Using structured and unstructured data to identify patients' need for services that address the social determinants of health☆

Joshua R. Vest[a,b,*], Shaun J. Grannis[b,c], Dawn P. Haut[d], Paul K. Halverson[a,b,c], Nir Menachemi[a,b]

[a] Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, IN, United States
[b] Regenstrief Institute, Indianapolis, IN, United States
[c] Indiana University School of Medicine, Indianapolis, IN, United States
[d] Eskenazi Health, Indianapolis, IN, United States

ARTICLE INFO

ABSTRACT

Introduction: Increasingly, health care providers are adopting population health management approaches that address the social determinants of health (SDH). However, effectively identifying patients needing services that address a SDH in primary care settings is challenging. The purpose of the current study is to explore how various data sources can identify adult primary care patients that are in need of services that address SDH.

Methods: A cross-sectional study described patients in need of SDH services offered by a safety-net hospital's federally qualified health center clinics. SDH services of social work, behavioral health, nutrition counseling, respiratory therapy, financial planning, medical-legal partnership assistance, patient navigation, and pharmacist consultation were offered on a co-located basis and were identified using structured billing and scheduling data, and unstructured electronic health record data. We report the prevalence of the eight different SDH service needs and the patient characteristics associated with service need. Moreover, characteristics of patients with SDH services need documented in structured data sources were compared with those documented by unstructured data sources.

Results: More than half (53%) of patients needed SDH services. Those in need of such services tended to be female, older, more medically complex, and higher utilizers of services. Structured and unstructured data sources exhibited poor agreement on patient SDH services need. Patients with SDH services need documented by unstructured data tended to be more complex.

Discussion: The need for SDH services among a safety-net population is high. Identifying patients in need of such services requires multiple data sources with structured and unstructured data.

1. Introduction

Recent changes to reimbursement policies in the United States (US) have actively incentivized preventative care, cost control, accountability for health, and a renewed focus on the quality of care [1]. In response, (US) providers have been adopting population health management approaches [2] that address the underlying behavioral, social, contextual, and environmental drivers of health status and health care utilization [3]. These factors, typically referred to as social determinants of health (SDH) [4,5], have typically been outside the scope of medical practice [6]. Nonetheless, health care organizations are increasingly offering the services of social workers [7], patient navigators [8], legal experts [9], behavioral health [10] and other professionals that directly address SDH to improve health and reduce costs in the population for which they care.

A key goal of successful population health management programs involves identifying patients at risk for developing poor outcomes due to SDH factors. As such, health care organizations' care delivery processes can benefit by better identifying patients in need of services that may address underlying SDH issues [11]. However, identifying such patients is challenging, particularly in the primary care setting. Documentation of both the need for, and actual delivery of, services that address SDH issues tends to occur within electronic health records (EHR) infrequently; and when it does, it is captured in a non-standard format [12]. Moreover, services that address a SDH are often not associated with any billing or diagnosis codes further limiting routine, structured collection in EHRs and other administrative systems [13]. To overcome this limitation, institutions have implemented patient surveys

to collect SDH factors [14–16] but these data, even when computerized, may not be interoperable with other information systems [17] or their use may be viewed by patients as controversial or stigmatizing [16]. In general, most providers do not have access to comprehensive information about patients' SDH services need [18,19].

The purpose of the current study is to explore how effectively various data sources support identifying adult primary care patients that need services that address SDH. Specifically, we examine the separate and joint contributions of structured and unstructured data in identifying patients' needs for SDH services. In addition, we characterized the patients in need for SDH services and how patient characteristics differed according to SDH services need determined by structured and unstructured data sources.

2. Methods

This cross-sectional study described patients in need of SDH services offered by a safety-net hospital's federally qualified health center (FQHC) clinics. SDH services offered on a co-located basis were identified using structured billing and scheduling data, and unstructured EHR data. We report the prevalence of different SDH service needs and the patient characteristics associated with need. Next, we compared the characteristics of patients with SDH services need documented by structured data sources with those documented by unstructured data sources.

