HEALTH POLICY

Sensitivity and Specificity of Real-World Social Factor Screening Approaches

Joshua R. Vest[1,4] · Wei Wu[2] · Eneida A. Mendonca[3,4]

Abstract

Health care organizations are increasingly documenting patients for social risk factors in structured data. Two main approaches to documentation, ICD-10 Z codes and screening questions, face limited adoption and conceptual challenges. This study compared estimates of social risk factors obtained via screening questions and ICD-10 Z diagnoses coding, as used in clinical practice, to estiamtes from validated survey instruments in a sample of adult primary care and emergency department patients at an urban safety-net health system. Financial strain, transportation barriers, food insecurity, and housing instability were independently assessed using instruments with published reliability and validity. These four social factors were also being collected by the health system in screening questions or could be mapped to ICD-10 Z code diagnosis code concepts. Neither the screening questions nor ICD-10 Z codes performed particularly well in terms of accuracy. For the screening questions, the Area Under the Curve (AUC) scores were 0.609 for financial strain, 0.703 for transportation, 0.698 for food insecurity, and 0.714 for housing instability. For the ICD-10 Z codes, AUC scores tended to be lower in the range of 0.523 to 0.535. For both screening questions and ICD-10 Z codes, the measures were much more specific than sensitive. Under real world conditions, ICD-10 Z codes and screening questions are at the minimal, or below, threshold for being diagnostically useful approaches to identifying patients' social risk factors. Data collection support through information technology or novel approaches combining data sources may be necessary to improve the usefulness of these data.

Keywords Social determinants · Survey · Validity · Safety-net

Introduction

Social risk factors include the array of nonclinical, contextual, and socioeconomic characteristics that negatively affect health and increase utilization and costs [1, 2]. Patients' social risk factor information can drive referrals to community partners [3] or be applied to population health management activities, like risk stratification [4]. In response to the potential value of these data and increased interest from payers and policymakers, the collection of patients' social risk factor information has become more common [5, 6].

While social risk factors may be documented as part of narrative notes within electronic health records (EHR), data are more useful when stored in a structured form (i.e., coded and not narrative text) that facilitates retrieval and analyses [7]. For organizations' wishing to collect patients' social factors as structured data, one option is to use one of the various multi-domain social factors screening questionnaires [8]. Provider organizations, payers, researchers, and technology vendors have designed and implemented a variety of short screening questions to reflect multiple social risk factors such as housing, financing, hunger, transportation, and other issues [8–11]. In addition, ICD-10 introduced "Z codes" to reflect patient factors other than disease, injury, or external causes as problems or diagnoses [12]. Using ICD-10 Z

✉ Joshua R. Vest
  joshvest@iu.edu

1 Department of Health Policy & Management, Indiana University Richard M. Fairbanks School of Public Health – Indianapolis, Indianapolis, IN, USA

2 Department of Psychology, Indiana University Purdue University Indianapolis, Indianapolis, IN, USA

3 Department of Pediatrics and Department of Biostatistics and Health Data Sciences, Indiana University School of Medicine, Indianapolis, IN, USA

4 Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, USA

codes, providers may document a variety of social factors, including housing, income, unemployment, and social relationships during the care of patients – just as they would for any clinical condition [13]. Innovative health care organizations have successfully integrated screening questions, as well as ICD-10 Z coding, into EHR systems and clinical workflows in order to facilitate structured data collection at the point of care [9, 14–16].

Nevertheless, these aforementioned social risk factor measurement approaches face limitations. For one, evidence suggests both ICD-10 Z codes and screening questions, even when integrated in the EHR, face limited adoption. The little available evidence on ICD-10 Z codes and earlier incarnations of non-medical diagnosis codes indicates substantial underusage and potentially biased usage [17–19]. Research and demonstration projects have successfully implemented screening questions in practice [3, 9, 15], but overall uptake remains low and inconsistent [20]. Moreover, usage of social risk factor screening questions faces general workflow and time constraint barriers [21, 22]. More importantly, however, the validity and reliability of both of these approaches are unknown. Systematic reviews conclude that the psychometric properties of many social factor screening questions have not been established [23, 24]. While no studies have specifically investigated the psychometric properties of Z codes, overall, the reliability and validity of ICD-10 diagnoses have been documented as low [25, 26], and concerns exist about the conceptual linkage between codes and social factors [27].

This study sought to compare estimates of social risk factors derived from screen questions and ICD-10 Z diagnoses coding, as used in clinical practice, against estimates derived using validated instruments. Estimates of social risk factor screening performance address a gap in the literature, and our use of validated instruments provides a strong reference standard. Moreover, establishing performance identifies opportunities for improved social factors measurement approaches.

Methods

We compared the performance of commonly used social risk factor screening questions and ICD-10 Z codes in identifying patients with financial strain, transportation, food insecurity, and housing instability against estimates obtained from validated survey instruments.

Setting & sample

The study sample included adult primary care and emergency department (ED) patients who sought care at an Indianapolis, IN safety-net provider during August and September of 2020. Patients were eligible for inclusion if they: were 18 years old or older; did not require an interpreter; were able to complete a self-administered survey unassisted; and were not marked as positive for COVID-19 symptoms in the EHR scheduling system. Data were collected via REDCap [28], and all patients received an incentive for participation.

