veradigm®

Below the Surface:
WHY MULTI-DIMENSIONAL SDoH DATA IS CRITICAL FOR CLINICAL RESEARCH

Cheryl Reifsnyder, PhD
Amanda Cohen, MPH
Lee Kallenbach, PhD, MPH

Obtaining and analyzing social determinant of health (SDoH) data relevant to a particular research question can be extremely challenging; but, ignoring SDoH data can have a significant impact on study results. That means that the lack of considering relevant SDoH information may put the study at risk.

In this eBook, we examine the various sources you can use to obtain SDoH data providing context for clinical research. After reading this eBook you will have a better understanding about why the breadth, depth, and validity of SDoH data is important for contextualizing the social risk variable in your clinical research.

Finally, you will learn how Veradigm® can support you to obtain and analyze SDoH data critical to accurately answering your clinical research questions.

Social determinants of health
and the impact on clinical research

SOCIAL DETERMINANTS OF HEALTH

Before diving into the details, this section provides a bit of a refresher on what the social determinants of health (SDoH) are and how they can impact clinical research.

An individual's health and well-being depend on more than simply good medical care. Health begins in social and physical environments—homes, workplaces, schools, neighborhoods, and communities. Socioeconomic and place-based factors in these environments have a significant influence on a person's health and well-being. These factors are known as social determinants of health (SDoH).[1,2] The national health and well-being-focused program, Healthy People 2030, divides SDoH into five key areas that include everything that affects day-to-day life:[1-3]

1   **ECONOMIC STABILITY**
Employment, poverty, and housing stability

2   **SOCIAL AND COMMUNITY CONTEXT**
Workplace conditions and civic activity

3   **ACCESS TO HEALTH CARE**
Accessibility and quality of healthcare as well as health insurance coverage

4   **AVAILABILITY OF EDUCATION**
Overall level of education, language, and literacy

5   **NEIGHBORHOOD AND BUILT ENVIRONMENT**
Access to transportation, air and water quality, and neighborhood crime and violence

Although these social characteristics may not directly cause poor health, they can create conditions that foster risk factors such as alcohol and drug use, lack of exercise, unhealthy diet, and weight gain. These risk factors, in turn, can lead to poor health and even death.[4] In fact, numerous studies suggest that SDoH are responsible for between 30% and 50% of health oucomes.[5]

Research shows that SDoH are important drivers of health inequities and disparities.[3,5,6] For example, one study examined the effects of SDoH on key health markers such as life expectancy, infant mortality, and prevention of chronic disease. Researchers found significant disparities in multiple health markers. For instance, both life expectancies and infant mortality rates varied greatly in relation to different social factors. Researchers also found a significant difference between the percentage of children diagnosed with asthma depending on a number of sociodemographic factors, such as race/ethnicity, family structure, language, income, and level of education.[4]

In another study, patient data from a Kaiser Permanente survey were analyzed to identify social factors with potential health impacts. Researchers then used electronic health record (EHR) data to determine which patients developed diabetes or hypertension in the three and a half years following the survey. Controlling for factors such as age, sex, and race/ethnicity, they analyzed the resulting data for any correlation between SDoH and onset of these chronic conditions. They found that patients with three or more risk factors were significantly more likely to develop these chronic diseases than patients with zero risk factors. The increased risk did not depend on which SDoH factors were observed.[7]

IMPACT ON CLINICAL RESEARCH

Because SDoH can cause such striking differences in health and disease outcomes, it is not surprising that they play a role in clinical research as well.

One study that illustrates the impact SDoH have on clinical trial results comes from the Yale School of Medicine. This study looked at a hypertension clinical trial comparing the effectiveness of three anti-hypertensive drugs.

After analyzing data from 13,000 trial participants, researchers found that the location where patients received care affected outcomes. This influence was observed regardless of which treatment the patient received. Patients who received care in the lowest income locations fared significantly worse than those who received care in higher income locations. These disparities were observed despite the use of standardized treatment protocols for all treatment locations.[8,9]

These findings illustrate the need for greater focus on SDoH when designing clinical trials.[8,9] This highlights that having an understanding of how SDoH contribute to clinical study results is growing in importance and will be essential for future work.

Traditionally, researchers obtain SDoH data about study participants by mining sources such as claims data, patient and disease registries, health surveys, peer-reviewed journal articles, Centers for Medicare and Medicaid Services (CMS) data, and EHR data. Researchers may also collect societal factor information as part of the study itself.[6,10,11] However, these data sources are not all equally reliable.

