SDOH machine learning model: A new way to reward providers and payers for preventive care

A Whitepaper

Abstract

Patient 360 refers to a comprehensive view of a patient's health data, including their lifestyle, medical history, and other social factors. According to the WHO, social factors alone contribute to 35–50 percent of health outcomes. These data are captured by physicians and transformed into electronic health records (EHRs). Manually extracting these social elements from physician notes is a time-consuming and error-prone process, which can limit the insights gained from the notes. Indium proposes a machine learning solution that uses NLP techniques to identify key social factors from the notes (unstructured text) and extract relevant data. Healthcare providers and payers can leverage the model to provide targeted preventive interventions, improve patient care, and reduce healthcare costs.

Highlights

- The goal of this whitepaper is to explore patient 360—SDOH and the need for ML in healthcare.
- The ML solution uses natural language processing to mine a huge volume of physician notes. It extracts SDOH elements from the notes and classifies them into one of six SDOH categories: healthcare access, education, social and community context, economic stability, food, and built environment.
- The paper highlights how the SDOH ML model rewards healthcare systems and payers in terms of quality care, reduced healthcare costs, and a healthy population.
- The study concludes that ML is essential in creating clinical decision support systems and individualized treatment plans to offer the greatest results.

Table of contents

Executive Summary

Social determinants of health (SDOH) are non-medical factors that influence health outcomes. These factors include socioeconomic status, education, neighborhood and physical environment, employment, and social support networks.

SDOH have a major impact on health, and they can contribute to health inequities. Healthcare outcomes are the results of healthcare interventions. These outcomes can include things like length of stay in the hospital, quality of life, and mortality. In fact, the WHO and CDC state that social determinants influence health outcomes more than lifestyle choices or genetic factors.

Indium proposes a solution to address the multifaceted health problem by integrating SDOH into the patient 360 framework, which is a critical step toward providing more holistic and effective patient care. This can help providers and payers develop more personalized treatment plans and connect patients with the resources they need to improve their health. Some of our solution highlights:

- First, it can help providers understand the full picture of a patient's health and identify the factors that may be contributing to their health problems.
- Second, it can help providers develop more personalized treatment plans that address the patient's SDOH needs and connect patients with the resources they need to improve their health.
- Most importantly, it saves the payers (healthcare insurance companies) from huge claim submissions and treatment costs. Insurers spend less than 1% of their net income on SDOH constituting a very small share of their profit.

"Do you know that six in ten adults in the US have a chronic disease and four in ten adults have two or more?"

Since socioeconomic factors alone account for 47 percent of health outcomes, the Centers for Medicare & Medicaid Services (CMS) has identified the following state flexibilities in the USA to address SDOH:

According to the CDC, 37.3 million Americans have diabetes, and 96 million adults are prediabetic. Diabetes costs $327 billion and is the eighth leading cause of death in the United States. For example, a diabetic patient may not be fully aware of the benefits of eating a healthy diet and leading an active lifestyle due to their socioeconomic situation.

By integrating SDOH into the patient 360 framework, the care management team can recognize these problems and connect the patient to a diabetes prevention care management program that includes active lifestyle interventions and ways to access nutritious meals through social security schemes established by the USDA.

Introduction

Preventive care is essential for improving healthcare outcomes and reducing costs. By identifying diseases at the onset, preventive care can help them from becoming more chronic, which frequently necessitates expensive treatment. For example, a study by the Centers for Disease Control and Prevention found that screening for cervical cancer can reduce the risk of death from the disease by up to 93 percent.