Measures: Validated measures of social risk factors

Figure 1 describes the construction of the analytic sample. Patients completed validated (e.g., published psychometric

Fig. 1 Combination of validated survey sample and existing electronic health record sources
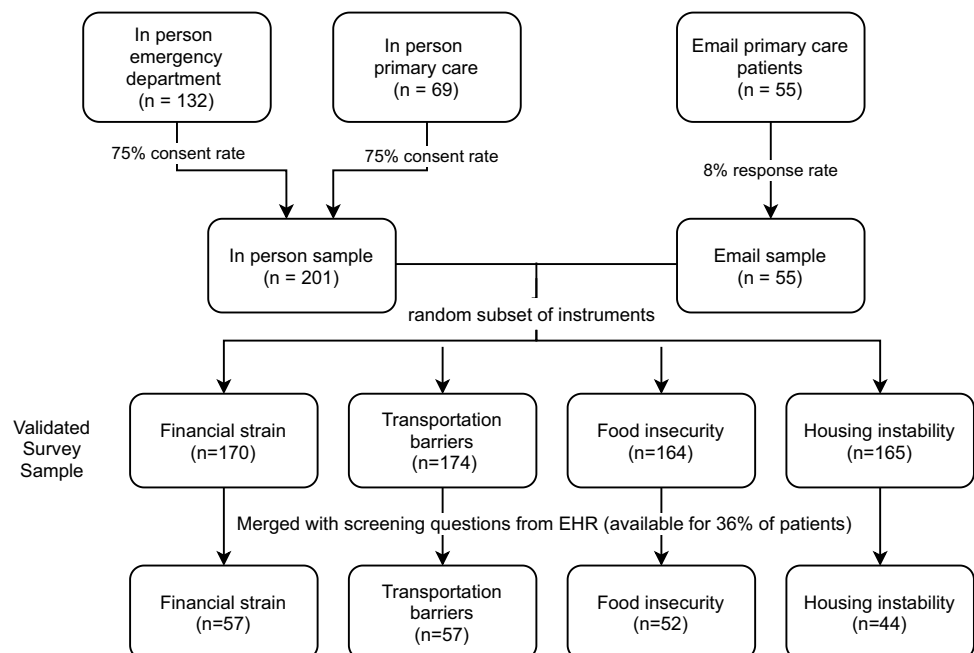
properties) social risk factor survey instruments covering: financial strain [29], transportation [30], food insecurity [31], housing instability [32], and other social risk factors. Due to the length of these individual instruments, patients were not administered all the validated survey instruments concurrently. Instead, we used a random number generator in REDCap to deliver a subset of instruments to the participants (e.g., 3 of the 4 domains) to minimize the response burden. The validated survey instruments were self-administered using tablets during health care encounters or via email. For those responding in-person using tablets, research staff approached patients during ED or primary care visits while the patient waited for the provider. Patients receiving in-person care at primary care sites that were physically too small to accommodate data collection staff (due to physical distancing requirements) were emailed a link to participate via email. No significant differences in social risk factors prevalence existed by mode or location of data collection. This approach resulted in 170 responses to the financial strain survey, 174 to the transportation survey, 164 to the food insecurity survey, and 165 to the housing instability survey among 256 unique patients.

All four validated survey instruments demonstrated levels of internal consistency (Cronbach's α) generally considered acceptable [33]. We created binary indicators for the presence of each social risk factor following recommendations in the literature. Financial strain was assessed using the comprehensive score for financial toxicity (COST) instrument ($\alpha = 0.75$) and defined by the median total scores [29]. The Transportation Barriers Measure assessed transportation barriers ($\alpha = 0.79$) and was also divided as the median total score [30]. Food insecurity was defined as a total food insecurity raw score of 3 or greater (low or very low food insecurity) on the U.S. Household Food Security Survey Module: Six-Item Short Form ($\alpha = 0.87$) [31, 34]. The Housing Instability Index was divided at the median to measure housing instability ($\alpha = 0.71$) [32].

Measures: Social risk factor screening questions

We merged responses from the validated survey instruments to the health system's EHR using patient identifiers to obtain responses to existing social risk factor screening questions and ICD-10 Z codes. The health system used the social risk factor screening questions included in the Epic EHR system [35]. For these screening questionnaires, the standard practice at the health system is for medical assistants to collect structured responses to these screening questions. The structured responses are retained and providers have the option of automatically incorporating the responses into clinical notes as text (called "smart phrases" within the EHR). However, these questions were not asked of all patients. We created binary indicators for positive screens from the screening

questions for financial strain, transportation, food insecurity, and housing instability following published guidelines [36].

Among the 256 patients in the sample (Fig. 1), 36% had been administered screening questions from at least one of the four social risks of interest. The combination of validated surveys and EHR survey questions resulted in: 57 observations for financial strain (mean days between validated survey and EHR screening question collection = 14.4; SD = 26.5); 57 observations for transportation barriers (mean days between = 11.1; SD = 24.9); 52 observations for food insecurity (mean days between = 7.8; SD = 16.3); and 44 observations for housing (mean days = 32.7; SD = 33.8).

