapolation to Underrepresented Populations. *ArXiv220108186 Cs* [Internet] 2022 Jan 20 [cited 2022 Jan 26]. Available from: <http://arxiv.org/abs/2201.08186>

**Correspondence to:**

Janna Hastings  
 Department of Clinical, Educational and Health Psychology  
 University College London  
 1-19 Torrington Place  
 London, WC1E 7HB, UK  
 E-mail: [j.hastings@ucl.ac.uk](mailto:j.hastings@ucl.ac.uk)