A Guide to Using Data from Johns Hopkins Epic Electronic Health Record for Behavioral, Social and Systems Science Research

Phase II:

Identifying Research Needs and Assessing the Availability of Behavioral/Social Variables

Authors:

- Hadi Kharrazi, *MD PhD* [1,2]
- Elham Hatef, *MD MPH* [1,2]
- Elyse Lasser, *MS* [1]
- Bonnie Woods, *MS* [2]
- Masoud Rouhizadeh, *PhD* [2]
- Julia Kim, *MD MPH* [2]
- Lisa DeCamp, *MD MSPH* [2]

[1] Johns Hopkins School of Public Health
[2] Johns Hopkins School of Medicine

Corresponding author:

Hadi Kharrazi
kharrazi@jhu.edu
Research Director & Assistant Professor
624 N Broadway, #606, Baltimore, MD 21205
Office: 443.287.8264 | Fax: 410.955.0470

Prepared for The Behavioral, Social, and Systems Science Translational Research Community
(Co-Leads Felicia Hill-Briggs, PhD, ABPP and Jill A. Marsteller, PhD, MPP)
of the Johns Hopkins Institute for Clinical and Translational Research (ICTR).

Sep 2018

Table of Content

- Executive Summary 3
  - Identifying Research Needs 3
  - Availability of SBDH in Epic's EHR 3
  - Overall Recommendations 5
- Appendix (A): BSSS Social-Epic Project – Phase II Tasks 6
- Appendix (B): Provider Perspectives 7
  - Introduction 7
  - Methods 7
  - Results 8
  - Recommendations and Next Steps 10
  - References 10
  - Interview Guide/Questions 11
- Appendix (C): Mockup of CCDA's Website 13
  - Socio-Behavioral Determinants of Health (SBDH) Data Catalog 13
  - Project Executive Summary 14
  - Data Availability Timeline 15
  - Data Variables 16
  - Address/Zip Code 17
  - Alcohol Use 19
  - Ethnicity 20
  - Income/Financial Issues 22
  - Housing Issues 24
  - Language 29
  - Race 30
  - Smoking Status 33
  - Social Support 35
  - References 40
- Appendix (D): NLP Application 42
  - Methods 42
  - Findings and Recommendations 42

EXECUTIVE SUMMARY

In 2017, the Johns Hopkins Institute for Clinical and Translational Research's (ICTR) Behavioral, Social and Systems Science (BSSS) advisory board – Translational Research Community (TRC) – funded this project to examine the availability of social and behavioral data (also known as, Socio-Behavioral Determinant of Health, or SBDH) within the Johns Hopkins Medical Institute's (JHMI) Electronic Health Record (EHR) system (i.e., Epic Health Record System).

In the first year of this project (phase-I), the project team developed a guide that can be used by JHU researchers to understand: (1) different types and frameworks of social and behavioral data; (2) learn from current and previous attempts to extract social/behavioral data from EPIC at JHMI: and (3) explore some aspects of the common social and behavioral data captured in Epic. Please see the A Guide to Using Data from EPIC, MyChart, and Cogito for Behavioral, Social and Systems Science Research for more information.

The second year of this project (i.e., phase-II) focused on two practical components: (1) further identifying the needs of JHMI's research community in collecting, capturing, accessing and using SBDH from Epic; and, (2) assessing the availability and completeness of the SBDH data in JHMI's Epic. See Appendix A for additional details of phase-II's tasks and deliverables.

- **Identifying Research Needs**

Interviews and focus groups were conducted to identify the research needs of JHU research community for collecting and accessing SBDH data in Epic. Common themes emerged around facilitators and barriers to accessing and using SBDH variables in Epic. Interviewed researchers had various wish lists of SBDH variables. All interviewees stated that their use of Epic to use and collect SBDH variables was limited. A major barrier to access SBDH variables was identified as unfamiliarity with how SBDH variables are captured in both the clinical workflow and Epic's EHR. Interviewee recommendations included: (1) Providing a standard approach to collect basic SHoH variables; (2) Conduct a university-wide survey to prioritize strategies for SBDH data collection; and, (3) Provide timely feedback on SBDH data collection practices to providers. See Appendix B for additional details of "provider perspectives".

- **Availability of SBDH in Epic's EHR**

The CCDA, collaborating with BSSS study team, analyzed Epic's clinical data to determine the existence, completeness rate, collection time range, and collection location of key SBDH variables for patients recorded in Epic[1].

For the purposes of this project, the following data variables were analyzed: patient address; race; ethnicity; alcohol use; smoking status; housing issues; financial issues; preferred language; and, social support.

Three methods were considered to extract SBDH from Epic's EHR: (1) analyzing the structured Epic data (both coded and custom flowsheets); (2) analyzing the unstructured data of Epic; and potentially, (3) generating SBDH variables based on patient addresses.

[1] Work began in April 2018 and ended on June 30, 2018 when the funding expired. The CCDA logged 472 hours on the project.

(1) Extracting SBDH from Structured Epic Data (including flowsheets)

The study team provided the prioritized list of SBDH data variables for inclusion. The study team also included ICD9&10[2], SNOMED[3] and LOINC[4] codes commonly associated with each of the identified SBDH variables, along with screening tools to assist in identifying matching data elements in Epic. Only ICD codes were used in this study as SNOMED and LOINC codes are not commonly collected at JHMI.

The CCDA and the study team developed a series of data collection metrics to capture information of interest to investigators, including:

- Most common collection method (structured [encoded or flowsheet] vs. unstructured)
- General completeness rate
- Collection date range
- Facility type (inpatient, outpatient)
- Provider type (physician, nurse, social worker, case manager)

For the data elements captured in Epic-provided data fields or existing in questionnaires, the CCDA queried SQL database fields to find mentions of the data variable (e.g., "race" or "housing") and recorded the findings in a spreadsheet. The SQL code created for the project was saved in the CCDA code repository so that it can be re-run on demand and re-used for other projects. The spreadsheet of metrics also included supporting patient counts by encounters, locations, and providers, along with detailed metadata for each variable, each questionnaire (group of questions) and each question.

For data variables associated with ICD-coded diagnoses, Epic's SlicerDicer tool was utilized to return counts of unique patients. These queries were also saved for future re-use.

See Appendix C for the "mockup" of CCDA's website that includes details on SBDH completeness rates using structured data (including both encoded fileds and flowsheets/survey data types).

(2) Extracting SBDH from Unstructured Epic Data (free-text) using NLP

Using natural language processing (NLP) techniques, keyword phrases indicating specific SBDH were extracted from clinic notes. We used hand-crafted linguistic patterns, developed by experts, utilizing sources like ICD10, SNOMED, LOINC, and Public Health Surveys, focusing on "Housing", "Finance", and "Social Support" domains. For example, based on the manual evaluation of NLP findings, the system could capture "homelessness" correctly about 85% of the time.

CCDA identified Clarity (Epic's SQL reporting database) as the best source for free text notes, and established a well-performing text processing pipeline using the SAFE virtual desktop and Jupyter Notebook within the PMAP environment. The developed pipeline and identified text sources can be modified and re-used for other projects.

See Appendix C for the "mockup" of CCDA's website that includes details on SBDH completeness rates using unstructured data (NLP results) for three domains of housing, finance and social support.

---

[2] International Classification of Diseases (version 9 and 10)
[3] Systematized Nomenclature of Medicine
[4] Logical Observation Identifiers Names and Codes

See Appendix D for additional details on the "NLP approach" used to extract SBDH from Epic's unstructured data (i.e., free-text).

(3) Generating SBDH using Patient's Address from Epic

The study team generated SQL tables that can be queried to extract Census-derived SBDH using latitude and longitude of the patients' residence. The study team will continue working with CCDA to generate Census-derived SBDH variables using JHMI patient addresses (this process was postponed due to data access complexities).

• Overall Recommendations

SBDH data are becoming an integral part of population health management efforts (e.g., value-based care) as well as research activities (e.g., ensuring diversity). Our findings show the lack of systematic efforts in collecting, encoding, and capturing SBDH in Epic at JHMI.

To our knowledge, this project is the first attempt by JHMI to provide an investigator-friendly data catalog of Epic variables to consult for the purpose of research. Technical data dictionaries exist for discretely captured data elements (e.g., race, ethnicity, preferred language, alcohol use), but the full scope of data availability is only attainable by considering all sources of data, including questionnaires and mentions in clinic notes.