Measures: ICD-10 Z codes for social factors

ICD-10 Z codes were available to providers to document any patients' primary or secondary diagnoses and were also assessed as implemented in practice at the health system. We adapted published mapping of ICD-10 Z codes to social factors [27] (see Appendix A). For the entire patient sample (n = 256), we considered the presence of the ICD-10 Z code within one year of the date of the validated survey administration as a positive screen and the absence as a negative screen for each social factor. We also created a single indicator reflecting either the presence of an ICD-10 Z code or a positive response to the screening questionnaires. One of the ICD-10 Z codes of interest was present in 15.2% of the study sample (mean days between validated survey and ICD-10 Z code diagnosis date = 103.6; SD = 113.3).

Analyses

We described the sample characteristics using frequencies and percents. For each social factor, we describe the percent positive according to each screening method. We compared the performance of screening questions and ICD-10 Z codes against the validated measures using sensitivity, specificity, positive predictive values (PPV), and the area under the curve (AUC) values. As a sensitivity check, we limited the sample to those with screening questions collected within 30 days of the validated survey. Additionally, we repeated the analyses using either a positive response to screening questions or the presence of ICD-10 Z code as an indicator of social risk for all respondents of the validated surveys. Differences in AUC were compared by the Delong test [37]. The study was approved by the IRB of Indiana University.

Results

The study sample was predominately female, under the age of 45, and covered by Medicaid (Table 1). The sample was also highly diverse, with less than one-third of patients

Table 1 Characteristics of the patient sample

	n	Percent
Gender		
Male	99	38.7
Female	157	61.3
Age		
18–34	97	37.9
35–44	40	15.6
45–64	101	39.5
>65	18	7.0
Race and ethnicity		
White non-Hispanic	77	30.1
African American non-Hispanic	145	56.6
Hispanic	23	9.0
Other / unknown	11	4.3
Insurance status		
Medicaid	141	55.1
Medicare	34	13.3
Private	46	18.0
Uninsured	35	13.7
Elixhauser score (mean, sd)	–	2.5 (2.4)
Questionnaire screening present	94	36.7
Z-code (any) present	39	15.2

identified as White and non-Hispanic. The percentage of patients with identified social factors was consistently highest per the validated questionnaires (Fig. 2). Only for food insecurity was the prevalence from screening questions similar to the validated surveys. Percentages were lowest using ICD-10 Z codes.

Neither the screening questions nor ICD-10 Z codes performed particularly well in terms of accuracy (Table 2). For

the screening questions, AUC scores ranged from 0.61 to 0.71, and were significantly higher than the ICD-10 Z codes for transportation, food insecurity, and housing instability. For the ICD-10 Z codes, AUC scores tended to be lower in the range of 0.52 to 0.53. In general, both measures were much more specific than sensitive. For ICD-10 Z codes, the specificity was greater than 90% for each social factor. For screening questions, specificity was greater than 90% for financial security, transportation barriers, and housing instability. Our sensitivity analysis, which limited the sample to screening questions collected within 30 days of the validated measures, did not change overall performance, except for food insecurity, where the AUC decreased significantly (see Appendix B).

When the available screening questions were added to the information from ICD-10 Z codes, only the AUC for housing instability increased significantly (Fig. 3). However, even with the additional information, all the AUC values remained below 0.60. While not statistically significant, the sensitivity for financial insecurity, transportation barriers, food insecurity, and housing instability all were higher when both data sources were used (see Appendix C).

Discussion

Current approaches to social factor measurement are very specific. When either an ICD-10 Z code is present or a screening question is positive, the social factor is very likely present. However, in real world conditions, these two screening approaches did not perform particularly well overall in this sample of safety-net patients. As identification of patients' social factors is critical to organizational operations, population health, and current health policy, current

Fig. 2 Percent of patients with social factors by screening approach. Note: EHR-based screening estimates were obtained from patients administered the screening questionnaire
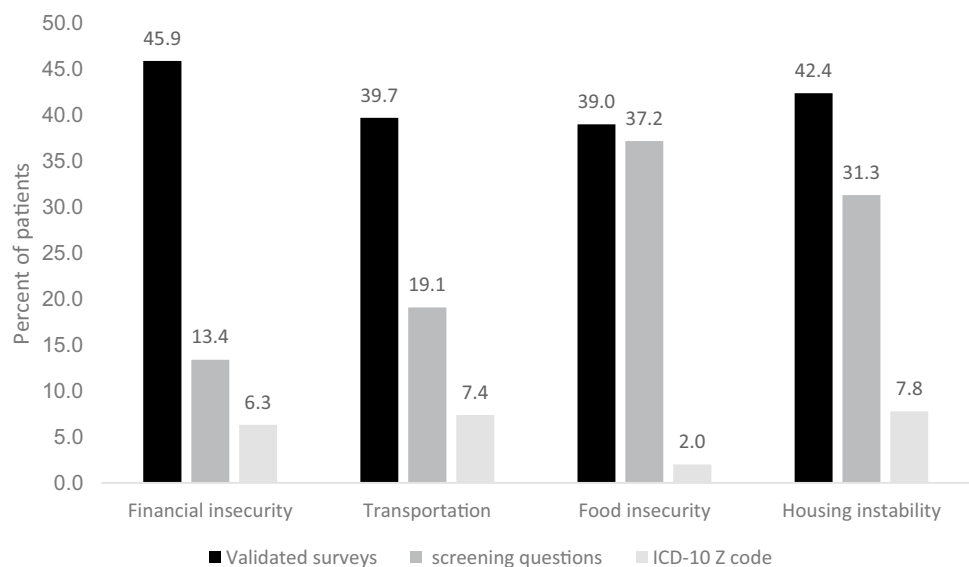
Table 2 Performance of screening questions and ICD-10 Z codes for social risk factor screening measurement compared to validated survey instruments