Obtaining and analyzing SDoH data relevant to a particular research question can be extremely challenging; but, obviously, lacking this SDoH data can have a significant impact on study outcomes. That means that the lack of relevant SDoH information will produce information gaps that may put the study at risk.

SDoH ARE RESPONSIBLE FOR BETWEEN *30% AND 50%* OF HEALTH OUTCOMES

The breadth, depth and validity of data—and why they're so important

THE BREADTH OF DATA

Surveying the full breadth of available data is essential to integrate adequate SDoH information into your clinical research. It is all too easy to include only certain segments of the population in a study and, as a result, fail to contextualize social risk in your study outcomes.

Researchers should carefully consider the makeup of the study population to avoid bias toward a particular population segment, as different population segments are associated with different social factors that can influence study outcomes. For example, one group of researchers found that social disparities such as race and ethnicity, education level, income, rural versus urban residence, and geographic location, were all connected to both life expectancy and infant mortality.[4]

In the study from the Yale School of Medicine mentioned earlier, researchers would have obtained different results if they had focused only on those patients receiving treatment in the lowest income study locations. Their results would have shown that patients were significantly more likely to die from any cause and significantly more likely to be hospitalized or die from heart failure.[8,9]

Studies clearly show that different SDoH can sway clinical research outcomes. In another study, researchers recently looked at the association of "neighborhood disadvantage" metrics (a scale known as the validated Area Deprivation Index) with the presence of Alzheimer disease neuropathology. They found that study participants who had lived in the most disadvantaged neighborhoods were more than twice as likely to display Alzheimer disease brain-related changes.[12]

It is clear that controlling for SDoH is a critical aspect of clinical study design, but have you considered that the breadth of your SDoH data sources matters just as much?

While it has long been standard practice to control for SDoH when creating study protocols and assembling cohorts, being too narrow in your sources of SDoH data can

cause additional problems. For example, data obtained from insurance claims will provide data only concerning study participants who are insured. Data collected via a survey mailed to study participants would exclude homeless people and people living in different geographic areas than that covered by the study. Data collected at a specific study site might exclude participants from a differing socioeconomic status.[6]

To fully appreciate the impact that SDoH may have on your research, you must analyze a broad set of data. Working with data scientists who understand how to access and analyze a breadth of real-world data can help ensure study accuracy and validity.

THE DEPTH OF DATA

Contextualizing social risk is essential if you want to identify the ways in which SDoH may be influencing a study outcome.

It is important that you obtain real-world data (RWD) from a broad selection of sources, but that's not the only key factor. Researchers also need to access data that is deep enough to provide meaningful information about the study participants. In this section, we examine the benefits and drawbacks of several different sources of SDoH data and provide insight into how best to achieve sufficient depth of data for your study.

✓ CLAIMS DATA

Until recently, researchers have relied heavily upon claims data as the primary source of digital RWD. CMS reimbursement information and insurance claims data both contain information that is required to determine reimbursement status, so these data are likely to provide a reliable record of the care received by the patient.[11] However, these data do not necessarily include SDoH information. Medicare data do contain demographic information, such as age, place of residence, and date of death, and this information is considered largely reliable. Beyond that, CMS data largely contains information covering patient diagnoses, treatment, equipment, and medications.[13]

While there are clearly many benefits to claims data, these data may lack longitudinal information about patients' health.[10] Instead, claims data tend to be limited to the specific timing of the event actually named in the claim, even if additional factors may have affected that event. Similarly, claims data will only contain data relating to services

covered by the insurer. Claims will not provide information about immunizations obtained at a grocery store clinic, or information about an experimental treatment the patient paid for out-of-pocket.[14]

Like claims data, CMS records contain limited information. CMS data lacks physiological measurements such as pulse, blood pressure, and the results of common laboratory tests.[11] CMS data may include information about the care patients receive, but they will not contain information about the care patients need if they did not actually receive that treatment.