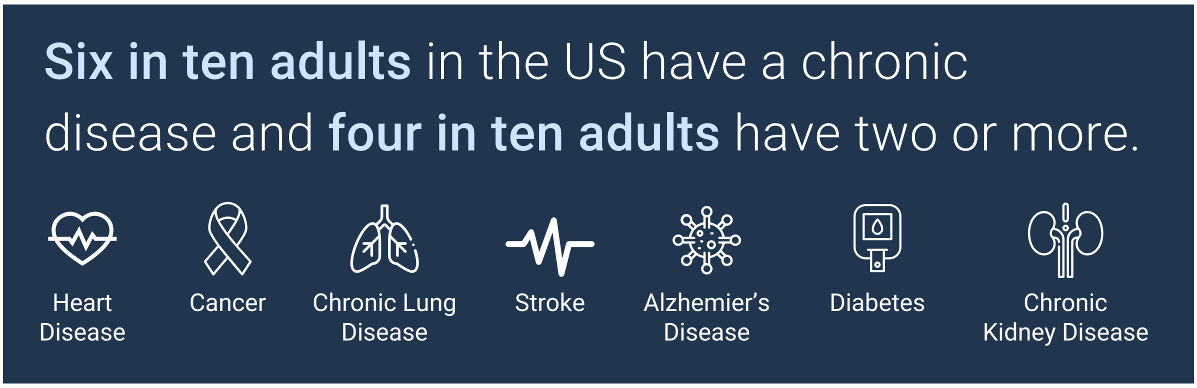


Image 1: Chronic diseases

Chronic diseases are a major public health crisis in the United States. They are the main contributor to the nation's $4.1 trillion yearly health care expenses and the top cause of death and disability.

Many chronic diseases are caused by a few key risk behaviors:

- Smoking tobacco and being in close proximity to smokers.
- Inactivity; poor nutrition, especially diets high in sodium and saturated fats and low in fruits and vegetables.
- Excessive alcohol abuse.

Since these risk behaviors are either directly or indirectly connected to SDOH indicators, learning about the patient's health status will be crucial in preventing them from needing chronic care.

Background – Unlock health equity

The objective of prevention is to reduce the probability of a person becoming ill or dying prematurely. There is no "one size fits all" approach to preventive healthcare. Specific goals should be developed by and for each person depending on their complete healthcare data and SDOH.

Since, the risk profile of a person, or their likelihood of contracting a disease, is greatly influenced by their age, sex, genetic background, lifestyle, and physical and social surroundings, their specific goals should be created considering these factors. Risk factors are variables that raise the risk. Age, sex, and family history are examples of risk variables that are outside of a person's control.

The likelihood of acquiring disorders may be reduced by changing other risk factors, such as a person's lifestyle, physical, and social surroundings. Additionally, risk can be minimized through good medical care and preventive measures.

Limitations of traditional approach

Traditional healthcare approaches often lack the ability to capture and utilize comprehensive patient data for proactive strategies. Moreover, they also don’t consider the impact of social determinants on patient health which are pivotal. This is because traditional healthcare systems are siloed—they do not share data between different providers. This can make it difficult to track a patient's overall health over time, and it can also make it difficult to identify patients who are at risk for developing certain diseases.

For example, a patient may see a primary care doctor for a routine checkup, but the doctor may not have access to the patient's medical records from other providers, such as a specialist or a hospital. This can make it difficult for the doctor to get a complete picture of the patient's health, and it can also make it difficult to identify any potential problems.