Social factor	n	Sensitivity	Specificity	Positive predictive value (PPV)	Area under the curve (AUC)
Financial strain					
screening questions[a]	57	28.0 (12.1, 49.4)	93.8 (79.2, 99.2)	77.8 (40.0, 47.4)	0.609 (0.509, 0.708)
ICD-10 Z codes[b]	170	10.3 (4.5, 19.2)	96.7 (90.8, 99.3)	72.7 (39.0, 94.0)	0.535 (0.496, 0.573)
Transportation					
screening questions	57	50.0 (23.0, 77.0)	90.7 (77.9, 97.4)	63.6 (30.8, 89.1)	0.703 (0.561, 0.846)*
ICD-10 Z codes	174	11.6 (5.1, 21.6)	94.3 (88.0, 97.9)	57.1 (28.9, 82.3)	0.529 (0.485, 0.574)
Food insecurity					
screening questions	52	66.7 (38.4, 88.2)	73.0 (55.9, 86.2)	50.0 (27.2, 72.8)	0.698 (0.555, 0.841)*
ICD-10 Z codes	164	4.7 (1.0, 13.1)	100.0 (96.4, 100.0)	100.0 (29.2, 100.0)	0.523 (0.487, 0.550)
Housing instability					
screening questions	44	50.0 (24.7, 75.3)	92.9 (76.5, 99.1)	80.0 (44.4, 97.5)	0.714 (0.579, 0.850)*
ICD-10 Z codes	165	10.0 (4.1, 19.5)	96.8 (91.0, 99.3)	70.0 (34.8, 93.3)	0.534 (0.495, 0.574)

*$p < 0.05$ comparison of AUC values

[a]Screening questions extracted from the EHR

[b]ICD-10 Z codes see Appendix

measurement strategies require substantial improvement or new innovative approaches.

The introduction of ICD-10 Z codes for social risks factors has been an exciting development for those interested in population health and an important step towards potential reimbursement for health care organizations [38]. Nevertheless, our findings add to the growing evidence that ICD-10 Z codes are significantly underutilized as a method of documenting diagnoses [13, 17–19]. In this sample, nearly 15% of patients had at least one ICD-10 Z code, whereas other studies have reported the prevalence of ICD-10 Z codes at around 2% [17, 39]. Even in this instance of more common usage, ICD-10 Z codes still underestimated the prevalence of social risk factors

and are limited in their potential to effectively infer the presence of a social factor for a patient. Tying ICD-10 Z codes to reimbursement would undoubtedly increase their usage. Nevertheless, changing documentation practices is always challenging. Technological solutions may improve data collection and presentation to providers. This site's use of automatic creation of text in the provider's notes through "smart phrases" is one such example that makes this information more accessible to providers. Also, some health systems have demonstrated success in linking ICD-10 Z codes to screening activities to improve data capture [14, 16]. Such automation could be combined with efforts to have patients self-report social factors with EHR portals [40] or tablets [41].

Fig. 3 Differences in area under the curve values for ICD-10 Z codes alone or in combination with available electronic health record screening questions
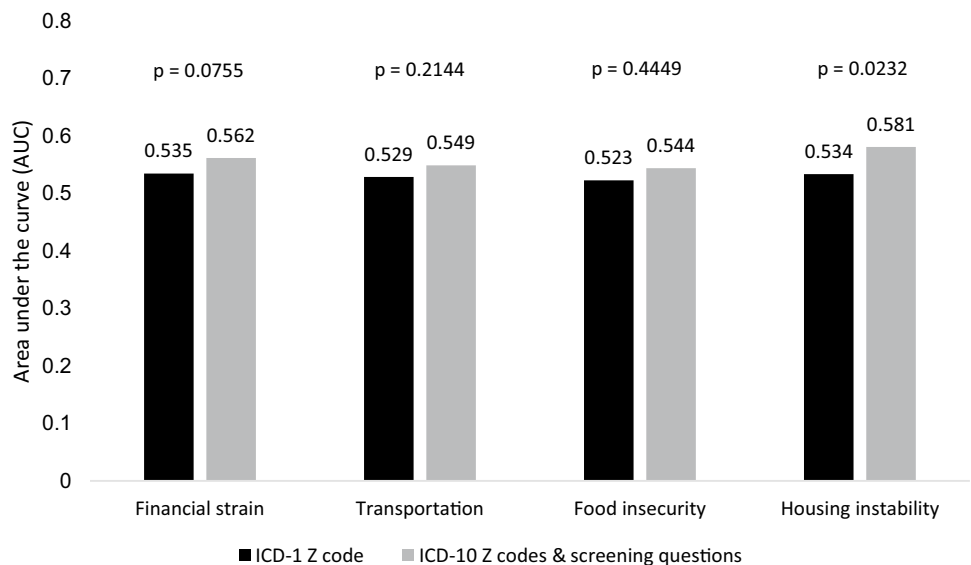
Additionally, no standard mapping of ICD-10 Z codes to social factors exists [27]. Individual codes could be mapped to multiple different social risks (e.g., Z59.8—Other problems related to housing and economic circumstances other problems mapped to housing and financial strain) [12] or could possibly reflect patient diagnoses not related to the social factor of interest (e.g., Z91.80 – other specified personal risk factors mapped to transportation [27], but may reflect other risks). As a result, those relying solely on ICD-10 Z codes may not be measuring the intended social construct. Unstructured data associated with the diagnosis code could clarify the factor being documented; although, that increases the difficulty of data extraction and analysis, and also undermines the advantages of structured data elements.