EHR DATA

EHRs provide a relatively new source of information concerning societal factors that might influence study results. Furthermore, EHRs contain detailed medical information on patients, such as diagnoses, procedures, and lab tests. This information is typically accurate because it comes directly from the patient's health care provider.[6]

EHRs, particularly ambulatory EHRs, may follow a patient's medical history for longer time periods allowing for more longitudinal data. They may also include sufficient depth of information to allow researchers to include SDoH in their analyses. Data collection, with the help of an EHR, is a routine part of most clinical encounters, so EHR data can cover a broader range than various other data types.[6]

EHR data captured in structured fields tends to be categoric, numeric, or coded and may be recorded, organized, and analyzed with relative ease. Unfortunately, much of the SDoH data captured in EHRs is housed in unstructured or semi-structured fields which is more difficult to mine and analyze for clinical research.[6]

As a result, it is best to work with an experienced team of data scientists who can help strategize with you the most effective ways to source and analyze SDoH data from EHRs.
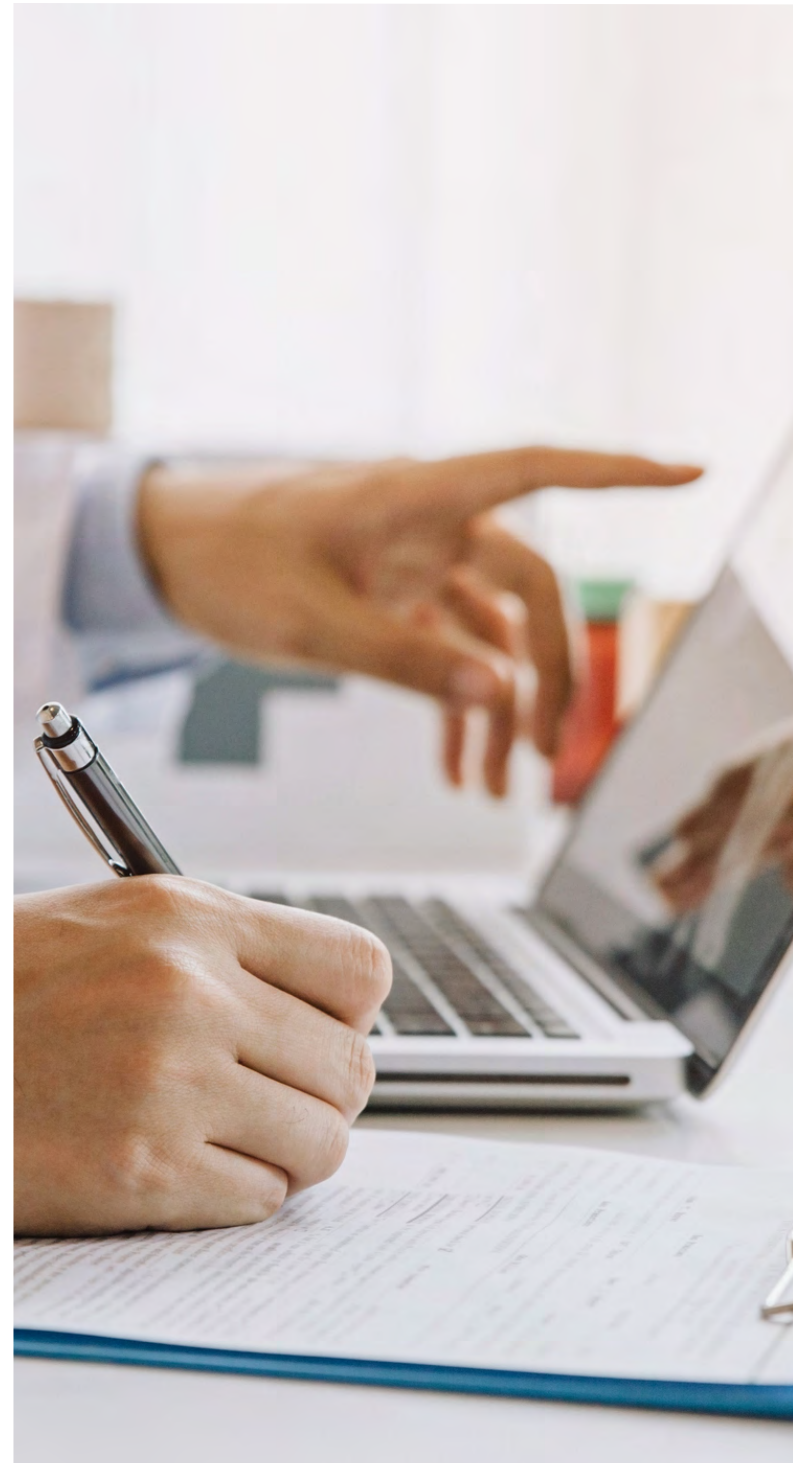
LINKING DATA SOURCES

