ABSTRACT Recent efforts in medical settings to identify social determinants of health have focused primarily on screening for the purpose of improving care for individual patients and getting standardized data into electronic health records (EHRs). Relatively little attention has been given to processes needed to extract data on social determinants of health out of medical records with adequate validity and efficiency to facilitate analysis across individual encounters to inform population health efforts relevant to the health care sector. In this article we describe the rationale for extracting data on social determinants of health from EHRs, including the potential influence of aggregated data on quality improvement activities and health care payment reform. We then discuss opportunities and challenges to pulling these data from EHRs to enable population-level applications, focusing on the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, as one potential data aggregation resource. Standardizing methods for extracting data on social determinants of health from EHRs will require understanding current challenges and refining existing translation tools.

Increased awareness of the influence of social determinants of health—factors outside of health care that affect individual and population health—has led the health care sector toward activities that strengthen the integration of medical and social services, including activities to incorporate the collection of data on social determinants of health into care delivery. Until recently, however, health care systems have lacked robust tools for screening patients on these determinants. In 2014 the National Academy of Medicine (formerly Institute of Medicine) issued strong recommendations for standardizing the collection of measures of social determinants of health in electronic health records (EHRs), although the proliferation of surveys has moved faster than standardization. During the past five years, several tools have emerged to capture information on patients’ social determinants of health. These tools have largely converged on the same determinants and are intended to facilitate the delivery of social information to providers as an initial step toward achieving better health outcomes for individual patients.

A growing body of evidence supports tailoring clinical interventions and referrals for individual patients based on social data. Providers can use social determinants of health data to identify patients who should receive specific health screenings (for example, cardiovascular screening), social interventions and care coordination programs (for example, patient navigator programs), or referrals (for example, food banks, shelters). Accurate information on social determinants of health can also be used to match an individual’s needs to the appropriate mode of care (for example, phone consultation instead...
of in-person visits or letter-based follow-up). A critical next step involves designing or modifying health information technology to ensure that data on social determinants of health affect not only health care quality for individual patients but also health care systems’ population health initiatives. In the health care sector, where the term “population” typically refers to patients served by a health care system, recent descriptions of population-level applications of data collection on social determinants of health have explored health care payment reform at the policy level (for example, using such data in risk adjustment for reimbursement and penalty assessments). Health system–relevant applications also can include using social data to guide panel management (sometimes called population management) efforts, in which clinics proactively assign patients to groups to target health care quality improvement activities. Both payment reform and population or panel management applications for social data will depend on the ability to extract information easily and efficiently across individual records and EHR systems using standardized medical terminology.

Responding to the need for extracting data on social determinants of health from EHRs, some social screening advocates already have attempted to link screening responses with existing medical coding tools. For instance, the National Association of Community Health Centers made preliminary links between its social determinants of health screening items and International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10), codes. Since 2015 the Department of Health and Human Services has required that the ICD-10 be used for EHR documentation. The ICD-10 provides an expanded set of codes reflecting patients’ social characteristics in the form of “z-codes.” Although z-codes are not yet used routinely by health care systems, together the mandate for ICD-10 adoption and its potential for coding and billing on social determinants of health make the ICD-10 an attractive option for a standardized approach to data aggregation.

In this article we first review health sector–relevant applications for coded data on social determinants of health. We then identify and discuss critical challenges that will need to be overcome in efforts to use existing medical coding tools such as the ICD-10 to extract these data from EHRs.

Using Social Data To Inform Population Health

Although linking social and medical data could inform policy decisions and health improvement strategies outside the health care sector, health systems are most likely to invest in this integration across patient encounters and health systems if the benefits of such activities are clear and meaningful in addressing internal challenges and priorities. Beyond individual patient care, health care sector activities that could be influenced by aggregated data on social determinants of health include panel management initiatives and value-based payment reform.

Improving Panel Management

The relevance of data on social determinants of health to population health initiatives is particularly pronounced in primary care. Maturation of the primary care patient-centered medical home and value-based care initiatives requires developing systems of care not only for individuals as they seek care but also for panels of patients for whom providers or health systems are accountable. Active panel management has been demonstrated to improve care quality and has become a central tenet of effective care transformation. To date, panel management has primarily involved stratifying patients based on chronic disease groups, such as asthma, diabetes, or heart failure, instead of on demographic factors beyond age and sex. Below, we outline how information on social determinants of health collected across patient groups could improve panel management activities.