Future funding should target:

(1) *JHMI-wide SBDH Questionnaire*: conducting an institution-wide questionnaire targeting all researchers and clinicians on their needs of SBDH data (not only Epic needs);

(2) *Epic SBDH Questionnaire Registry*: the creation of a report to capture data about existing Epic questionnaires, the questions associated with the questionnaires, and the answers to the questions;

(3) *Generate Population-level Geo-driven SBDH*: complete the geo-derived SBDH data attachment using patient addresses for the entire JHMI population denominator; and,

(4) *Publication and Seminars*: disseminate and share the findings internally at JHMI, with other academic medical centers, and the larger behavioral/social sciences community (e.g., manuscripts, webinars, seminars, and ICTR/BSSS workshops).

▪ APPENDIX (A): BSSS SOCIAL-EPIC PROJECT – PHASE II TASKS

Task	Status	Deliverables
• **Finalize FY1's report based on feedback received** o Feedback from CCDA o Feedback from individual faculty members o Feedback from BSSS members	Completed	FY1 (phase-I) report
• **Disseminate Y1's report** o Post FY1's report on CCDA/ICTR's website o Convert the report into a web-based resources	Completed	CCDA's website
• **Assess needs of researchers interested in SBDH** o Interview JHU faculty/researchers o Interview non-JHU researchers/executives	Completed	Interview Report (Appendix B)
• **Expand the technical resources for SBDH** o Maintain a data catalogue of SBDH ▪ Individual level using structured EHR fields/data ▪ Individual level using unstructured data (free-text) ▪ Geographical/aggregate level SBDH using location ▪ List the data quality specs of SBDH variables o Develop technical documents to use Epic's SBDH o Catalogue new SBDH variables using existing codes	Completed	CCDA's SBDH data catalogue website (Appendix C) Technical documents; and, SQL&NLP code (Appendix D)
	Geo-analysis in-progress	*Geo-Census tables (pending)*
• **Disseminate FY2's report** o Update the online data catalogue o Publish the new findings in peer-reviewed venues	Completed	CCDA SBDH catalogue (Appendix C)
	Manuscripts in-progress	Authors assigned but may require additional funding to complete

▪ APPENDIX (B): PROVIDER PERSPECTIVES

Elyse Lasser, Julia Kim, Lisa DeCamp

• Introduction

Increasing understanding on how social and behavioral data may affect health outcomes have put an increase focus on how these variables are collected and utilized in the electronic medical records (Gottlieb 2016). Socio-behavioral determinants of health (SBDH) encompass both individual level information such as education, income, and employment status and environmental information such as neighborhood walkability and air quality (Cantor 2018). Collecting SBDH information is challenging. Electronic health records (EHR) are not set up to capture information in a standard way, there are no clear clinical standards for capturing SBDH information and documenting the information is not part of the current clinical workflow (Cantor 2018; Kharrazi 2017). Research is only starting to show the impact that SBDH have on health outcomes, so it is important to understand what variables are needed for research and how those variables are collected at point of care. Collecting individual and community level SBDH information have separate challenges.

To begin to understand the complex issue of how individual SBDH information is captured at Johns Hopkins and how those data can be used for research or clinical purposes, we conducted interviews of faculty and staff to discuss and obtain feedback on what social and behavioral variables in Epic have been used, how providers are documenting the information, and how the information may be better documented in the future. The focus of this phase was to inform the team on how to improve access to Epic data at Johns Hopkins for the purposes of behavioral and social sciences research.

The goals of this task were to: (1) Understand the barriers and facilitators to accessing SBDH data in EPIC, and (2) Understand the reasons and needs for accessing the social and behavioral data for research and for clinical care.

• Methods

We contacted 15 researchers and clinicians who are known to have an interest in or are currently using social and behavioral information in their work. We conducted 4 one on one interviews and 2 focus groups. Focus groups contained 5-8 people in each. Participants were from a variety of departments including pediatric and adolescent health, pulmonology, general internal medicine and gastroenterology. Snowball sampling was done to help identify more participants.

We conducted semi-structured interviews that lasted approximately 45-60 minutes. The interview guide was adapted after one focus group to provide more information and clarity around the goals of this project. A copy of the interview guide can be found at the end of this appendix/report. Participants were promised anonymity to allow for open and honest communication about their experiences both positive and negative with utilizing or considering use of Epic a source for SBDH data.

Participants were asked to identify facilitators and barriers to collecting and accessing social and behavioral variables from Epic for research and clinical care. We also elicited more information around the reason these social and behavioral variables are used and how they may be collected in a more comprehensive manner in the future.

- **Results**

Common themes emerged around facilitators and barriers to accessing and using SBDH variables in Epic. A review of the notes from the interviews in phase I revealed similar themes suggesting that thematic saturation on this topic has been reached. Clinicians, who also conducted research, indicated they had different needs to use social and behavioral variables based on if they were performing clinical care or a research project. Type of SBDH variables requested also varied by clinician and the type of research making it difficult to get consensus on how to modify the EHR or modify provider behavior to collect the information which would improve data access in the future.

Overall, Epic has been designed for patient care and respondents posited whether expanding the focus of Epic as a source of socio-behavioral data for research presents a competing interest. Competing interests include the burden of collecting high fidelity, standardized data that would be appropriate for research. At present, the information collected is not consistent and there is no standard way to check accuracy or capture all needed social and behavioral variables. Most providers believe adding more information to be collected in structured manner would increase the clinical burden on themselves or staff. In addition, the clinicians presented a strong preference for patient-reported data but were unclear how to standardize that or ensure that data collected are patient self-report. Suggestions included ways to create a structure that ensures patient self-report of SBDH to help decrease the burden on clinical care providers and could provide researchers with more consistent data quality because they would know who provided the information and how it was collected. Detailed information gleaned from the interviews and categorized using themes extracted from the interviews/focus-groups can be found below:

Common Variables

Interviewed researchers had various wish lists of SBDH variables. Their needs varied based on the type of patients they were providing care for, the type of provider they are, or if they were conducting research. For example, many academic pediatricians wanted to know more information about school absences, care givers, and foster care information. General internal medicine and primary care physicians were interested in behavioral information such as smoking status, physical activity, nutrition, food insecurity, and drug use.

Barriers to Access and Use

All interviewees stated that their use of Epic to use and collect SBDH variables was limited. Social and behavioral variables are not consistently captured and those that are captured are not often captured in a structured manner, requiring researchers to conduct manual reviews of notes to pull all required variables. This is an issue affecting many different health systems and not limited to Johns Hopkins Medical Institute (Senteio 2018).

Another major barrier to access SBDH variables is unfamiliarity with how SBDH variables are captured in both the clinical workflow and Epic's EHR. The Epic EHR's fields that are stored in a structured field (e.g., gender, age, race, smoking status) may be captured by a patient reporting the information, a medical assistant recording what they see, or a combination. It is not known how the information is collected and therefore the researcher who needs the data does not often trust the accuracy of the data. For more complex measures such as physical activity, food insecurity, and

behavioral information such as depression, there are a number of different ways to capture the information such as hours of activity in a week, month or day or being asked if a patient has access to fresh food or if they have enough money to buy food by the end of the week. The variability of approaches to capture SBDH variables makes it difficult for researchers to utilize EHR-derived/extracted SBDH information in a meaningful way.

It is often difficult for clinicians to understand where information should be captured. There is a social history table in Epic, but providers have found it difficult to click between various flowcharts to capture everything in their structured fields. Most clinicians primarily documented these types of variables in the free-text notes of Epic. In some cases, providers did not document specific SBDH variables to improve the patient and provider relationship. One provider reported a patient not wanting their drug-use to be documented in the EHR for fear it would be seen by other providers when it is not necessary for them to know. Others suggested that some sensitive information needs to be documented in code or specific ways in order to protect their patients from retaliation or abuse when the record is viewable to others such as a parent or caregiver.

