



D2 Project Final Report

Prepared for the Asociación de Salud Primaria de Puerto Rico

By Srini Rao, PhD
and Feygele Jacobs, DrPH

May 2021

D2 PROJECT FINAL REPORT

1. EXECUTIVE SUMMARY

RCHN COMMUNITY HEALTH FOUNDATION, INC. (RCHN CHF) is a not-for-profit foundation dedicated to supporting community health centers across America. RCHN CHF magnifies the work of community health centers, developing and facilitating the sharing of important resources, tools, and information, and helping to broaden, in scope and intensity, the essential work of health center providers.

Path2Analytics was a project developed in 2014 by RCHN CHF with the goal of helping community health centers to adopt data-driven decision-making practices. A key part of the project involved assisting health centers in developing the processes and tools to maintain consistent data and productively use contemporary analytics derived from the data. In 2019, the Foundation undertook an extension of this project, referred to as “Level-Up” and later renamed “D2”, with Puerto Rico’s primary care association, Asociación de Salud Primaria de Puerto Rico, Inc. (ASPPR), and its affiliate, Puerto Rico Primary Care Association Network (PR PCAN).

The primary objective of the D2 project was to support the development and reinforcement of a robust data culture at the health centers in Puerto Rico (PR), in collaboration with ASPPR and PRPCAN. Engaging in the continuous use of data on an ongoing basis, by enhancing the use of dashboards and other data summaries, can broaden the access to and appreciation of data across the organizations to all internal stakeholders – analysts, clinicians, management, administrative and front office staff, as well as board members – and ultimately, improve patient outcomes.

In order to support this strategic objective, the project activities focused around two broad areas. The first was to identify strengths, weaknesses, and variations in data collection and documentation practices among the health centers that participated in the study. The second area of activity was to develop sample analytics and dashboards and demonstrate their benefit in clinical, financial, and operational decision-making.

The approach taken to carry out the two broad areas of activity - data quality analysis and development of dashboards - was to extract patient data from the Electronic Health Records (EHR) systems of each of the centers in the form of *Continuity of Care Documents* (CCDs) and ingest it into the Health Gorilla data warehouse platform for analysis and dashboards development.

While past Path2Analytics analyses had more direct access to EHR databases, the D2 project was limited to snapshots of patient data in the form of CCDs created at different times and frequencies (monthly, quarterly) at the three participating centers, providing a non-uniform and more limited view of the data. However, since the goal of the project was not the development of comprehensive quantitative reports, the impact of the data limitations on the overall project was

limited. The available data was sufficient to analyze and highlight data capture and quality aspects, and to illustrate the potential benefits of accurate data dashboards.

A good portion of the encounters were categorized as being Ambulatory/Outpatient, as expected in a health center scenario; encounter type information was not documented in a majority of the patient data records, across all the three centers. Thus, it was not possible to distinguish lab visits, emergency, home or hospital-based ambulatory services.

Analysis of the incidence of common chronic conditions revealed both commonalities and differences among the populations of the centers, with hypertension or obesity being the most common among the three centers. The analysis also highlighted inconsistencies in the configuration of the EHRs that could make it challenging for the providers to consistently document patients' conditions.

Social and economic circumstances are being recognized as influential factors in determining an individual's health risks and treatment outcomes. The availability and utilization of social determinants of health (SDOH) data are essential for developing social and medical interventions to address healthcare inequalities. Social determinants are particularly relevant for Puerto Rico where 34% of adults report fair or poor general health compared to 18% in the 50 states and D.C., and 46% of the population was below the Federal Poverty Level in 2016 compared to 15% in the 50 states and D.C.

Several organizations are working to standardize social health assessment tools and clinical documentation. EHR systems are beginning to support the emerging standardized documentation for capturing the SDOH, and providers are being trained in leveraging these tools for documentation and coding of SDOH. These nationwide trends seem to be mirrored in the partial and incomplete capture of social determinants information in the EHRs of the three CHCs in this study.

In conclusion, while the data quality and status of data capture practices is likely better than what was reflected in the limited data views available for analysis during this study, the D2 project highlights the strengths and differences among the centers in their approaches to handling data. Recommendations are provided to further enhance the quality and uniformity of patient data across all health centers. Best practices and lessons learned from this study are also being shared with all the centers.

2. D2 (LEVEL-UP) PROJECT - BACKGROUND

RCHN COMMUNITY HEALTH FOUNDATION, INC. (RCHN CHF) is a New York based not-for-profit organization, whose mission is to support and benefit community health centers and the patients and communities they serve through strategic investment, grant making research, coalition building, education and outreach. RCHN CHF works to enhance and magnify the community health center program by developing and facilitating the sharing of important resources, tools, and information, and helping to broaden, in scope and intensity, the essential work of health center providers.

Path2Analytics (P2A) was a project developed in 2014 by RCHN with the goal of helping community health centers to adopt data-driven decision-making practices. A key part of the project involved assisting health centers in developing the processes and tools to maintain consistent data and productively use contemporary analytics derived from the data. The project was led by the late Dr. David Hartzband, then the Foundation's Director of Technology Research, and Foundation CEO Dr. Feygele Jacobs, with the technology and analytics developed by Dr. Srinivasa Rao and team at Datycs Inc. using the big data Hadoop platform. Two urban and one rural health center in different states, as well as a large Primary Care Association with more than seventy member health centers, participated in the Path2Analytics project.

In late 2019, the Foundation began an extension of this project, referred to as "Level-Up", with Puerto Rico's primary care association (PCA), Asociación de Salud Primaria de Puerto Rico, Inc. (ASPPR) and its affiliated Health Center Controlled Network, Puerto Rico Primary Care Association Network (PRPCAN). While Puerto Rico's community health centers have performed well with respect to the Quality Improvement Grant Awards administered by Health Resources and Services Administration (HRSA), including recognition for Advancing Health Information Technology (HIT) for Quality, the PCA was determined to accelerate the pace of improvement and help all member organizations create a strong IT foundation. The PRPCAN teamed recognized the limitations inherent in the UDS-focused data cycle, which at best provided only an annual snapshot of information. Accordingly, the project was initiated under the guidance of PCA Executive Director Mrs. Alicia Suarez, and again led by Dr. David Hartzband. Following the sudden demise of Dr. Hartzband soon after the Level-Up project was initiated, Dr. Rao joined the team to lead the project through its conclusion. The project was renamed by the team as the **D2 project** (for David & Data) in fond remembrance of Dr. Hartzband.

The primary objective of the D2 project was to support the development and reinforcement of a robust data culture at the health centers in PR, in collaboration with ASPPR and PRPCAN. Enhancing the use of dashboards and other data summaries can broaden the access to and appreciation of data to all internal stakeholders in the organizations – analysts, clinicians, administrative, front office staff as well as executive leadership and Board members. Ultimately, more robust data and successfully applied analytics tools can pave the way for smoother operations, improved workflow and patient care practices, and better patient outcomes. An additional overarching goal, closely related to the concept of a strong data culture, was to encourage the ongoing and continuous assessment and sharing of data as a means to improve operations and quality, and drive policy.

In order to support these strategic objectives, the project activities focused around two broad areas. The first was to identify strengths, weaknesses, and variations in data collection and documentation practices among the health centers that participated in the study. Recommendations were developed to further enhance the quality and uniformity of patient data across all health centers. Best practices and lessons learned from this exercise are also being shared with all the centers. The second area of activity was to develop sample analytics and dashboards and demonstrate their benefit in clinical, financial, and operational decision-making.

3. APPROACH

PRPCAN is a membership organization in which all of Puerto Rico's community health centers participate. All members were advised of the project and invited to join. The PRPCAN team had preliminary discussions with five Community Health Centers that were interested in participating. A survey document designed to understand the status of information technology (IT) infrastructure and current data handling practices at each center was developed and shared with interested centers. [Refer to Appendix B for a copy of the survey document.]

