



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Data gaps in electronic health record (EHR) systems: An audit of problem list completeness during the COVID-19 pandemic

Jordan Poulos^{a,b}, Leilei Zhu^{b,c}, Anoop D. Shah^{c,*}

^a UCL Medical School, University College London, Gower Street, London, WC1E 6BT, UK

^b EHRS Directorate, University College London Hospitals NHS Foundation Trust, 250 Euston Rd, London, NW1 2PG, UK

^c Clinical and Research Informatics Unit, UCL/UCLH NIHR Biomedical Research Centre, UCL Institute of Health Informatics, 222 Euston Road, London, NW1 2DA, UK

ARTICLE INFO

Keywords:

EHR systems
Electronic health record
Problem lists
Health information technology
COVID-19
Data completeness

ABSTRACT

Objective: To evaluate the completeness of diagnosis recording in problem lists in a hospital electronic health record (EHR) system during the COVID-19 pandemic.

Design: Retrospective chart review with manual review of free text electronic case notes.

Setting: Major teaching hospital trust in London, one year after the launch of a comprehensive EHR system (Epic), during the first peak of the COVID-19 pandemic in the UK.

Participants: 516 patients with suspected or confirmed COVID-19.

Main outcome measures: Percentage of diagnoses already included in the structured problem list.

Results: Prior to review, these patients had a combined total of 2841 diagnoses recorded in their EHR problem lists. 1722 additional diagnoses were identified, increasing the mean number of recorded problems per patient from 5.51 to 8.84. The overall percentage of diagnoses originally included in the problem list was 62.3% (2841 / 4563, 95% confidence interval 60.8%, 63.7%).

Conclusions: Diagnoses and other clinical information stored in a structured way in electronic health records is extremely useful for supporting clinical decisions, improving patient care and enabling better research. However, recording of medical diagnoses on the structured problem list for inpatients is incomplete, with almost 40% of important diagnoses mentioned only in the free text notes.

1. Introduction

The problem list is a feature of electronic health records (EHR) which provides a persistent summary of diagnoses and other health issues, in order to facilitate handovers and continuity of care [1,2]. Dr Lawrence Weed originally envisioned the problem list as an index, containing a “complete list of all the patient’s problems, including both clearly established diagnoses and all other unexplained findings that are not yet clear manifestations of a specific diagnosis, such as abnormal physical findings or symptoms” [3].

Recording information about diagnoses in a structured way can potentially enable decision support such as medication alerts, treatment suggestions and differential diagnoses [4]. Problem lists terms can be coded using a terminology system, such as SNOMED CT concepts [5], and used as valuable resource to support health research and informatics [6,7].

Previous studies of problem list completeness have based estimates

on selections of high prevalence conditions, using other data in the EHR as a gold standard (Table 1). Improving the completeness of problem lists, and ensuring their accuracy, is critical to patient safety, medical education and clinical communication in the era of health digitalisation [8]. Methods that have been shown to increase problem list completeness include problem orientated charting [9], problem list integration throughout the EHR [10], clinician alters [11], self-reporting of conditions from patients [12–15] and automatic population of the problem list from other areas of the EHR [16] or via natural language processing (NLP) [17–20] (further information in Appendix).

This study sought to assess the completeness of recoding of problem list entries during the COVID-19 pandemic, one year after the installation of a comprehensive EHR system (Epic, May 2019 edition) at UCLH (University College London Hospitals) Trust. Epic is a widely used EHR system internationally, with an estimated 29% market share in the US [21], and is currently live in 3 other NHS Trusts and being prepared for installation in at least 4 additional NHS Trusts. The EHR included a

* Corresponding author at: Room 403, UCL Institute of Health Informatics, 222 Euston Road, London, NW1 2 DA, UK.

E-mail address: anoop@doctors.org.uk (A.D. Shah).

<https://doi.org/10.1016/j.ijmedinf.2021.104452>

Received 6 March 2021; Accepted 29 March 2021

Available online 1 April 2021

1386-5056/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1
Estimates of problem list completion rates in the existing literature.