Facilitators to Access and Use

Researchers expressed some positive experience accessing SBDH variables in Epic. Epic can provide researchers with information to identify patients that would fit their study through specific clinical inclusion and exclusion criteria. Information to contact patients for study recruitment has been successful using data pulled from Epic. Other SBDH data are often captured through patient surveys and questionnaires or collected for specific studies. SBDH surveys are usually limited to individual studies and are not conducted for the entire JHMI patient population. Thus, survey-based SBDH are often collected just for the patients enrolled in limited studies and are often not integrated into Epic automatically. Researchers who have utilized ICTR-CCDA services found the services useful and the team very knowledgeable. For example, the CCDA was able to pull Epic's SBDH data needed for a study including patients contact information, allowing the researcher to consent patients more efficiently.

Uses of Social and Behavioral Variables

Several approaches are used by the researchers and administrators to use SBDH variables (extracted from various data sources). For research, the data found in Epic has been used to identify patients who would qualify for specific interventions or studies. Contact information has then been used to inform patients and register them in various studies. In some cases, SBDH variables have been used to conduct trend analysis and report on clinical performance.

Clinicians have expressed the need for SBDH information to draw a complete picture of their patients. Social and behavioral information can be very useful to help identify barriers to care or provide an understanding on how a patient is handling a treatment plan (i.e., losing weight). These variables may also be very useful for care management and health assessment. SBDH variables, collected over a longer timeframe, may help clinicians to identify patients who are "going off-track" of their treatment plan.

There are no incentives for providers to capture the information so at this time how it can be used for research and clinical care is highly speculative. Clinicians capture information that is important to them

and their interests and their patient population without necessarily thinking how this information can also be used for research (e.g., to understand the health outcomes of individuals and populations).

- **Recommendations and Next Steps**

The interviews in phase I and II of this project offered a clear image that there are significant gaps in the type of SBDH variables collected in Epic, how they are collected, and how they should be utilized. Clinicians and researchers often have competing interests in their needs, making it difficult for Epic to fit all needs. A few recommendations for next phases to help address some of these issues are:

- Provide a standard/streamlined approach to collect basic SBDH variables: Most clinicians and researchers would rather have SBDH information being collected from patients directly, but it is often unclear how the information is collected in Epic. Creating a standard method to collect the data (either at patient check-in or through a web portal such as MyChart) will alleviate the ambiguity of how data has been collected and provide a way to prepopulate specific variables in the EHR.
- Conduct a university-wide survey to prioritize strategies for SBDH data collection: Palacio et al conducted a survey to understand provider perspectives on collecting SBDH information. Analysis found that there were some differences between the perspectives of non-minority and minority providers and the need and use for collecting SBDH. The study also dove into concerns about collecting the information and how the collection should be prioritized. A similar survey may help JHMI fully understand how best to proceed to capture the information needed by both clinical providers and researchers.
- Provide timely feedback on SBDH data collection practices to providers: Some providers suggested if they were provided a "report-card" on how often they are collecting or had incentives to capture and collect SBDH variables, they would be more likely to document the information in the future. It was suggested that documenting the information in free-text notes, rather than structured templates or smart forms, is due to perceived burden, the number of clicks required to do so in the available templates or forgetting to move information from one section to another (e.g., moving from the visit problem list to the main problem list). If providers were given more information on their recording practices it may improve data collection which can then improve access for research.

- **References**

- *Cantor MN and Lorna T. "Integrating data on social determinants of health into electronic health records" Health Affairs (2018): 37(4) 585-590.*
- *Gottlieb L, et al. "Integrating social and medical data to improve population health: opportunities and barriers." Health Affairs (2016): 35(11) 2116-2123.*
- *Kharrazi H, et al. "A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop." J Am Med Inform Assoc. (2017): 24 (1) 2-12*
- *Palacio A, et al. "Provider perspectives on the collection of social determinants of health" Pop Health Manage (2018): Epub.*
- *Senteio C, et al. "Physicians' perceptions of the impact of the EHR on the collection and retrieval of psychosocial information in outpatient diabetes care." Intern J of Med Informatics 113 (2018): 9-16*

- **Interview Guide/Questions**
  - Goal 1: Understand the barriers and facilitators to accessing SBDH data in EPIC
  - Goal 2: Understand the reason/needs to access SBDH data

Domain	Question	Probes
Grand Question	1. Tell me about a recent research project that has incorporated data from Epic?	N/A
What data is requested, accessed, or used?	2. What type of social/behavioral variables have you needed for your work? For your research?	• What has or has not been available to you?
	3. For data that hasn't been easily available in Epic, how have you found it (if at all)? Or how has not having the information impacted your work?	N/A
	4. Do you know how social/behavioral data is collected? Do you personally collect or document the variables in your setting?	• Is it collected at point of care? Intake? Patient reported? • Where is the data stored? (a specific place in an EHR; or a new tool)
	5. When you think about how social and behavioral data is collected in your setting, how does that impact your research?	N/A
How has the data been collected or accessed?	6. Where have you looked for the needed data?	• Do you search the EHR? • Epic data warehouse? • Use a new survey or tool to collect the needed variable?
	7. What has worked well to access the social and behavioral data in Epic?	• Do you work with CCDA? • Do you search for it on your own?
	8. How has the data been structured when you receive it? (to ask the analyst/technical people)	• Is it structured? • From the free-text? • Categorical or yes/no?
	9. What challenges have you faced to access the data?	• Variables requested are not available? • Variables are limited to specific populations? • What do you do if the variables are not available?
Why is the data needed?	10. What is the main reason you search for or use social and behavioral data?	• Is it to identify a denominator population? • Is it for stratification or adjustment purposes • Other reasons?

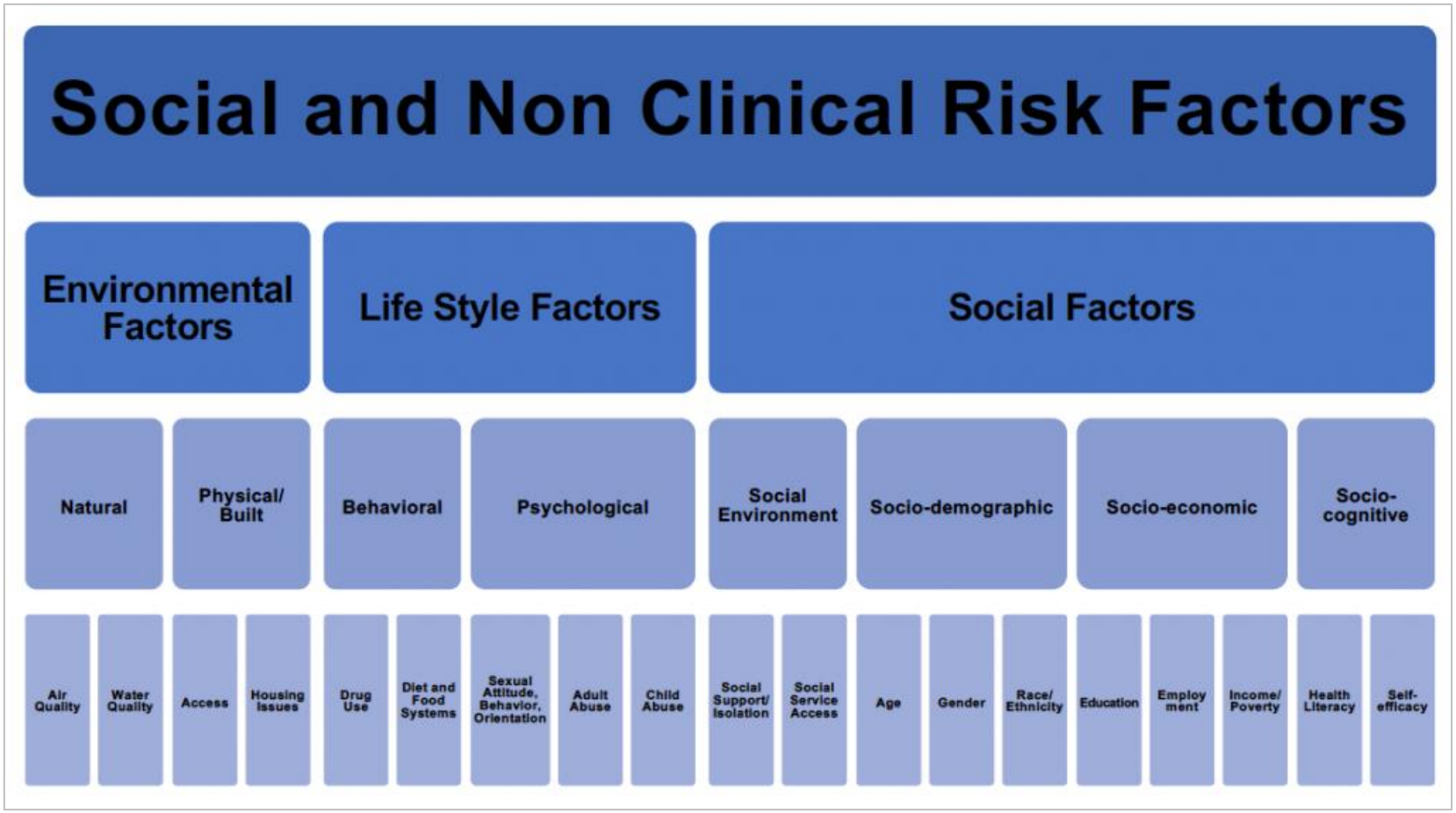
What can it be used for?	11. How else besides research have you used social and behavioral data?	• Clinical care?
	12. How would treatment or care change if you had access to the dream list of social/behavioral variables? (for providers)	• Would the treatment plan change? • Would you prescribe more social services or be able to provide more education material?
	13. Do you see the access to social/behavioral data as a way to improve patient safety? If so how? If not, why or could you see that in the future?	N/A