In general, social risk factor screening questions were also very specific, but had higher sensitivity than ICD-10 Z codes. The screening questions performed better than ICD-10 Z codes, but with qualifications. The transportation barriers and housing instability just reached commonly accepted thresholds for acceptable performance, but financial strain and food insecurity did not reach the level of being diagnostically useful [42, 43]. This better, but still not overwhelming, performance could be due to both characteristics of the questions and workflow issues. First, screening questionnaires, as in this study, typically draw single or pairs of items from existing tools [8]. This practice results in strongly worded and consistent screening questions, whereas interpretation of ICD-10 Z codes could result from very different and inconsistently worded questioning by providers [44]. Hence, a possible reason for better performance. Nevertheless, adopted or adapted items do not inherit the psychometric properties of the original validated instruments [45], which may explain the middling to poor performance. Any resulting questionnaire that adapts or adopts a subset of items would require independent psychometric evaluation to assess reliability and validity. Although, in the case of food insecurity, the screening questions included as part of the Epic EHR come from a well-tested short-screening tool [46], which likely explains the stronger performance. Second, the administration of screening questions is at risk for selective screening practices and incomplete implementation [6]. While the evidence suggests that most providers and health systems are favorably disposed to screening for social risk factors, implementation is inconsistent and faces barriers of time and workflow integration [20, 22, 47]. Related to concerns about data collection time burdens, the social risk factors screening questions were much shorter than the validated instruments, e.g., 3 questions versus 10 for housing stability and 2 compared to up to 16 for transportation in the validated instrument). Nevertheless, even with these less time consuming approaches, screening via questionnaires was less than universal in practice.

The context of these findings is important; the health care organization had integrated data collection into the EHR, adjusted workflows, and made social risk factors a priority. The limited performance of screening questionnaires and ICD-10 Z codes was observed even with such efforts to address data collection challenges. As such, more innovative and alternative methods of social factor measurement may be needed to achieve higher screening rates and more accurate measurements. Recently, experts have suggested alternative means of representing patient-level social factors such as computable phenotypes [48]. In particular, computable phenotypes are composites of characteristics defined through single data elements or a collection of data elements, observations or events, to identify relevant patients [49–51]. This study indicated that single measurement approaches, as currently used in practice, are insufficient. However, in support of the phenotype idea, the combination of screening questions and ICD-10 Z codes resulted in small improvements in performance. Work has already indicated that social risk factor identification may be improved using textual and structured data sources [52]. While this study focused on structural data, health care providers have historically recorded their patients' social risk factors within patient records as text [53, 54] with social history as a standard portion of health records [55]. As a result, unstructured text constitutes an important and rich source for information [56], and multiple researchers and institutions are increasingly leveraging natural language processing (NLP) to extract a variety of social risk factors into structured data elements [57–60]. Future work would be necessary to see if leveraging the existing screening questions and ICD-10 Z code data collected by health systems could yield informative social risk factor measures.

Limitations

First, this was an analysis of social screening as conducted in practice. Thus, the potential for selection or non-response bias in the administration of screening questions or the documentation of ICD-10 Z codes cannot be assessed. Studies that control both for all three different processes of social risk factor screening may result in alternative measures of performance. Second, performance may be due in part to inconsistent construct definitions. Social factors as measured in the validated survey instruments, screening questions, and the individual provider's diagnosis coding may reflect slightly different concepts (e.g., housing instability vs. homelessness). Also, differences may be due to temporal issues, as we did not collect all data points simultaneously, and social factors change over time. This study was also limited by the COVID-19 pandemic. Limited patients access meant the sample reflected different care settings and a smaller sample size. In particular, our sample sizes only provided sufficient power to detect AUC differences greater

than 0.10 [61]. Thus, our study is unpowered to detect smaller AUC differences. The prevalence of financial insecurity and transportation barriers were higher among ED patients, but the sample size prohibited stratified analyses. Lastly, the findings are limited in terms of generalizability both for ICD-10 Z codes and screening questions. These data reflect workflows and processes of our clinical partner, which may be different from other locations. Specifically, our health system partner was using ICD-10 Z codes at a much higher rate than national reports [17, 18]. Likewise, no single standard multi-domain social risk factor screening questionnaire exists [23], and while many share common questions, the EHR-based screening questions in this study may not generalize completely to all other tools.

