Patient Engagement in Clinical Data Research Networks: The 2015 Greater Plains Collaborative Health and Medical Research Family Survey

### Audit Trail

This is not part of the actual paper. This is here so that if anybody has questions for me about the data, it will help me exactly reproduce this version of the output.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | time | type | name | value | hash |
| 0002 | 2018-04-12 10:45:11 | this\_script | RUN\_FROM\_INTERACTIVE\_SESSION | afb\_feb2018 | 84c279d |
| 0003 | 2018-04-12 10:45:11 | rdata | datafile | obesityPaper01.rdata | f65d8e23e6b106ebc014dd8cf9d8c28e |
| 0003.0002 | 2018-04-12 10:44:17 | this\_script | prepPaper01.R | afb\_feb2018 | TEST\_OUTPUT\_DO\_NOT\_USE |
| 0003.0003 | 2018-04-12 10:44:18 | rdata | datafile | survProcessed.rdata | 23af788c363443061b7c5a4bb9c30b04 |
| 0003.0003.0002 | 2018-04-12 10:43:01 | this\_script | ObesityScript.R | afb\_feb2018 | 84c279d |
| 0003.0003.0003 | 2018-04-12 10:43:01 | seed | rseed | 6062016 |  |
| 0003.0003.0004 | 2018-04-12 10:43:05 | file | input\_data | testoutput.csv | 7aebfc4c7cd0529c5f2d78fbf72ef1a9 |
| 0003.0003.0005 | 2018-04-12 10:43:08 | seed | rseed | 6062016 |  |
| 0003.0003.0007 | 2018-04-12 10:43:08 | save | save | tsave(.workenv, obd, rseed, obd.backup, samp, file = “survProcessed.rdata”) |  |
| 0003.0004 | 2018-04-12 10:44:19 | file | noticefile | coauthors\_notice.md | f1261475bf88f3be855f0fc5af0f4742 |
| 0003.0005 | 2018-04-12 10:44:19 | seed | tb$d01.seed <- rseed | 6062016 |  |
| 0003.0007 | 2018-04-12 10:44:55 | save | save | tsave(.workenv, df\_unilogist, dct0, obd, tb, file = “obesityPaper01.rdata”) |  |

NOTICE:

Dear Co-Authors:

This document incorporates all the latest data updates/fixes into the tables and into the parts of the text where values from those tables are referenced. This has two goals:

1. Being able to see the tables all in one place, in the same order they are referenced in the text.
2. To have one place from which to copy-paste the version of the text that has the up-to-date numbers.

However, please note this is **not** the official draft of the HMRFS paper. Currently, not all sections are included here. Moreover, this document is automatically generated by scripts and will be updated on a regular basis so any edits you make to it will be lost. Please continue treating the document [MOST\_CURRENT\_VERSION.docx](https://cmh.box.com/s/a30oj1vg0p3zv3bcqoin0nfrowbejvyb) as the working draft.

While I have your attention– if the long ugly variable names in the tables are giving you a headache, now you can help solve this problem. Please go to the file [var\_level\_names.xls.gsheet](https://cmh.box.com/s/ignwc8jo05nd7dmr358c4xseftrom7rg), and click **Open** in the upper right. It is a spreadsheet and you can write into the **description** column what you think we should standardize on as the published name/s of the variable/s you are thinking about to the left of that column.

– Sincerely, Alex F. Bokov

# Abstract

**Objective:** The aim of this paper is to report on the initial obesity project from the Greater Plains Collaborative (GPC), a Clinical Data Research Network (CDRN) that is part of the Patient Centered Outcomes Research Institute (PCORI) National Patient Centered Clinical Research Network (PCORNet).

**Materials and Methods:** The electronic health records of 10 academic medical centers in 8 Midwestern states were coordinated in order to allow for a prospective multi-site obesity focused research project to be conducted. A survey was designed to assess participant willingness for themselves, and their children (if appropriate) to participate in research, as well as the use of their health record data for research.The survey was composed of 7 questions regarding interest in medical research, as well as basic demographic and health related information. Each site had a goal of 1,000 initial positive survey responses and 100 completed surveys.

**Results:** A cohort of 54,269 patients was created across the participating sites. The eligible patient population was 57.02% female, 89.33% Caucasian, 28.37% self-pay, with mean age of 27.5 (18) years with 27.19% overweight and 30.36% obese. In general, a third of respondents had participated in research previously and were positive about being contacted for additional research both for themselves and for their children. Responses did vary by site.

# Background and Significance

In 2013, The Greater Plains Collaborative (GPC) was established as Clinical Data Research Network (CDRN), funded by the Patient-Centered Outcomes Research Institute (PCORI) to securely collect and organize patient health information obtained during routine care in its member health systems. To date, 13 such CDRNS have been funded, creating a national “network of networks”. These networks are organized by a coordinating center and overseen by the National Patient-Centered Clinical Research Network (PCORnet)[1].

The purpose of CDRNs and PCORnet is to support efficient clinical research by creating centralized access to the de-identified data of millions of patients across the country. Each CDRN is responsible for harmonizing patient data across its member systems, and for creating streamlined governance and procedures to facilitate researcher access. Importantly, CDRNs actively involve a variety of stakeholders, including patients, clinicians, healthcare system leaders, and other stakeholders to build and oversee CDRN activities. To test each CDRN’s ability to identify and recruit patients with a particular condition, and to test the ability to harmonize data elements within a network, each CDRN was required to create three cohorts: one of a common disease, one of a rare disease and one concerning height and weight; the GPC height weight cohort is the one described herein.

## GPC Member Sites

Placeholder text, map graphic?

In Phase I of funding, GPC was comprised of ten member health systems, with the data of roughly 6 million people across 7 states, north and south across the Great Plains region. Member institutions included Children’s Mercy Hospital (Kansas City, MO), University of Kansas Medical Center, Marshfield Clinic, Medical College of Wisconsin, University of Iowa Healthcare, University of Minnesota, University of Nebraska Medical Center, University of Texas Health Science Center at San Antonio, University of Texas Southwestern Medical Center, and University of Wisconsin - Madison.

Covering more than 1300 miles, the broad reach of the GPC network encompasses large swaths of rural populations as well as multiple urban centers. Four systems in the GPC have established significant relationships with Native American populations. Two health systems located in Texas serve heavily Spanish-speaking populations. Of the ten member health systems participating in the height weight cohort’s health and Medical Research Family Survey (HMRFS), all provide comprehensive adult and pediatric care, with the exception of Children’s Mercy Kansas City, which exclusively serves children. In Phase II of the GPC, largely following the work described in this paper, two additional members were added – University of Missouri and Indiana University.

|  |  |
| --- | --- |
|  | Table 1, Survey Questions |
|  | (placeholder) |

# Objective

The overall purpose of the HMRFS was to conduct a demonstration survey across all participating GPC sites focused on the topic of pediatric height/weight, or more specifically, pediatric overweight and obesity. The expected outcome of the project was to understand the practical challenges and operational details of a large semi-interconnected system such as the GPC for conducting collaborative research focused on pediatric obesity. Aims of the project were to: 1) estimate the willingness of individuals to be contacted about research activities; 2) obtain information on the attitudes of parents and adults of child bearing age about research, including participation of their child/ren; 3) gain insight into participant attitudes about the use of gathered data for both local and national research; 4) explore the impact of various demographic factors on survey outcomes; 5) examine if there are differences between individuals across various weight classes; 5) determine if there are regional variations in all of the above. Although previous studies have been published on adult obesity using a PCORI funded CDRN [2], these studies were retrospective in nature reporting on the number of patients in the network who met certain criteria. In contrast, the GPC HMRFS not only gathered retrospective data on individuals who met certain specific inclusion criteria but also contacted a random sub-sample at each site with a survey invitation. The goal for all participating sites was to collect 1000 responses each to “Survey 1” and 100 responses each to “Survey 2” Survey 1 was a single question asking whether or not the individual was willing to be contacted for research, and survey 2 was a more detailed assessment (see Table 1).