Three out of the five centers ultimately decided to participate in the project. These will be referred to as CHC 1, CHC 2, and CHC 3 in this report. Each of these centers has received HRSA recognition for Advancing Health Information Technology (HIT) for Quality in FY 2020.

While the initial group of five interested centers had two different Electronic Health Records (EHR aka EMR) vendor systems in use among them, it turned out that the final three participating centers use the same EHR vendor platform. This made the task of data extraction, ingestion into the data warehouse and analysis easier; however, it was also a missed opportunity to understand the differences in data practices, capabilities and limitations across different EHR systems in use at the health centers.

The approach taken to carry out the two broad areas of activity - data quality analysis and development of dashboards - was first, to extract data from the EHR systems of each of the centers and ingest it into a third-party data warehouse platform. After conducting a competitive analysis of various platform choices, the PRPCAN team selected Health Gorilla as the solution that best addressed the needs of the network. A skilled PRPCAN data analyst was assigned to coordinate all aspects of the project, serve as a dedicated resource, and support the engagement of the health centers. The data analysis and dashboard development were subsequently performed on the Health Gorilla platform. Health Gorilla staff supported the analysis and performed the dashboard development during this key phase of the project.

Health Gorilla is a secure interoperability solution that enables patients, payers, providers, digital health solutions, and labs to share health data and aggregate patients' clinical history in a unified view. With enterprise-grade clinical data Application Programming Interfaces (APIs), HIPAA-compliant user authentication, and a master patient index, the Health Gorilla network allows providers to retrieve their patient's information from any clinical records system.

One of the key differences between the earlier Path2Analytics project and the D2 project was in the early step of getting access to and extraction of demographic and clinical data from the EHRs. While the P2A projects had full access to daily backups of the EHR databases, the D2 team was limited to dealing with snapshots of data created at different times and frequencies (monthly, quarterly) at the three participating centers. These snapshots provided a more limited view of the data, due to the nature of the available standards-based mechanism that was used to generate them. The non-uniformity and the limited windows of time that these snapshots covered posed some difficulties during the final stages of analysis and will be described further in the following sections.

The snapshots of data extracted from the EHRs were in the form of *Continuity of Care Documents* (CCDs) (ONC, 2013). CCDs are electronically generated, patient-specific clinical summary documents. The purpose of a CCD is to improve communication between health care providers during a transition of care – when a patient is being referred to another provider or coming back to their primary care provider after a hospital stay, for example. CCDs are among the document templates standardized by HL7 Consolidated Clinical Document Architecture (C-CDA). Other document templates include, for example, Consultation Notes, Discharge Summaries, and Progress Notes.

These document templates together cover various clinical workflow scenarios such as Transition of Care (between Primary Care Provider (PCP) and a Consultant, or Admitting Hospital to a Skilled Nursing Facility), Diagnostic Imaging, and Annual physical visit to a PCP. CCDs as well as other C-CDA document templates can contain both structured and unstructured information; the template format is comprised of a header that defines the patient and the provider, and a structured body. Document headers provide the context that facilitates compilation of a patient’s individual clinical documents into a complete electronic patient record. The body of the CCD incorporates narrative descriptions for presenting problems, procedures, family and social history as well as electronically coded entries for clinical indicators including medications, problems, allergies and immunizations, and lab results. Figure 1 lists other sections that are generally present in CCDs.

The CMS Meaningful Use requirements for EHR certification led to tighter standardization of clinical data and document representations. While managing the CCDs would still be a burden on the analyst and IT staff of the health centers, it is anticipated that this interim necessity of handling CCDs will soon disappear, with the integration of the EHRs at the PR health centers into a common data warehouse platform such as Health Gorilla, using automated APIs for real-time (or near-real-time) data extraction. Fast Health Interoperability Resources (FHIR) (HL7 International, 2019) is the next-generation standard developed by HL7 that offers open APIs for exchange of healthcare data between various systems. FHIR Bulk Data Access API is another that is expected to be adopted as a more convenient and modern alternative to ingest data into a data warehouse platform for the development of dashboard and analytics (Mandl, 2020).

4. ADMINISTRATIVE AND CLINICAL DATA ANALYSIS

This section describes two broad areas of effort - data quality analysis and the development of data dashboards.

As described in the previous section, patient data were extracted from the EHRs in the form of *Continuity of Care Documents* (CCDs) by the staff at the three participating centers and shared with the D2 project team at various times. These CCDs were then ingested into the Health Gorilla platform for analysis and development of dashboards. The generation and management of CCDs involves significant manual intervention; hence care must be taken to institute appropriate processes for their handling.

4.1 Source Data Challenges and Limitations

As noted in the previous section, the formats and transmission method of source EHR data for the D2 project differed significantly from that of the earlier Path2Analytics projects. The P2A projects had direct access to daily backup copies of the EHR databases and leveraged the granularity of direct SQL queries, while the D2 team was limited to dealing with collections of CCD files. These CCDs covered various durations of clinical care - one month per file, one quarter, an entire year, or partial months in some cases.

HRSA Uniform Data System (UDS) defines countable visits (or qualifying encounters) as those that are “documented, individual, face-to-face or virtual contacts between a patient and a licensed or credentialed provider” (Bureau of Primary Health Care, HRSA, 2020). Determination of qualifying encounters forms the basis of calculating several key statistics required for UDS reporting purposes. For example, the patient count for each year at a CHC is the count of patients who had at least one qualifying encounter during that year.

The encounter type field value is one source that can be used to determine whether an encounter is qualified or not, another being a documented CPT code. However as further described in Section 4.2 below, the encounter type information was available only in a small subset of the analyzed CCDs.

Because the time spans covered by the data was not precisely labeled or known, it was not until later phases of analysis, when discrepancies were detected, that the team went back and analyzed data month-by-month and uncovered the issue of varying amounts of missing data across the three centers. For example, for the year 2019, CHC 1 had data missing for November and December, CHC 2 had data missing for part of December, and CHC 3 had data missing for the last five months of the year. One of the centers transitioned from one EHR vendor platform to another in 2018, resulting in the availability of only partial information for that year.

Due to these issues with the available data, the following analysis will focus mostly on qualitative aspects and limit quantitative comparisons, for example to UDS-reported data which is generated directly from EHR data more rigorously by analysts at the centers.

If the time period covered in each (batch) of the CCDs is not managed carefully, there is a potential for overlap of time periods resulting in the duplication of CCDs content such as encounters, diagnoses, medications, and procedures. If the time periods are identical (whether generated inadvertently by the same person or different staff members), or the patients had no visits during the overlap period, the resulting CCDs can be identical.

The total number of CCDs received from each of the centers during 2018 and 2019 ranged from 70,000 to nearly 200,000. The number of duplicates detected ranged from a few hundred in one of the centers to a few thousand at another, and more than 15,000 at another center. The Health Gorilla platform also has the functionality to analyze the CCDs for completeness of individual sections of the document. Figure 1 below shows an example of the CCD completeness summary from one of the centers.

Another example of data quality review included the tracking of number of medications that were missing coding in the RxNorm standard terminology. Medications not encoded using RxNorm may lead to analysis challenges or interoperability issues when shared with other health information systems. For the same center for which the CCD section completeness is shown in Figure 1, nearly 50% of the patients had some medications missing RxNorm codes. Overall, about 18% of the medications did not have the RxNorm codes.