Conclusions

Under real world conditions, ICD-10 Z codes and screening questions did not perform very well in accurately identifying patients' social risk factors. Data collection support through information technology or novel approaches combining data sources may be necessary to improve the usefulness of these data.

Declarations

Ethics approval The study was approved by the IRB of Indiana University.

Consent to participate All subjects provided active, written consent to participate.

Consent for publication Not applicable – no identifying information presented.

Conflicts of interest/competing interests Joshua Vest is a founder and equity holder in Uppstroms, a health technology company. Drs. Wu and Mendonca have nothing to disclose.

References

1. Alderwick H, Gottlieb LM. Meanings and Misunderstandings: A Social Determinants of Health Lexicon for Health Care Systems. *The Milbank Quarterly* 2019;**97**:1–13. https://doi.org/10.1111/1468-0009.12390

2. Pruitt Z, Emechebe N, Quast T, *et al.* Expenditure Reductions Associated with a Social Service Referral Program. *Population Health Management* 2018;**21**:469–76. https://doi.org/10.1089/pop.2017.0199

3. Gottlieb LM, Hessler D, Long D, *et al.* Effects of Social Needs Screening and In-Person Service Navigation on Child Health: A Randomized Clinical Trial. *JAMA pediatrics* 2016;:e162521–e162521. https://doi.org/10.1001/jamapediatrics.2016.2521

4. Hao S, Wang Y, Jin B, *et al.* Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange. *PloS one* 2015;**10**:e0140271–e0140271. https://doi.org/10.1371/journal.pone.0140271

5. Fraze TK, Brewster AL, Lewis VA, *et al.* Prevalence of Screening for Food Insecurity, Housing Instability, Utility Needs, Transportation Needs, and Interpersonal Violence by US Physician Practices and Hospitals. *JAMA Netw Open* 2019;**2**:e1911514–e1911514. https://doi.org/10.1001/jamanetworkopen.2019.11514

6. Lee J, Korba C. Social determinants of health : How are hospitals and health systems investing in and addressing social needs?. 2017. https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/addressing-social-determinants-of-health-hospitals-survey.html

7. Institute of Medicine. *Capturing social and behavioral domains in electronic health records: Phase 2.* Washington, DC: The National Academies Press 2014.

8. LaForge K, Gold R, Cottrell E, *et al.* How 6 Organizations Developed Tools and Processes for Social Determinants of Health Screening in Primary Care: An Overview. *The Journal of ambulatory care management* 2018;**41**:2–14. https://doi.org/10.1097/jac.0000000000000221

9. Weir RC, Proser M, Jester M, *et al.* Collecting Social Determinants of Health Data in the Clinical Setting: Findings from National PRAPARE Implementation. *Journal of Health Care for the Poor and Underserved* 2020;**31**:1018–35. https://doi.org/10.1353/hpu.2020.0075

10. Health Leads. The Health Leads Screening Toolkit. Health Leads. 2018.https://healthleadsusa.org/resources/the-health-leads-screening-toolkit/ (Accessed 5 Feb 2020).

11. Alley DE, Asomugha CN, Conway PH, *et al.* Accountable Health Communities — Addressing Social Needs through Medicare and Medicaid. *New England Journal of Medicine* 2016;**374**:8–11. https://doi.org/10.1056/nejmp1512532

12. World Health Organization. ICD-10 Version:2019. Chapter XXI Factors influencing health status and contact with health services (Z00-Z99). 2021.https://icd.who.int/browse10/2019/en#/XXI (Accessed 21 Jan 2021).

13. American Hospital Association. ICD-10-CM Coding for Social Determinants of Health. 2020.https://www.aha.org/system/files/2018-04/value-initiative-icd-10-code-social-determinants-of-health.pdf (Accessed 23 Sep 2020).

14. Buitron de la Vega P, Losi S, Sprague Martinez L, *et al.* Implementing an EHR-based Screening and Referral System to Address Social Determinants of Health in Primary Care. *Medical Care* 2019;**57**:S133–9. https://doi.org/10.1097/mlr.0000000000001029

15. Gold R, Bunce A, Cowburn S, *et al.* Adoption of social determinants of health EHR tools by community health centers. *Annals of Family Medicine* 2018;**16**:399–407. https://doi.org/10.1370/afm.2275

16. Friedman NL, Banegas MP. Toward Addressing Social Determinants of Health: A Health Care System Strategy. *Perm J* 2018;**22**. https://doi.org/10.7812/TPP/18-095

17. Truong HP, Luke AA, Hammond G, *et al.* Utilization of Social Determinants of Health ICD-10 Z-Codes Among Hospitalized Patients in the United States, 2016-2017. *Med Care* 2020;**58**:1037–43. https://doi.org/10.1097/MLR.0000000000001418

18. Matthew J, Hodge C, Khau M. Z Codes Utilization among Medicare Fee-for-Service (FFS) Beneficiaries in 2017. Baltimore, MD: : CMS Office of Minority Health 2019. https://www.cms.gov/files/document/cms-omh-january2020-zcode-data-highlightpdf.pdf (Accessed 19 Jan 2021).