Study	Setting	Condition(s) studied	Method for comparison ('gold standard')	EHR system used	Percentage of diagnoses included in problem list
Wang et al., 2020 [25]	383,404 patients in Partners Healthcare System, USA	Asthma, Crohn's disease, depression, diabetes, epilepsy, hypertension, schizophrenia, ulcerative colitis	Algorithm based on ICD-10 diagnoses and prescriptions	Epic	72.9–93.5% (unweighted mean 85.5%)
Wright et al., 2015 [26]	160,341 patients in 10 healthcare organisations in USA, UK and Argentina	Diabetes	Haemoglobin A1c $\geq 7.0\%$	Mixture of EHR Systems: Epic (n = 3), Allscripts (n = 2), EMIS (n = 1) and self-developed EHR systems (n = 5)	60.2–99.4% (weighted mean 78.2%)
Wright et al., 2011 [27]	100,000 patients at a single hospital in USA	17 medical conditions	Associations with medication and laboratory results	N/A	4.7–76.2% (unweighted mean 51.7%)
Polubriaginof et al., 2016 [15]	1472 patients in a single hospital in USA	59 medical conditions	Self-reported past medical history using a tablet questionnaire	N/A	54.2%
Hoffman et al., 2002 [28]	148 patients attending a general medicine clinic at a university affiliated Veterans Affairs hospital	9 diagnoses relevant to the choice of drug therapy for hypertension	Sensitivity and specificity of diagnoses recorded in the problem list of electronic records, with medical charts as the standard for comparison	VISTA (Veterans Health Information Systems and Technology Architecture)	42–81% (unweighted mean 62.4%)
Current study	516 inpatients with COVID-19 in a London teaching hospital	Any medical condition relevant to ongoing care	Manual review of free text medical records	Epic	62.3%

structured data field for COVID-19 which was used consistently, but other information (such as diagnoses recorded in the problem list) were commonly recorded only as free-text electronic notes. Comprehensive retrospective chart review of this specific cohort of patients, and recovery of problem list data was required for EHR-derived COVID-19 datasets (see [Appendix](#)) and to support research at the trust to evaluate prognostic models for COVID-19 [22].

The aim of our audit was to assess whether information on key diagnoses was included in the problem list or stored only as unstructured free text notes.

2. Materials and methods

2.1. Study type

This is a retrospective EHR system-based chart review during the COVID-19 pandemic.

2.2. Participants

All inpatients with confirmed or clinically suspected COVID-19 infection were identified by an EHR data warehouse search, with 516 patients included based on the following criteria:

2.3. Inclusion criteria

Patients with a 'suspected' or 'confirmed' COVID-19 infection flag in the EHR prior to 2nd June 2020 were included in this case note review. The COVID-19 flag was set by the infectious diseases team, according to an overall clinical assessment including virology testing and the clinical picture.

2.4. Exclusion criteria

Patients were excluded if their hospital admission during the same period was unrelated to a suspected or confirmed COVID-19 infection.

2.5. EHR system

The EHR System deployed at UCLH, Epic (May 2019 version),

includes electronic documentation, order communications, clinical workflow with decision support and knowledge management to develop an evidence-based care pathway.

2.6. Data collection

Medical students recruited to the EHR department, under close supervision of the Trust Clinical Data Standards Lead/Advisor (LZ), undertook an assessment of pre-specified data fields in each patient's EHR in accordance with a clinically signed off Standard Operating Procedure (see [Appendix](#)). Primary data collection was undertaken in May 2020.

Problems were considered 'missing' if there was evidence in the text notes that the patient had a medical condition (either a new diagnosis or past medical history) that was not included on the problem list, and it was important enough that it would be considered good practice to include it. Good practice was defined to include all on-going chronic medical conditions and major new diagnoses, particularly those that require ongoing treatment or monitoring, as per recent guidance on the use of problem lists [1]. Recommended practice at UCLH is for all patients to have an up-to-date and complete problem list, and this is incorporated into EHR system training for clinicians [23]. The judgement on whether a medical condition should be included on the problem list was supervised by a consultant clinician with experience in problem list management (ADS) [1].

2.7. Statistical analysis

An estimate of the problem list 'data gap' between all the clinically relevant problems which should be recorded in the problem list, and those that were actually recorded during the admission, was assessed using the EHR audit trail.