CDA Completeness

Total Uploaded: 57010

Skipped as failed: 13

Skipped as duplicate: 3109

Clinical Document Percent Average Completeness 45.78

| Clinical Document Section | Breakdown of data present | Percentage |
|--|---------------------------|------------|
| Allergies, adverse reactions, alerts | 4637/57010 | 8.13 |
| Medications | 33193/57010 | 58.22 |
| Problem List | 42954/57010 | 75.34 |
| Relevant diagnostic tests and/or laboratory data | 30249/57010 | 53.06 |
| Social History | 57010/57010 | 100.00 |
| Vital Signs | 45053/57010 | 79.03 |
| History of Procedures | 4949/57010 | 8.68 |
| Encounters | 55304/55304 | 100.00 |
| Family History | 0/0 | 0.00 |
| Functional Status | 57010/57010 | 100.00 |
| Immunizations | 18273/57010 | 32.05 |
| Medical Equipment | 57010/57010 | 100.00 |
| Payers | 50654/50654 | 100.00 |
| Plan of Treatment | 57010/57010 | 100.00 |
| Mental Status | 0/57010 | 0.00 |
| Nutrition | 0/0 | 0.00 |
| Advance Directives | 0/0 | 0.00 |
| Goals | 0/57010 | 0.00 |
| Health Concerns Document | 0/57010 | 0.00 |
| Interventions | 0/0 | 0.00 |
| Health Status Evaluations and Outcomes | 0/0 | 0.00 |
| History of Present Illness | 0/0 | 0.00 |
| Assessment | 0/57010 | 0.00 |
| Assessment and Plans | 0/0 | 0.00 |
| Review of Systems | 0/0 | 0.00 |
| Chief Complaint | 0/0 | 0.00 |

Figure 1 CCD Quality Summary - Section Completeness

4.2 Administrative Data Analysis

The available data in the form of CCDs have been analyzed to extract administrative and clinical measures. This study did not include the review of any financial data from billing or claims platforms.

Administrative data review included the analysis of three basic entities - the number of patients, number of encounters, and types of encounters at each center, during the years 2018 and 2019. A review of locations of service and provider information was also carried out but is not described in this report because the data available within the CCDs was not found to be sufficiently comprehensive. Thus, while dashboards with administrative data such as providers and locations of service would be beneficial, at this time those data are unreliable for dashboard purposes. Going forward, and as part of the general quality review processes, it is suggested that the health centers review and consider options for capturing location-and provider-specific information.

4.2.1 Number of Patients

Total patient count for each center in a given year is the number of unique patients who had at least one qualifying encounter at the center during the specified year. Figure 2 shows the relative scales of the number of patients at the three centers in 2018 and 2019. As noted above, CHC 3 data for 2019 only covered the first seven months of the year, which is reflected in the figure as a lower patient count than for the year 2018. In fact, it is this surprising anomaly in the dashboard that triggered the analysis of monthly data counts which then led to the finding of the missing data.

A key lesson learned here is that even the simplest of dashboards can reveal interesting and useful patterns of data that are accessible by a wider group of stakeholders and may also reveal inconsistencies of the underlying data. It also highlights the importance of careful data preparation and data governance upstream, as the number of users who utilize and depend on these dashboards in their daily workflow increases.

Due to the lack of availability of encounter type information in the data as described earlier, the number of encounters calculated (and shown in Figure 3 below) include many potentially non-qualifying encounters according to UDS definition. The number of encounters, and as a result, also the number of patients shown below are higher than the corresponding figures from the UDS reports. For example, the patient counts for CHC 1 and CHC 2 for the year 2019 turned out to be 10% and 21% higher than the corresponding figures from the UDS reports, respectively.

While UDS reporting requirements for electronic Clinical Quality Improvement (eCQM) measures allow the use of a *random* sample of 70 patient records (Bureau of Primary Health Care, HRSA, 2020), the deployment of a data warehouse that can aggregate data across all the CHCs can ease the burden of defining and maintaining standardized measures and queries. It will also have the additional advantage of various clinical and administrative measures being accessible to a wider group of stakeholders on-demand throughout the year, as opposed to waiting for the compilation of annual reports by the analysts.

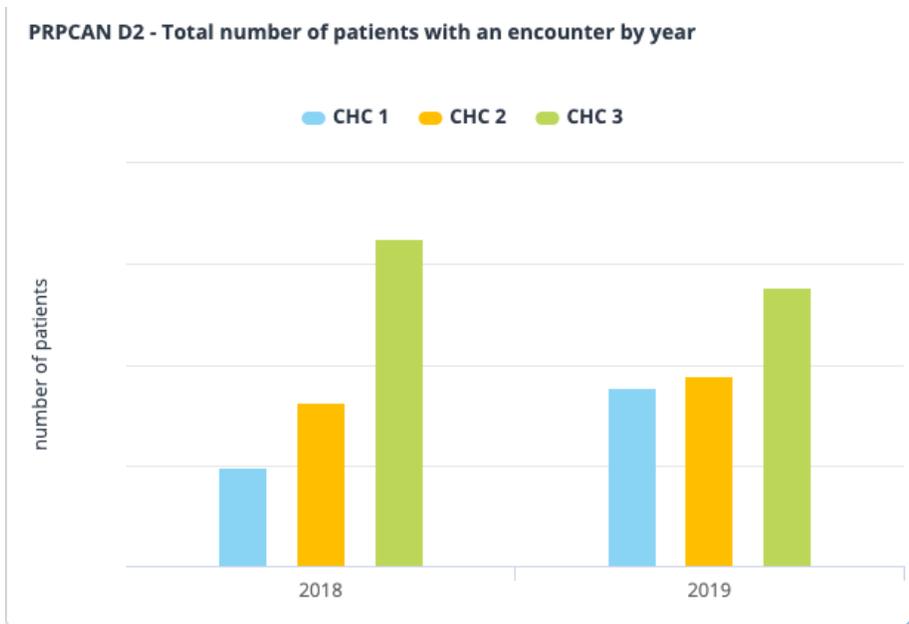


Figure 2 Number of patients with an encounter by year

4.2.2 Number of Encounters

Total number of encounters (by year)

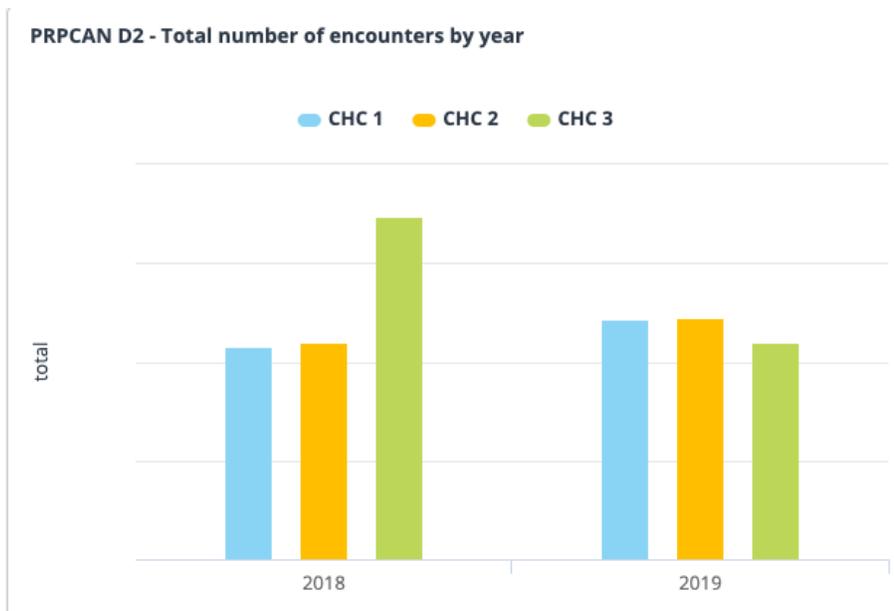


Figure 3 Total Number of Encounters by Year

4.2.3 Encounter Types

As noted above, encounter type information was not documented in most of the patient records in the data. In a few cases, the description of the encounter was documented in fields other than the encounter type. All undocumented (and mis-documented) encounter types are categorized in the *Unknown* class in Figure 4 below. The trend showing reduced number of encounters in the *Unknown* category in 2019 versus 2018 may be due to an improvement in the documentation of

this field. The *Other* category includes Emergency, Home visit, Inpatient, and Lab visit encounters, and as can be seen, very few encounters are documented with these types.