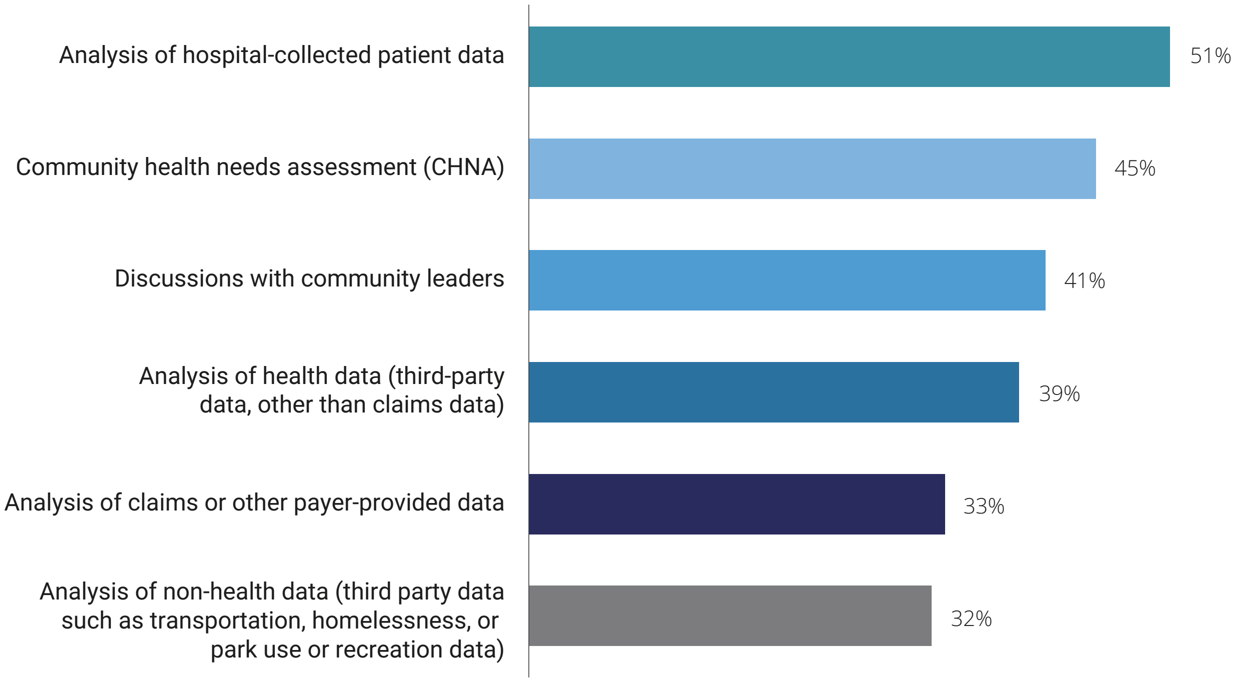
Even though the social determinants of health (SDOH) account for 80 percent of a person's health, the healthcare industry is only at the inception of understanding the interrelationships of these factors with preventive care. In fact, integrating social factors into routine healthcare is still in its infancy.

Problem statement

Physician notes comprise of a patient's medical history, including their symptoms, diagnosis, treatment plan, progress, and social elements. Manually extracting the SDOH parameters and flagging them under their respective categories from the physician notes can be a tedious process. It has several other limitations as well:

- **Time-consuming:** Manual analysis can take hours or even days to extract the data from a single note, and this process is often error-prone.
- **Subjective:** The interpretation of physician notes is subjective—different people may interpret the same note in different ways, leading to inaccurate or incomplete data.
- **Inefficient:** Manual extraction and analysis cannot be easily applied to large datasets of notes.
- **Complexity:** Physician notes can be complex, containing a lot of information, making it difficult to extract the relevant data timely and accurately.
- **Error-prone:** Manual data extraction can cause transcription errors, coding errors, or interpretation errors, which can impact the quality of care.

Methods used for assessing social needs in hospitals' target population

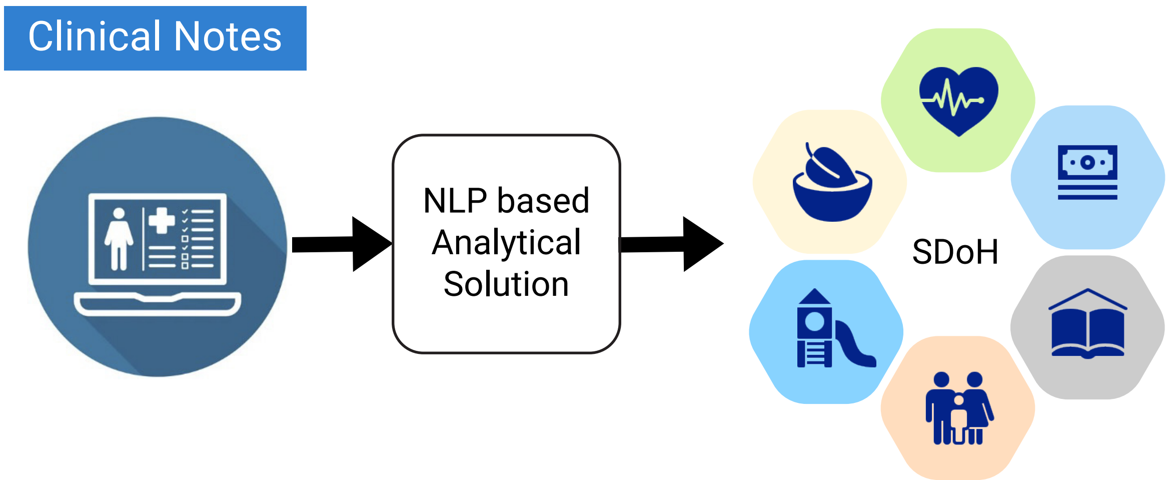


Source: Deloitte Center for Health Solutions' 2017 Social Determinants of Health Hospital Survey