Expanding the Definition of ‘Quality Improvement’

Information on social determinants of health could be used to identify and develop a wider scope of quality improvement activities relevant to a patient panel or health care system’s population. For instance, for low-income populations whose hypoglycemia-related admissions have been found to increase predictably at month’s end as food budgets are exhausted, quality improvement efforts to decrease hospital readmissions could focus on food access interventions. Quality improvement activities that rely on data on patients’ social determinants of health are not purely theoretical: At Kaiser Permanente, panel information collected on food insecurity has resulted in clinic projects focused on improving a partnership with Hunger Free Colorado.
minants of health could help ensure that teams have sufficient expertise on a wider range of social domains relevant to population health. For instance, the Cincinnati Children’s Hospital Medical Center aggregated patient data on housing at the population level. This enabled lawyers on a multidisciplinary team to address housing code violations in the community, which reduced emergency department use by children with asthma.18

**Adjusting Provider Panel Sizes** Calculations to determine provider panel sizes differ across clinical settings but primarily are based on provider factors (for example, number of days in clinic) and staffing ratios. Panel size calculations are also based on patients’ medical conditions, which help to estimate patient complexity and predict visit frequency.19 Alongside those more traditional methods of determining panel size, easily extracted data on social determinants of health could enable clinics to better tailor panel sizes by accounting for differences in the time it takes to care for patients with specific social needs. The Veterans Health Administration, for instance, requires panel sizes for its Homeless Patient Aligned Care Teams to be half the size of nonspecialized primary care panels to accommodate increased frequency of appointments, longer appointment times, open-access scheduling, case management, and care coordination activities necessary for high-quality health care.20

**Refining Value-Based Payment Models**

Additionally, data on social determinants of health could have important implications for future payment models. Value-based payment is becoming a cornerstone of health care payment. Both the Centers for Medicare and Medicaid Services (CMS) and the Health Care Payment Learning and Action Network, a national network of public and private health care stakeholders working together to drive the adoption of effective alternative payment models, have set a goal that by 2018 50 percent of all health care payments will be in alternative payment methodologies.21 Data on social determinants of health can inform the design of two distinct types of alternative payment models: performance-based incentive payments and penalties, and population-based payments.22

**Performance-Based Incentive Payments and Penalties** Many current provider payment contracts already incorporate performance-based payments, which are tied to quality processes and outcomes such as resource use (for example, emergency department visits, inpatient admissions and readmissions). In some of these pay-for-performance calculations, CMS allows adjustment for sex, age, comorbidities, and medical frailty but has not yet included adjustments for social determinants of health.23 Paying providers based on cost and quality targets that do not account for social determinants of health might unfairly penalize providers that care for a greater proportion of patients facing social and economic adversity.24,25 For example, starting in 2012 the ACA established readmissions penalties that disproportionately affected safety-net hospitals. A report projecting differences in hospital readmissions penalties found safety-net hospitals 30 percent more likely to incur thirty-day hospital readmission penalties for common conditions than the national average.26 Readmission and mortality prediction models leveraging social determinants factors could significantly improve readmission model performance.27 A 2014 National Quality Forum (NQF) report recommended that performance assessments risk-adjust for social determinants factors that influence clinical quality outcomes. This could prevent unfair performance penalties on hospitals or providers caring for more patients with low socioeconomic status or other sociodemographic factors.25 Readily available data on social determinants of health would help pay-for-performance initiatives incorporate the NQF recommendations. For example, if readmissions are more likely for individuals with inadequate housing, then data on patients’ housing status could enable risk adjustments of payment penalties based on the number of patients “diagnosed” with that condition.

**Population-Based Payments** Population-based payments, including capitation and partial capitation models, are increasingly common in Medicare and Medicaid and are likely to constitute a substantial portion of future provider payments. These new models could motivate health care systems to pay attention to social determi-
nants of health, but with few exceptions, existing payment models do not yet allow risk adjustment for social determinants.

Although the NQF recommended some risk adjustment for social determinants of health in pay-for-performance, it made no specific recommendations related to population-based payment risk adjustment. Data on social determinants of health could inform more accurate models. For instance, in Medicare Advantage, a capitation Medicare insurance product, payment rates for health plans are risk-adjusted for the number and severity of medical diagnoses in the population using the CMS Hierarchical Condition Category model. In 2014 this model began to risk-adjust for poverty using Medicaid eligibility as a poverty proxy, resulting in higher payments for beneficiaries dually eligible for Medicare and Medicaid. Nonetheless, the current model underpredicts costs for 30–40 percent of the community-dwelling dual-eligible population who have low medical acuity but who likely suffer from high social acuity. For example, a homeless and food-insecure individual with prediabetes and no other medical conditions might be defined as having low medical acuity yet high social acuity. A more robust set of codes reflecting social acuity in revisions of Hierarchical Condition Category coding models could be used to better predict costs.