There were very few documented telemedicine visits in the CCD data. According to the 2019 UDS report, only 0.4% of visits were reported to be telehealth visits (telemedicine visits refer to remote clinical services, while telehealth refers to broader set of remote healthcare services; check if this figure was across PR or nationwide) according to UDS reports (Geiger Gibson / RCHN Community Health Foundation Research Collaborative, September 2020). That figure is expected to increase significantly in 2020, due to the impact of coronavirus. HRSA's most recent COVID-19 survey reports 29.55% of health center visits nationally were conducted virtually during the first week of January 2021.

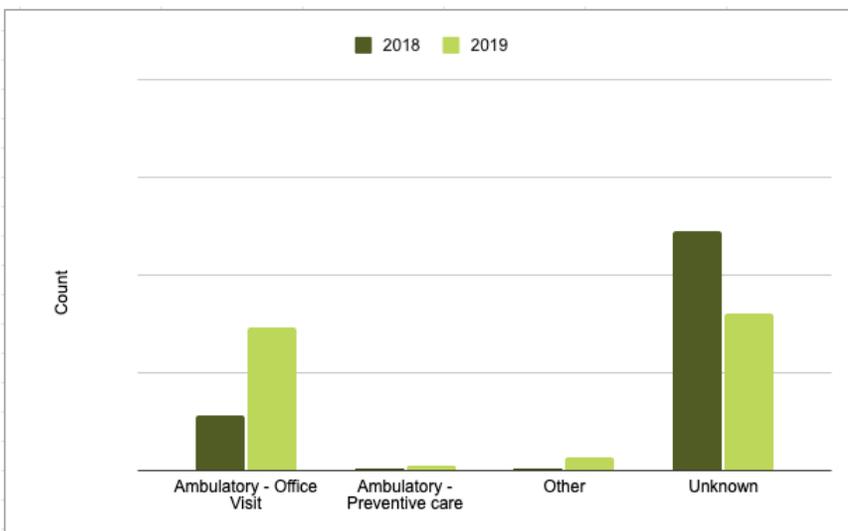
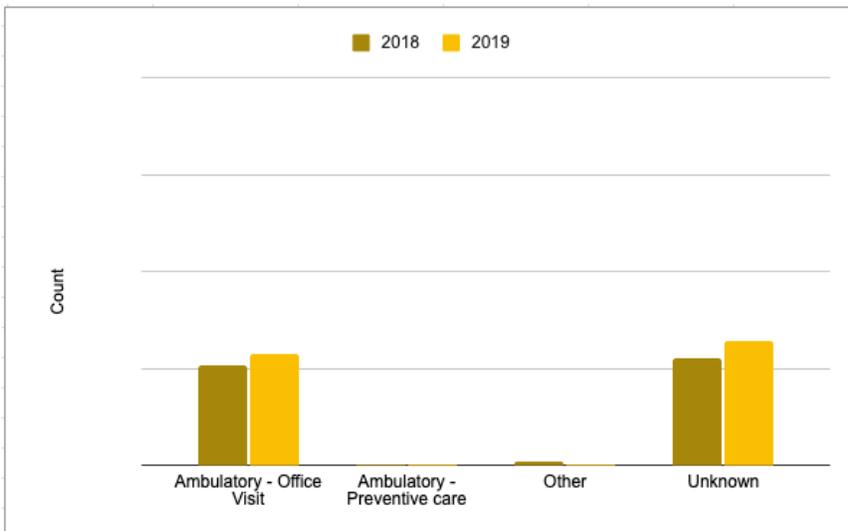
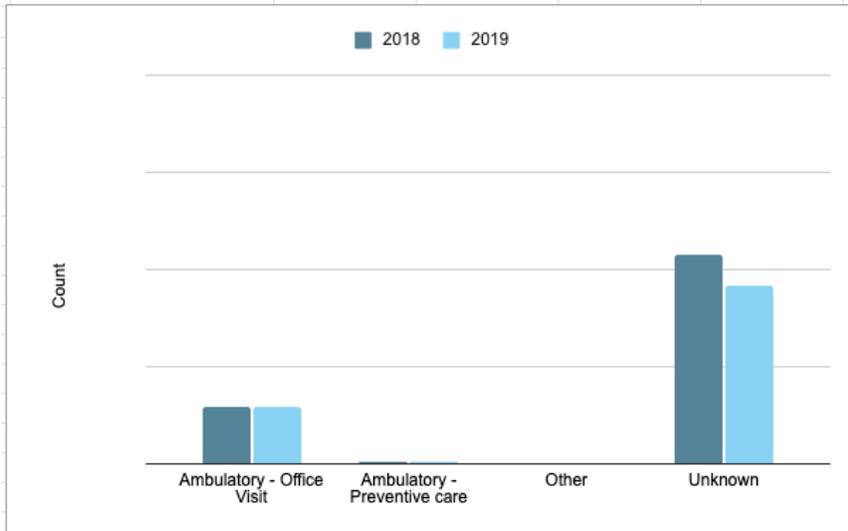


Figure 4 Number of Encounters of Various Types, CHC 1, CHC 2, CHC 3

Most common encounter type, which is the same across all 3 centers, is the Office Visit Established Patient Level 3.

4.3 Clinical Data Analysis

Chronic disease management at the population level involves measuring health outcomes related to chronic conditions such as hypertension, diabetes, and heart disease. Other areas covered by HRSA electronic clinical quality measures (eCQMs) include screening and preventive care, and maternal and children's health.

The following figures show the top diagnoses (descriptive names and corresponding ICD-10 codes) in 2019, and the percentage of patients who had the diagnosis documented for at least one encounter that year. Hypertension is the most frequent diagnosis for the CHC 2 and CHC 3 patient populations, while obesity and overweight-related diagnoses seem to be the most prevalent in the CHC 1 patient population.