# Materials and Methods

The height weight cohort team composed of representatives from all sites began regular meetings in January of 2014. Based on discussion and collective interest, the group quickly decided to develop its cohort and survey around a pediatric population. Weekly working group calls established an interest in characterizing the cohort around data elements that would be attainable for the nascent GPC network, and creating a survey that could be used as a building block for future GPC and healthy weight cohort work. Thus, the HMRFS focused on respondents’ willingness to take part in future clinical research as well as key demographic and health related issues theorized to impact these responses.

## Institutional Review Board Process

Through its efforts to streamline governance, the GPC IRB Consortium was established to facilitate IRB review and approval. The consortium, including all the GPC sites, signed a common IRB reliance agreement and adopted standard operating procedures which would govern the reliance process. The University of Texas Health Science Center at San Antonio (UTHSCSA) served as the reviewing IRB site for the HMRFS across the GPC network. The HMRFS team of investigators and staff developed the necessary IRB documents which were submitted to the IRB at UTHSCSA. Once the documents were reviewed and approved by the reviewing IRB, the documents were shared with all other participating sites’ IRBs, and these documents were approved under the existing overall IRB reliance agreement. Currently, the GPC participates in the SMART IRB platform with over 175 participating institutions designed to facilitate multisite research and implement the NIH Single IRB Review Policy.

|  |  |
| --- | --- |
|  | Table 2, Eligibility Query |
|  | (placeholder) |

## Data Harmonization

This study combined electronic chart review with the results of a survey administered to patients, or in the case of minors, to the patients’ guardians. For the electronic chart review, data was extracted from an open source data warehouse platform called Integrating Informatics from Bench to Bedside (i2b2) [3]. Each of the participating sites had an i2b2 instance deployed, where they stored a de-identified version of all structured data from their respective EMR systems. As a result, no site had to transmit any identifying information to any other site.

Though not all the sites used the same type of EMR system, because of i2b2’s very flexible star-schema design the task of combining data from disparate sources was made more tractable. In fact, each site was completely autonomous in what software it used and how it organized its data internally as long as they had some means of running i2b2 queries. This decentralized, flexible approach is in keeping with the best practices of the sofware development field **???**– “specify standards, not implementations”. It would have been burdensome, expensive, and inappropriate stewardship of public research funds to impose a single vendor’s database software on all participating sites. I2b2 itself does not have a vendor— it is a free, open-source product originally developed under an NIH grant **???**. Though the lead site at KUMC has an extract-transform-load (ETL) process for extracting data from the EMR, deidentifying it, and loading it into i2b2. This system, called HERON **???** was made available to all GPC sites but its use was not mandatory and only KUMC, UTHSCSA, and UNMC used it in its entirety. Most other sites adapted parts of it to their existing processes as needed.

The other required component besides i2b2 was REDCap [4] or Research Electronic Data Capture. This is not an open source product but it is non-commercial software developed at Vanderbilt University for the purposes of conveniently capturing research data, especially surveys. All sites had their surveys on REDCap. At sites where paper and telephone responses were accepted, survey personnel manually filled in REDCap surveys on behalf of the respondents.

Data was extracted from the EMR using DataBuilder [5] collated into an analyzable tabular form using DataFinisher [6]. Results of analyzing the combined survey and EMR datasets will be reported in a forthcoming paper [manuscript in preparation]. The survey data and the EMR data extracted from i2b2 could be merged together because each record is keyed to the same non-informative index across the two data sources.

## Cohort Identification

Cohort selection was also done using i2b2, using the selection criteria shown in Table 2. Though GPC is currently deploying SHRINE **???**, at the time there was no query federation in place. The inclusion criteria had to be manually replicated at each site to the equivalent local codes. [placholder where I will write about how we came up with the criteria]. Once a query ran at a given site, that site’s i2b2 server would generate a “patient set”– a list of de-identified patient numbers. The informatics teams that run their respective i2b2 servers have crosswalk files matching the de-identified patient numbers to actual medical record numbers (MRN) or to database keys. Neither were shared externally, nor was there any reason to do so. The local informatics teams would use these identifiers to obtain names and contact information from their EMR systems. These contact lists were securely transmitted to the respective local HMRFS leads who uploaded them into REDCap and proceeded with recruitment. [placeholder to talk about the difficulties of using REDCap for purposes that are different from its intended use and the misconceptions surrounding what it’s intended use actually is].

It became necessary to send out a supplementary request to the study sites to acquire from their EMRs information about race, ethnicity, and class of insurance provider (private, Medicaid, employer, government, self-pay, etc.). Together with these data elements, information about income was requested. Since not all health systems record patient income, we had to rely on median household income for the census block group in which each patient’s address was located, as obtained from the 2013 American Community Survey **???**.

## Contacting Participants - A Natural Experiment

Individuals who were sent the survey invitation were contacted via a variety of means, including postal service, email, and through the patient portal feature of certain sites’ electronic medical record systems (EMR). The method of contact was selected by each HMRFS site PI based on logistics and local policy requirements. Allowing each site to select their own contact method allowed for a natural experiment where recruitment rates and other factors are compared for different methods of contact. For a detailed list of site and contact method see Table 3. Different contact methods posed different challenges. Traditional mail was costly in terms of staff time needed to validate addresses, stuff envelopes in addition to printing and postage costs. While mail would be sent in waves it was cheaper to send a larger sample all at once. Sites doing email recruitment had difficulties with invalid or missing email addresses with some sites simply not having enough email addresses in the EMR to accrue a reasonable sample. For most sites the eligibility cohort was selected from children’s records but with the goal of contacting the parents. This raised the concern of whether the email truly belonged to the parent rather than to a child. Sites were also affected by a time crunch secondary to IRB delays.

Table 3: Detailed list of site, adult/pediatric cohort, and contact method.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | MCW | UNMC | WISC | KUMC | MCRF | CMH | UTSW | UMN | IOWA | UTHSCSA |
| Cohort Makeup1 | A | A | A | P | P | P | P | P | P | P |
| Contact Method2 | Email | Email | Patient Portal | Email | Email | Email | Email | USPS | USPS | USPS |

1 Cohort makeup: A = adult only, P = pediatric only.

2 Contact method: USPS = United States Postal Service, Email = electronic mail on file, Portal = patient portal feature of the electronic medical record system.

Table 4a: Counts, Age, and BMI: Adult Index Patients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | n | Survey.1.or.2 = TRUE (%) | Survey.2 = TRUE (%) | Age (mean (sd)) | pat\_bmi\_raw (mean (sd)) |
| MCW | 6000 | 560 (9.3) | 560 (9.3) | 36.15 (7.74) | 28.98 (7.29) |
| UNMC | 9849 | 360 (3.7) | 360 (3.7) | 35.37 (8.40) | 30.02 (7.82) |
| WISC | 4999 | 728 (14.6) | 728 (14.6) | 47.82 (15.10) | 30.35 (7.27) |

Table 4b: Counts, Age, and BMI: Pediatric Index Patients

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | n | Survey.1.or.2 = TRUE (%) | Survey.2 = TRUE (%) | Age (mean (sd)) | pat\_bmi\_raw (mean (sd)) |
| CMH | 3411 | 235 (6.9) | 222 (6.5) | 9.86 (5.14) | 21.99 (6.04) |
| KUMC | 6705 | 396 (5.9) | 396 (5.9) | 14.83 (5.91) | 22.63 (6.40) |
| MCRF | 5010 | 405 (8.1) | 404 (8.1) | 9.47 (5.01) | 20.30 (5.24) |
| IOWA | 5016 | 205 (4.1) | 178 (3.5) | 11.61 (5.97) | 23.12 (6.82) |
| UMN | 4992 | 81 (1.6) | 81 (1.6) | 11.27 (5.91) | 22.01 (6.20) |
| UTHSCSA | 4328 | 174 (4.0) | 115 (2.7) | 12.33 (4.75) | 22.21 (6.23) |
| UTSW | 3959 | 95 (2.4) | 95 (2.4) | 12.02 (4.85) | 24.10 (7.84) |

## Analysis plan

# Results

The cohort flow chart with all sites combined is displayed in Figure 1. There were over xxxxx eligible individuals throughout the CDRN who (describe our flow chart here – who was included and who was excluded with representative sample sizes)￼. The final cohort includes 54,269 individuals who met inclusion/exclusion criteria. Cohort demographics (￼Table 5) indicate that across sites our index patients had a slight majority of females (53.48%), predominantly Caucasian (76.58%) and privately insured (47.67%).