▪ APPENDIX (C): MOCKUP OF CCDA'S WEBSITE

Data Variables Quick Search:

- **Socio-Behavioral Determinants of Health (SBDH) Data Catalog**

Socio-behavioral determinants of health (SBDH) – a combination of behavioral, social, economic, environmental, and occupational factors – are powerful drivers of morbidity, mortality, and future well-being of individuals and communities, yet they mostly lie outside domain of the conventional medical care delivery system. Modifiable behaviors and exposures that occur in the community play a significant role in 60% of preventable deaths in the U.S. [1,2].



Socio-Behavioral and Non-Clinical Risk Factors[2]

- **Project Executive Summary**

In 2017, the Institute for Clinical and Translational Research's (ICTR) Behavioral, Social and Systems Science (BSSS) Translational Research Community (TRC) advisory board funded this project to examine the availability of social and behavioral data in JHMI's EPIC EMR/PHR systems. Both researchers and administrators recognize that patients' socio-behavioral determinants play a critical role their care experiences and outcomes. Being located in Maryland with its global budgets and population-based reimbursement scheme, it is advantageous for JHU/JHMI to find cost-effective, community-level solutions that improve the population's health status. The vision of the BSSS TRC board is to enable JHU researchers to utilize social and behavioral data collected from JHMI patients and stored in various data sources such as EPIC.

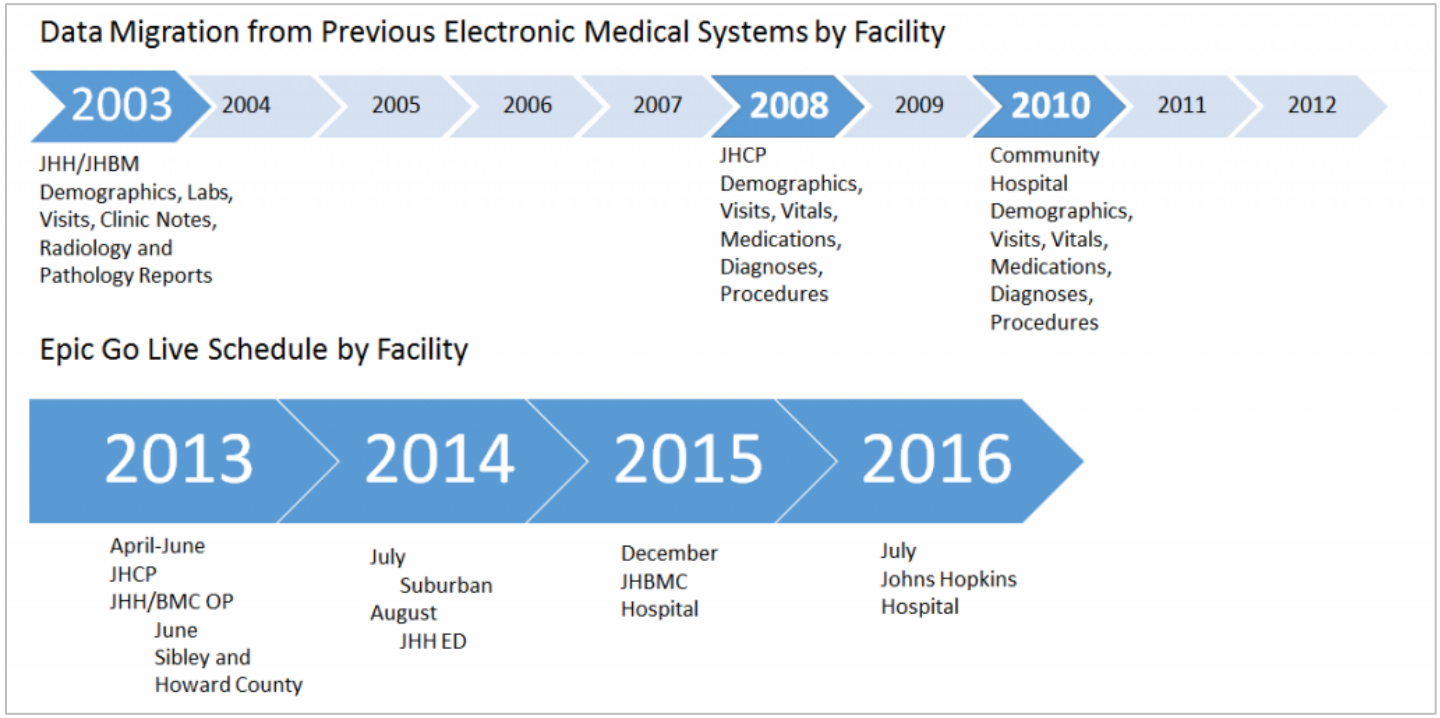
EHR data (i.e., Epic) will play an instrumental role in population health management efforts of value-based providers such as JHMI [3-9]. EHR-extracted socio-behavioral determinants data can potentially help to coordinate care and risk stratification efforts [10]. However, certain challenges will remain while extracting socio-behavior factors from JHMI's Epic such as: data quality issues and missing socio-behavioral data [11-12], ongoing immaturity of EHR's advanced functionality across various providers [13], and the need for complex methods to extract socio-behavioral determinants from EHR's unstructured text [14].

Given the increasing alignment of population and public health efforts [15-18], identifying socio-behavioral factors of high-risk JHMI patients will be key in addressing underlying disparities within JHMI's population as Maryland's all-payer waiver program is already piloting population-level outcome measures [19-20]. Using non-Epic data sources, such as CRISP information (i.e., Maryland's health information exchange), can also be used to extract socio-behavioral determinant data [21].

For the purposes of this project, Epic is the sole data source. No legacy or ancillary systems were queried. As of 7/1/2018, there are approximately 5.4 million unique patients in Epic. The appendix of this report may be reviewed and updated in the future.

- **Data Availability Timeline**

The following diagram depicts the approximate timeline for when data was migrated into Epic or when Epic began capturing data, segmented by data type and facility.



Epic Data Availability Timeline

- **Data Variables**
  - Address/Zip Code
  - Alcohol Use
  - Ethnicity
  - Income/Financial Issues
  - Housing Issues
  - Language
  - Race
  - Smoking Status
  - Social Support

- **Address/Zip Code**

  - ***Most Common Collection Method***

The most common method to collect patient permanent address is upon registration of each encounter. Address is defined as a street address/number, an optional line for apartment or other information, a city, a state or province, and a zip code.

  - ***General Completeness Rate***

95% – out of approximately 5.4 million patients, 5.2 million patients have indicated a permanent address using this collection method.

  - ***Collection Date Range***

Permanent address was migrated into Epic beginning in 2003 and is currently collected.

  - ***Facility Type (Inpatient or Outpatient)***

Permanent address is collected at all facilities at the time of registration.

  - ***Address History***

In 66% of the 5.4 million patients in Epic, there are address change records available. Effective start and end dates are included to track address changes over time. Address change records were migrated into Epic beginning in 2003 and are currently tracked.