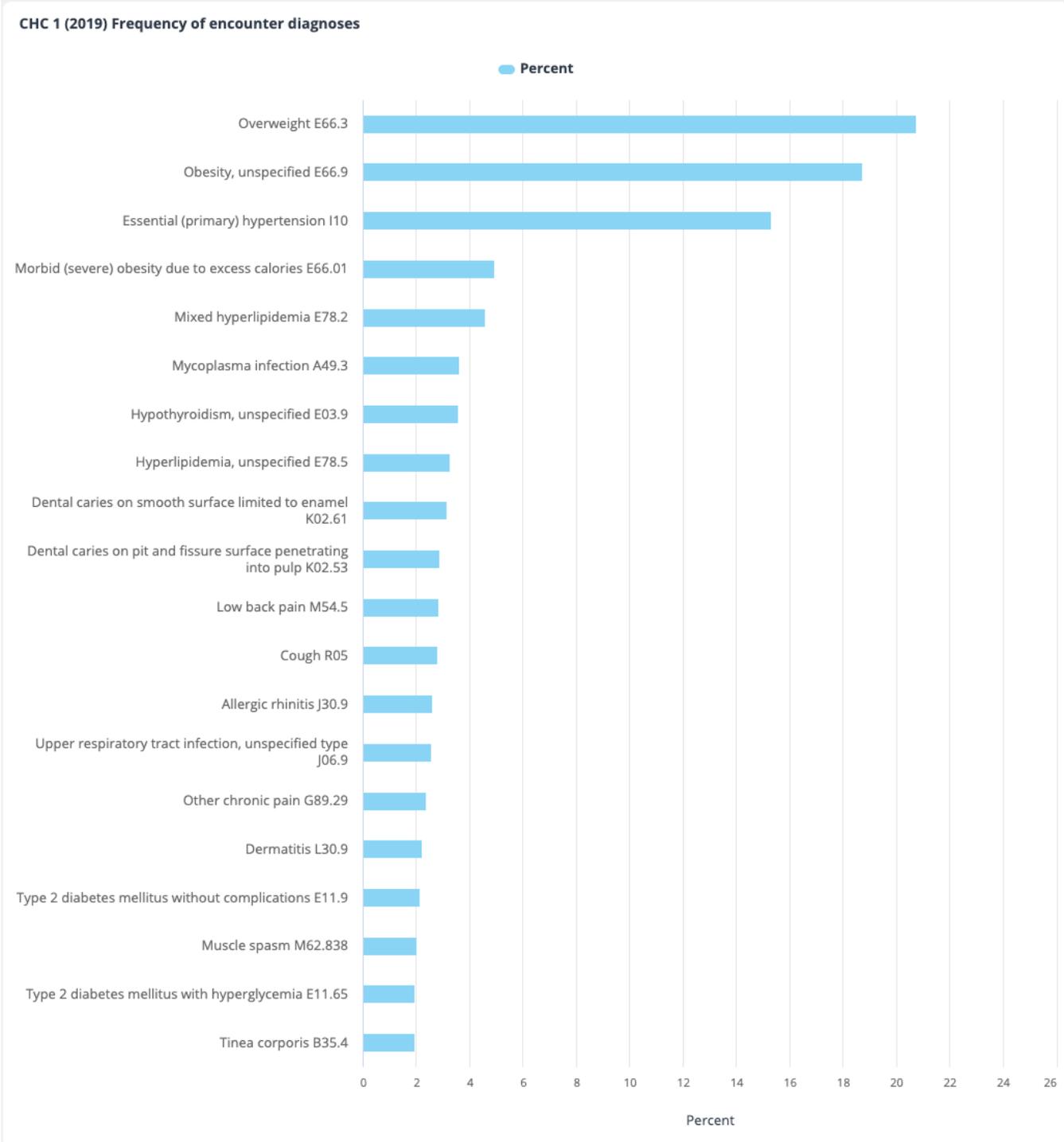
Note that the same hypertension ICD-10 code (I10) is documented with different, synonymous descriptive names and appears multiple times in these figures. For example, CHC 1 data included these nine different descriptions in the order of decreasing frequency - Essential (primary) hypertension, Essential hypertension, Benign essential HTN, Benign essential hypertension, Asymptomatic hypertension, Hypertension, unspecified type, Uncontrolled stage 2 hypertension, Benign diastolic hypertension, and Accelerated essential hypertension. CHC 2 data similarly included these 11 descriptions - Essential (primary) hypertension, Benign essential HTN, Essential hypertension, Hypertension, HBP (high blood pressure), Benign hypertension, Hypertension, unspecified type, HTN (hypertension), Accelerated hypertension, Malignant essential hypertension, and Asymptomatic hypertension.

The total number of patients with an encounter in 2019 that had a diagnosis of hypertension were calculated for CHC 1 and CHC 2. These figures were compared with corresponding figures from the 2019 UDS report for these two centers. HTN percentage figures in UDS report uses "hypertensive adults as a percent of estimated adult medical patients of ages 18-85". We noted earlier that the total patient counts in our estimates for these two centers were higher than those in the UDS report. Similarly, the number of hypertension patients in our estimates is also higher than that in the UDS report. However, using total number of patients in the denominator (without the age qualification), HTN percentage figures are quite close to the corresponding UDS figures.

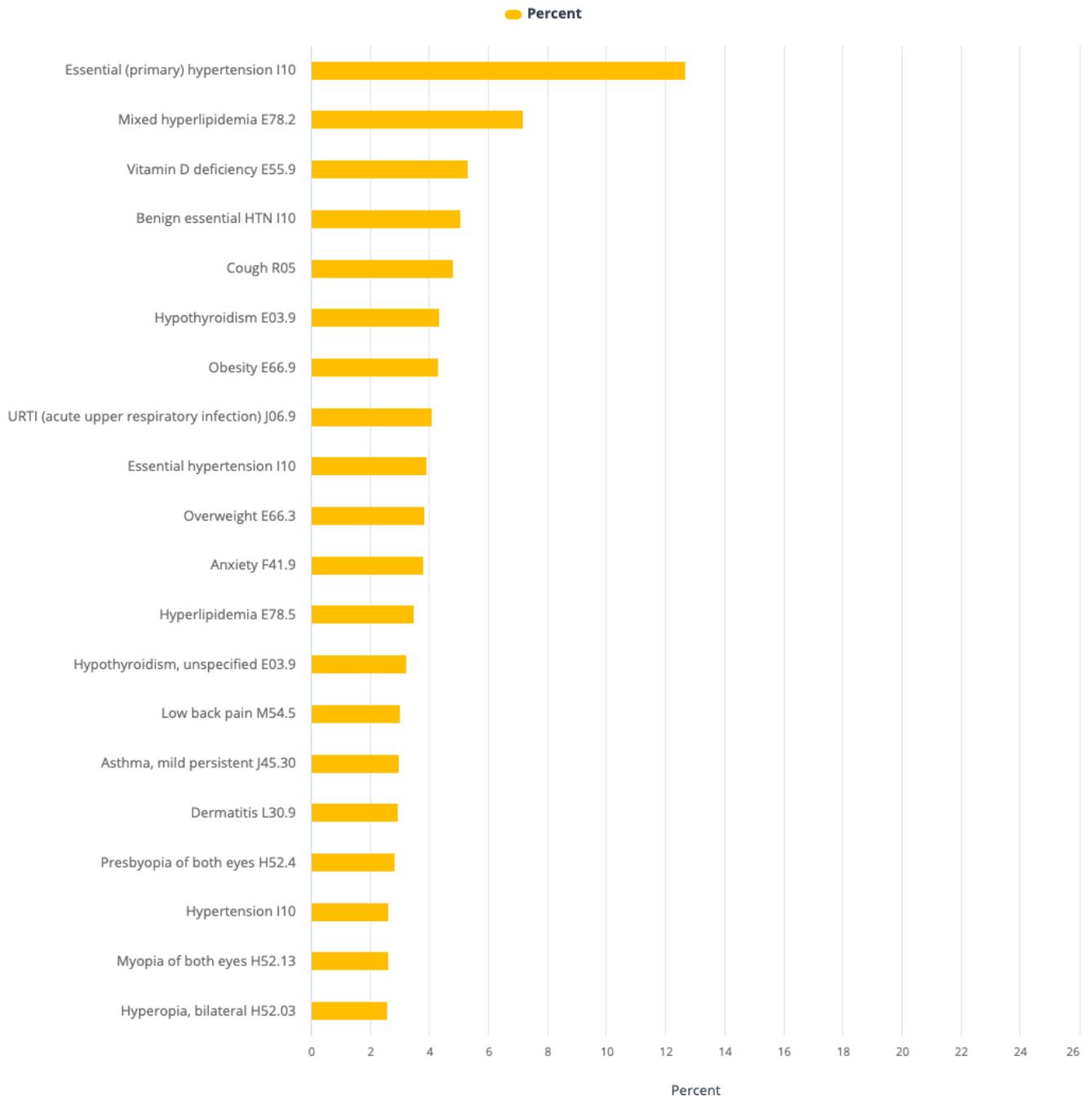
Lack of consistency in configuring and provisioning the document templates and pull-down menus for structured data elements in the EHR can make it more difficult and time consuming for the clinicians to select the correct choice. It can also lead to challenges in the development of analytics and dashboards. While the analysts may be able to take into account such variations, with increased effort, for annual reporting purposes, it would pose a challenge for streamlining and automating the generation of dashboards on a regular basis.