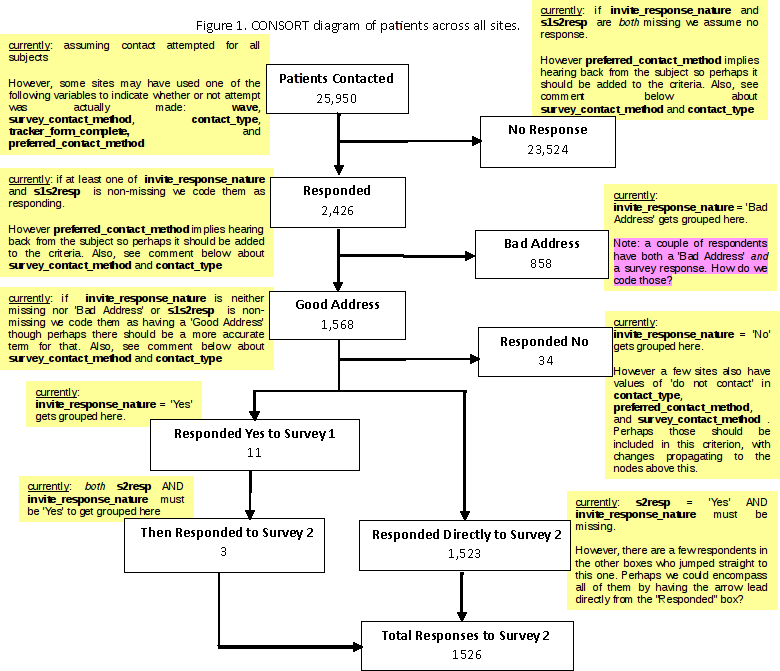


Figure 1: Current draft of consort diagram with comments.

Table 5: Cohort, Survey 1 and Survey 2 demographics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Cohort** | **Survey 1 or 2** | **Survey 2** | **p** |
| **n** | 54269 | 3239 | 3139 |  |
| **Sex = Male (%)** | 25169 (46.5) | 1404 ( 43.3) | 1349 ( 43.0) | \* |
| **Race (%)** |  |  |  | \* \* |
| Caucasian | 40434 (76.6) | 2808 ( 88.3) | 2753 ( 89.3) |  |
| African   American | 4437 ( 8.4) | 83 ( 2.6) | 72 ( 2.3) |  |
| Native   American | 250 ( 0.5) | 12 ( 0.4) | 11 ( 0.4) |  |
| Asian | 1396 ( 2.6) | 63 ( 2.0) | 62 ( 2.0) |  |
| Other | 4552 ( 8.6) | 170 ( 5.3) | 139 ( 4.5) |  |
| No Answer | 1730 ( 3.3) | 45 ( 1.4) | 45 ( 1.5) |  |
| **Hispanic = TRUE (%)** | 5290 ( 9.7) | 145 ( 4.5) | 118 ( 3.8) | \* \* |
| **Financial.Class (%)** |  |  |  | \* \* |
| Unknown | 5030 ( 9.5) | 299 ( 9.4) | 289 ( 9.4) |  |
| Medicare | 1018 ( 1.9) | 87 ( 2.7) | 87 ( 2.8) |  |
| Medicaid | 9031 (17.1) | 312 ( 9.8) | 284 ( 9.2) |  |
| Self-Pay | 11210 (21.2) | 895 ( 28.1) | 875 ( 28.4) |  |
| Private   Insurance | 25206 (47.7) | 1547 ( 48.6) | 1511 ( 49.0) |  |
| Other | 1385 ( 2.6) | 43 ( 1.4) | 38 ( 1.2) |  |
| **Income (mean (sd))** | 58.20 (25.03) | 61.52 (24.16) | 61.97 (23.98) | \* \* |
| **Age (mean (sd))** | 22.04 (15.43) | 27.00 (18.29) | 27.50 (18.35) | \* \* |
| **BMI (%)** |  |  |  | NS |
| Underweight | 130 ( 0.2) | 7 ( 0.2) | 7 ( 0.2) |  |
| Normal | 22360 (41.7) | 1335 ( 42.2) | 1292 ( 42.2) |  |
| Overweight | 15107 (28.2) | 854 ( 27.0) | 832 ( 27.2) |  |
| Obese | 15978 (29.8) | 964 ( 30.5) | 929 ( 30.4) |  |
| **Survey.1.or.2 = TRUE (%)** | 3239 ( 6.0) | 3239 (100.0) | 3139 (100.0) |  |
| **Survey.2 = TRUE (%)** | 3139 ( 5.8) | 3139 ( 96.9) | 3139 (100.0) |  |

…text omitted…

Table 6a: Participant demographics by site (Cohort)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CMH** | **KUMC** | **MCRF** | **MCW** | **IOWA** | **UMN** | **UNMC** | **UTHSCSA** | **UTSW** | **WISC** | **p** |
| **n** | 3411 | 6705 | 5010 | 6000 | 5016 | 4992 | 9849 | 4328 | 3959 | 4999 |  |
| **Sex = Male (%)** | 1727  (50.6) | 3252  (48.5) | 2535  (50.6) | 2999  (50.0) | 2508  (50.0) | 2496  (50.0) | 3523  (35.8) | 2393  (55.3) | 1911  (48.3) | 1825  (37.7) | \* \* |
| **Race (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Caucasian | 1876  (63.4) | 4584  (68.4) | 4587  (91.6) | 5070  (84.7) | 3875  (77.3) | 3589  (71.9) | 7483  (76.7) | 2744  (64.3) | 2108  (64.4) | 4518  (93.4) |  |
| African   American | 518  (17.5) | 772  (11.5) | 50  ( 1.0) | 331  ( 5.5) | 324  ( 6.5) | 532  (10.7) | 1280  (13.1) | 129  ( 3.0) | 393  (12.0) | 108  ( 2.2) |  |
| Native   American | 14  ( 0.5) | 17  ( 0.3) | 32  ( 0.6) | 28  ( 0.5) | 10  ( 0.2) | 57  ( 1.1) | 56  ( 0.6) | 2  ( 0.0) | 13  ( 0.4) | 21  ( 0.4) |  |
| Asian | 44  ( 1.5) | 126  ( 1.9) | 136  ( 2.7) | 251  ( 4.2) | 172  ( 3.4) | 227  ( 4.5) | 218  ( 2.2) | 43  ( 1.0) | 85  ( 2.6) | 94  ( 1.9) |  |
| Other | 492  (16.6) | 1063  (15.9) | 119  ( 2.4) | 295  ( 4.9) | 537  (10.7) | 62  ( 1.2) | 601  ( 6.2) | 1352  (31.7) | 26  ( 0.8) | 5  ( 0.1) |  |
| No Answer | 14  ( 0.5) | 136  ( 2.0) | 86  ( 1.7) | 9  ( 0.2) | 98  ( 2.0) | 525  (10.5) | 121  ( 1.2) | 0  ( 0.0) | 648  (19.8) | 93  ( 1.9) |  |
| **Hispanic = TRUE (%)** | 290  ( 8.5) | 602  ( 9.0) | 160  ( 3.2) | 239  ( 4.0) | 363  ( 7.2) | 200  ( 4.0) | 612  ( 6.2) | 1576  (36.4) | 1154  (29.1) | 94  ( 1.9) | \* \* |
| **Financial.Class (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Unknown | 46  ( 1.6) | 3170  (47.3) | 81  ( 1.6) | 0  ( 0.0) | 9  ( 0.2) | 0  ( 0.0) | 776  ( 7.9) | 604  (14.1) | 0  ( 0.0) | 344  ( 7.1) |  |
| Medicare | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 82  ( 1.4) | 0  ( 0.0) | 0  ( 0.0) | 468  ( 4.8) | 3  ( 0.1) | 23  ( 0.7) | 442  ( 9.1) |  |
| Medicaid | 1300  (43.9) | 1177  (17.6) | 1624  (32.4) | 204  ( 3.4) | 554  (11.0) | 194  ( 3.9) | 1073  (10.9) | 1509  (35.3) | 1294  (39.5) | 102  ( 2.1) |  |
| Self-Pay | 216  ( 7.3) | 0  ( 0.0) | 1  ( 0.0) | 3195  (53.4) | 2665  (53.1) | 400  ( 8.0) | 1233  (12.5) | 125  ( 2.9) | 390  (11.9) | 2985  (61.7) |  |
| Private   Insurance | 1310  (44.3) | 2351  (35.1) | 3304  (65.9) | 2456  (41.0) | 1764  (35.2) | 4398  (88.1) | 5530  (56.2) | 1685  (39.5) | 1442  (44.1) | 966  (20.0) |  |
| Other | 86  ( 2.9) | 0  ( 0.0) | 0  ( 0.0) | 47  ( 0.8) | 24  ( 0.5) | 0  ( 0.0) | 760  ( 7.7) | 344  ( 8.1) | 124  ( 3.8) | 0  ( 0.0) |  |
| **Income (mean (sd))** | 55.79  (26.66) | 63.01  (27.63) | 42.77  (15.32) | 62.89  (22.86) | NaN  (NA) | 67.02  (24.34) | 55.69  (24.34) | 51.69  (27.07) | 63.06  (26.69) | 65.30  (22.84) | \* \* |
| **Age (mean (sd))** | 9.86  (5.14) | 14.83  (5.91) | 9.47  (5.01) | 36.15  (7.74) | 11.61  (5.97) | 11.27  (5.91) | 35.37  (8.40) | 12.33  (4.75) | 12.02  (4.85) | 47.82  (15.10) | \* \* |
| **BMI (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Underweight | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 1  ( 0.0) | 23  ( 0.5) | 0  ( 0.0) | 16  ( 0.2) | 0  ( 0.0) | 76  ( 1.9) | 14  ( 0.3) |  |
| Normal | 1275  (37.4) | 4629  (69.0) | 2633  (52.6) | 2000  (33.3) | 1737  (34.6) | 1664  (33.3) | 2893  (29.4) | 2598  (60.0) | 1877  (47.4) | 1054  (24.5) |  |
| Overweight | 1060  (31.1) | 1061  (15.8) | 1377  (27.5) | 2001  (33.4) | 1554  (31.0) | 2282  (45.7) | 2832  (28.8) | 763  (17.6) | 828  (20.9) | 1349  (31.3) |  |
| Obese | 1076  (31.5) | 1015  (15.1) | 1000  (20.0) | 1997  (33.3) | 1700  (33.9) | 1046  (21.0) | 4108  (41.7) | 967  (22.3) | 1176  (29.7) | 1893  (43.9) |  |
| **Survey.1.or.2 = TRUE (%)** | 235  ( 6.9) | 396  ( 5.9) | 405  ( 8.1) | 560  ( 9.3) | 205  ( 4.1) | 81  ( 1.6) | 360  ( 3.7) | 174  ( 4.0) | 95  ( 2.4) | 728  (14.6) | \* \* |
| **Survey.2 = TRUE (%)** | 222  ( 6.5) | 396  ( 5.9) | 404  ( 8.1) | 560  ( 9.3) | 178  ( 3.5) | 81  ( 1.6) | 360  ( 3.7) | 115  ( 2.7) | 95  ( 2.4) | 728  (14.6) | \* \* |