Problems in UCLH EHR are entered using a proprietary terminology, which are mapped in the background to SNOMED CT and ICD-10. For ease of reporting categories of problems in this audit, problems were aggregated using ICD-10. Confidence intervals for proportions were calculated using the binomial distribution. Statistical analysis was carried out using R, version 3.4.4 [24].

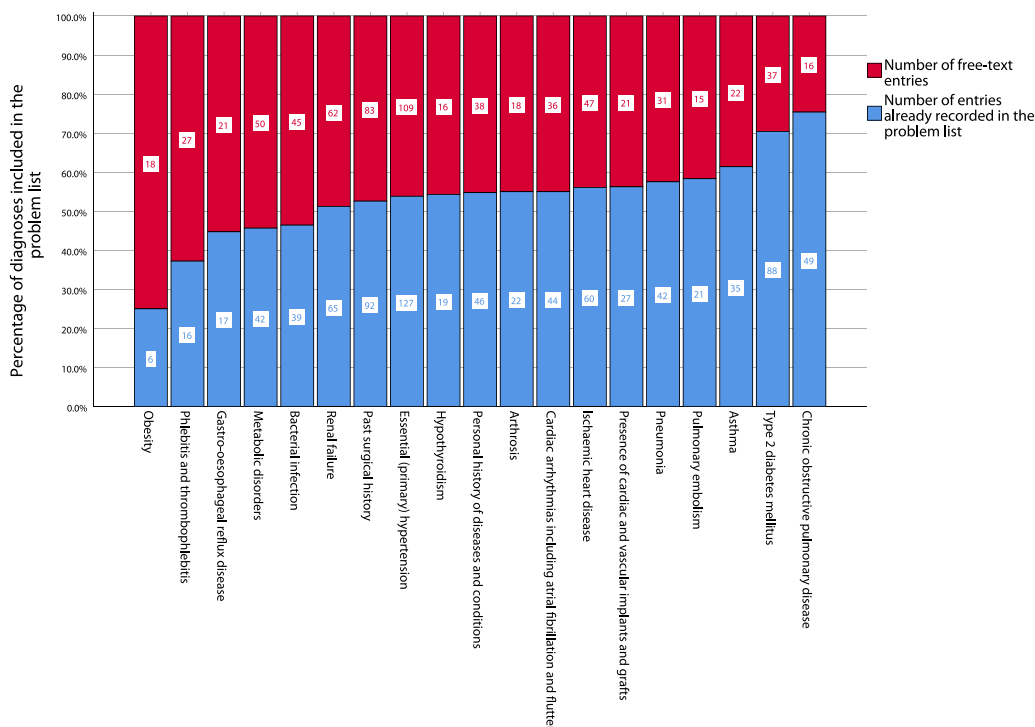


Fig. 1. Proportion of diagnoses recorded on the problem list, aggregated at the ICD-10 block level, for conditions with at least 15 additional entries from manual review of electronic notes.

2.8. Patient and public involvement

Patients were not directly involved in the design of this study. However patients are involved within a broader programme of work led by the senior author to improve recording of diagnoses and problems [23].

3. Results

We reviewed the problem list of 516 inpatients with suspected or confirmed COVID-19. The patients included 336 men and 180 women, with median age 65 years (interquartile range 53, 78). The majority (290) were of White ethnicity, 72 were Black, 73 were South Asian, 58 were of mixed or other ethnicity and 23 had no ethnicity recorded. Prior to review, these patients had a combined total of 2841 diagnoses recorded in their EHR problem lists. 1722 additional diagnoses were identified as free text in electronic patient notes and transcribed into the problem list, increasing the mean number of recorded problems per patient from 5.51 to 8.14. The overall percentage of diagnoses originally included in the problem list was 62.3% (2841 / 4563, 95% confidence interval (CI) 60.8%, 63.7%), with variation by disease area. Chronic obstructive pulmonary disease was included on the problem list in 75.4% of patients with the condition (49 / 65, 95% CI 62.9%, 84.9%), and type 2 diabetes in 70.4% of cases (88 / 125, 95% CI 61.5%, 78.1%), but hypertension was recorded in only 53.8% of cases (127 / 236, 95% CI 47.2%, 60.3%), as shown in Fig. 1.