Also note that the diagnoses shown here are from the patient's "Problem List" (and not from diagnoses made during the encounter). The diagnoses and codes are not included in the CCDs as

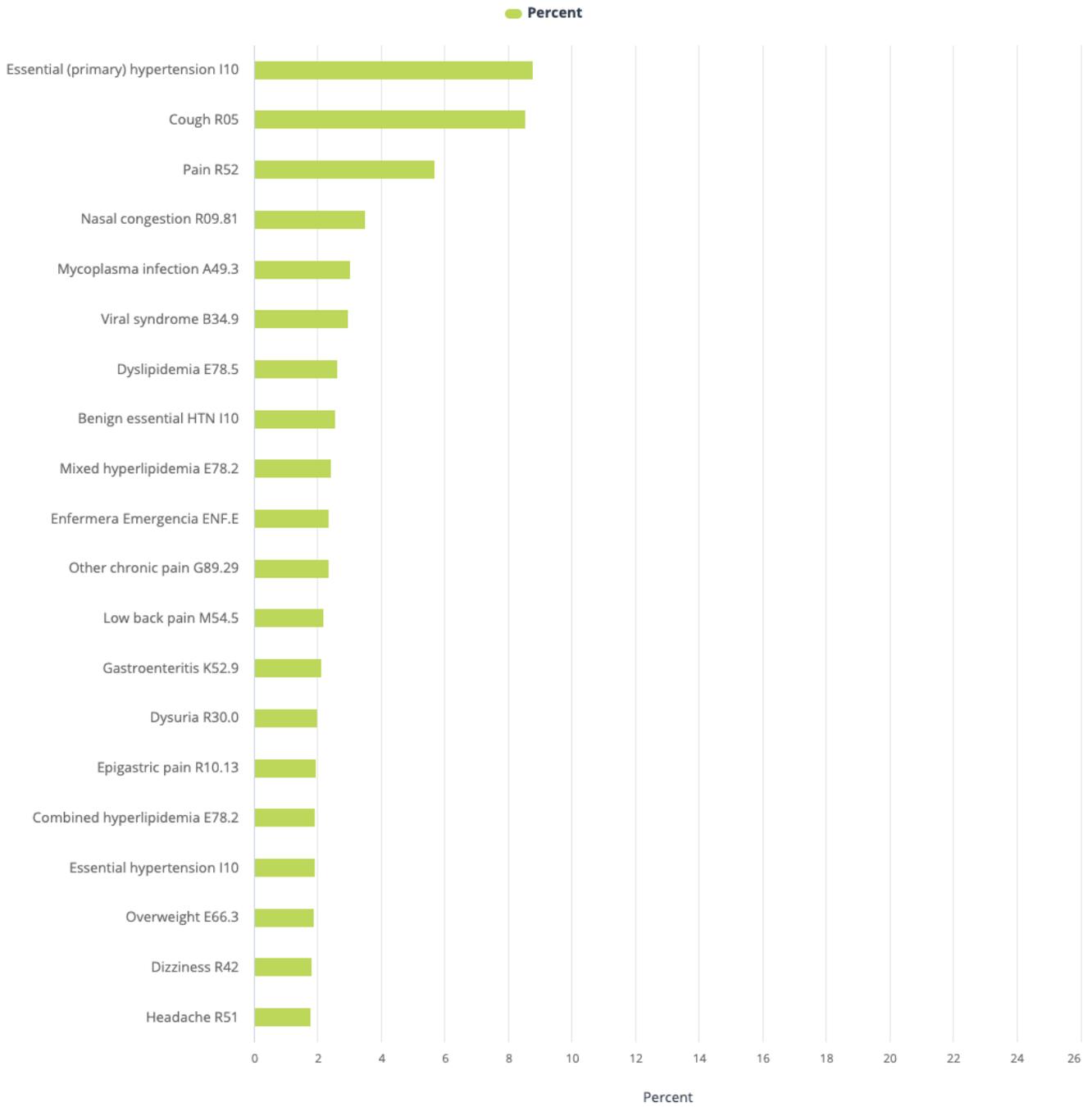
part of each of the encounters. As long as the problem lists are maintained accurately, this approach of including active conditions during a visit, is in accordance with UDS quality measure reporting (Bureau of Primary Health Care, HRSA, 2020).



CHC 2 (2019) Frequency of encounter diagnoses



CHC 3 (2019) Frequency of encounter diagnoses



5. SOCIAL DETERMINANTS OF HEALTH (SDOH)

This section covers the approach used to extract and analyze SDOH related information from the EHR data extracts (CCDs) used in the project from the three CHCs.

5.1 Background

Even before the ravages of Hurricane Maria and island-wide earthquakes, Puerto Rico's working-age residents were more than twice as likely to be unemployed (10% versus 4%) as mainland residents. Forty six percent of Puerto Rico's population is below the Federal Poverty Level of \$20,420 for a family of three (Office of the Assistant Secretary for Planning and Evaluation, 2017) as compared to fifteen percent, and the median household income (in 2016) was \$20,078 compared \$57,617 in the 50 states and D.C. (Peter Shin, 2017). The percentage of adults who report fair or poor general health in Puerto Rico is 34% compared to 18% in the 50 states and D.C. (Kaiser Family Foundation, 2017). The island's residents also report higher rates of chronic but manageable diseases, such as diabetes and heart disease. While those served by community health centers in Puerto Rico are similar to health center patients across the country, there are some striking differences, with 86% of health center patients in Puerto Rico reporting below-poverty income, compared to 70% of patients served by health centers in the 50 states and D.C.19 (Shin, et al 2017).

Over the last few years, there has been increasing recognition of the influence of social and economic circumstances on an individual's health risks and treatment outcomes. The social determinants of health (SDOH) are defined as the "conditions in which people are born, grow, live, work and age that shape their health." (Samantha Artiga, 2018) Screening for and mitigating SDOH is seen as a major tool for addressing healthcare inequalities. The ongoing transition to value-based care, and increased health system and provider accountability have also created incentives to address the social determinants of health. Examples of the SDOH related domains include Access to health care, Education, Employment, Food, General (non-specific), Housing, Social Connections, and Transportation.

The Department of Health and Human Services (HHS) announced a website (<https://sdoh4covid.crowdicity.com/>) focused on using SDOH data to fight COVID-19. The core objective of the effort is to identify actions HHS can implement that will help the department understand risk patterns and develop social and medical interventions.

Availability of uniform data about SDOH is a prerequisite for screening, assessment, and social intervention. Several organizations are working with health systems and other stake holders to standardize social health assessment tools and clinical documentation.

The Social Interventions Research and Evaluation Network (SIREN) at University of California San Francisco promotes interventions addressing patients' SDOH in health care delivery. The team documented 20 SDOH-related domains used in 6 different social health assessment tools such as *HealthLeads*, *SEEK*, and *PRAPARE*, the tool developed by National Association of Community Health Centers in conjunction with AAPCHO an Oregon Primary Care Association, and others.

The SIREN team identified existing documentation tools in common medical coding systems (ICD-10-CM, SNOMED CT, CPT, and LOINC) that reflect emerging clinical practices related to identification, diagnosis, and intervention of patients’ social risks in the context of health care delivery. They (Bureau of Primary Health Care, HRSA, 2020) developed and published a spreadsheet to help all stakeholders, including clinical and social service providers, identify the set of existing codes related to social risk factors and SDOH (Arons A, 2018).

5.2 Summary of CHC Population SDOH Data

The summary analysis presented here is based on identification of SDOH-related ICD-10 codes in the Problem Lists of patients at each of the three health centers. The SDOH-related ICD-10 codes are derived from the SIREN group spreadsheet referred to in the previous section. While there might be additional SDOH-related data documented in the centers’ EHRs that is coded in CPT or LOINC, the CCD data used in this project primarily contained only the ICD-10 information. SDOH related codes are represented in ICD-10 code categories Z55 - Z65 to describe conditions such as poverty, homelessness, unemployment etc.