The data found in insurance claims are relatively complete, so researchers have used claims data for the past 20 years to study a variety of topics such as drug safety. However, claims data tend to lack key clinical information, such as test results. By linking claims data to other types of data sources—such as EHR data—researchers can access a much deeper selection of patient information. Claims data may lack records of patients' lab work and test results, but EHR records, which are much more likely to contain that information, can fill in the gaps.[15]

Similarly, claims data is generally limited to the timing of the specific event named in the claim and so may lack longitudinal information about the diagnosis.[10] By linking claims data to data from an EHR, researchers can provide additional context beyond what is recorded in an insurance claim.

The EHR will generally provide a deeper, more complete set of diagnoses and other conditions affecting the patient.[15]

THE VALIDITY OF DATA

Not all sources of information provide data that is equally trustworthy. Sometimes the data is incomplete, so you cannot properly contextualize social risk in your study design. Other times, the data is simply incorrect. As a result, invalid data may point you in an entirely wrong direction. In this section, we examine the benefits and drawbacks of several different sources of SDoH data as they relate to data validity.

Once you have the ability to analyze data that is both broad enough and takes a deep enough look at SDoH, you have another hurdle to cross: You must verify that the data is actually valid.

HEALTH SURVEYS

Data derived from a health survey, for instance, may be less reliable than data from other sources for several reasons. It can be difficult to obtain detailed information in self-reported surveys. Surveys can have a low response rate, so the data they yield is only based on a subset of the population.

Alternatively, people may choose not to answer difficult questions or may fail to remember important details in the answers they do provide. In these cases, the survey data will not be broad enough or deep enough to adequately complement the study results.[6]

CLAIMS DATA

Some claims data may be less reliable because some services are not covered by insurance and, therefore, are not included in records. Claims data can also cause difficulties because different care settings use different coding systems to record procedures (for example, ICD-10 versus CPT codes for medical procedures).

The use of multiple different coding systems can make accurate interpretation of the data difficult.[6] Claims data also only cover care the patient received, but does not account for care the patient may have needed but did not receive.[11]

EHR DATA

Because EHR data comes directly from the health care provider's notes and interactions with the patient, information derived from EHRs tends to be more reliable than information derived from other sources. However, data in the EHR can be limited by the types of tests and procedures that are covered by insurance. Information that is needed for patients' insurance claims is more likely to be recorded in the EHR than treatments or treatment components that are not eligible for payment, or for which reimbursement rates are very low.[10,11]

Data captured in structured EHR fields tend to be categoric, numeric, or coded and may be recorded, organized, and analyzed with relative ease. The data housed in unstructured or semi-structured fields, however, are more difficult to analyze but may represent a significant source of understanding, particularly when contextualizing social risk.

Data scientists have begun enriching EHR data through natural language processing (NLP) and machine learning (ML) to make this SDoH data accessible for use in clinical research.

LINKED DATA SOURCES

Just as linking data sources can give you access to a deeper selection of study data than a single source, it is also a better way to ensure that your data is valid. Studies show that medical information recorded in the EHR is not always captured by claims data. For instance, in one study researchers identified 507 children and adolescents (3.6% of the study's participants) who had hypertension based on the blood pressure measurements recorded in their EHR during three well-child visits. Of these 507, only 131 (26%) had a diagnosis of hypertension or elevated blood pressure recorded.

Identifying all 507 participants with hypertension required an in-depth look at information in the EHR.[16] In this sort of situation, the only way to identify which data is valid would be to link the data from the EHR and insurance claims.

When it comes to data validity, above all, it is critical that you work with an expert team of data scientists who can help you think through your findings to best contextualize social risk.

Adding another dimension with surveillance and observational data

Surveillance or observational data come from the continuing, systematic collection and analysis of information, coupled with a system for communicating these data to those people who are in charge of preventing and controlling disease, illness, and injury.[10]

The Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO), and other medical institutions provide electronic reporting systems and operate databases with which to track and monitor outbreaks of certain diseases.[6]

Patient and disease registries are other types of public health surveillance that record health and demographic information about patients who came down with those diseases.[10] Observational data from patient and/or disease registries have some advantages as information sources for research studies.