19. Torres JM, Lawlor J, Colvin JD, *et al.* ICD Social Codes: An Underutilized Resource for Tracking Social Needs. *Medical Care* 2017;**55**:810–6. https://doi.org/10.1097/MLR.0000000000000764

20. Cottrell EK, Dambrun K, Cowburn S, *et al.* Variation in Electronic Health Record Documentation of Social Determinants of Health Across a National Network of Community Health Centers. *American Journal of Preventive Medicine* 2019;**57**:S65–73. https://doi.org/10.1016/j.amepre.2019.07.014

21. Schickedanz A, Hamity C, Rogers A, *et al.* Clinician Experiences and Attitudes Regarding Screening for Social Determinants of Health in a Large Integrated Health System. *Med Care* 2019;**57 Suppl 6 Suppl 2**:S197–201. https://doi.org/10.1097/mlr.0000000000001051

22. Greenwood-Ericksen M, DeJonckheere M, Syed F, *et al.* Implementation of Health-Related Social Needs Screening at Michigan Health Centers: A Qualitative Study. *The Annals of Family Medicine* 2021;**19**:310–7. https://doi.org/10.1370/afm.2690

23. Henrikson NB, Blasi PR, Dorsey CN, *et al.* Psychometric and Pragmatic Properties of Social Risk Screening Tools: A Systematic Review. *American Journal of Preventive Medicine* 2019;**57**:S13–24. https://doi.org/10.1016/j.amepre.2019.07.012

24. O'Brien KH. Social determinants of health: the how, why, and where screenings are occurring; a systematic review. *Social Work in Health Care* 2019;**58**:719–45. https://doi.org/10.1080/00981389.2019.1645795

25. Stausberg J, Lehmann N, Kaczmarek D, *et al.* Reliability of diagnoses coding with ICD-10. *International Journal of Medical Informatics* 2008;**77**:50–7. https://doi.org/10.1016/j.ijmedinf.2006.11.005

26. Hennessy DA, Quan H, Faris PD, *et al.* Do coder characteristics influence validity of ICD-10 hospital discharge data? *BMC Health Serv Res* 2010;**10**:99. https://doi.org/10.1186/1472-6963-10-99

27. Olson D, Oldfield B, Morales Navarro S. Standardizing Social Determinants Of Health Assessments | Health Affairs Blog. Health Affairs Blog. 2019.https://www.healthaffairs.org/do/https://doi.org/10.1377/hblog20190311.823116/full/ (Accessed 4 Feb 2021).

28. Harris PA, Taylor R, Minor BL, *et al.* The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* 2019;**95**:103208. https://doi.org/10.1016/j.jbi.2019.103208

29. de Souza JA, Yap BJ, Hlubocky FJ, *et al.* The development of a financial toxicity patient-reported outcome in cancer: The COST measure. *Cancer* 2014;**120**:3245–53. https://doi.org/10.1002/cncr.28814

30. Locatelli SM, Sharp LK, Syed ST, *et al.* Measuring Health-related Transportation Barriers in Urban Settings. *J Appl Meas* 2017;**18**:178–93.

31. Economic Research Service, USDA. U.S. Household Food Security Survey Module: Six-Item Short Form. 2012.https://www.ers.usda.gov/media/8282/short2012.pdf (Accessed 10 Mar 2020).

32. Rollins C, Glass NE, Perrin NA, *et al.* Housing Instability Is as Strong a Predictor of Poor Health Outcomes as Level of Danger in an Abusive Relationship: Findings From the SHARE Study. *J Interpers Violence* 2012;**27**:623–43. https://doi.org/10.1177/0886260511423241

33. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *International Journal of Medical Education* 2011;**2**:53–5. https://doi.org/10.5116/ijme.4dfb.8dfd

34. Coleman-Jensen A, Rabbitt M, Gregory C. Examining an "Experimental" Food-Security-Status Classification Method for Households With Children. USDA: Economic Research Service 2017.

35. Colorado Health Institute. Implementation Guidance. For Screening for Social Determinants of Health in an Electornic Health Record. 2021.https://oehi.colorado.gov/sites/oehi/files/documents/S-HIE%20Screening%20Implementation%20Guidance.pdf (accessed 20 Sep 2021).

36. Gold R, Cottrell E, Bunce A, *et al.* Developing Electronic Health Record (EHR) Strategies Related to Health Center Patients' Social Determinants of Health. *Journal of the American Board of Family Medicine : JABFM* 2017;**30**:428–47. https://doi.org/10.3122/jabfm.2017.04.170046

37. StataCorp. *Stata Base Reference Manual: Release 14*. College Station, TX: StataCorp LP 2015.

38. Gottlieb L, Tobey R, Cantor J, *et al.* Integrating Social And Medical Data To Improve Population Health: Opportunities And Barriers. *Health Affairs* 2016;**35**:2116–23. https://doi.org/10.1377/hlthaff.2016.0723

39. Guo Y, Chen Z, Xu K, *et al.* International Classification of Diseases, Tenth Revision, Clinical Modification social determinants of health codes are poorly used in electronic health records. *Medicine (Baltimore)* 2020;**99**:e23818. https://doi.org/10.1097/MD.0000000000023818

40. Nitkin K. A New Way to Document Social Determinants of Health. BestPractice For and about members of the Office of Johns Hopkins Physicians. 2019.https://www.hopkinsmedicine.org/office-of-johns-hopkins-physicians/best-practice-news/a-new-way-to-document-social-determinants-of-health (Accessed 13 Feb 2021).