By ICD-10 chapter, diagnoses in chapters XIX and XX (injuries, poisoning and external causes of morbidity) were most likely to be included on the problem list, followed by neoplasms in chapter II, as shown in Table 2.

Note that the COVID-19 infection flag used to identify the cohort of patients for this project is separate to the problem list; thus there is more than one location where a formal COVID-19 diagnosis can be recorded in the EHR. Only 250 of the 516 patients (48.4%, 95% CI 44.1%, 52.9%) had a problem list entry for suspected or confirmed COVID-19 prior to review (some patients had more than one entry, hence there were 360

COVID-19 related problem list entries in total).

4. Discussion

Overall, only 62.3% of diagnoses were recorded on the problem list for patients in this study, with considerable variation by condition. These estimates of problem list completion should be considered in the context in which the study was undertaken, such as the time since the EHR system launch (1 year), the EHR system employed (Epic) and associated EHR training, and the effects of the pandemic. The level of data incompleteness of problem lists in our study is similar to that from previous studies using electronic health records, despite differences in methodologies [15,25–28] (Table 1).

Further study is warranted to understand the causes of variation with problem list completion by ICD-10 chapters. The desire amongst clinicians to avoid cluttering problem list [29], and the uncertainty surrounding which specialities are responsible for updating and maintaining the problem list [30,31], may explain why certain acute presentations are less likely to be recorded on the problem list. Inter-rater agreement between clinicians as to what should, and what should not be added to the problem list is especially poor for secondary diagnoses and complications that have arisen from primary problem list terms [32–34]. In combination with other studies [25,27], we found that chronic conditions such as type 2 diabetes mellitus and asthma were more likely to be recorded on the problem list.

Injury, poisoning and external causes of morbidity tend to directly cause an inpatient presentation and admission to hospital. Related problem list terms therefore may be more likely to be identified as the primary diagnosis and placed on the problem list.

4.1. Bridging the gap

Quality improvement initiatives aiming to improve the completeness of problem lists should consider a) the organisational and non-organisational factors at the individual healthcare trust b) inter-clinician agreement as to the role and scope of the problem list in

Table 2
Percentage of patients with diagnosis already recorded in problem list, by ICD-10 chapters. ICD-10 chapters with fewer than 10 entries were omitted.

ICD-10 chapter	Total number of entries (in problem list and free text)	Number of entries already in problem list	Percentage of diagnoses included in problem list (95% confidence interval)
I Certain infectious and parasitic diseases	153	83	54.2 (46.0, 62.3)
II Neoplasms	198	149	75.3 (68.5, 81.0)
III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	74	47	63.5 (51.5, 74.2)
IV Endocrine, nutritional and metabolic diseases	386	231	59.8 (54.7, 64.7)
V Mental and behavioural disorders	189	123	65.1 (57.8, 71.8)
VI Diseases of the nervous system	129	84	65.1 (56.2, 73.2)
VII Diseases of the eye and adnexa	37	20	54.1 (37.1, 70.2)
VIII Diseases of the ear and mastoid process	13	8	61.5 (32.3, 84.9)
IX Diseases of the circulatory system	739	421	57.0 (53.3, 60.6)
X Diseases of the respiratory system	362	245	67.7 (62.6, 72.5)
XI Diseases of the digestive system	252	162	64.3 (58.0, 70.1)
XII Diseases of the skin and subcutaneous tissue	67	39	58.2 (45.5, 69.9)
XIII Diseases of the musculoskeletal system and connective tissue	217	142	65.4 (58.7, 71.7)
XIV Diseases of the genitourinary system	239	135	56.5 (49.9, 62.8)
XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	329	222	67.5 (62.1, 72.5)
XIX Injury, poisoning and certain other consequences of external causes	111	90	81.1 (72.3, 87.7)
XX External causes of morbidity and mortality	41	37	90.2 (76.9, 97.3)
XXI Factors influencing health status and contact with health services	379	225	59.4 (54.2, 64.3)
* Suspected or confirmed COVID-19	627	360	57.4 (53.4, 61.3)