Image 2: Traditional methods used to assess social factors

Proposed solution – Upstream with paddle

Our SDOH ML engine is a powerful tool that can be used to overcome the limitations of traditional approaches to extracting and analyzing data from physician notes. We have trained our ML model to extract data from the unstructured physician notes automatically, which can be used to identify SDOH attributes and flag them under their respective buckets. This can lead to more accurate, scalable, and flexible data extraction and analysis.



HEALTHCARE
- Access to Healthcare
- Coverage
- Quality of Care
- Provider cultural and linguistic competency

ECONOMIC STABILITY
- Employment
- Income (Poverty)
- Expenses
- Debt
- Medical Bills
- Housing Instability

EDUCATION
- Literacy
- Early Childhood Education
- Higher Education
- Health Literacy
- Vocational Training

PHYSICAL ENVIRONMENT
- Housing
- Transportation
- Safety
- Parks
- Playgrounds
- Walkability

SOCIAL CONTEXT
- Civic Participation
- Discrimination
- Incarceration
- Social Cohesion
- Community Engagement
- Discrimination

FOOD
- Hunger
- Access to Healthy Food
- Food Insecurity

Image 3 – ML model extracts data from physician notes, identifies SDOH attributes, and flags them under specific SDOH bracket

Our SDOH model undergoes a dynamic process of unsupervised learning that allows the model to continuously learn and change its behavior autonomously with the increasing amount of data input while retaining previously gained knowledge. Similar to how the human clinicians learn, our engine can incrementally learn and fine-tune its performance.