Table 6b: Participant demographics by site (Responders)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CMH** | **KUMC** | **MCRF** | **MCW** | **IOWA** | **UMN** | **UNMC** | **UTHSCSA** | **UTSW** | **WISC** | **p** |
| **n** | 235 | 396 | 405 | 560 | 205 | 81 | 360 | 174 | 95 | 728 |  |
| **Sex = Male (%)** | 127  (54.0) | 193  ( 48.7) | 218  (53.8) | 241  ( 43.0) | 94  (45.9) | 47  ( 58.0) | 91  ( 25.3) | 95  (54.6) | 43  ( 45.3) | 255  ( 35.0) | \* \* |
| **Race (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Caucasian | 156  (78.8) | 319  ( 80.6) | 387  (95.6) | 515  ( 92.0) | 182  (88.8) | 69  ( 85.2) | 335  ( 93.6) | 102  (59.0) | 63  ( 81.8) | 680  ( 93.4) |  |
| African   American | 18  ( 9.1) | 24  ( 6.1) | 4  ( 1.0) | 7  ( 1.2) | 7  ( 3.4) | 0  ( 0.0) | 6  ( 1.7) | 4  ( 2.3) | 2  ( 2.6) | 11  ( 1.5) |  |
| Native   American | 0  ( 0.0) | 0  ( 0.0) | 1  ( 0.2) | 5  ( 0.9) | 1  ( 0.5) | 0  ( 0.0) | 0  ( 0.0) | 1  ( 0.6) | 0  ( 0.0) | 4  ( 0.5) |  |
| Asian | 2  ( 1.0) | 7  ( 1.8) | 4  ( 1.0) | 18  ( 3.2) | 4  ( 2.0) | 2  ( 2.5) | 5  ( 1.4) | 2  ( 1.2) | 2  ( 2.6) | 17  ( 2.3) |  |
| Other | 22  (11.1) | 42  ( 10.6) | 8  ( 2.0) | 13  ( 2.3) | 8  ( 3.9) | 0  ( 0.0) | 12  ( 3.4) | 64  (37.0) | 0  ( 0.0) | 1  ( 0.1) |  |
| No Answer | 0  ( 0.0) | 4  ( 1.0) | 1  ( 0.2) | 2  ( 0.4) | 3  ( 1.5) | 10  ( 12.3) | 0  ( 0.0) | 0  ( 0.0) | 10  ( 13.0) | 15  ( 2.1) |  |
| **Hispanic = TRUE (%)** | 12  ( 5.1) | 18  ( 4.5) | 10  ( 2.5) | 14  ( 2.5) | 7  ( 3.4) | 1  ( 1.2) | 8  ( 2.2) | 52  (29.9) | 7  ( 7.4) | 16  ( 2.2) | \* \* |
| **Financial.Class (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Unknown | 2  ( 1.0) | 189  ( 47.7) | 4  ( 1.0) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 20  ( 5.6) | 26  (15.0) | 0  ( 0.0) | 58  ( 8.0) |  |
| Medicare | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 8  ( 1.4) | 0  ( 0.0) | 0  ( 0.0) | 8  ( 2.2) | 0  ( 0.0) | 0  ( 0.0) | 71  ( 9.8) |  |
| Medicaid | 68  (34.3) | 46  ( 11.6) | 84  (20.7) | 12  ( 2.1) | 10  ( 4.9) | 0  ( 0.0) | 13  ( 3.6) | 46  (26.6) | 16  ( 20.8) | 17  ( 2.3) |  |
| Self-Pay | 16  ( 8.1) | 0  ( 0.0) | 0  ( 0.0) | 296  ( 52.9) | 110  (53.7) | 6  ( 7.4) | 26  ( 7.2) | 1  ( 0.6) | 7  ( 9.1) | 433  ( 59.5) |  |
| Private   Insurance | 108  (54.5) | 161  ( 40.7) | 317  (78.3) | 241  ( 43.0) | 84  (41.0) | 75  ( 92.6) | 275  ( 76.4) | 88  (50.9) | 49  ( 63.6) | 149  ( 20.5) |  |
| Other | 4  ( 2.0) | 0  ( 0.0) | 0  ( 0.0) | 3  ( 0.5) | 1  ( 0.5) | 0  ( 0.0) | 18  ( 5.0) | 12  ( 6.9) | 5  ( 6.5) | 0  ( 0.0) |  |
| **Income (mean (sd))** | 61.84  (28.97) | 70.59  (26.33) | 44.92  (13.15) | 64.57  (22.02) | NaN  (NA) | 75.49  (20.59) | 61.76  (24.13) | 55.75  (29.49) | 72.86  (27.22) | 65.32  (22.99) | \* \* |
| **Age (mean (sd))** | 9.69  (5.07) | 13.90  (5.67) | 9.51  (4.80) | 37.32  (7.64) | 11.22  (5.65) | 9.80  (6.29) | 37.12  (8.29) | 12.82  (4.77) | 11.81  (3.88) | 48.26  (15.00) | \* \* |
| **BMI (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Underweight | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 3  ( 1.5) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 1  ( 1.1) | 3  ( 0.5) |  |
| Normal | 83  (35.3) | 291  ( 73.5) | 230  (56.8) | 186  ( 33.2) | 92  (44.9) | 26  ( 32.1) | 120  ( 33.3) | 108  (62.1) | 49  ( 51.6) | 150  ( 23.1) |  |
| Overweight | 80  (34.0) | 54  ( 13.6) | 102  (25.2) | 178  ( 31.8) | 58  (28.3) | 42  ( 51.9) | 88  ( 24.4) | 34  (19.5) | 11  ( 11.6) | 207  ( 31.9) |  |
| Obese | 72  (30.6) | 51  ( 12.9) | 73  (18.0) | 196  ( 35.0) | 52  (25.4) | 13  ( 16.0) | 152  ( 42.2) | 32  (18.4) | 34  ( 35.8) | 289  ( 44.5) |  |
| **Survey.2 = TRUE (%)** | 222  (94.5) | 396  (100.0) | 404  (99.8) | 560  (100.0) | 178  (86.8) | 81  (100.0) | 360  (100.0) | 115  (66.1) | 95  (100.0) | 728  (100.0) | \* \* |