* Suspected and confirmed COVID-19, though not an ICD-10 chapter, is included in this table to demonstrate the contribution of these problem list terms to overall estimates of problem list completion. Some patients had more than one COVID-19 problem list entry, e.g. a general COVID-19 term and a term for a complication of COVID-19, hence the total number of entries (627) is greater than the number of patients in the study (516).

practice and c) factors behind successful problem list culture at other healthcare trusts

4.1.1. Organisational and non-organisational factors

Educational approaches are necessary for staff need to become aware of the aspects of the EHR that lie outside their day to day use of the system, as well as the value of structured data in health informatics [10]. At UCLH trust, there was no culture of structured data use prior to implementation of an EHR. We found that free text may became the established way of working with a new EHR, and this was difficult to

shift later. We suggest that a parallel effort to support adoption of structured data approaches such as problem listing should exist alongside technical EHR training, and ideally commence before EHR system implementation.

One year after implementation, many of the secondary data returns from the Trust still rely on ICD-10 codes entered retrospectively by coding staff rather than diagnoses recorded in problem lists by clinicians at the point of care, as use of the problem list is not yet systemic and consistent. The EHR department has sought to close this gap from multiple angles, including the publication of problem list leader boards by department and the dissemination of tip sheets to standardise EHR practice. A competition within the Acute Medicine department in which the problem list statistics for each individual clinician was published resulted in a threefold increase in recording of problem list items over this period.

Improvements to the user interface may also help to encourage clinicians to record information in a structured way. UCLH is commencing a project funded by the National Institute of Health Research to develop natural language processing technology to convert diagnoses entered in text into coded terms in real time, enabling clinicians to validate the entries before they are committed to the record.

4.1.2. The role and scope of the problem list

Poor problem list practice can increase fragmentation of problems [8] and propagate inaccuracies [35–37] at the expense of disrupting the patient narrative [38]. Variability in problem list practice can diminish trust in the problem list as an objective source of clinical truth [34].

The key requirement of a problem list is that it should be useful for clinical care [30]. For example, the problem list term ‘childhood asthma’ is of relevance for a young healthy patient with no other past medical history, but of less importance for an elderly patient with several co-morbidities. Clinicians often find it useful to add free text comments to provide additional detail or a description of the problem, in order to supplement the coded term [39]. Pigeonholing problems as either ‘active’ or ‘resolved’ also does not reflect the trajectory of health care problems [40].

These practices suggest a more successful problem list interface, which supports varied definitions of the role of the problem list and allows for individual clinician preferences, would be far less restrictive and more akin to Weed’s original conception of the problem list as an index. Recommendations from the literature include: the creation of a past medical history as a separate section within the problem list [39], clear guidelines as to when to undertake problem list review [30] and better means of avoiding data duplication in problem lists [41].

Structured data approaches such as problem listing, are only more efficient if there is a conscious and concerted effort to use the EHR system to its full capacity beyond electronic storage of notes. Structured data fields need to be used appropriately, as they are not necessarily the best option for all health data – the patient’s story and qualitative observations can be more faithfully recorded in free text, and excessive requirements for structured data may the system onerous to use and lead to poor quality or incomplete data [42].

Previous studies have identified a number of factors associated with improved problem list charting, including financial incentives, gap reporting, shared responsibility, usability, training, supportive policies and organisational culture [26,43].

4.2. Moving forward

The trust now has a foothold using the Epic EHR effectively, but several key questions remain moving forward: Who has the responsibility for recording and maintaining information that persists between healthcare encounters? Who should review problem list data for accuracy? How can we identify and plug gaps in the problem list?

Clinical coders and clinicians should seek to work closer in tandem during an inpatient hospital admission to create more accurate real-time

Summary table

What is already known on this topic

- Diagnoses and other clinical information stored in a structured way in electronic health records is extremely useful for supporting clinical decisions, improving patient care and enabling better research.
- There is evidence in the literature that the recording of diagnoses on the structured problem list for patients is often incomplete, with many diagnoses mentioned only in free text notes.
- Previous studies had assessed problem list completeness only for specific conditions in comparison with other structured data, and there is an absence of such studies in UK hospitals.