  - ***Other Collection Methods***

In addition to the most commonly collected method, patient address is captured in the following cases:

Collection Method	Completeness Rate	Facility/Provider	Dates of Collection
Hospital Account	73% of the 7.3 million hospital accounts	Inpatient facilities, primarily in imaging and PT services at Howard County General and Sibley Memorial Hospital.	2013 – current
Temporary Address	.3% out of 5.4 million patients	All facilities at the time of registration	2003 – current

Billing Address	99% of 2.1 million patient accounts where the guarantor is the patient	All facilities at the time of registration	2003 – current
Claims Processing Address	100% of 4.9 million claims where the patient was listed	N/A	2013 – current
Insurance Coverage Address	92% of 3.1 million coverage records where the subscriber is also the patient	All facilities at the time of registration	2006 – current
Home Health Encounters and Episodes	99% of the 41,604 home care encounters	At the time of the Home Health scheduling encounter	Sep 2015 – current
MyChart Patient Proxy Addresses	79% of the 30,410 MyChart proxy records	MyChart proxies are MyChart users who log in to MyChart on behalf of a patient. Proxies are not patients. Address is not complete for all records.	
Communications for Specific Encounters	100% of the 747,616 encounter-based communications	Clinicians have the ability in Epic to track communications associated with specific encounters. The most common provider types are physicians and medical assistants.	2013 – current
Social Work Assessment Questionnaire: Primary Caregiver Address Question	7% of the 14,129 times that the questionnaire was completed, address was provided.	Sibley and Bayview inpatient emergency services, by social workers	Jul 2014 – current
Social Work Assessment Questionnaire: Patient Address	7% of the 14,129 times the template was used and the question was answered	This questionnaire focuses on the patient's social, economic, safety, and psychological well-being. Captured at inpatient admissions by social workers.	Jun 2014 – current
Maltreatment Assessment: Father's address/Mother's address	7% of the 14,129 times the template was used and the question was answered	JHH Pediatric ED, other pediatric clinics, by social workers.	Apr 2013 – current

- **Alcohol Use**

  - ***Most Common Collection Method***

Alcohol use is captured in the Social History portion of Epic during a patient encounter, whether in-person or non in-person encounters (telephone, MyChart, documentation). For the purposes of this data catalog, the variable of focus is “Alcohol Drinks Per Week”.

  - ***Collection Date Range***

Social History has been collected since April 2013 and is currently collected.

  - ***General Completeness Rate***

Of the approximately 5.4 million patients, 490,348 (.09%) patients reported having any value (including 0 alcoholic drinks per week) in social history. 178,789 (.03%) reported having one or more alcoholic drinks per week.

  - ***Facility Type (Inpatient or Outpatient)***

Alcohol use is collected at all facilities at the time of an encounter.

  - ***Other Collection Methods***

For future revisions of this data catalog, alcohol use data collected from flowsheets, questionnaires, and clinical notes will be included.

- **Ethnicity**

  Epic captures a patient's ethnicity separate from a patient's race, and ethnicity is defined as either Hispanic or Not Hispanic.

  - ***Most Common Collection Method***

  The most common method to collect patient ethnicity is upon registration of each encounter.

  - ***General Completeness Rate***

  50% – out of approximately 5.4 million patients, 2.7 million patients have indicated ethnicity using this collection method.

  - ***Collection Date Range***

  Ethnicity was migrated into Epic beginning in 2003 and is currently collected.

  - ***Facility Type (Inpatient or Outpatient)***

  Ethnicity is collected at all facilities at the time of registration.

  - ***Other Collection Methods***

  In addition to the most commonly collected method, ethnicity is captured in the following cases:

Collection Method	Completeness Rate	Facility/Provider	Dates of Collection
Transplant Organ Donors	50% of the 4,647 human transplant donors in Epic	N/A	Aug 1996 – current
Ethnicity Questionnaire	In 20% of the 2,471 times that the questionnaire was completed, the question "What is your race/ethnicity" was answered. The choices for selection are "Black", "Hispanic", and "Neither Black Nor Hispanic".	Inpatient encounters at Johns Hopkins Hospital by registered nurses and case workers.	Jan 2016 – Oct 2016

Ethnicity Origin Questionnaire	In 5% of the 3,558 times that the questionnaire was completed, the questions "Father's Ethnic Origin" and "Mother's Ethnic Origin" were answered. The choices for selection are: o Caucasian o Hispanic o Black o Asian o Middle Eastern o Other	Two ophthalmology clinics by technicians	Apr 2013 – Sep 2017
Pregnancy/Delivery Episodes: Father's Ethnicity	.1% of the 109,574 combined pregnancy episodes The choices for selection are: o Not Hispanic or Latino o Hispanic or Latino o Unknown o Patient Refused	Two JHCP OB/GYN outpatient clinics	2015 – current

- **Income/Financial Issues**

Patients with income/financial issues are patients in deteriorated financial status, financial hardship, or in poverty. They are unable to afford the basics of life and/or medical interventions. They are in need and eligible for any benefit or enrolment in financial assistant programs.

➢ ***Collection Methods***

In order to determine the population and distribution of patients having income/financial issues, several collection methods were utilized.

Diagnosis Codes (ICD-10)*

Searching Epic for patients having ICD-10 coded diagnoses on the problem list, billing codes, or recorded at the time of an encounter yielded the following results:

ICD-10 Code	Patient Count
Z59.7 Insufficient Social Insurance and Welfare Support	46
Z59.8 Other Problems Related to Housing and Economic Circumstances	3,357
Z59.5 Extreme Poverty	68
Z59.6 Low Income	72

* Patients with ICD-9 coded diagnoses are also included in this query.
The timeframe is January 1, 2003 through June 26, 2018.

- ***Clinical Notes***

Using Natural Language Processing (NLP) techniques, keyword phrases indicating specific SBDH were extracted from 23 million clinic notes authored by specific provider types between July 1, 2016 and May 31, 2018. The notes represent 1,188,202 unique patients and 9,066,508 unique encounters.

We pre-processed the notes to mark the line boundaries and remove non-Unicode characters. To identify notes containing SBDH, we used hand-crafted linguistic patterns developed by a team of experts based on SBDH descriptions in ICD10, SNOMED, LOINC, Public Health Surveys and Instruments (ACS, American Housing Survey, NHANE, etc.), phrases from the literature review and other studies, and manual tagging.

To craft the linguistic patterns, the expert team focused on three domains: Housing, Finance, and Social Support.

Findings for Financial Issues:

- Of the 1,188,202 unique patients 1% had at least one note containing mentions of financial issues.
- Of the 20,219 notes containing mentions of financial issues, about 52% were authored by physicians, 37% were authored by social workers, 8% were authored by nurses, 3% were authored by case managers and case coordinators.
- The top note types for mentions of social support are progress notes (46%), H&P notes (8%), treatment plans (6%), consults (6%), with the remaining note types dispersed among committee review, patient instructions, discharge summaries, and plan of care notes.

- **Housing Issues**

Housing issues are categorized to those related to housing instability or insecurity, homelessness, and characteristics of the house and a number of sub-categories as follows:

- Housing Instability
    1. Having a House with Problems
    2. Prone to Homelessness
- Homelessness
    1. Currently Homeless
    2. Homelessness Being Addressed
- Housing Characteristics
    1. Quality of Building
    2. Characteristics of Building
    3. Building Environmental Health Hazards
        - Air Quality (Including Mold)
        - Infestation
        - Old Paint
        - Hazardous Material and Fire Protection
        - Problems with House Amenities
            - Water
            - Sewage and Disposal
            - Fuel and Heating/ Cooling System
            - Electricity and Internet
            - Age of Building
            - Quality Problems Being Addressed

- ***Collection Methods***

In order to determine the population and distribution of patients having a housing insecurity situation, several collection methods were utilized.

Diagnosis Codes (ICD-10)*

Searching Epic for patients having ICD-10 coded diagnoses on the problem list, billing codes, or recorded at the time of an encounter yielded the following results:

ICD-10 Code	Patient Count
Z59.0 (Homelessness)	7,022
Z59.1 (Inadequate Housing)	120
Z59.8 (Other problems related to housing and economic circumstances)	3,291

* Patients with ICD-9 coded diagnoses are also included in this query.
The timeframe is January 1, 2003 through May 31, 2018.