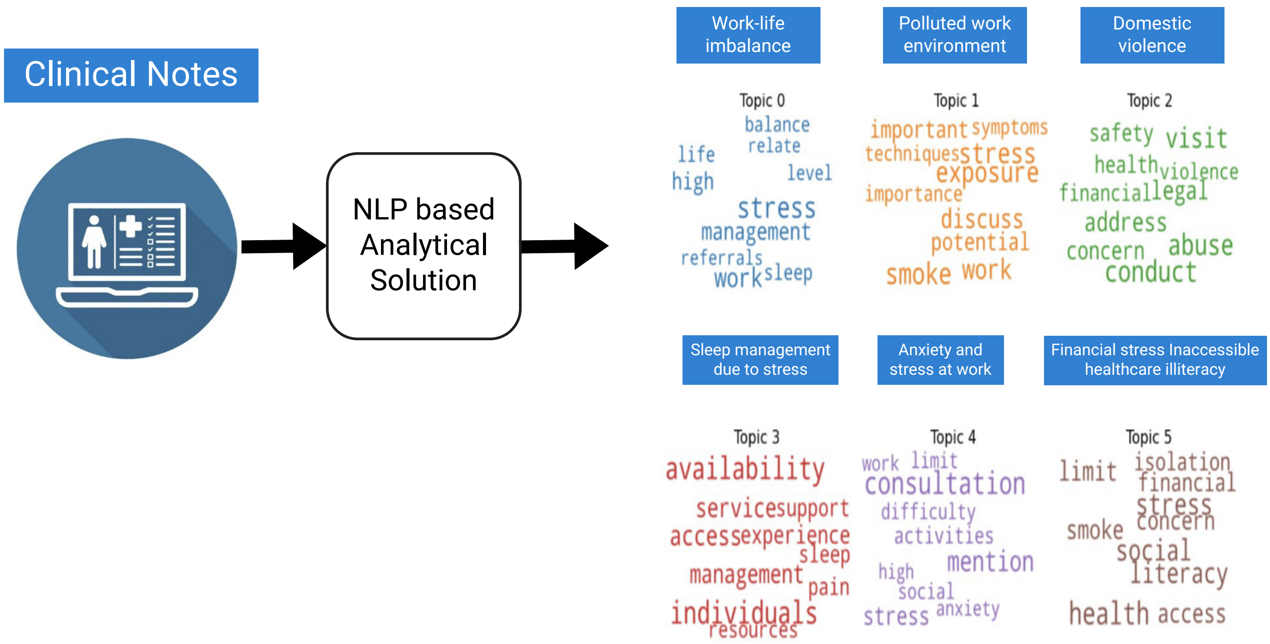


Image 4 – This is a deep dive SDOH model classification

For example, our machine learning model can identify patients who are at risk for heart disease by analyzing their medical history, lifestyle habits, and social determinants of health.

The model can then be used to recommend interventions that could help reduce the patient's risk of heart disease, considering SDOH factors such as stress, smoking, work, or sleep.

Process flow

Here is the process of our machine learning model that can be incorporated into healthcare systems to identify SDOH attributes:

- **Data collection:** We collect large datasets of patient records from electronic health records (EHRs) or electronic medical records (EMRs) in the form of csv, word, or text documents.

- **Data import:** We upload the patient's medical records into the ML model for Natural Language Processing (NLP).

- **NLP processing:** The model scans the patient records to look for SDOH-relevant terms and identifies and classifies them as shown in image

- **Output:** The output will be in the form of a grid, Excel, or database as shown in image 5, with accurate segregation of the SDOH parameters.

Member ID	Economic Stability	Physical Environment	Education	Food	Social Context	Healthcare
XYZ1	Yes - Unemployment	-	-	-	Yes - Loneliness	-
XYZ2	-	Yes – Homelessness	-	Yes - Food Insecurity	-	-
XYZ3	-	-	Yes – Attended High School only	-	-	-
XYZ4	Yes - Poverty	Yes – Homelessness	-	Yes - Malnutrition	-	-
XYZ5	-	-	Yes – No Access to School	-	Yes - Social Isolation	-
XYZ6	-	Yes – Safety	-	Yes - Malnutrition	-	Yes – Limited Access to healthcare

Image 5 – Output of the ML model

Scope of the solution

Input: The input to the model can be a text, word, or a CSV file. The text or word should encapsulate SDOH elements for the model to extract and classify.

Output: The output of the model can be stored in the cloud or downloaded in a simple Excel file. It could be incorporated into any population health management or care management system, allowing care coordinators to follow up with patients and provide prompt care. The findings may be sent directly to other third-party systems, and the output can also be integrated into such systems.

Analytical modeling

We built an NLP solution for extracting SDoH categories by gathering a significant number of clinical notes with SDoH-related data in various formats.

Data pre-processing

The quality of the data used to train the model is critical to its accuracy. We used a variety of pre-processing techniques to ensure the quality of our data, including:

- **Data cleansing:** Removed any incomplete or noisydata from the dataset.
- **Tokenization:** Separated the text into discrete tokens for NLP models to process.
- **Stop word removal:** Eliminated common and frequently occurring words across the documents to prevent model biasing.
- **Lemmatization:** Created root forms for each token by grouping together inflected forms of words.
- **Vectorization/Word embeddings:** Mapped the tokens to corresponding numerical vectors using techniques such as Word2Vec, GloVe, and FastText. The NLP receives these vectors as input.

Model training

We developed an ensemble model that uses three different techniques to predict Social Determinants of Health (SDoH) categories. These techniques are:

- **Topic modeling:** It locates clusters or groupings of related terms in a text corpus.
- **Semantic search:** This method identifies words that are conceptually connected to a specific query.
- **Multiclass classification:** This technique places a document in one of many predetermined groups.

We trained each of these techniques on the entire text corpus and then used the mode of the individual predictions from each technique to make the final prediction. The mode is the most frequent value in a set of data.