What this study adds

- This study is the first to undertake a manual electronic notes review of a cohort of acute hospital inpatients in the UK to ascertain problem list sensitivity for all medical conditions relevant to ongoing care.

clinical data instead of retrospectively plugging ‘data gaps’ after the point of discharge. EHR systems have not rendered clinical coding obsolete. Rather, clinical coders could be utilised to flag missing clinically relevant data of current inpatients, such as an incomplete problem list, for clinical review. Improved and automated interoperability with the GP record may reduce such a workload if certain data fields can be imported directly into the hospital EHR. Problem lists should be reviewed at key stages in the patient journey, such as on admission and discharge. When faced with the problem list of a discharged patient, a clinician-coder team should ask themselves both a) is this problem list accurate and complete? and b) what information is relevant now the patient has been discharged? The former is a more clear-cut question than the latter, which requires some level of standardisation to an inherently subjective issue.

4.3. Future study

This study focuses on one data field only (problem lists), and further study is needed to assess engagement and completeness for other structured data practices in EHR systems. Other areas of the EHR, such as social history, medication lists and family history can similarly suffer from incompleteness or inaccuracy. Though we have been able to accurately characterise problem list usage, and identified variation in problem list completion by ICD-10 chapter, we are only able to speculate as to the causes of incompleteness and variation. Thematic analysis, clinical surveys, and observations of problem list practice may help form a picture as to why these shortcomings occur. Further study is also warranted to assess if recommendations made by this study, such as the use of clinician-coder teams to regularly review problem lists at discharge, can be successful in practice.

4.4. Limitations

There is uncertainty amongst clinicians over exactly which set of conditions should ideally be included on a problem list, so the size of the discrepancy between free-text electronic notes and the problem list assessed in this study may partly depend on the judgement of the clinicians involved. Our method for estimating the completion rates for problem list terms also rests on the assumption that all the patients in the cohort were clerked comprehensively, without omissions of relevant past medical history from their electronic health record. Organisational factors at UCLH and the NHS, as well as the EHR system used (Epic), will impact and limit the generalisability of the results.

5. Conclusion

Diagnoses and other clinical information stored in a structured way

in electronic health records is extremely useful for supporting clinical decisions, improving patient care and enabling research. However, one year after implementation of a comprehensive electronic health record in a major teaching hospital, recording of medical history on the structured problem list for inpatients is incomplete, with almost 40% of important diagnoses mentioned only in the free text notes.

Contributions

LZ and ADS conceived the idea of the study. LZ and JP conducted the data collection, with clinical supervision from ADS. Analysis was performed by JP, ADS and LZ. All authors contributed to the interpretation and appraisal of the results, with JP writing the manuscript. ADS is the guarantor.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. ADS is supported by a postdoctoral fellowship (PD-2018-01-004) funded by the Health Foundation’s grant to the University of Cambridge for The Healthcare Improvement Studies Institute.

Ethical approval

This is a quality improvement project carried out to improve data quality for operational reporting. The project was approved by the UCLH Data Access Committee for COVID-19 studies, who classified it as an audit, and therefore it did not require review by an ethics committee.

Data sharing

This work was carried out as an internal audit with the University College London NHS Trust. As the individual data used in this study has the potential to be patient identifiable, they are not available for sharing.

Transparency

The manuscript’s guarantor (ADS) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Acknowledgements

We would like to acknowledge Alfio Missaglia for undertaking the manual review of patient notes and problem lists for this study, Qifang

Deng for mapping the problem list terms to ICD-10 nomenclature and Gary Philippo for Epic analytical support. This work uses data provided by patients and collected by the NHS as part of their care and support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.ijmedinf.2021.104452>.