- ***Clinical Notes***

Using Natural Language Processing (NLP) techniques, keyword phrases indicating specific SBDH were extracted from 23 million clinic notes authored by specific provider types between July 1, 2016 and May 31, 2018. The notes represent 1,188,202 unique patients and 9,066,508 unique encounters.

We pre-processed the notes to mark the line boundaries and remove non-Unicode characters. To identify notes containing SBDH, we used hand-crafted linguistic patterns developed by a team of experts based on SBDH descriptions in ICD10, SNOMED, LOINC, Public Health Surveys and Instruments (ACS, American Housing Survey, NHANE, etc.), phrases from the literature review and other studies, and manual tagging.

To craft the linguistic patterns, the expert team focused on three domains: Housing, Finance, and Social Support. We delved into the housing domain by created patterns for its subdomains, including homelessness, prone to homelessness, homelessness being addressed, housing instability, housing characteristics – characteristics of the building, housing characteristics – the quality of the building.

- ***Findings***
  - Of the 1,188,202 unique patients, 3% have at least one note containing mentions of housing issues.
  - Of the 108,439 notes containing mentions of housing issues, about 53% were authored by physicians, 32% were authored by social workers, 10% were authored by nurses, 5% were authored by case managers.
  - The top note types for mentions of housing issues are progress notes (46%), ED provider notes (13%), H&P notes (8%), with the remaining note types dispersed among discharge summaries, consults, and plan of care notes.
  - Manual annotation of 100 randomly selected notes containing the phrase "homeless" revealed:
    - Out of 100 notes, there are 130 mentions of the word "homeless".
    - In 2 out of 100 notes, there were conflicting true positives and false positives within the same note.
    - In 20 out of 100 notes, there were true negatives (the note was derived from a SmartPhrase), and the answer to Homeless Y/N was No.
  - From the remaining notes, 64% were true positives, and 14% were false positives. The other notes were true negatives or unclear.
  - The following numbers show the count of notes containing specific variables from the housing subdomains:
    - Homelessness Being Addressed: 82
    - Housing Instability: 5,653
    - Housing Characteristics – Characteristics of the Building:148
    - Housing Characteristics – Quality of the Building: 2,011

- ***Questionnaires***

Several questionnaires exist in Epic that capture data related to housing issues. Some are described briefly below, including frequency, general completeness rate, and facility type where the questionnaire answers were completed. Please note that data provided on general completeness reflects questions that were available to an Epic user *and* were answered. The list below is not all-inclusive but represents the majority of questionnaires available as of July 1, 2018.

Questionnaire template	**Question**	**Completeness Rate**	**Facility/Provider**	**Dates of Collection**
Housing/Utility Voucher-Moore Clinic	Housing Assistance Provided?	Of the 217 times the times that the questionnaire was completed, 97 (44%) indicated that housing financial assistance was awarded.	Bartlett HIV Clinics by social workers and case managers	Mar 2017 – current
Abuse/Neglect Screen	Homeless	Of the 12,058 times that the questionnaire was completed, 96% answered the question of homelessness (yes or no).	The top four facilities that answer the question are the infusion center at Sibley Memorial Hospital, the Breast Clinic at Howard County General Hospital, Physical Therapy at Howard County General Hospital, and the Howard County General Wound Center. The top providers completing the form at these facilities are registered nurses, physical therapists, and wound ostomy and continence nurses.	Jun 2013 – current
Screening, Brief Intervention, and Referral to Treatment (SBIRT) Social History	Need Help Finding Housing?	Of the 1,900 times that the questionnaire was completed, 96% answered the question (yes or no).	The top two facilities that answer the question are Bayview Emergency Services and the BMC Chemical Dependency Unit. Peer recovery coaches complete this	Jul 2017 – current

Questionnaire			question.	
ED Triage Abuse Indicators and Resource Planning	Patient indicators, resource planning, and outcomes for shelter, transportation, and clothing	Of the 713,702 times that the questionnaire was completed, 5.5% answered the questions.	Inpatient, mostly in emergency units, by registered nurses.	May 2013 – current
Sibley Chemical Dependence Unit Admission Screen	Homeless	Of the 15,056 times that the questionnaire was completed, 15% answered the question.	The top facility that answers the question is the Sibley Memorial Hospital Clinical Decision Unit, by registered nurses.	Jul 2014 – present
Ambulatory Priority Access Primary Care (PAPC) Screen	What is Your Housing Situation?	Of the 1,116 times that the questionnaire was completed, 7% answered the question.	JHCP Internal Medicine EBMC is the only facility that captured this information, by physicians and medical assistants	Apr 2015 – May 2017
Adult Admission General Intake Form	Homeless	Of the 77,230 times that the questionnaire was completed, 35% answered the question (yes/no).	Top facilities where the template was used are inpatient units at Sibley Memorial Hospital and Howard County General Hospital, by registered nurses.	May 2013 – April 2016
Pediatric/Newborn General Intake Form	Homeless	Of the 1,067 times that the questionnaire was completed, 55% answered the question (yes/no).	Top facilities are inpatient units at Johns Hopkins Hospital and Bayview Medical Center, by social workers.	2016 – current
JHM Psychiatry Social Work Assessment	Living Arrangement	Of the 4,913 times that the questionnaire was completed, 90% answered the question.	The top facility is BMC inpatient Psychiatry, by social workers.	Sep 2015 – current

- **Language**

➢ ***Most Common Collection Method***

Patient preferred language is captured in Epic at the time of admission.

➢ ***Collection Date Range***

Preferred language was migrated into Epic beginning in 2003 and is currently collected.

➢ ***General Completeness Rate***

Of the approximately 5.4 million patients, 2,718,416 patients (49%) have indicated a preferred language in Epic. Of patients indicating a preferred language, 2,624,122 indicated English as their preferred language.

The top 7 preferred languages, by unique patient count are as follows:

- No language reported (null) - 2,804,973 (51%)
- English - 2,626,379 (48.5%)
- Spanish - 53,446 (1%)
- Arabic - 7,317 (.1%)
- Unknown (a valid value in Epic, different from an empty record) - 5,936 (.1%)
- Chinese (Mandarin) - 4,036 (.1%)
- Korean - 3,168 (.1%)

➢ ***Facility Type (Inpatient or Outpatient)***

Preferred language is collected at all facilities at the time of an encounter.

➢ ***Other Collection Methods***

For future revisions of this data catalog, preferred language collected from flowsheets, questionnaires, and clinical notes will be included.

- **Race**

➢ ***Most Common Collection Method***

The most common method to collect patient race is upon registration of each encounter. Patients can select multiple races to self-identify:

- American Indian or Alaska Native
- Asian
- Black or African American
- Declined to Answer
- Native Hawaiian or Other Pacific Islander
- Other
- Unknown
- White or Caucasian

➢ ***General Completeness Rate***

90% – out of approximately 5.4 million patients, 4.9 million patients have indicated at least one race using this collection method.

➢ ***Collection Date Range***

Race was migrated into Epic beginning in 2003 and is currently collected.

➢ ***Facility Type (Inpatient or Outpatient)***

Race is collected at all facilities at the time of registration.

➢ ***Other Collection Methods***

In addition to the most commonly collected method, race is captured in the following cases:

Collection Method	Completeness Rate	Facility/Provider	Dates of Collection
Home Health	84% of the 40,301 Home Health episodes in Epic.	For Home Health encounters only, patients can select multiple races to self-identify. The selections differ from the most common collection method: ● American Indian or Alaska Native ● Asian ● Black or African-American ● Hispanic or Latino ● Native Hawaiian or Pacific Islander ● White	2004 – current
BCRA Race	.05% of the 113 million BCRA encounters	Race is also collected at outpatient and inpatient encounters at breast imaging, ultrasound, and MRI clinics throughout the institution. One race per patient is selected, and race selection is restricted to the following: ● White ● African American ● Hispanic ● Chinese American ● Japanese American ● Filipino American ● Hawaiian ● Pacific Islander ● Other Asian American	May 2013 – current
Transplant Organ Donors	44% of the 4,602 organ donors	Multiple races are captured for each donor from the same selections as available during the patient registration process.	1996 – current

Race Questionnaire	7.4% of the 1,748 times the questionnaire was used at an HIV clinic encounter	Race is documented for patients seen in the JHH HIV Clinics, specifically collected by social workers and case managers. Only one race can be selected. Selections are restricted to the following: ● White ● Black or African American ● Asian ● Native Hawaiian or Other Pacific Islander ● American Indian or Alaska Native ● Other (Comment)	April 2013 – current

- **Smoking Status**

➢ ***Most Common Collection Method***

Smoking status is captured in the Social History portion of Epic during a patient encounter, whether in-person or non in-person encounters (telephone, MyChart, documentation). For the purposes of this data catalog, the variable of focus is "Smoking Status".