Of the 20 SDOH-related domains documented in the SIREN spreadsheet, some level of ICD-10 code documentation corresponding to six of the domains was present in one or more health center’s data. Table 1 lists these six domains, and number of patients (as a percentage of 2019 population at each of the centers) for whom information related to each of these domains was documented.

| | Education | Employment | Finances | General | Housing | Social Connections/ Isolation |
|--------------|------------------|-------------------|-----------------|----------------|----------------|--------------------------------------|
| CHC 1 | 0.36% | < 0.1% | < 0.1% | 0.72% | < 0.1% | 0.42% |
| CHC 2 | < 0.1% | < 0.1% | < 0.1% | 0.55% | 6.6% | 0.34% |
| CHC 3 | 0.31% | 0.83% | 3.9% | 1.1% | < 0.1% | 1.0% |

Table 1. Number of patients (as a percentages of 2019 population at each center) that had social determinants in one of six domains documented.

As seen from Table 1, the documentation of SDOH data in the EHRs is quite sparse and uneven among the social determinant categories. It is not surprising considering that social determinants have been gaining attention only in the last few years, and providers and other staff are still being trained on emerging standards for documentation and coding of SDOH. EHR systems are also beginning to support the appropriate standardized documentation.

While these figures are incomplete due to partial capture of social determinants information in the EHRs, they are shown here to illustrate the potential use of comprehensive documentation of social determinants in screening, assessment and diagnosis, and treatment and intervention.

Among the three centers, CHC 3 data captured the most documentation of SDOH information across the six domains, while CHC 1 data had the least information related to social determinants. Of the 6.6% of CHC 2 patients documented with housing concerns, 6.4% reported homelessness, 0.2% reported exposure to lead, and there was only one mention of inadequate housing. Under the finances domain, 3.8% of CHC 3 population reported extreme poverty, while less than 0.1% of patients had a documentation of low income.

In the General (non-specific) domain, 0.23% of CHC 1 patients had 'Other symptoms and signs involving appearance and behavior', 0.22% had 'Other problems related to lifestyle', 0.11% had 'Other symptoms and signs involving emotional state', and fewer people had 'Problem related to social environment', and 'Problem related to unspecified psychosocial circumstances' in their records. The top four codes in the General domain of CHC 3 patient population included 'Problem related to unspecified psychosocial circumstances' (0.43%), 'Other symptoms and signs involving appearance and behavior' (0.29%), 'Other personal risk factors' (0.16%), and 'Other symptoms and signs involving emotional state' (0.13%).

The most common conditions documented under the Social Connections/Isolation domain include 'Problems related to primary support group', 'Problems related to living alone', and 'Disruption of family by separation and divorce'.

This analysis shows that useful information about social determinants can be obtained through dashboard aggregation. Investing additional resources and focus on SDOH-related training, data capture and documentation can be beneficial in leveraging the patient EHR data to address the prevalent health inequities and improve health care delivery. Documenting SDOH is also an essential consideration for policy; only by continuously documenting disparities can we hope to secure the resources needed to address them.

6. CONCLUSION/FUTURE DIRECTIONS

One of the key lessons learned in this study is that in order to develop complete and accurate real-time views of the data residing primarily in the Electronic Health Record (EHR) systems being used by the Health Centers, the data extraction should utilize more modern approaches that are more robust and less burdensome. Use of data extracted in the form of *Continuity of Care Documents* (CCDs) (ONC, 2013) resulted in several limitations in this study. Fast Health Interoperability Resources (FHIR) (HL7 International, 2019) is the next generation interoperability standard mandated by CMS that offers open interfaces for exchange of healthcare data between various systems. It is anticipated that in the near future, the EHRs at the Health Centers will be integrated with a data warehouse application using the FHIR standard for data extraction.

The deployment of a common data warehouse and analytics platform across all the CHCs can ease the burden of defining and maintaining standardized measures and queries. It will also have the advantage of various clinical and administrative measures being accessible to all internal stakeholders on-demand throughout the year. Such continuous and ongoing data review is essential for health centers to succeed.

Another important lesson learned is that even the simplest of dashboards can reveal interesting and useful patterns of data that are accessible by a wider group of stakeholders. These simple dashboards may also reveal inconsistencies of the underlying data, as they did during the course of this study, including incomplete sections of the records and missing coded information in standard clinical terminologies. It also underscores the importance of careful data preparation and data governance upstream, as the number of users who utilize and depend on these dashboards in their daily workflow increases.

Configuration and provisioning of document templates, and pull-down menus for structured data elements in the EHRs should be reviewed to ensure adoption of uniform definitions, specificity and consistency. Lack of consistency can make it more difficult and time consuming for the clinicians to select correct choices in their documentation, and lead to challenges in the development of analytics and dashboards.

There is an increasing focus on screening for and mitigating social determinants of health as an approach for addressing healthcare inequalities and improving health outcomes. Several organizations are developing standards for social health risk assessment tools and SDOH documentation. It is important for the EHR systems to support these evolving standards and for the clinical and front office staff to be trained on utilizing the risk assessment tools in their workflow and adopting the standardized documentation methods. Another industry-wide effort is attempting to streamline and standardize the current ad hoc workflows for social care referrals by bringing together provider, payers, Community-Based Organizations (CBOs), and health IT companies (Center for Open Data Enterprise (CODE) and the Office of the CTO at the HHS, 2019).

Some of the sparse information available in the data extracts from the three centers spanning six domains of social determinants (education, employment, finances, general, housing, and social connections/isolation) has been reviewed and summarized in this study. The summary with

partial information points to the promise of the role that comprehensive data can play in addressing the social needs of the patient population.

We explored some of the available data on chronic conditions, and social determinants of health in this study. Long-standing data show a relationship between social determinants and general health outcomes. Recent studies have established a relationship between social determinants of health such as employment status, income and nutrition, and the incidence of chronic conditions (Center for Health Analytics, Research & Transformation at NJHA, 2019) (William C. Cockerham, 2017). It would be beneficial to explore the correlation between the incidence of some chronic conditions and SDOH factors in PR in a future study.

Despite the limitations of the existing data, an important take-away is the relevance of continuous review of the data. First, only by ongoing processes to review the data for quality and completeness can gaps or omissions be detected in a timely manner. Second, in order to improve outcomes and quality, it is essential that health center staff engage in routine analysis of the available data. If deployed consistently, data dashboards provide the functionality that is necessary to support ongoing and near real-time analysis.

The ASPPR is to be commended for initiating this study. Through its leadership of PRPCAN, the Association has demonstrated its commitment to strengthening crucial IT and analytics capabilities for Puerto Rico's community health centers. All of the island's 20 health centers now participate in the HCCN and with the assistance of PRPCAN, all 20 centers adopted and implemented Certified Electronic Health Record Technology and made significant advances with respect to Meaningful Use Stage 2 modified CMS measures including clinical decision support, e-prescribing, computerized order entry and other core requirements. Many of these achievements were attained despite the considerable damage and ongoing challenges resulting from the devastation wrought by Hurricanes Irma and Maria, and the subsequent earthquakes. Indeed, this project was successfully completed despite the challenges of operating in a pandemic environment.

Through this D2 initiative and other projects, PRPCAN has turned its attention to enhancing capabilities for data exchange and analysis as well as data sharing. This work is both timely and essential. Successfully participating in a rapidly evolving delivery system requires a concrete strategy. Strategy, in turn, requires reliable information. Useable information that can be leveraged for strategic purposes requires robust data. Finally, information needs to be shared across the network. Health centers should be encouraged to participate, through available secure mechanisms, in data sharing and aggregation. The value of such information cannot be overemphasized and indeed, crisis situations such as widespread hurricanes and the current pandemic underscore the merits of information sharing, as well as the considerable value of dedicated ASPPR resources to support data collection and assessment.

A continuous process to improve data validation, aggregation, analysis and transparency will ensure that the available data becomes as a key strategic asset that ultimately serves to improve the quality of health center services.

History and summary conclusions of the Path2Analytics project that is a precursor to Level-Up.

Path2Analytics and the Level-Up Project

The Level-Up process was originally developed as a part of the Path2Analytics project carried out by the RCHN Community Health Foundation (RCHN CHF). It focused, in part, on the effect of organizational structure and culture on data quality¹. The initial purpose of the P2A project was to introduce Hadoop-based analytics² into a small number of CHCs and a PCA.