Table 6c: Participant demographics by site (Completers)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CMH** | **KUMC** | **MCRF** | **MCW** | **IOWA** | **UMN** | **UNMC** | **UTHSCSA** | **UTSW** | **WISC** | **p** |
| **n** | 222 | 396 | 404 | 560 | 178 | 81 | 360 | 115 | 95 | 728 |  |
| **Sex = Male (%)** | 120  (54.1) | 193  (48.7) | 217  (53.7) | 241  (43.0) | 81  (45.5) | 47  (58.0) | 91  (25.3) | 61  (53.0) | 43  (45.3) | 255  (35.0) | \* \* |
| **Race (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Caucasian | 148  (79.6) | 319  (80.6) | 386  (95.5) | 515  (92.0) | 163  (91.6) | 69  (85.2) | 335  (93.6) | 75  (65.8) | 63  (81.8) | 680  (93.4) |  |
| African   American | 14  ( 7.5) | 24  ( 6.1) | 4  ( 1.0) | 7  ( 1.2) | 4  ( 2.2) | 0  ( 0.0) | 6  ( 1.7) | 0  ( 0.0) | 2  ( 2.6) | 11  ( 1.5) |  |
| Native   American | 0  ( 0.0) | 0  ( 0.0) | 1  ( 0.2) | 5  ( 0.9) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 1  ( 0.9) | 0  ( 0.0) | 4  ( 0.5) |  |
| Asian | 2  ( 1.1) | 7  ( 1.8) | 4  ( 1.0) | 18  ( 3.2) | 3  ( 1.7) | 2  ( 2.5) | 5  ( 1.4) | 2  ( 1.8) | 2  ( 2.6) | 17  ( 2.3) |  |
| Other | 22  (11.8) | 42  (10.6) | 8  ( 2.0) | 13  ( 2.3) | 5  ( 2.8) | 0  ( 0.0) | 12  ( 3.4) | 36  (31.6) | 0  ( 0.0) | 1  ( 0.1) |  |
| No Answer | 0  ( 0.0) | 4  ( 1.0) | 1  ( 0.2) | 2  ( 0.4) | 3  ( 1.7) | 10  (12.3) | 0  ( 0.0) | 0  ( 0.0) | 10  (13.0) | 15  ( 2.1) |  |
| **Hispanic = TRUE (%)** | 12  ( 5.4) | 18  ( 4.5) | 9  ( 2.2) | 14  ( 2.5) | 4  ( 2.2) | 1  ( 1.2) | 8  ( 2.2) | 29  (25.2) | 7  ( 7.4) | 16  ( 2.2) | \* \* |
| **Financial.Class (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Unknown | 2  ( 1.1) | 189  (47.7) | 4  ( 1.0) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 20  ( 5.6) | 16  (14.0) | 0  ( 0.0) | 58  ( 8.0) |  |
| Medicare | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 8  ( 1.4) | 0  ( 0.0) | 0  ( 0.0) | 8  ( 2.2) | 0  ( 0.0) | 0  ( 0.0) | 71  ( 9.8) |  |
| Medicaid | 62  (33.3) | 46  (11.6) | 83  (20.5) | 12  ( 2.1) | 8  ( 4.5) | 0  ( 0.0) | 13  ( 3.6) | 27  (23.7) | 16  (20.8) | 17  ( 2.3) |  |
| Self-Pay | 14  ( 7.5) | 0  ( 0.0) | 0  ( 0.0) | 296  (52.9) | 92  (51.7) | 6  ( 7.4) | 26  ( 7.2) | 1  ( 0.9) | 7  ( 9.1) | 433  (59.5) |  |
| Private   Insurance | 104  (55.9) | 161  (40.7) | 317  (78.5) | 241  (43.0) | 77  (43.3) | 75  (92.6) | 275  (76.4) | 63  (55.3) | 49  (63.6) | 149  (20.5) |  |
| Other | 4  ( 2.2) | 0  ( 0.0) | 0  ( 0.0) | 3  ( 0.5) | 1  ( 0.6) | 0  ( 0.0) | 18  ( 5.0) | 7  ( 6.1) | 5  ( 6.5) | 0  ( 0.0) |  |
| **Income (mean (sd))** | 63.83  (28.65) | 70.59  (26.33) | 45.01  (13.05) | 64.57  (22.02) | NaN  (NA) | 75.49  (20.59) | 61.76  (24.13) | 60.00  (30.64) | 72.86  (27.22) | 65.32  (22.99) | \* \* |
| **Age (mean (sd))** | 9.68  (5.08) | 13.90  (5.67) | 9.51  (4.81) | 37.32  (7.64) | 11.42  (5.68) | 9.80  (6.29) | 37.12  (8.29) | 12.91  (4.84) | 11.81  (3.88) | 48.26  (15.00) | \* \* |
| **BMI (%)** |  |  |  |  |  |  |  |  |  |  | \* \* |
| Underweight | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 3  ( 1.7) | 0  ( 0.0) | 0  ( 0.0) | 0  ( 0.0) | 1  ( 1.1) | 3  ( 0.5) |  |
| Normal | 81  (36.5) | 291  (73.5) | 230  (56.9) | 186  (33.2) | 85  (47.8) | 26  (32.1) | 120  (33.3) | 74  (64.3) | 49  (51.6) | 150  (23.1) |  |
| Overweight | 77  (34.7) | 54  (13.6) | 102  (25.2) | 178  (31.8) | 49  (27.5) | 42  (51.9) | 88  (24.4) | 24  (20.9) | 11  (11.6) | 207  (31.9) |  |
| Obese | 64  (28.8) | 51  (12.9) | 72  (17.8) | 196  (35.0) | 41  (23.0) | 13  (16.0) | 152  (42.2) | 17  (14.8) | 34  (35.8) | 289  (44.5) |  |

Sample size and demographics varied by site (Table 6a). Sample size by site varied from a maximum of 9,849 (UNMC) to a minimum of 3,411 (CMH). The proportion of the cohort that was male at each site varied from 55.30% to 35.80%. The proportions of African Americans (17.50% - 1.00%), Native Americans (1.10% - 0.00%), and Asians (4.50% - 1.00%) varied as well. The most common insurance categories were self-pay (61.70% - 0.00%), Medicaid (43.90% - 2.10%) and private insurance (88.10% - 20.00%). Estimated median household income ranged from $85,400 to $98,400 and mean age for pediatric index patients ranged from 9.5 to 14.8 years old (Table 4b). Three sites targeted adults as they did not have sufficient children in their systems. The mean age of their index patients ranged from 35.4 to 47.8 years old (Table 4a). The median body mass index (BMI) for these adult patients (as obtained from the most recent EMR record prior to initiation of recruitment) ranged from 27.3 to 29.1.