References

- [1] A.D. Shah, N.J. Quinn, A. Chaudhry, et al., Recording problems and diagnoses in clinical care: developing guidance for healthcare professionals and system designers, *BMJ Health Care Inform.* 26 (1) (2019) e100106, <https://doi.org/10.1136/bmjhci-2019-100106>.
- [2] J. Meredith, I. McNicoll, N. Whitehead, et al., Defining the contextual problem list, *Stud. Health Technol. Inform.* 270 (June) (2020) 567–571.
- [3] L.L. Weed, Medical records that guide and teach, *N. Engl. J. Med.* 278 (11) (1968) 593–600, <https://doi.org/10.1056/NEJM196803142781105>.
- [4] R.A. Greenes, *Clinical Decision Support: the Road to Broad Adoption*, Elsevier, London, 2014.
- [5] SNOMED International. <https://www.snomed.org/> (Accessed 26 Nov 2020).
- [6] L. Chu, V. Kannan, M.A. Basit, et al., SNOMED CT concept hierarchies for computable clinical phenotypes from electronic health record data: comparison of Intensional versus extensional value sets, *JMIR Med. Inform.* 7 (January 1) (2019), e11487.
- [7] S.J. Wang, D.W. Bates, H.C. Chueh, et al., Automated coded ambulatory problem lists: evaluation of a vocabulary and a data entry tool, *Int. J. Med. Inform.* 72 (December 1–3) (2003) 17–28.
- [8] S.M. Chowdhry, R.G. Mishuris, D. Mann, Problem-oriented charting: a review, *Int. J. Med. Inform.* 103 (July) (2017) 95–102.
- [9] R.C. Li, T. Garg, T. Cun, et al., Impact of problem-based charting on the utilization and accuracy of the electronic problem list, *J. Am. Med. Inform. Assoc.* 25 (May 5) (2018) 548–554.
- [10] P. Rajbhandari, M. Auron, S. Worley, et al., Improving documentation of inpatient problem list in electronic health record: a quality improvement project, *J. Patient Saf.* (2018), <https://doi.org/10.1097/PTS.0000000000000490>. Apr 19.
- [11] W.L. Galanter, D.B. Hier, C. Jao, et al., Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance, *Int. J. Med. Inform.* 79 (May 5) (2010) 332–338.
- [12] S. Natarajan, S.R. Lipsitz, P.J. Nietert, Self-report of high cholesterol: determinants of validity in US adults, *Am. J. Prev. Med.* 23 (July 1) (2002) 13–21.
- [13] N. Goldman, I.F. Lin, M. Weinstein, et al., Evaluating the quality of self-reports of hypertension and diabetes, *J. Clin. Epidemiol.* 56 (February 2) (2003) 148–154.
- [14] A. Goto, A. Morita, M. Goto, et al., Validity of diabetes self-reports in the Saku diabetes study, *J. Epidemiol.* (July) (2013), JE20120221.
- [15] C.G. Fernanda Polubriaginof, G. Paulo Pastore, B. Drohan, et al., Comparing patient-reported medical problems with the electronic health record problem list, *Gen. Med. Los Angel. (Los Angel)* 4 (258) (2016) 2.
- [16] M.S. Fossier, M.A. Gaiera, M.C. Otero, et al., Automatic loading of problems using a COMORBIDITIES subset, one step to organize and maintain the patient's problem list, *Ann. Fam. Med.* 4 (2006), 101–3.10.
- [17] S. Meystre, P. Haug, Improving the sensitivity of the problem list in an intensive care unit by using natural language processing, in: *American Medical Informatics Association AMIA Annual Symposium Proceedings*, Vol. 2006, 2006, p. 554.
- [18] S. Meystre, P.J. Haug, Automation of a problem list using natural language processing, *BMC Med. Inform. Decis. Mak.* 5 (December 1) (2005) 1–4.
- [19] F. Plazzotta, C. Otero, D. Luna, et al., Natural language processing and inference rules as strategies for updating problem list in an electronic health record, *Stud. Health Technol. Inform.* 192 (January) (2013) 1163.
- [20] M.V. Devarakonda, N. Mehta, C.H. Tsou, et al., Automated problem list generation and physicians perspective from a pilot study, *Int. J. Med. Inform.* 105 (September) (2017) 121–129.
- [21] KLAS Research, US Hospital EMR Market Share, 2020 (Accessed 28th Feb 2021), www.klasresearch.com/report/us-hospital-emr-market-share-2020/1616.
- [22] R.K. Gupta, M. Marks, T.H. Samuels, et al., Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: an observational cohort study, *Eur. Respir. J.* 56 (December 6) (2020).