Model Evaluation

Once we had trained our model, we evaluated it to ensure that it was accurate and reliable. We used a variety of techniques to evaluate the model, including:

- **Holdout validation:** We split the dataset into two parts: a training set and a validation set. We trained the model on the training set and then evaluated it on the validation set. This allowed us to assess the model's accuracy on data that it had not seen before.

- **Cross-validation:** We repeated the holdout validation process multiple times, using different splits of the dataset each time. This allowed us to get a more accurate estimate of the model's accuracy.

Model deployment

The model is now ready for deployment. It can be used in healthcare systems to analyze patient data and accurately identify SDOH attributes.

Benefits - Get more than what you expect

Based on the POC, it is observed that the NLP-based SDOH model finds 70% more potential patients than the traditional methods of going through clinical records. Treatment costs and administrative overheads in end-to-end claim processing are likely to be reduced by 30–40%.

Healthcare system

- Early intervention
- Quality care
- Short-term treatment

Patient

- Patient satisfaction
- Reduced disparities

Payers

- Automated risk assessment
- Targeted interventions
- Less treatment costs
- Fewer claims processing

Healthcare system benefits

Improved quality of care: By identifying patients who are at risk for certain diseases or health conditions, healthcare providers can intervene early and prevent these diseases from developing.

Reduced healthcare costs: Early intervention can help prevent costly emergency room visits and hospitalizations. It can also help to reduce the need for long-term care.

Improved patient outcomes: Patients who receive care that is tailored to their SDOH are more likely to have better health outcomes. This is because the care is more likely to address the underlying factors that are contributing to their health problems.

Patient benefits

Increased patient satisfaction: Patients who feel that their healthcare providers understand their SDOH are more likely to be satisfied with their care. This is because they feel that they are being treated as a whole person, not just as a collection of symptoms.

Reduced disparities in healthcare: Our machine learning model can help reduce disparities in healthcare by ensuring that all patients have access to the care that they need, regardless of their socioeconomic status or race/ethnicity.

Improved population health: The ML model, when harnessed to identify and track trends in SDOH, can be used to develop interventions that improve the health of the entire population.

Healthcare insurance companies and payers benefits

Reduced administrative costs: Payers can use our SDOH ML model to automate the process of identifying patients who are at risk for certain diseases or health conditions. This can help reduce the administrative costs associated with managing these patients.

Improved risk assessment: Payers can use our tool to improve their risk assessment models. This can help them better predict which patients are at risk for high healthcare costs.

Targeted interventions: Payers can use targeted interventions to reach patients who are most likely to benefit from them. This can help to improve the effectiveness of these interventions and reduce the overall cost of healthcare.

Increased compliance: Payers can use our SDOH model to monitor patients' adherence to treatment plans. This can help to ensure that patients are receiving the care that they need and are not at risk for complications.

Identify patients who are at risk for chronic diseases: Payers can use our machine learning model to identify patients who are at risk for chronic diseases, such as heart disease, diabetes, and stroke. This information can be used to enrol these patients in preventive care programs that can help prevent or delay the onset of these diseases.

Enrol patients in wellness programs: Payers can use our model to enrol patients in wellness programs that are tailored to their specific needs. For example, patients who are at risk for obesity might be enrolled in a program that provides them with resources to help them lose weight.

Provide financial assistance: Payers can leverage our model to identify patients who are struggling to afford their healthcare costs. This information can be used to provide these patients with financial assistance, such as subsidies or discounts.

Conclusion

In this whitepaper, we propose an innovative approach to leveraging patient information for preventive care management by combining patient 360 and social determinants of health (SDOH). We introduce the SDOH machine learning (ML) solution to mine meaningful data from physician notes, enabling healthcare organizations to gain deeper insights into patients' health conditions and address social factors that impact their well-being.

The SDOH ML model has the potential to revolutionize the way preventive care is delivered and drive proactive healthcare strategies. The model can be trained to assist healthcare organizations in early intervention and the prevention of these disorders by identifying individuals early in the care management spectrum who are at risk for specific diseases or health conditions. This can lead to improved patient outcomes, reduced healthcare costs, and a healthier population.

One of the Healthy People 2030 initiative's overarching goals is to "promote health and well-being for all by creating social, physical, and economic environments that support it."