2.1. Sample & setting

We identified needs for services addressing SDH in a sample of 73,085 adult (> 18 years) patients seeking care from Eskenazi Health between April 2012 and July 2016. Eskenazi Health is the Indianapolis, IN metropolitan area public safety-net provider with a 315-bed hospital and 10 FQHC sites. Eskenazi Health offered SDH services on a co-located basis at the primary care sites during the study period, i.e. Eskenazi Health was not relying on referrals to external providers and agencies for these services. All patients had at least one outpatient encounter during the study period.

2.2. Data

Information reflected multiple information systems linked through common patient identifiers. The primary data sources were Eskenazi Health's home grown EHR and the local community health information exchange (HIE). We also accessed appointment data from four different outpatient registration and scheduling systems.

2.3. Measuring the need for SDH services

SDH services of interest were: social work, behavioral health, nutrition counseling, respiratory therapy, financial planning, medical-legal partnership assistance, patient navigation, and pharmacist consultation. Realizing that patients' need for services can be met or unmet, we operationalized need broadly as any documentation that a health care professional judged that any of the aforementioned services were appropriate, recommended, or potentially beneficial to the patient. Therefore, we considered any received services as indicative of need as well as scheduled appointments regardless of whether they were kept or not. Need for each SDH service was operationalized as binary yes or no variables.

The eight SDH services had different workflows and documentation processes, which required searching multiple information sources to identify patients in need of services. First, we examined registration and scheduling systems for any patient appointments for nutritional counseling, behavioral health, respiratory therapy services, patient navigation, and financial counseling. The appointment data reflected all kept, canceled, and "no show" visits. Second, we queried diagnosis and

billing codes from the EHR and the HIE for ICD-9, ICD-10, and CPT procedure codes associated with behavioral health, nutritionist, respiratory therapy, and pharmacist consultation. Third, we reviewed the EHR's unstructured data (i.e. orders and notes) for additional documentation of SDH services need. We searched provider orders for all eight of SDH services using keywords informed by a review of the literature and professional society's documentation recommendations (see Appendix for sources and a tabular list of SDH by data source). Additionally, because the EHR progress notes contained a specific section for social worker documentation, we used natural language processing to identify instances of social worker contact with patients (see Appendix). We excluded any notes associated with inpatient admission periods and limited the notes to only those signed by social workers practicing in the outpatient setting. For each service, the source of documented need could be structured (i.e. billing and appointments), unstructured (i.e. orders and notes), or documented by both. Additionally, we created a summary variable of any documented need based on the presence of need for any of the eight SDH services.

2.4. Additional measures

The EHR and HIE data provided patient demographics such as age, race/ethnicity, and gender. We also created binary indicators for the 20 most common chronic conditions [20] and tobacco use [21], as well calculated the Charlson comorbidity index using diagnosis codes [22]. We also counted each patient's total number of emergency department encounters, primary care visits, and hospitalizations during the entire study period. For the secondary analysis of patients with SDH services need (see below), we limited the above measures to the data available prior to the earliest date of documented need. This alternative method of measurement enabled comparisons of prior patient history between the different sources of information (i.e. structured and unstructured) available at the time of SDH services need documentation.

2.5. Analyses

We described the sources of information on SDH services, prevalence of need, and patient characteristics using frequencies and means. We compared patient characteristics in need of SDH services to those without an identified need using X^2 and *t*-tests. To assess the level of agreement between structured and unstructured information sources, we calculated kappa coefficients for documented need of any service by individual SDH services.

As a secondary analysis, we compared the characteristics of patients with SDH services need documented by structured and unstructured data. In a multinomial probit model, we examined the patient factors associated with having SDH service need documented by unstructured data, and by both structured and unstructured data, in comparison to need documented by structured data only. A best fitting model was identified using a backward elimination approach using the Bayesian Information Criterion with regression coefficients expressed as marginal effects.

3. Results

Overall, 53% of patients (n = 38,563) were in need of at least one SDH service during the study period (Table 1). Patients identified needing SDH services tended to be female, older, more medically complex, and higher utilizers of services. Specifically, the prevalence of nearly all chronic conditions examined was higher among those documented to be in need of services that addressed a SDH. For example, diagnoses of hypertension (53% vs. 35%), diabetes (32% vs. 14%), and depression (31% vs. 12%) were significantly higher in the in patients in need of SDH services (all p < 0.05). In addition, the patients with documented SDH services need also more often had a history of substance abuse (20% vs. 14%; p < 0.0001) and tobacco usage (29% vs.