➢ ***Collection Date Range***

Smoking status has been collected since April 2013 and is currently collected.

➢ ***General Completeness Rate***

Of the approximately 5.4 million patients, 1,728,749 patients (32%) reported having any value smoking status in social history. Smoking Quit Date is also populated, but only in 137,958 (3%) of the time. The status breakdown, with collection rate, is as follows:

Smoking Status selection options include:

- Current Every Day Smoker (114,566 – 2.1%)
- Current Some Day Smoker (28,547 – .5%)
- Former Smoker – (297,099 – 5.5%)
- Heavy Tobacco Smoker (3,111 – .1%)
- Light Tobacco Smoker (12,857 – .2%)
- Never Assessed (302,631 – 5.6%)
- Never Smoker – (952,636 – 17.7)
- Passive Smoke Exposure – Never Smoker – (4,274 – .1%)
- Smoker, Current Status Unknown – (1,133 – < .01%)
- Unknown if Ever Smoked – (11,915 – < .2%)

➢ ***Facility Type (Inpatient or Outpatient)***

Smoking status is collected at all facilities at the time of an encounter.

- ***Other Collection Methods***

  For future revisions of this data catalog, alcohol use data collected from flowsheets, questionnaires, and clinical notes will be included.

- **Social Support**

"Lack of Social Support" (Social Isolation) or "At Risk for Social Isolation" is referred to patients in deteriorated aloneness state with lack of interaction with others, feel lonely, detached, and isolated with no help and support system. They have restricted social participation and can't maintain social relationships. They are unable to communicate with others and have difficulty visiting friends, attending clubs, meetings, and going to parties. They might be the target of perceived adverse discrimination or persecution.

- ***Collection Methods***

In order to determine the population and distribution of patients lacking social support, several collection methods were utilized.

Diagnosis Codes (ICD-10) *

Searching Epic for patients having ICD-10 coded diagnoses on the problem list, billing codes, or recorded at the time of an encounter yielded the following results:

ICD-10 Code	Patient Count
R45.8 Other symptoms and signs involving emotional state	3,340
Z60.4 Social Exclusion and Rejection	223
Z60.2 Problems related to living alone	1,222
Z63.0 Problems in relationship with spouse or partner	852
Z63.5 Disruption of family by separation and divorce	548
Z63.8 Other specified problems related to primary support group	2,230
Z63.9 Problem related to primary support group, unspecified	3,247
Z65.9 Problems related to unspecified psychosocial circumstances	938

Z73.4 Inadequate social skills, not elsewhere classified	81
Z91.89 Other specified personal risk factors, not elsewhere classified	18,947

* Patients with ICD-9 coded diagnoses are also included in this query.
The timeframe is January 1, 2003 through May 31, 2018.

- ***Clinical Notes***

Using Natural Language Processing (NLP) techniques, keyword phrases indicating specific SBDH were extracted from 23 million clinic notes authored by specific provider types between July 1, 2016 and May 31, 2018. The notes represent 1,188,202 unique patients and 9,066,508 unique encounters.[3]

We pre-processed the notes to mark the line boundaries and remove non-Unicode characters. To identify notes containing SBDH, we used hand-crafted linguistic patterns developed by a team of experts based on SBDH descriptions in ICD10, SNOMED, LOINC, Public Health Surveys and Instruments (ACS, American Housing Survey, NHANE, etc.), phrases from the literature review and other studies, and manual tagging.

To craft the linguistic patterns, the expert team focused on three domains: Housing, Finance, and Social Support.

- ***Findings***

  - Of the 1,188,202 unique patients 2.6% have at least one note containing mentions of social support.
  - Of the 63,185 notes containing mentions of social support, about 70% were authored by physicians, 15% were authored by nurses, 13% were authored by social workers, 2% were authored by case managers and case coordinators.
  - The top note types for mentions of social support are progress notes (46%), H&P notes (8%), treatment plans (6%), consults (6%), with the remaining note types dispersed among committee review, patient instructions, discharge summaries, and plan of care.

- ***Questionnaires***

Several questionnaires exist in Epic that capture data related to social support. Some are described briefly below, including frequency, general completeness rate, and facility type where the questionnaire answers were completed. Please note that data provided on general completeness reflects questions that were available to an Epic user *and* were answered. The list below is not all-inclusive but represents the majority of questionnaires available as of July 1, 2018.

Questionnaire template	Question	Completeness Rate	Facility/Provider	Dates of Collection
Nursing Assessment – Psychosocial	Psychosocial (WDL)	Of the 1,026,988 times that the questionnaire was completed, 92% answered the question.	Top units are inpatient units at Howard County General and Suburban Hospital by registered nurses, licensed practical nurses, and case managers	May 2013 – current
ED Assess Head to Toe	Psychosocial (WDL)	Of the 237,143 times that the questionnaire was completed, 39% answered the question.	The top site is the inpatient SH Clinical Decision Unit, Suburban Hospital by registered nurses	Oct 2013 – current
T AD ED Nursing Assessment	Psychosocial (WDL)	Of the 217,954 times that the questionnaire was completed, 94% answered the question.	The top units are emergency medicine inpatient units at Johns Hopkins Hospital and Bayview Medical Center by registered nurses	Dec 2015 – current
T CD ED Nursing Assessment	Psychosocial (WDL)	Of the 278,084 times that the questionnaire was completed, 61% answered the question.	The top units are emergency units at Howard County General Hospital, Sibley Memorial Hospital, and Suburban Hospital by registered nurses	Dec 2015 – current
T AD/CD ED PEDS Assessment	Psychosocial (WDL)	Of the 131,134 times the questionnaire was completed, 71% answered the question.	The top units are pediatric emergency departments at Johns Hopkins Hospital, Bayview Medical Center, Sibley Memorial Hospital, and Howard County General Hospital by registered nurses	Dec 2015 – current

T JHM OR AD PACU Flowsheet	Psychosocial (WDL)	Of the 147,694 times the questionnaire was completed, 56% answered the question.	Inpatient post-anesthesia units at Johns Hopkins Hospital by registered nurses	Sep 2015 – current
JHM IP OT New Home Setup	Social support available at discharge	Of the 131,948 times the questionnaire was completed, 36% answered the question.	Inpatient orthopedic and surgical floors at Bayview Medical Center and Suburban Hospital by occupational therapists	Jun 2015 – current
IP OB Postpartum	Recent loss, or change in status? (Includes loss of social status, job, divorce, death, demotion, etc.) / Has multiple friends/family members with several close confidants?	Of the 135,587 times the questionnaire was completed, 89% answered the question.	The top units were Howard County General and Sibley Memorial inpatient units by registered nurses	May 2013 – current
IP Spiritual Care Interventions	Spiritual/Social Network	Of the 116,719 times the questionnaire was completed, 59% answered the question.	The top units are Howard County General and Sibley Memorial inpatient units by chaplains.	May 2013
PED Screening	Personal-Social / 22. Relationship with peers / Is there a recent stress on the family or child such as birth of a child, moving, divorce or separation, death of a close relative / 23. Relationship with parents / 24. Relationship with siblings /	Of the 144,659 times that the questionnaire was completed, 59% answered the question.	The top clinics are outpatient JHCP pediatric clinics by medical assistants, physicians, and nurse practitioners	Apr 2013 – current

	25. Relationship with peers / Socially withdrawn-decreased interaction with others / Socially withdrawn-decreased interaction with others			
JHM SBIRT Social History	Marital Status / Need to improve relationships with family? / Social Network / Social Network Comments / Participation in other social activities? / Social Activities Comments	Of the 2,015 times that the questionnaire was completed, 99% answered the question.	Inpatient units at Bayview Medical Center by peer recovery coaches	Jul 2017 – current
JHM ED SW Suicide/Homicide Assessment	Suicide Risk: Single, Widowed, or Divorced / Suicide Risk: No Social Support	Of the 15,101 times that the questionnaire was completed, 97% answered the question.	The top units are inpatient units at Sibley Memorial Hospital and Suburban Hospital by social workers and therapists	Jun 2013 – current
JHM ED Social Work Patient Info	Support System's Name(s) / Support System Contact / Other Support System Information / Support System	Of the 14,481 times that the questionnaire was completed, 88% answered the question.	The top units are emergency departments at Suburban Hospital and Bayview Medical Center by social workers and therapists	Jul 2014 – current