First, this type of data is valuable because the databases usually work in partnership with a broad spectrum of health care providers (labs, hospitals, and private health care providers.)

Second, this data is generally more valid than data obtained from surveys, for example, because they come directly from lab results, diagnoses, and other patient records.

The registries in which this data is stored also makes it easier to supply and analyze this data.[6]

However, observational data cannot stand on its own because it is completely dependent on doctors and hospitals for its trustworthiness. If a doctor, lab, hospital, or other health care provider neglects to report an occurrence of the disease in question, then the data will have gaps where this information is lacking.[6] Ideally, surveillance data and observational data should be used to supplement information gained from the EHR.

Through the mining and analysis of broad, deep, and valid data combined with additional insights gained from surveillance and observational data, you can begin to properly contextualize social risk in your clinical research. In the next section, we demonstrate how Veradigm® data scientists did just that.

Case study: using SDoH data to contextualize social risk

A group of Veradigm researchers performed a retrospective cohort analysis to investigate how RWD collected during routine clinical care could provide data about SDoH as they related to patients with chronic conditions.

Researchers used de-identified RWD obtained from Practice Fusion, an ambulatory EHR platform provided by Veradigm. Practice Fusion is a large cloud-based EHR system focused on the U.S. It provides tools for more than 20,000 medical practices to care for five million patients every month.[17] This study assessed the relationship between a selection of social determinants and the chronic conditions: diabetes, asthma, chronic obstructive pulmonary disease (COPD), cancer, mental illness, HIV/AIDS, and cardiovascular disease. Previous research shows that these chronic conditions are often influenced by social determinants.

THE SOCIAL DETERMINANTS ASSESSED WERE GROUPED INTO THE FIVE BROAD CATEGORIES IDENTIFIED BY HEALTHY PEOPLE 2030

(1) ECONOMIC STABILITY

(2) SOCIAL AND COMMUNITY CONTEXT

(3) HEALTH AND HEALTH CARE

(4) EDUCATION

(5) NEIGHBORHOOD AND BUILT ENVIRONMENT[1]

Eligible patients in this study had at least one new social determinant code during the intake for the study and at least one healthcare interaction (such as a phone consult, lab test, or office visit.)[17]

This study confirmed a clear connection between SDoH and the chronic conditions outlined above. The findings demonstrate that social determinant data can be effectively captured in the EHR and mined and analyzed for clinical research. These data may be used for tasks such as identification of individuals who are at risk for illness; identification of individuals in need of social service referrals and outreach; providing additional context to assist in clinical decision-making; and gaining insight into overall population health.[17]

This SDoH data is also vital in the context of clinical research because they may enable researchers to accurately contextualize social risk in their findings. Studies, such as this one, that leverage RWD from EHR platforms may provide insight into behavioral, socioeconomic, and place-based factors that create barriers to treatment access.[17]

Working together for more robust clinical research

All researchers strive to construct robust and unbiased studies. One way to ensure this outcome is to properly contextualize social risk by accessing and analyzing patient-level SDoH data.

These data must be broad and deep enough while, at the same time, accurate and valid. That is a tall order for study teams to take upon themselves. Fortunately, Veradigm provides world-class research consulting, offering subject matter expertise in RWD and multiple therapeutic areas.

Veradigm's point-of-care presence provides a large patient footprint for biopharma companies. Our current network encompasses approximately 20% of U.S. outpatient providers, from small independent practices to larger, multi-state health systems. Veradigm's research database contains nearly 130 million patients and provides a healthcare data and analytics platform to efficiently generate real-world evidence.

Adding another dimension to Veradigm's data and analytics capabilities is our partnership with the American College of Cardiology's PINNACLE and Diabetes Collaborative registries which provide access to outpatient cardiometabolic registries.

We know that accessing and analyzing SDoH data is a particular challenge as much of these data live in unstructured or semi-structured fields within the EHR. Veradigm's team of expert data scientists have developed proprietary technology to access and analyze this valuable data more efficiently. We are using artificial intelligence tools such as natural language processing and machine learning to extract social determinant information from unstructured and semi-structured fields in our EHR products.[17] Extracting and standardizing social determinant data can help make SDoH information available for use when contextualizing social risk in clinical research. In this way, EHR data may provide novel insight into barriers that interfere with treatment or enable non-adherence.[17]

Veradigm's integrated research platform, StudySource, directly connects physicians and patients to research opportunities. StudySource helps life science research sponsors gain access to RWD to advance their research efforts.[17]

If you are interested in working with Veradigm's expert team of data scientists to help answer your research questions, visit veradigm.com.