41. Kosowan L, Katz A, Halas G, *et al.* Using Information Technology to Assess Patient Risk Factors in Primary Care Clinics: Pragmatic Evaluation. *JMIR Formative Research* 2021;**5**:e24382. https://doi.org/10.2196/24382

42. Swets JA. Measuring the accuracy of diagnostic systems. *Science (New York, NY)* 1988;**240**:1285–93. https://doi.org/10.1126/science.3287615

43. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* 2010;**5**:1315–6. https://doi.org/10.1097/JTO.0b013e3181ec173d

44. Theiss J, Regenstein M. Facing the Need: Screening Practices for the Social Determinants of Health. *The Journal of Law, Medicine & Ethics* 2017;**45**:431–41. https://doi.org/10.1177/1073110517737543

45. Rothman M, Burke L, Erickson P, *et al.* Use of Existing Patient-Reported Outcome (PRO) Instruments and Their Modification: The ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value in Health* 2009;**12**:1075–83. https://doi.org/10.1111/j.1524-4733.2009.00603.x

46. Gundersen C, Engelhard EE, Crumbaugh AS, *et al.* Brief assessment of food insecurity accurately identifies high-risk US adults. *Public Health Nutrition* 2017;**20**:1367–71. https://doi.org/10.1017/s1368980017000180

47. Kovach KA, Reid K, Grandmont J, *et al.* How Engaged Are Family Physicians in Addressing the Social Determinants of Health? A Survey Supporting the American Academy of Family Physician's Health Equity Environmental Scan. *Health Equity* 2019;**3**:449–57. https://doi.org/10.1089/heq.2019.0022

48. Parikh RB, Jain SH, Navathe AS. The Sociobehavioral Phenotype: Applying a Precision Medicine Framework to Social Determinants of Health. *AJMC* 2019;**25**:421–3.

49. Frey LJ, Lenert L, Lopez-Campos G. EHR Big Data Deep Phenotyping. Contribution of the IMIA Genomic Medicine Working Group. *Yearb Med Inform* 2014;**9**:206–11. https://doi.org/10.15265/iy-2014-0006

50. Verchinina L, Ferguson L, Flynn A, *et al.* Computable Phenotypes: Standardized Ways to Classify People Using Electronic Health Record Data. *Perspectives in Health Information Management* 2018;**Fall**:1=6.

51. Richesson RL, Hammond WE, Nahm M, *et al.* Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc* 2013;**20**:e226–31. https://doi.org/10.1136/amiajnl-2013-001926

52. Feller DJ, Iv OJBDW, Zucker J, *et al.* Detecting Social and Behavioral Determinants of Health with Structured and Free-Text Clinical Data. *Appl Clin Inform* 2020;**11**:172–81. https://doi.org/10.1055/s-0040-1702214

53. Weed LL. Medical Records That Guide and Teach. *New England Journal of Medicine* 1968;**278**:652–7. https://doi.org/10.1056/nejm196803212781204

54. Zander LI. Recording family and social history. *J R Coll Gen Pract* 1977;**27**:518–20.

55. Podder V, Lew V, Ghassemzadeh S. SOAP Notes. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing 2020. http://www.ncbi.nlm.nih.gov/books/NBK482263/ (Accessed 6 Nov 2020).

56. Chen ES, Manaktala S, Sarkar IN, *et al.* A Multi-Site Content Analysis of Social History Information in Clinical Notes. *AMIA Annu Symp Proc* 2011;**2011**:227–36.

57. Lybarger K, Ostendorf M, Yetisgen M. Annotating social determinants of health using active learning, and characterizing determinants using neural event extraction. *Journal of Biomedical Informatics* 2021;**113**:103631. https://doi.org/10.1016/j.jbi.2020.103631

58. Skaljic M, Patel IH, Pellegrini AM, *et al.* Prevalence of Financial Considerations Documented in Primary Care Encounters as Identified by Natural Language Processing Methods. *JAMA Network Open* 2019;**2**:e1910399–e1910399. https://doi.org/10.1001/jamanetworkopen.2019.10399

59. Oreskovic NM, Maniates J, Weilburg J, *et al.* Optimizing the Use of Electronic Health Records to Identify High-Risk Psychosocial Determinants of Health. *JMIR Medical Informatics* 2017;**5**:e25. https://doi.org/10.2196/medinform.8240

60. Bako AT, Walter-McCabe H, Kasthurirathne SN, *et al.* Reasons for Social Work Referrals in an Urban Safety-Net Population: A Natural Language Processing and Market Basket Analysis Approach. *Journal of Social Service Research* 2020;1:12. https://doi.org/10.1080/01488376.2020.1817834

61. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193–204 10/f5254d.