- **References**

1. Bazemore AW, Cottrell EK, Gold R, et al. "Community vital signs": Incorporating geocoded social determinants into electronic records to promote patient and population health. *J Am Med Inform Assoc*. 2016;23(2):407-412
2. McGinnis JM, Williams-Russo P, Knickman JR. The case for more active policy attention to health promotion. *Health Aff (Millwood)*. 2002;21(2):78-93
3. Kharrazi H, Lasser E, Yasnoff WA, Loonsk J, Advani A, Lehmann H, Chin D, Weiner JP. A proposed national research and development agenda for population health informatics: summary recommendations from a national expert workshop. *J Am Med Inform Assoc*. 2017; 24 (1):2-12
4. Kharrazi H, Chi W, Chang HY, Richards TM, Gallagher JM, Knudson SM, Weiner JP. Comparing population-based risk-stratification model performance using data extracted from electronic health records versus administrative claims. *Med Care*. 2017; 55 (8): 789-796
5. Chang HY, Richards TM, Shermock KM, Elder-Dalpoas S, Kan H, Alexander CG, Weiner JP, Kharrazi H. Evaluating the impact of prescription fill rates on risk stratification model performance. *Med Care*. 2017; 55 (12): 1052-1060
6. Lemke K, Gudzune KA, Kharrazi H, Weiner JP. Assessing markers from ambulatory laboratory tests for predicting high-risk patients. *Am J Manag Care*. 2018; 24(6): e190-e195
7. Kharrazi H, Weiner JP. A practical comparison between the predictive power of population-based risk stratification models using data from electronic health records versus administrative claims: setting a baseline for future EHR-derived risk stratification models. *Med Care*, 2017; 56(2), 202-203
8. Kharrazi H, Chang HY, Heins S, Weiner JP, Gudzune K. Enhancing the prediction of healthcare costs and utilization by including outpatient BMI values to diagnosis-based risk models. *Med Care*. 2018; 56 (12): 1042-1050
9. Kan H, Kharrazi H, Leff B, Boyd C, Davison A, Chang H-Y, Kimura J, Wu S, Anzaldi LJ, Richards T, Lasser E, Weiner JP. Defining and assessing geriatric risk and associated health care utilization among elderly patients using claims and electronic health records. *Med Care*. 2018; 56(3): 233-239
10. Hatef E, Searle KM, Predmore Z, Lasser EC, Kharrazi H, Nelson K, Sylling P, Curtis I, Fihn S, Weiner JP. The impact of social determinants of health on hospitalization in the Veterans Health Administration. *Am J of Prev Med*. In-press.
11. Kharrazi H, Wang C, Scharfstein D. Prospective EHR-based clinical trials: the challenge of missing data. *J Gen Intern Med*. 2014; 29 (7): 976-978
12. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013; 20(1): 144–151
13. Kharrazi H, Gonzalez CP, Lowe KB, Huerta TR, Ford EW. Forecasting the maturation of electronic health record functions among US hospitals: retrospective analysis and predictive model. J Med Internet Res. 2018; 20(8): e10458
14. Kharrazi H, Anzaldi L, Hernandez L, Davison A, Boyd CM, Leff B, Kimura J, Weiner JP. Measuring the value of electronic health record's free text in identification of geriatric syndromes. J Am Geriatr Soc. 2018; 66(1) 1499-1507
15. Dixon B, Kharrazi H, Lehman H. Public health and epidemiology informatics: recent research and events. Yearb Med Inform. 2015; 10 (1): 199-206

16. Kharrazi H, Weiner JP. IT-enabled community health interventions: challenges, opportunities, and future directions. Generating Evidence & Methods to Improve Patient Outcomes (eGEMs). 2014; 2 (3): 1-9
17. Gamache R, Kharrazi H, Weiner JP. Public health and population health informatics: the bridging of big data to benefit communities. Yearb Med Inform. 2018; 27(1): 199-206
18. Dixon B, Pina J, Kharrazi H, Gharghabi F, Richards J. What's past is prologue: a scoping review of recent public and global health informatics literature. Online J Public Health Inform. 2015; 7 (2) e1-31
19. Hatef E, Lasser EC, Kharrazi H, Perman C, Montgomery R, Weiner JP. A population health measurement framework: evidence-based metrics for assessing community-level population health in the global budget context. Popul Health Manag. 2017; 21(4): 261-270
20. Hatef E, Kharrazi H, VanBaak E, Falcone M, Ferris L, Mertz K, Perman C, Bauman A, Lasser EC, Weiner JP. A state-wide health IT infrastructure for population health: building a community-wide electronic platform for Maryland's all-payer global budget. Online J Public Health Inform. 2017; 9(3): e195
21. Kharrazi H, Horrocks D, Weiner JP. Use of HIEs for value-based care delivery: a case study of Maryland's HIE. In Dixon B (Ed.) Health Information Exchange: Navigating and Managing a Network of Health Information Systems. 2016; 313-332. Cambridge, MA: Academic Press Elsevier

Notes

Disclaimer about NLP findings: The reported NLP findings are based on the occurrences/mentions of specific linguistic patterns (aka, keywords or key phrases) within a given clinical note. At this stage of the project, we did not perform a manual quality check but we think occurrences of variable-specific patterns could give us a good proxy of the variable in question.

■ APPENDIX (D): NLP APPLICATION

The purpose of this part of the project was exploring the use of natural language processing (NLP) and text mining techniques to determine SBDH from Epic's clinical notes. Keyword phrases indicating specific SBDH were extracted from 23 million clinic notes authored by specific provider types between July 1, 2016 and May 31, 2018. The notes represent 1,188,202 unique patients and 9,066,508 unique encounters.

- **Methods**

We pre-processed the notes to mark the line boundaries and remove non-Unicode characters. To identify notes containing SBDH, we used hand-crafted linguistic patterns developed by a team of experts based on SBDH descriptions in ICD10, SNOMED, LOINC, Public Health Surveys and Instruments (e.g., ACS, American Housing Survey, NHANE), phrases from the literature review and other studies, and manually tagged free-text.

To craft the linguistic patterns, the expert team focused on three domains: Housing, Finance, and Social Support. We delved into the housing domain by created patterns for its subdomains, including homelessness, prone to homelessness, homelessness being addressed, housing instability, housing characteristics - characteristics of the building, housing characteristics - the quality of the building.

- **Findings and Recommendations**
  - *Provider types*: The majority of notes containing SBDH were authored by physicals, followed by nurses and social workers.
  - *Note types*: The top note types for mentions of SBDH include progress notes, followed by H&P notes, treatment plan, and ED provider notes.
  - *Quality of findings*: In this phase of the project, we focused on determining SBDH mentions which could include false positives such as negated forms. By manual annotation of 100 randomly selected notes containing the phrase "homeless", we found 20 true negatives (the note was derived from a SmartPhrase, and the answer to Homeless Y/N was No), and 14 false positives (the patent was not homeless).
  - *Note sources*: We found that some of the text markers such as newline characters are removed from the text blobs in EDW. Clarity, on the other hand, has some of those markers including some of the newline boundaries. In our investigations, we found that Chronicles is the best source to get text metadata.
  - *Processing environment*: We successfully started to use the PMAP's powerful computing environment combined with the SAFE desktop to extract SBDH-related linguist patterns from free texts and post-process the extracted text. We will use this pipeline for future large text processing projects.
  - *Manual annotation process*: We have been working with the Data Trust team to provide a heavily logged environment for domain experts to read and mark up the clinical notes. We have

not used large scale manual annotation in the project specifically, but we have made substantial progress in creating the annotation pipeline.

- *Data discovery for machine learning*: We found valuable SBDH data in the structured fields (mainly flowsheets) to create SBDH-specific patient cohorts. These cohorts can be used for machine learning experiments to automatically extract highly associate linguistic patterns from the notes of specific SBDH cohorts.