The SDOH ML model will become an increasingly valuable tool for healthcare organizations. We believe that the model has the potential to change the way healthcare is delivered and improve the lives of millions of people.

Here are some examples of how healthcare systems can be rewarded for using SDOH machine learning models for preventive care:

- In California, the state's Medicaid program, Medi-Cal, rewards providers for providing preventive care to patients with SDOH risk factors.
- In Massachusetts, the state's health insurance exchange, MassHealth, is offering bundled payments to providers who provide comprehensive care to patients with complex needs, including those with SDOH risk factors.
- The Blue Cross Blue Shield Association is working with a number of providers to develop risk-adjusted payment models that take into account SDOH risk factors.

References

1. World Health Organization. (n.d.). Social Determinants of Health. Retrieved from https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

2. World Health Organization. (2021). Social determinants of health. Retrieved from https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1

3. Centers for Disease Control and Prevention. (2021). Addressing social determinants of health. Retrieved from https://www.cdc.gov/about/sdoh/addressing-sdoh.html

4. Becker's Payer Issues. (2019). Private payers spent $1.87B on social determinants of health over 5 years: 5 things to know. Retrieved from https://www.beckerspayer.com/payer/private-payers-spent-1-87b-on-social-determinants-of-health-over-5-years-5-things-to-know.html

5. American Society for Policy Analysis and Research. (2022). Addressing Social Determinants of Health: Examples of Successful Evidence-Based Strategies and Current Federal Efforts. Retrieved from https://aspe.hhs.gov/sites/default/files/documents/e2b650cd64cf84aae8ff0fae7474af82/SDOH-Evidence-Review.pdf

6. CDC. "Cervical Cancer." Vital Signs, Centers for Disease Control and Prevention, 13 Sept. 2022, www.cdc.gov/vitalsigns/cervical-cancer/index.html.

7. Centers for Disease Control and Prevention. "Flu Vaccines Work: Vaccine Effectiveness." Centers for Disease Control and Prevention, 2023, https://www.cdc.gov/flu/vaccines-work/vaccineeffect.htm.

8. Centers for Disease Control and Prevention. (2023, January 10). Chronic Diseases: About. Retrieved from https://www.cdc.gov/chronicdisease/about/index.htm.

9. M.D.S. Manuals. (2022, November 29). Overview of Preventive Care. Retrieved from https://www.msdmanuals.com/en-in/home/fundamentals/prevention/overview-of-preventive-care

10. Bazzoli, Fred. "Unlocking health equity: The promise and challenges of integrating social determinants of health data." Health Data Management, June 2023, https://www.healthdatamanagement.com/articles/unlocking-health-equity-the-promise-and-challenges-of-integrating-social-determinants-of-health-data/.

11. U.S. Department of Health and Human Services. "Healthy People 2030 Framework." Healthy People, 2023, https://health.gov/healthypeople/about/healthy-people-2030-framework.

12. California Health Care Foundation. "MediCal Explained: How Health Centers Are Paid." California Health Care Foundation, May 2022, https://www.chcf.org/wp-content/uploads/2022/05/MediCalExplainedHealthCentersPaid.pdf.

13. Ayache, N. (2016). Digital health care: Perspectives, applications, and cases. JAMA Internal Medicine, 177(10), 1424–1430. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5710209/

14. Minemeyer, P. (2018, March 15). Blue Cross Blue Shield Association, BCBS Institute Take Aim at Social Determinants of Health. Fierce Healthcare. Retrieved July 14, 2023, from https://www.fiercehealthcare.com/payer/blue-cross-blue-shield-association-bcbs-institute-social-determinants-health-lyft-cvs

indium

Make Technology Work

USA	INDIA	UK	SINGAPORE
Cupertino | Princeton	Chennai | Bengaluru | Mumbai | Hyderabad	London	Singapore
Toll-free: +1-888-207-5969	Toll-free: 1800-123-1191	Ph: +44 1420 300014	Ph: +65 6812 7888

www.indiumsoftware.com

For Sales Inquiries
sales@indiumsoftware.com

For General Inquiries
info@indiumsoftware.com