Responders to Survey 1 or 2 (Table 5) were also primarily female (56.65%), Caucasian (88.27%), and privately insured (48.60%) with a mean age of 27 (SD:18.3) years and 42.25% had a BMI in the normal range. Survey 2 responders were 45.00% female, 55.56% Caucasian, and 36.36% privately insured, with a mean age of 11.5 (SD:5.1) years and 43.00% had a normal BMI.

## EHR Predictors of Participation

Prior to analysis the 54,269 records were randomly assigned to a development (N=10,751), validation (N=10,748), and test subsets (N=32,770) in order to avoid overfitting and bias due to within-sample testing. All analysis decisions were made based on the developmental subset. The first goal was to identify candidate predictors for survey participation from among the variables available for all members of the cohort (i.e. those listed in Table 5).

Accordingly, for each candidate predictor, a separate logistic regression model was fit to the developmental subset with responder status as the outcome. Discrete variables with multiple levels were broken up into an equal number of indicator variables. This not only informed the decision of which predictor and random-effect variables but also in some cases how condense their levels. Table 8 shows the results of these 85 tests.

Table 7: Preliminary assessment of variables via univariate logistic regression.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | estimate | conf.low | conf.high | p.value |
| **Adult Patient** | 1.5851 | 1.3483 | 1.864 | \* \* \* |
| **Age** | 1.0189 | 1.0141 | 1.0237 | \* \* \* |
| **BMI.NA** | 3.0422 | 1.8688 | 4.7287 | \* \* \* |
| BMI.Normal | 0.90692 | 0.76825 | 1.0689 |  |
| BMI.Obese | 1.053 | 0.88136 | 1.2534 |  |
| BMI.Overweight | 0.95851 | 0.79642 | 1.1482 |  |
| BMI percentile | 0.99565 | 0.99161 | 0.99982 |  |
| BMI raw | 1.0093 | 0.99897 | 1.0194 |  |
| BMI.Underweight | 7.6322e-06 | NA | 0.29075 |  |
| **Hispanic** | 0.4162 | 0.27489 | 0.60291 | \* \* \* |
| **Income** | 1.0061 | 1.0025 | 1.0098 | \* |
| **Insurance.Medicaid** | 0.4661 | 0.3471 | 0.61303 | \* \* \* |
| Insurance.Medicare | 1.6309 | 0.97848 | 2.5619 |  |
| Insurance.NA | 0.65398 | 0.33477 | 1.1436 |  |
| Insurance.Other | 0.30243 | 0.10749 | 0.66027 |  |
| Insurance.Private.Insurance | 1.0389 | 0.88353 | 1.2211 |  |
| **Insurance.Self.Pay** | 1.6527 | 1.3777 | 1.9749 | \* \* \* |
| Insurance.Unknown | 0.9637 | 0.72081 | 1.2641 |  |
| Male | 1.0079 | 0.85688 | 1.1848 |  |
| **Pediatric Site** | 0.59429 | 0.5054 | 0.69871 | \* \* \* |
| **Race.African.American** | 0.3099 | 0.1831 | 0.48802 | \* \* \* |
| Race.Asian | 0.85263 | 0.47242 | 1.413 |  |
| **Race.Caucasian** | 2.1993 | 1.7524 | 2.7942 | \* \* \* |
| Race.NA | 0.61639 | 0.3157 | 1.0771 |  |
| Race.Native.American | 2.8022e-06 | 9.8513e-38 | 0.054829 |  |
| Race.No.Answer | 0.57108 | 0.29268 | 0.99694 |  |
| Race.Other | 0.57775 | 0.39434 | 0.8161 |  |
| Recruitment.email | 0.86115 | 0.73026 | 1.0173 |  |
| **Recruitment.mychart** | 3.2633 | 2.6699 | 3.9679 | \* \* \* |
| **Recruitment.post** | 0.50482 | 0.40334 | 0.62549 | \* \* \* |
| Site.CMH | 1.1751 | 0.84128 | 1.5997 |  |
| Site.IOWA | 0.82455 | 0.60402 | 1.1003 |  |
| Site.KUMC | 0.8295 | 0.63177 | 1.0712 |  |
| **Site.MCRF** | 1.5015 | 1.1745 | 1.8969 | \* |
| **Site.MCW** | 1.7232 | 1.3789 | 2.1346 | \* \* \* |
| **Site.UMN** | 0.2497 | 0.14512 | 0.39776 | \* \* \* |
| **Site.UNMC** | 0.5009 | 0.38221 | 0.64552 | \* \* \* |
| Site.UTHSCSA | 0.6368 | 0.43943 | 0.89186 |  |
| **Site.UTSW** | 0.35361 | 0.20882 | 0.55725 | \* \* \* |
| **Site.WISC** | 3.2633 | 2.6699 | 3.9679 | \* \* \* |

## HMRFS Survey

3,139 individuals completed the HMRFS survey across all sites. Of these respondents, most self-reported as female (78.17%), Caucasian (92.04%), and having private insurance (78.60%). The modal range of self-reported household income was approximately $50,000-$100,000/year and average age was 42.2 years with a baseline mean BMI of 28.85 (7.48). These data indicate the sample that responded to the HMRFS survey was similar to the overall cohort of eligible patients. Responses indicate that across all sites (Table 8a), 29.44% of respondents had participated previously in medical research. Respondents were generally open to investigators contacting them about possible participation in studies (42.71% yes; 38.93% maybe). Key factors in making the decision to participate in medical research included: topic of research (62.98%), time (48.20%), doctor recommendation (22.05%), reimbursement (17.23%), and child involvement (23.89%). Of respondents who had children (64.13%), they were also generally positive about possibly allowing their child to take part in medical research (27.21% yes; 27.21% maybe) (Table 8b). The factors for deciding whether their children could participate in research followed the same order of priority as above. Most participants were in favor of their medical information being used for research (35.32% fantastic idea, 48.61% good idea) and even more so if the data being used for medical research is de-identified (32.45% fantastic idea, 49.55% good idea).

Table 8a: Survey responses.