- [23] Professional Records Standard Body, *Diagnosis Recording Final Report*, 2019. <https://theprsb.org/standards/diagnoserecording/>.
- [24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, 2018 <https://www.r-project.org/> (Accessed 28th Feb 2021).
- [25] E.C. Wang, A. Wright, Characterizing outpatient problem list completeness and duplications in the electronic health record, *J. Am. Med. Inform. Assoc.* 27 (8) (2020) 1190–1197.
- [26] A. Wright, A.B. McCoy, T.T. Hickman, et al., Problem list completeness in electronic health records: a multi-site study and assessment of success factors, *Int. J. Med. Inform.* 84 (10) (2015) 784–790.
- [27] A. Wright, J. Pang, Jc Feblowitz, et al., A method and knowledge base for automated inference of patient problems from structured data in an electronic medical record, *J. Am. Med. Inform. Assoc.* 18 (6) (2011) 859–867.
- [28] B.B. Hoffman, M.K. Goldstein, Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic, *Am. J. Manag. Care* 8 (January) (2002) 37–43.
- [29] C.M. Hodge, S.P. Narus, Electronic problem lists: a thematic analysis of a systematic literature review to identify aspects critical to success, *J. Am. Med. Inform. Assoc.* 25 (5) (2018) 603–613.
- [30] C. Holmes, M. Brown, D. St Hilaire, et al., Healthcare provider attitudes towards the problem list in an electronic health record: a mixed-methods qualitative study, *BMC Med. Inform. Decis. Mak.* 12 (December 1) (2012) 1–7.
- [31] A. Wright, J. Feblowitz, F.L. Maloney, et al., Use of an electronic problem list by primary care providers and specialists, *J. Gen. Intern. Med.* 27 (August 8) (2012) 968–973.
- [32] A.S. Rothschild, H.P. Lehmann, G. Hripsak, Inter-rater agreement in physician-coded problem lists, in: *American Medical Informatics Association AMIA Annual Symposium Proceedings*, Vol. 2005, 2005, p. 644.
- [33] A. Wright, F.L. Maloney, J.C. Feblowitz, Clinician attitudes toward and use of electronic problem lists: a thematic analysis, *BMC Med. Inform. Decis. Mak.* 11 (December 1) (2011), 1–0.
- [34] J.C. Krauss, P.S. Boonstra, A.V. Vantsevich, et al., Is the problem list in the eye of the beholder? An exploration of consistency across physicians, *J. Am. Med. Inform. Assoc.* 23 (September 5) (2016) 859–865.
- [35] E.L. Siegler, R. Adelman, Copy and paste: a remediable hazard of electronic health records, *Am. J. Med.* 122 (6) (2009) 495–496.
- [36] K.W. Hammond, S.T. Helbig, C.C. Benson, et al., Are electronic medical records trustworthy? Observations on copying, pasting and duplication, in: *American Medical Informatics Association AMIA Annual Symposium Proceedings*, Vol. 2003, 2003, p. 269.
- [37] I. Kushner, P.J. Greco, P.K. Saha, S. Gaitonde, The trivialization of diagnosis, *J. Hosp. Med.* 5 (February 2) (2010) 116–119.
- [38] D.A. Moros, The electronic medical record and the loss of narrative, *Camb. Q. Healthc. Ethics* 26 (2017) 328.
- [39] T.T. Van Vleck, A. Wilcox, P.D. Stetson, et al., Content and structure of clinical problem lists: a corpus analysis, in: *American Medical Informatics Association AMIA Annual Symposium Proceedings*, Vol. 2008, 2008, p. 753.
- [40] P. Salmon, A. Rappaport, M. Bainbridge, G. Hayes, J. Williams, Taking the problem oriented medical record forward, in: *Proceedings of the AMIA Annual Fall Symposium*, American Medical Informatics Association, 1996, p. 463.
- [41] J. Mantas, Caveats for the use of the active problem list as ground truth for decision support, *Decis. Supp. Syst. Educ.* 255 (October) (2018) 10.
- [42] D. Kalra, B. Fernando, Approaches to enhancing the validity of coded data in electronic medical records, *Prim. Care Resp. J.* 20 (March 1) (2011) 4.
- [43] S.M. Simons, F.H. Gillessen, J.A. Hazelzet, Determinants of a successful problem list to support the implementation of the problem-oriented medical record according to recent literature, *BMC Med. Inform. Decis. Mak.* 16 (December 1) (2016) 1–9.