Data on social determinants of health also can inform capitation payment arrangements. For example, alternative payment methodology pilot programs in Oregon and California translate volume-based payments to federally qualified health centers into flexible funding that can be used to provide care to Medicaid beneficiaries. Data on social determinants of health collected across pilot programs in both states, for instance, could help justify nontraditional medical providers and service delivery and more accurately predict plan and provider care costs for Medicaid populations.

In both Medicare and Medicaid, new types of payments designed to reimburse and thus promote care management and case coordination activities could be refined using social data. In Medicare’s newest alternative payment methodology experiment, the Comprehensive Primary Care Plus initiative, providers receive a care management fee tiered by a patient’s Hierarchical Condition Category coding score (a score based on a patient’s medical diagnoses). If those scores accounted for social diagnoses, care management fees could more accurately reflect the costs of caring for patients with medical or social acuity, or both.

Using Medical Codes To Translate Patients’ Social Information
The ultimate goals of increasing social screening in clinical settings are to improve health outcomes for individuals and for populations. The data needed to accomplish each goal differ. In caring for individuals, providers need access to patient-level data on social determinants of health that are shareable with all members of provider teams, ideally including referral agencies. This goal can be met with the adoption of patient- and provider-facing social screening tools in health information technology platforms.

Population health activities, in contrast, require extraction of data on social determinants of health across patient encounters to inform a health system’s population management activities and across systems to inform broader health-sector initiatives (such as payment reform). Using food security as an example, Exhibit 1 gives an overview of how data on social determinants of health might be used at the patient and population levels.

This population-level aggregation is more challenging since, despite recommendations from the National Academy of Medicine about standardized measures for capturing data on social determinants of health, health care providers and organizations continue to rely on a wide variety of tools that they can customize for specific patient populations.

To ensure that EHR data on social determinants of health inform health sector–based population health activities—including billing and claims—translation tools should be rooted in existing medical documentation systems such as the ICD-10, which reflects diagnosis codes (what diagnoses and characteristics patients have that are relevant to their health); Current Procedural Terminology, Fourth Edition (what procedures and services were provided during the medical visit); and other data sets that provide common terminology for use in EHR systems.

To explore possibilities for aggregation using existing medical documentation databases, we mapped several existing social screening tools to codes available in the ICD-10. As the most common medical diagnosis coding system in the United States and internationally, the ICD-10 has the advantages of maximizing intersystem operability; having the potential to be used for billing if and when social determinants of health become billable diagnoses; and already linking with other medical terminology data sets (including the Logical Observation Identifiers Names and Codes [LOINC] system and the Systematized Nomenclature of Medicine—Clinical Terms [SNOMED CT]). This enabled us to iden-
tify opportunities and challenges to connecting “front-end” social screening tools with a “back-end,” standardized, universally available data set.

The ICD-10 successfully matched the majority of social domains covered in the selected screening tools, including economic circumstances, housing, food insecurity, safety, transportation, and legal issues. For instance, the ICD-10 Z56 category “Problems related to employment and unemployment” includes a code for unemployment (Z56.0) that maps directly to questions about an individual’s employment status. Questions related to social isolation could map to the ICD-10 category Z60, which includes specific codes for problems related to living alone (Z60.2) and social exclusion and rejection (Z60.4); and questions about refugee status could be linked directly to category Z65 “Problems related to other psychosocial circumstances,” which includes items such as exposure to disaster, war, and other hostilities (Z65.5).

Several important challenges emerged in trying to link front-end social determinants of health screening questions to back-end data. These included instances where there was no ICD-10 code equivalent for a social screening item, items where there was more than one ICD-10 code equivalent, and items where meaning was lost in the conversion to the most equivalent ICD-10 code choices. Below, we provide examples highlighting these challenges and suggest ways to overcome them to facilitate adoption of standardized medical documentation systems as data aggregation tools for social screening. (See the online Appendix for detailed examples of challenges to translating social determinants of health screening to ICD-10 codes.)

**LACK OF ICD-10 CODE** Several social determinants of health screening domains and items used in existing tools have no clear ICD-10 code equivalent. For example, “Do you have any concerns about safety in your neighborhood?” from the HealthBegins Upstream Risks Screening Tool does not have an exact correlate in the ICD-10. The National Association of Community Health Centers recently reported the problem of matching items when attempting to link them in its social screening EHR templates to both ICD-10 z-codes and LOINC, noting that it needed to create its own codes for some domains where there were not yet code equivalents. Responding to these code gaps could also inform ICD-11 development. One way to create these codes could include assigning specific uses for the category “other” that is associated with many of the z-code domains of social determinants of health, such as “Other problems related to social environment” (Z60.8).