|  |  |
| --- | --- |
|  | Overall |
| **n** | 3139 |
| **research (%)** |  |
| PreferNotAnswer | 3 ( 0.1) |
| No | 1838 (58.6) |
| Unsure | 374 (11.9) |
| Yes | 924 (29.4) |
| **possible\_research (%)** |  |
| PreferNotAnswer | 69 ( 2.2) |
| No | 504 (16.1) |
| Maybe | 1215 (38.9) |
| Yes | 1333 (42.7) |
| **PR\_Me\_DependsAbout = TRUE (%)** | 1977 (63.0) |
| **PR\_Me\_If\_Spec = TRUE (%)** | 240 ( 7.6) |
| **PR\_Me\_Time = TRUE (%)** | 1513 (48.2) |
| **PR\_Me\_Doctor\_Op = TRUE (%)** | 692 (22.0) |
| **PR\_Me\_Compensation = TRUE (%)** | 541 (17.2) |
| **PR\_Me\_Involve\_Child = TRUE (%)** | 750 (23.9) |
| **PR\_Me\_Other = TRUE (%)** | 70 ( 2.2) |
| **children\_in\_home (%)** |  |
| PreferNotAnswer | 13 ( 0.4) |
| No | 1011 (32.4) |
| None | 95 ( 3.0) |
| Yes | 2001 (64.1) |
| **res\_talk\_family (%)** |  |
| PreferNotAnswer | 35 ( 1.1) |
| No | 1102 (35.5) |
| Unsure | 967 (31.2) |
| Yes | 1000 (32.2) |
| **research\_feeling (%)** |  |
| Unsure | 331 (10.7) |
| PreferNotAnswer | 23 ( 0.7) |
| NotGoodIdea | 30 ( 1.0) |
| Fantastic | 1095 (35.3) |
| GoodIdea | 1507 (48.6) |
| Terrible | 12 ( 0.4) |
| Other | 102 ( 3.3) |
| **deid\_data (%)** |  |
| Unsure | 379 (12.2) |
| PreferNotAnswer | 21 ( 0.7) |
| Fantastic | 1004 (32.4) |
| GoodIdea | 1533 (49.5) |
| Terrible | 28 ( 0.9) |
| NotGoodIdea | 72 ( 2.3) |
| Other | 57 ( 1.8) |
| **height\_in (mean (sd))** | 66.26 (4.80) |
| **weight\_value\_lbs (mean (sd))** | 181.12 (57.02) |
| **bp\_hypten\_self (%)** |  |
| PreferNotAnswer | 9 ( 0.3) |
| No | 2479 (80.8) |
| Unsure | 64 ( 2.1) |
| Yes | 516 (16.8) |
| **chole\_triigl\_hyperlip\_self (%)** |  |
| PreferNotAnswer | 11 ( 0.4) |
| No | 2281 (74.6) |
| Unsure | 159 ( 5.2) |
| Yes | 605 (19.8) |
| **bloodsugar\_diabetes\_self (%)** |  |
| PreferNotAnswer | 9 ( 0.3) |
| No | 2670 (87.5) |
| Unsure | 95 ( 3.1) |
| Yes | 276 ( 9.0) |
| **cancer\_anytype\_self (%)** |  |
| Unsure | 80 ( 2.6) |
| PreferNotAnswer | 13 ( 0.4) |
| No | 2691 (88.8) |
| None | 95 ( 3.1) |
| Yes | 150 ( 5.0) |
| **hpb\_hprtnsn (%)** |  |
| PreferNotAnswer | 10 ( 0.3) |
| No | 888 (29.0) |
| Unsure | 200 ( 6.5) |
| Yes | 1967 (64.2) |
| **chole\_trig\_hyperlip (%)** |  |
| PreferNotAnswer | 11 ( 0.4) |
| No | 973 (32.0) |
| Unsure | 341 (11.2) |
| Yes | 1718 (56.5) |
| **elev\_bs\_diabetes (%)** |  |
| PreferNotAnswer | 10 ( 0.3) |
| No | 1488 (49.1) |
| Unsure | 188 ( 6.2) |
| Yes | 1344 (44.4) |
| **cancer\_anytype (%)** |  |
| Unsure | 140 ( 4.6) |
| PreferNotAnswer | 12 ( 0.4) |
| No | 1388 (45.7) |
| None | 95 ( 3.1) |
| Yes | 1402 (46.2) |
| **sex (%)** |  |
| PreferNotAnswer | 12 ( 0.4) |
| Male | 662 (21.4) |
| Female | 2421 (78.2) |
| Other | 2 ( 0.1) |
| **age (mean (sd))** | 42.23 (12.19) |
| **latino\_origin (%)** |  |
| 5 | 4 ( 0.1) |
| PreferNotAnswer | 25 ( 0.8) |
| No | 2880 (94.1) |
| Yes | 150 ( 4.9) |
| **White = TRUE (%)** | 2889 (92.0) |
| **Black = TRUE (%)** | 76 ( 2.4) |
| **American\_Indian = TRUE (%)** | 37 ( 1.2) |
| **Asian = TRUE (%)** | 68 ( 2.2) |
| **OtherRace = TRUE (%)** | 39 ( 1.2) |
| **PreferNotAnswer = TRUE (%)** | 28 ( 0.9) |
| **income (%)** |  |
| 0-$24999 | 274 ( 8.9) |
| $25000-$49999 | 481 (15.6) |
| $50000-$99999 | 1002 (32.4) |
| $100000-$199999 | 765 (24.7) |
| $200000+ | 182 ( 5.9) |
| PreferNotAnswer | 359 (11.6) |
| Unsure | 30 ( 1.0) |
| **insurance (%)** |  |
| Unsure | 28 ( 0.9) |
| Uninsured | 62 ( 2.0) |
| PreferNotAnswer | 55 ( 1.8) |
| MedicareMedicaid | 322 (10.4) |
| Other | 194 ( 6.3) |
| Private | 2428 (78.6) |
| **education (%)** |  |
| HighSchool | 233 ( 7.5) |
| Bachelors | 1028 (33.2) |
| Doctorate | 218 ( 7.0) |
| Masters | 559 (18.1) |
| Associates | 385 (12.4) |
| SomeCollege | 518 (16.7) |
| SomeHighSchool | 43 ( 1.4) |
| NoHighSchool | 33 ( 1.1) |
| PreferNotAnswer | 34 ( 1.1) |
| Other | 43 ( 1.4) |
| **household (%)** |  |
| I live alone | 208 ( 6.7) |
| 1-2 | 1259 (40.7) |
| 3-4 | 1200 (38.8) |
| 5-6 | 361 (11.7) |
| 7+ | 43 ( 1.4) |
| PreferNotAnswer | 21 ( 0.7) |
| **language = Yes (%)** | 179 ( 6.0) |
| **a\_bmi (mean (sd))** | 28.85 (7.48) |

Table 8b: Survey responses regarding children.

|  |  |
| --- | --- |
|  | Overall |
| **n** | 2001 |
| **children\_research (%)** |  |
| Maybe | 797 (40.0) |
| No | 613 (30.8) |
| No\_Kids | 4 ( 0.2) |
| PreferNotAnswer | 36 ( 1.8) |
| Yes | 542 (27.2) |
| **PR\_Child\_DependsAbout = TRUE (%)** | 1132 (56.6) |
| **PR\_Child\_If\_Spec = TRUE (%)** | 133 ( 6.6) |
| **PR\_Child\_TiChild = TRUE (%)** | 771 (38.5) |
| **PR\_Child\_Doctor\_Op = TRUE (%)** | 529 (26.4) |
| **PR\_Child\_Compensation = TRUE (%)** | 247 (12.3) |
| **PR\_Child\_Involve\_Child = TRUE (%)** | 275 (13.7) |
| **PR\_Child\_Other = TRUE (%)** | 58 ( 2.9) |

# Discussion

## Lessons Learned

The goal of this experiment was to show whether and how we could use the electronic record across institutions to reach out to patients to answer research questions. Variation in some methodologies (such as method of contact) by site was allowed to facilitate the process and test various approaches the problem (a natural experiment). Through this process we learned many lessons to be applied in the future.

## Different contact methods posed different challenges.

For a detailed list of site and contact method see Table 3.

1. Traditional mail. Traditional mail was costly in terms of staff time needed to validate addresses, stuff envelopes in addition to printing and postage costs. Mail surveys are traditionally sent in waves to the same sample with diminishing returns with each follow up. The original design was then to send two to three contacts to each individual within stratified groups. Then if certain group were not filled addition potential participants within that strata would be contacted. This process however was time consuming and costly. When paying others to stuff and address envelopes, there were setup costs each time surveys were sent. Thus it was cheaper and more expedient to send a larger sample all at once rather than smaller samples. As a result, some sites chose to send more survey initially to a larger sample rather than repeated survey to a smaller sample in order to get the sample number of responses. However, this has a negative effect on response rates with a larger denominator. Incorrect addresses within the system remained a difficulty.
2. Email. Sites doing email recruitment had difficulties with invalid or missing email addresses. Some sites simply did not have enough email addresses in the EMR to accrue a reasonable sample and thus chose to use traditional mail. For most sites the eligibility cohort was selected from children’s records but with the goal of contacting the parents. This raised the concern of whether the email belonged to the parent rather than to a child.
3. Contact from within the EMR (e.g. My Chart). One site sent out messages inviting participants to connect to the survey through the electronic medical record (a form of secure email that is retained in the medical record). This required many more levels of permission. The institution initially said no. However it was the patient advisory board who got it approved as they believed the surveys were part of what should be done in a research institution. Potential benefits are that this form of communication is seen as more reputable and not likely to be dismissed as a phishing attempt. However patients may not check their account in the medical record frequently and often lose their password if they have not been seen recent or used the system recently.