Appendix 1. Brief Description, Path2Analytics (P2A) Project

The Path-2-Analytics Project (P2A) was initiated by the RCHN Community Health Foundation and carried out in the 2014-2016 time period. The goal of the project was to assess the readiness of community health centers (CHCs) and Primary Care Associations (PCAs) to deploy and use contemporary analytics infrastructure and methods. This is of strategic importance as health centers have to deal with more complex decisions using more varied types and greater quantities of data.

The project consisted of several elements including:

- Education of IT and all CHC departments in the technology and methods of contemporary analytics.
- Focus on involving the executive staff of the organization in the project.
- Jointly deploying an open source analytic stack (based on the Cloudera Hadoop stack) including management & SQL-based query facilities.
- Execution of a data quality exercise (Level-Up) using both the health center's infrastructure (EHR, BI or reporting tool, etc.) and the Hadoop-based stack (Hadoop Distributed File System, Yarn (Map Reduce 2), Impala (massively parallel SQL-based query) etc.
- Interpretation of data quality exercise & evaluation of use of analytic stack by health center personnel.

Results were reported in Hartzband and Jacobs, *op. cit.*

¹ Lateral Alignment in Complex Systems: Project Description. Engineering Systems Division. Massachusetts Institute of Technology. 1/2009.

² *c.f.* https://en.wikipedia.org/wiki/Apache_Hadoop

Copy of questionnaire distributed to the Centers by PRPCAN to gather information regarding their IT infrastructure and data handling practices.



PR PCAN – RCHN Community Health Foundation HIT & Data Survey

General

Respondent:

Organization Name

Contact Information

Number of Sites

Name of Person Compiling Survey

Title

Telephone Number

Email Address

EHR

Who is your EHR Vendor?

Product Name and Version currently implemented – if applicable, please include build number

When is your next scheduled upgrade? Is your software up to date?

What is the underlying Database for this system? (Oracle, SQL Server, other commercial, proprietary)?

Location of EHR Server(s): On Premise? Cloud-based?

Who manages these servers? Organization Staff? Vendor?

Is this EHR Meaningful Use 2 Certified?

If yes, what MU version is this Certified?

Is your organization currently participating in the Meaningful Use incentive program? If so, what level of participation have you reached?

How long has this EHR been implemented?

Reports

Does the Center have the capability to generate the requested reports? (CQMs & MU)

How many reports generated by the Center have been shared with PRPCAN?

What types of reports & what specific reports (including content where applicable) have been shared with PRPCAN?

Who is responsible at the Center for generating these reports? Please give contact information.

What date were these reports shared with PRPCAN?

Who is responsible for sending the reports to PRPCAN? Contact information.

How are reports sent to PRPCAN (fax? email? FTP? Other electronic? Other?)

Who at the Center is responsible for the handling of data & content information? Contact information.

Practice Management

Is your Practice Management system from the same vendor as your Electronic Health Record system?
If not, which Practice Management System are you using?
Are your Practice Management & EHR systems integrated (exchange data)?

User-Defined Fields

Does your current EHR permit user-defined fields?
Are you currently using this feature for any clinical assessments?
Are you able to search the data in these user-defined fields?

Structured vs. Unstructured Data

How much of the data in your (EHR) system is in unstructured text (notes etc.)?
More than 30%, 30% 20% 10%
Can this data be searched?
Do you use natural language processing to facilitate use of this data electronically? If so, what program?

Electronic Data Exchange

What types of Electronic Data Interchange (EDI) is your system currently capable of?
HL7v2 Messages, 837/835 billing forms, 270/271
Can your EHR send an HL7v2 ADT (Admit, Discharge, Transfer) message?
Can your EHR receive HL7v2 ADT (Admit, Discharge, Transfer) message?
Can your EHR send an HL7v2 C-CDA as discrete data?

Database Access & Use

Do you ever access the EHR database separately from the EHR interface (*i.e.* for *ad hoc* queries, customized reports etc.)?
If so, how is this data accessed? Export to Excel or CSV file, Direct SQL query? Other?
Do you use of a BI tool or Report Writer?
If so which one?
How is this data accessed?
Which roles within your organization access the data?

Project Data

Does your Center run any projects that generate data not entered into your EHR? Population Health projects? Home visitation projects? Housing or other social determinants projects?
Do you capture this data separately? In a separate database, Excel file etc.
Is this data added to your EHR in some way? If so, how? If not, how is it used?

Analytic & BI Tools

Are you using an analytic or visualization application such as SPSS, SAS or Tableau?
Which one?
Who in the organization uses it?
Can providers or others request a specific analysis?

Other HIT Software

Are you using any other HIT software? Population Health? Disease Registries? Etc.?

ACKNOWLEDGEMENTS

The RCHN Community Health Foundation wishes to acknowledge Srini Rao, PhD as lead author of this white paper and recognize the major contributions of past and current colleagues at the ASPPR, PRPCAN and the RCHN CHF and at Health Gorilla.

In addition, the Foundation acknowledges the health centers whose participation, engagement and data sharing formed the basis for this analysis.

This report is dedicated to the memory of David J. Hartzband, DSc.

BIBLIOGRAPHY

- Arons A, D. S. (2018). *Compendium of Medical Terminology Codes for Social Risk Factors*. Retrieved from Social Interventions Research and Evaluation Network: <http://sirenetwork.ucsf.edu/tools-resources/mmi/compendium-medical-terminology-codes-social-risk-factors>.
- Bureau of Primary Health Care, HRSA. (2020). *Uniform Data System Reporting Instructions for 2020 Health Center Data*.
- Center for Health Analytics, Research & Transformation at NJHA. (2019). *CHART-Chronic-Conditions.pdf*. Retrieved from NJHA New Jersey Hospital Association: <http://www.njha.com/media/546633/CHART-Chronic-Conditions.pdf>
- Center for Open Data Enterprise (CODE) and the Office of the CTO at the HHS. (2019). *Leveraging Data on the Social Determinants of Health Roundtable Report*.
- Geiger Gibson / RCHN Community Health Foundation Research Collaborative. (September 2020). *Community Health Centers on the Eve of the COVID-19 Pandemic: An Overview of Findings from the 2019 Uniform Data System*.
- HL7 International. (2019, October 30). *HL7 FHIR Release 4*. Retrieved from HL7 FHIR: <http://hl7.org/fhir/>
- Kaiser Family Foundation. (2017, October). *Puerto Rico: Fast Facts*. Retrieved from KFF: <https://www.kff.org/racial-equity-and-health-policy/fact-sheet/puerto-rico-fast-facts/>
- Mandl, K. G. (2020, November). *Push Button Population Health: The SMART/HL7 FHIR Bulk Data Access Application Programming Interface*. Retrieved from nature.com: <https://doi.org/10.1038/s41746-020-00358-4>
- Office of the Assistant Secretary for Planning and Evaluation. (2017). *2017 Poverty Guidelines for the 48 Contiguous States and the District of Columbia*. Retrieved from U.S. Department of Health and Human Services: <https://aspe.hhs.gov/poverty-guidelines>
- ONC. (2013, April 5). *Implementing Consolidated-Clinical Document Architecture (C-CDA) for Meaningful Use Stage 2*. Retrieved from https://www.healthit.gov/sites/default/files/c-cda_and_meaningfulusecertification.pdf
- Peter Shin, P. J. (2017). *Puerto Rico's Community Health Centers: Struggling to Recover in the Wake of Hurricane Maria*. Geiger Gibson / RCHN Community Health Foundation Research Collaborative.
- Samantha Artiga, E. H. (2018, May). *Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity*. Retrieved from KFF: <https://www.kff.org/racial-equity-and-health-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/>
- William C. Cockerham, P. B. (2017). The Social Determinants of Chronic Disease. *Am J Prev Med*.