**MULTIPLE ICD-10 CODES** Some social screening items map to more than one ICD-10 code. For instance, most existing social screening tools contain questions regarding challenges to paying for food, housing, medical care, or heat. These could map to multiple Z59 codes, includ-
The health care sector is increasingly interested in integrating social and medical care.

ing extreme poverty (Z59.5), low income (Z59.6), and insufficient social insurance and welfare support (Z59.8). Similarly, questions focused on housing security can be mapped to ICD-10 category Z59, which includes specific codes for homelessness (Z59.0) and inadequate housing (Z59.1), but the difference between these two codes is unclear. In these cases, it will be necessary to build consensus on one-to-one mapping relationships.

The field could follow a recent precedent for building consensus around mapping ICD-10 data to medical terminology. In linking the ICD-10 and SNOMED CT, which is used in EHRs to semi-automatically generate ICD-10 codes from clinical data, the World Health Organization and the International Health Technology Standards Development Organization developed mapping methodologies to ensure that each term linked to only one ICD-10 code.30

**Loss Of Meaning** A consensus-building process will also be needed to resolve instances when there is a loss of meaning or nuance that occurred by linking to a z-code. For instance, across multiple tools, food insecurity questions could be mapped with the code for lack of adequate food and safe drinking water (Z59.4). But aggregated across patients, these data would not actually distinguish between food adequacy and access to safe water or, further, whether food-specific concerns were related to lack of resources to purchase food or lack of access to healthy food. Collapsing multiple social issues into a single code might therefore hinder relevance to population-level interventions. The ICD-10 was not initially designed to ensure sufficient specificity to inform action, whether at the individual or population level. Similar to the situation in which no code equivalent exists, algorithms should be designed that delineate when it is appropriate to use existing codes versus create new ones. In developing rules to overcome these challenges, links between screening responses and the ICD-10 or other codes should be programmed into EHRs to prevent providers from using codes differently across settings as much as possible.

**Discussion**

The health care sector is increasingly interested in integrating social and medical care.31,32 The enthusiasm for such integration unfortunately has fostered the proliferation of multiple social screening tools. Ongoing efforts to standardize metrics on social determinants of health that are entered into EHRs are critical.1 But to maximally inform health sector–relevant population health initiatives, front-end social screening efforts will need to be accompanied by back-end efforts focused on meaningfully extracting data on social determinants of health from EHRs. Using social determinants of health to guide initiatives to improve quality, revise team-based care models, adjust panel sizes, and reform payment will depend on being able to aggregate these data using medical codes. A short window of opportunity exists for developing consensus on how to most effectively aggregate data on social determinants of health in EHRs using existing coding systems, since similar to the proliferation of front-end screening tools, site-specific data translation efforts are already emerging. Proliferation of these without interoperability standards will limit shared learning around transforming care.

While ICD-10 z-codes are not a panacea, they offer important lessons from which to conceptualize social and medical data integration in any documentation system. We chose the ICD-10 because, while other systems might offer more granularity to match social determinants of health screening items, the ICD-10 provides actionable diagnostic aggregation that might eventually be relevant to both clinical management and reimbursement. Procedural codes (whether in inpatient or outpatient systems) might also eventually help to capture both screening and service activities related to ICD-10 diagnosis codes.

Understanding how to most effectively develop a crosswalk between social screening and medical codes will require collaboration that should be initiated by key stakeholders—such as the NQF, the Institute for Healthcare Improvement, and the Office of the National Coordinator for Health Information Technology. These efforts will need to involve medical documentation experts, researchers in health services social interventions, and organizations that have advanced social screening tools.

The history of “e-codes”—patient injury data collected by hospitals—illustrates potential population health impacts of aggregated data. In 1991 the National Committee on Vital and Health...
Statistics recommended that external cause-of-injury codes be included in hospital discharge data sets. In 1992 a hospital claims form was approved that included a labeled space for such codes. Today, more than half of US states require routine collection of e-codes. E-codes adoption has resulted in improved research and informed policy around injury prevention. In California, e-code data have been used to inform legislation to decrease the number of small children that drown in pools and spas; help establish mandates for firearm safety in response to youth accidents and suicides; and create programs addressing preventable injuries associated with bikes, motorcycles, senior falls, and child abuse. mirroring the influence of e-codes, high-quality, easily extractable data on social determinants of health can expand the health sector’s role in broadening the definition of health.

**Conclusion**

Identifying a clear process for collecting and aggregating data on social determinants of health is an important next step toward transforming health care, refining value-based payment, and ultimately influencing both health- and non-health-sector strategies to improve population health.

The authors gratefully acknowledge the manuscript editing contributions of Stephanie Chernitsky. Support for this work was provided by the Robert Wood Johnson Foundation and Kaiser Permanente.

**NOTES**