## Using a Central IRB

## Obtaining data from the EMR – adapting to site differences

# Conclusion

# Supplementary Materials

Table S1: Participant demographics by recruitment method (Cohort)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **email** | **post** | **mychart** | **p** |
| **n** | 34934 | 14336 | 4999 |  |
| **Sex = Male (%)** | 15947 (45.7) | 7397 (51.6) | 1825 (37.7) | \* \* |
| **Race (%)** |  |  |  | \* \* |
| Caucasian | 25708 (76.3) | 10208 (71.5) | 4518 (93.4) |  |
| African   American | 3344 ( 9.9) | 985 ( 6.9) | 108 ( 2.2) |  |
| Native   American | 160 ( 0.5) | 69 ( 0.5) | 21 ( 0.4) |  |
| Asian | 860 ( 2.6) | 442 ( 3.1) | 94 ( 1.9) |  |
| Other | 2596 ( 7.7) | 1951 (13.7) | 5 ( 0.1) |  |
| No Answer | 1014 ( 3.0) | 623 ( 4.4) | 93 ( 1.9) |  |
| **Hispanic = TRUE (%)** | 3057 ( 8.8) | 2139 (14.9) | 94 ( 1.9) | \* \* |
| **Financial.Class (%)** |  |  |  | \* \* |
| Unknown | 4073 (12.1) | 613 ( 4.3) | 344 ( 7.1) |  |
| Medicare | 573 ( 1.7) | 3 ( 0.0) | 442 ( 9.1) |  |
| Medicaid | 6672 (19.8) | 2257 (15.8) | 102 ( 2.1) |  |
| Self-Pay | 5035 (14.9) | 3190 (22.3) | 2985 (61.7) |  |
| Private   Insurance | 16393 (48.6) | 7847 (55.0) | 966 (20.0) |  |
| Other | 1017 ( 3.0) | 368 ( 2.6) | 0 ( 0.0) |  |
| **Income (mean (sd))** | 56.33 (24.51) | 59.95 (26.75) | 65.30 (22.84) | \* \* |
| **Age (mean (sd))** | 22.71 (13.71) | 11.71 (5.62) | 47.82 (15.10) | \* \* |
| **BMI (%)** |  |  |  | \* \* |
| Underweight | 93 ( 0.3) | 23 ( 0.2) | 14 ( 0.3) |  |
| Normal | 15307 (43.8) | 5999 (41.9) | 1054 (24.5) |  |
| Overweight | 9159 (26.2) | 4599 (32.1) | 1349 (31.3) |  |
| Obese | 10372 (29.7) | 3713 (25.9) | 1893 (43.9) |  |
| **Survey.1.or.2 = TRUE (%)** | 2051 ( 5.9) | 460 ( 3.2) | 728 (14.6) | \* \* |
| **Survey.2 = TRUE (%)** | 2037 ( 5.8) | 374 ( 2.6) | 728 (14.6) | \* \* |

Table S2: Variables for consort diagram.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **CMH** | **KUMC** | **MCRF** | **MCW** | **IOWA** | **UMN** | **UNMC** | **UTHSCSA** | **UTSW** | **WISC** |
| **n** | 3411 | 6705 | 5010 | 6000 | 5016 | 4992 | 9849 | 4328 | 3959 | 4999 |
| **contact\_type** |  |  |  |  |  |  |  |  |  |  |
| BILLING | 0 | 437 | 4985 | 0 | 0 | 0 | 0 | 2781 | 0 | 0 |
| do not   contact | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| email | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1234 | 0 |
| EMERG1 | 0 | 1362 | 0 | 0 | 0 | 0 | 0 | 262 | 0 | 0 |
| EMERG2 | 0 | 114 | 0 | 0 | 0 | 0 | 0 | 61 | 0 | 0 |
| FATHER | 0 | 287 | 7 | 0 | 0 | 0 | 0 | 130 | 0 | 0 |
| GUARDIA | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GUARDIAN | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |
| mail | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2723 | 0 |
| MOTHER | 0 | 853 | 13 | 0 | 0 | 0 | 0 | 764 | 0 | 0 |
| None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4999 |
| PAT | 0 | 3649 | 4 | 0 | 0 | 0 | 9849 | 326 | 0 | 0 |
| NA | 3411 | 0 | 0 | 6000 | 5016 | 4992 | 0 | 0 | 0 | 0 |
| **wave** |  |  |  |  |  |  |  |  |  |  |
| 8/29/15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1226 | 0 |
| None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4999 |
| NA | 3411 | 6705 | 5010 | 6000 | 5016 | 4992 | 9849 | 4328 | 2733 | 0 |
| **invite\_response\_nature** |  |  |  |  |  |  |  |  |  |  |
| Bad Address | 183 | 1319 | 310 | 72 | 452 | 0 | 316 | 754 | 67 | 0 |
| None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4999 |
| No | 3 | 3 | 2 | 9 | 44 | 0 | 18 | 63 | 5 | 0 |
| Yes | 13 | 0 | 4 | 0 | 42 | 0 | 0 | 110 | 0 | 0 |
| NA | 3212 | 5383 | 4694 | 5919 | 4478 | 4992 | 9515 | 3401 | 3887 | 0 |
| **preferred\_contact\_method** |  |  |  |  |  |  |  |  |  |  |
| do not   contact | 0 | 0 | 0 | 0 | 24 | 0 | 0 | 38 | 0 | 0 |
| email | 10 | 0 | 0 | 0 | 25 | 0 | 0 | 69 | 0 | 0 |
| mail | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4999 |
| telephone | 0 | 0 | 3 | 0 | 16 | 0 | 0 | 34 | 0 | 0 |
| NA | 3399 | 6705 | 5007 | 6000 | 4951 | 4992 | 9849 | 4186 | 3959 | 0 |
| **survey\_contact\_method** |  |  |  |  |  |  |  |  |  |  |
| do not   contact | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 5 | 0 | 0 |
| email | 0 | 0 | 1 | 0 | 22 | 0 | 0 | 44 | 0 | 0 |
| mail | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 64 | 0 | 0 |
| None | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4999 |
| telephone | 0 | 0 | 3 | 0 | 16 | 0 | 0 | 20 | 0 | 0 |
| NA | 3411 | 6705 | 5006 | 6000 | 4977 | 4992 | 9849 | 4195 | 3959 | 0 |
| **tracker\_form\_complete** |  |  |  |  |  |  |  |  |  |  |
| Partial | 0 | 0 | 0 | 0 | 1 | 4992 | 0 | 1 | 3 | 0 |
| Complete | 3411 | 0 | 5 | 0 | 536 | 0 | 0 | 587 | 2 | 0 |
| NA | 0 | 6705 | 5005 | 6000 | 4479 | 0 | 9849 | 3740 | 3954 | 4999 |
| **s2resp = Yes** | 222 | 396 | 404 | 560 | 178 | 81 | 360 | 115 | 95 | 728 |
| **s1s2resp** |  |  |  |  |  |  |  |  |  |  |
| No | 184 | 1322 | 312 | 79 | 496 | 0 | 334 | 817 | 72 | 4271 |
| Yes | 235 | 396 | 405 | 560 | 205 | 81 | 360 | 174 | 95 | 728 |
| NA | 2992 | 4987 | 4293 | 5361 | 4315 | 4911 | 9155 | 3337 | 3792 | 0 |

# References

[1] R. L. Fleurence, L. H. Curtis, R. M. Califf, R. Platt, J. V. Selby, and J. S. Brown, “Launching PCORnet, a national patient-centered clinical research network,” *Journal of the American Medical Informatics Association*, vol. 21, no. 4, pp. 578–582, Jul. 2014.

[2] D. R. Young *et al.*, “The Patient Outcomes Research To Advance Learning (PORTAL) Network Adult Overweight and Obesity Cohort: Development and Description,” *JMIR Research Protocols*, vol. 5, no. 2, p. e87, Jun. 2016.

[3] S. Murphy *et al.*, “Instrumenting the health care enterprise for discovery research in the genomic era,” *Genome research*, vol. 19, no. 9, pp. 1675–1681, 2009.

[4] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, “Research electronic data capture (REDCap) metadata-driven methodology and workflow process for providing translational research informatics support,” *Journal of Biomedical Informatics*, vol. 42, no. 2, pp. 377–381, Apr. 2009.

[5] B. Adagarla *et al.*, “SEINE: Methods for Electronic Data Capture and Integrated Data Repository Synthesis with Patient Registry Use Cases,” 2015.

[6] A. Bokov, L. Manuel, C. Cheng, A. Bos, and A. Tirado-Ramos, “Denormalize and Delimit: How not to Make Data Extraction for Analysis More Complex than Necessary,” *Procedia Computer Science*, vol. 80, pp. 1033–1041, 2016.