REFERENCES:

1. Social Determinants of Health. Office of Disease Prevention and Health Promotion (ODPHP). Updated October 8, 2020. Accessed January 8, 2021, https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health.

2. Centers for Disease Control and Prevention. About Social Determinants of Health (SDOH). Updated August 19, 2020. Accessed January 13, 2021, https://www.cdc.gov/socialdeterminants/about.html.

3. Office of Disease Prevention and Health Promotion. Social Determinants of Health Literature Summaries. Healthy People 2030 website. Accessed January 13, 2021, https://health.gov/healthypeople/objectives-and-data/social-determinants-health/literature-summaries.

4. Singh GK, Daus GP, Allender M, et al. Social Determinants of Health in the United States: Addressing Major Health Inequality Trends for the Nation, 1935-2016. *Int J MCH AIDS*. 2017;6(2):139-164.

5. WHO. Social determinants of health. World Health Organization website. Accessed January 13, 2021, https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1.

6. National Information Center on Health Services Research & Health Care Technology (NICHSR). Finding and Using Health Statistics. NIH - National Library of Medicine website. Updated April 3, 2019. Accessed January 10, 2021, https://www.nlm.nih.gov/nichsr/stats_tutorial/cover.html.

7. Pantell MS, Prather AA, Downing JM. Association of Social and Behavioral Risk Factors With Earlier Onset of Adult Hypertension and Diabetes. *JAMA Network Open*. 2019;2(5)doi:doi:10.1001/jamanetworkopen.2019.3933

8. Heath S. Should Clinical Trials Account for Social Determinants of Health? Xtelligent Healthcare Media. Updated August 1, 2019. Accessed January 8, 2021, https://patientengagementhit.com/news/should-clinical-trials-account-for-social-determinants-of-health.

9. Shahu A, Herrin J, Dhruva SS, et al. Disparities in Socioeconomic Context and Association With Blood Pressure Control and Cardiovascular Outcomes in ALLHAT. *J Am Heart Assoc*. 2019;8(15):e012277. DOI: 10.1161/JAHA.119.012277.

10. Data Resources in the Health Sciences. Health Sciences Library - University of Washington website. Accessed January 9, 2021, https://guides.lib.uw.edu/c.php?g=99209&p=642709#12207218.

11. Virnig B. Strengths and Limitations of CMS Administrative Data in Research. Research Data Assistance Center website. Updated January 10, 2018. Accessed January 10, 2021, https://www.resdac.org/articles/strengths-and-limitations-cms-administrative-data-research.

12. Powell RW, Buckingham WR, Larson JL, et al. Association of Neighborhood-Level Disadvantage With Alzheimer Disease Neuropathology. *JAMA Network Open*. 2020;3(6):e207559. doi:10.1001/jamanetworkopen.2020.7559.

13. Siedelman L. Diagnosis and Procedure Coding Resources. Research Data Assistance Center website. Updated July 14, 2017. Accessed January 10, 2021, https://www.resdac.org/articles/diagnosis-and-procedure-coding-resources.

14. Siwicki B. EHR vendor Veradigm and partners creating new shared data tools for researchers. HealthcareITNews website. Updated October 22, 2019. Accessed January 9, 2021, https://www.healthcareitnews.com/news/ehr-vendor-veradigm-and-partners-creating-new-shared-data-tools-researchers.

15. West SL, Johnson W, Visscher W, Kluckman M, Qin Y, Larsen A. The challenges of linking health insurer claims with electronic medical records. *Health Informatics Journal*. February 18, 2014 2014;20(1):22-34.

16. Hansen ML, MD, Gunn P, BS, DKaelber D, MD, PhD. Underdiagnosis of Hypertension in Children and Adolescents. *JAMA Network Open*. August 22, 2007 298(8):874-879.

17. Farah JM, Martinez E, Fischer L, et al. Social Determinants of Health, Chronic Conditions, and Provider Specialties: A Retrospective Electronic Health Record-Derived Analysis. Veradigm website. https://veradigm.com/img/resource-social-determinants-of-health-a-retrospective-analysis.pdf.