Analysis Plan

Project Name: Increasing Patient use of Electronic Health Records for Blood Glucose Monitoring

Project Code: 1729

Date Finalized: 4/27/2018

This document serves as a basis for distinguishing between planned (confirmatory) analysis and any unplanned (exploratory) analysis that might be conducted on project data. This is crucial to ensuring that results of statistical tests will be properly interpreted and reported. In order that the Analysis Plan fulfill this purpose, it is essential that it be finalized and date-stamped before we begin looking at the data — ideally, before we take possession of the data. Once this plan is finalized, a date is entered above, and the document is posted publicly on our team website.

Data and Data Structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

The table below gives a comprehensive list of raw data that will be available for analysis.

DATASET- Variable Names	*Data Level*- Variable Descriptions	**DATASET**- Variable Names	*Data Level*- Variable Descriptions
(1) ACTIVE MEDS	*Medication level*	**(4) FLOWSHEET ORDERS**	*Order level*
pat_ID	patient ID	PAT_ID	patient ID
Most_Recent_Contact_Date	Most recent appointment date	Description	Description of Order type
PAT_ENC_CSN_ID	encounter ID of appointment	Ordering Date	Ordering Date
CURRENT_MED_ID	current medication list at time of appointment	Authrzing_PROV_ID	Provider authorizing order
IS_ACTIVE_YN	whether medication is active	**(5) FLOWSHEET READINGS**	Flowsheet entry level
description	description of medication	PAT_ID	patient id
(2) OVERALL REGISTRY REPORT	*Patient level*	entry date	date of glucose entry
PAT_ID	patient ID	entry time	time of glucose entry
Last Initial	last initial of patient name	MEAS_VALUE	value of glucose entry
Provider ID	primary care provider ID	FLO_MEAS_NAME	category of glucose entry
birth date	birthdate	**(6) MYCHART MESSAGES TO PATIENT**	*Message level*
sex	sex	MESSAGE_ID	message ID
ethnicity	ethnicity	recipient ID	patient ID
HBA1C_LAST	value of most recent A1c test	senderID	sender ID
HBA1C_LAST_DT	date of last A1c	message date	message date
last office visit	date of last office visit	message time	message time
OFF_VIS_PROV_ID	ID of last office visit	Read/Unread	whether message has been read at time of data pull
Activation date	date MyChart Activated	**(7) MYCHART MESSAGES FROM PATIENT**	*Message level*
(3) PRESCRIPTION ORDERS	*Order level*	MESSAGE_ID	message ID
ORDER_MED_ID	Order ID	recipient ID	recipient ID
PAT_ID	patient ID	senderID	patient ID
Description	description of Medication	message date	message date
dose	dose amount	message time	message time
measurement	measurement of dose	Read/Unread	whether message has been read at time of data pull
QUANTITY	quantity of doses	**(8) ENCOUNTERS**	*Encounter level*
FREQ_NAME	frequency medication prescribed	PAT_ID	patient ID
Ordering Date	date medication ordered	VISIT_PROV_ID	visit provider ID
		visit date	date of encounter
		PAT_ENC_CSN_ID	encounter ID of appointment
		NAME	in person vs telephone encounter

Each of the eight datasets will be produced monthly from the baseline period (3 months prior to first enrollment of patients) through the implementation and follow-up periods. The "Active Meds" dataset lists active medications associated with a patient's most recent encounter. For this dataset, we will have access to encounters going back to July 2017.

Outcome Variables to Be Analyzed:

ID	Outcome(s) Description	Type	Measurement Variable	Analysis Metric	Method of Aggregation	Time Point	Explantion of Clinical Relevance
1	Flowsheet use, Extensive	Primary	Whether patient enter data to an electronic glucose flowsheet during the measurement period	Occurrence over time period	Binary (proportion)	(0-14) weeks after initial practice orientation meeting	See (i)
2	Patient HbA1c	Primary	A1c test value	Most recent test value at timepoint	Mean	26 weeks after initial practice orientation meeting	See (ii)
3	Flowsheet use, Extensive	Secondary	Whether patient enter data to an electronic glucose flowsheet during the measurement period	Occurrence over time period	Binary (proportion)	(14-26) weeks after initial practice orientation meeting	See (i)
4	Flowsheet use, Total	Secondary	Patient total days of entry to an electronic glucose flowsheet during the measurement period	Number of entries over time period	Mean	(0-14), (14-26) weeks after initial practice orientation meeting	See (i)
5	Flowsheet Orders	Secondary	Whether patient has open physician order for electronic flowsheet	Value at endpoint	Binary (proportion)	(0-14), (14-26) weeks after initial practice orientation meeting	See (i)
6	Patient HbA1c	Secondary	A1c test value	Most recent test value at timepoint	Quantile regression analysis (4 quartiles)	14, 26 weeks after initial practice orientation meeting	See (ii)
7	Patient HbA1c	Secondary	A1c test value	Most recent test value at timepoint	Mean	14 weeks after initial practice orientation meeting	See (ii)
8	Improvement in Patient HbA1c	Secondary	A1c test value	Reduction from baseline	Binary (proportion)	14, 26 weeks after initial practice orientation meeting	See (ii)
9	Patient HbA1c below benchmark	Secondary	A1c test value	Most recent test value at timepoint below 7	Binary (proportion)	14, 26 weeks after initial practice orientation meeting	See (ii)
10	Total secure messages sent by patient	Secondary	Total number of MyChart messages sent by patient during the measurement period	Total number of messages over time period	Mean	(0-14), (14-26) weeks after initial practice orientation meeting	See (iii)
11	Total secure messages sent by patient to PCP	Secondary	Total number of MyChart messages sent by patient to the PCP during the measurement period	Total number of messages over time period	Mean	(0-14), (14-26) weeks after initial practice orientation meeting	See (iii)
12	Total secure messages sent by PCP to patient	Secondary	Total number of MyChart messages sent by PCP to the patient during the measurement period	Total number of messages over time period	Mean	(0-14), (14-26) weeks after initial practice orientation meeting	See (iii)
13	Total number of patient phone appointments	Secondary	Total number of patient phone appointments during the measurement period	Total appointments over time period	Mean	(0-14), (0-26) weeks after initial practice orientation meeting	See (iii)
14	Total number of patient in-person appointments	Secondary	Total number of patient in-person appointments during the measurement period	Total appointments over time period	Mean	(0-14), (0-26) weeks after initial practice orientation meeting	See (iii)
15	Change to patient active medications	Secondary	Change (Any; Addition;Removal) to patient list of active medications during measurement period	Change (Any; Addition; Removal) from beginning to end point	Binary (proportion)	(0-14), (0-26) weeks after initial practice orientation meeting	See (iv)
16	Prescription Orders	Secondary	Number of prescription orders for patient during measurement period (Total all; Total new/non-refill; Total diabetes related)	Total number of orders over time period (All, Non-Refill, Diabetes Related)	Mean	(0-14), (0-26) weeks after initial practice orientation meeting	See (iv)
17	Flowsheet Entry Value	Secondary	Value of blood glucose entered into flowsheet	Descriptive analysis of flowsheet entries	10th 25th 50th 75th and 90th percentile flowsheet entries	(2, 4, 6, 10, 12, 14, 18, 22, 26) weeks after initial practice orientation meeting	See (i)

(i) Self monitoring of blood glucose causally associated with decreases in HbA1c for patients with type II diabetes (Zhu et al 2016), critical factor in reducing risk of complications for patients for type I diabetes (Diabetes Control and Complications Trial Research Group 1993). (ii) Improved average blood sugar control (as measured by A1c levels) is associated with significant decreases in the probability of complications from diabetes (ADA 2016). (iii) Physician communication is significantly positively correlated with patient adherence (Zolnierek 2009). (iv) People with type I diabetes must use insulin, and oral medications can help those with type II diabetes reach target blood glucose levels (ADA 2016).

Transformations of Variables:

Raw data will be aggregated according to the table above (see the analysis metric, method of aggregation, and time point columns). Multiple baseline Active Medications files will be aggregated to a single list of most recent active medications (based on most recent associated appointment date), which will be used as the baseline for the outcome "Change to patient active medications".

Imported Variables:

A file corresponding physician IDs to clinics, treatment assignment, and clinic size strata used for random assignment of clinics will be merged into the data described above.

Transformations of Data Structure:
After outcomes have been aggregated as indicated, they can be merged with treatment assignment status and covariates from the Overall Registry Report file using the patient ID variable.

Data Exclusion:
Only obvious data recording errors (e.g. values outside of medical feasibility) will be excluded, after assessing for any relation with treatment assignment.

Treatment of Missing Data:
The only anticipated treatment of missing data will be for covariates such as age or ethnicity which may be missing in the Diabetes Registry dataset. For specifications that include these covariates, missing values of continuous variables will be re-coded to a fixed value equal to the mean of that covariate and controlled for flexibly using dummy variable indicating that the observation has a missing value for the covariate. For categorical covariates, missing values will be coded as an additional category/dummy variable.

Statistical Models & Hypothesis Tests

This section describes the statistical models and hypothesis tests that will make up the analysis — including any follow-ups on effects in the main statistical model and any exploratory analyses that can be anticipated prior to analysis.

Statistical Models:
For all models below that indicate use of covariates for increased precision, the following list will be used: Patient Age (quadratic), Sex (categorical), ethnicity (categorical), value of most recent baseline A1c test result (linear), days since most recent baseline A1c test result (linear), days since most recent appointment at baseline (linear).

Research Question 1: Will interfacing with primary care practices to encourage physicians to implement bulk online orders of blood glucose flowsheets and informational messaging for all patients with diabetes increase patient adoption?
Outcome Measures: Comparison of individuals between treatment and control practices. Outcomes 1, 3, 4, and 5.
Specification: OLS with Lin covariate adjustment, CR2 standard errors clustered at practice level, Y=outcome, T=treatment indicator, D= doctor fixed effects , X= covariates, S= strata fixed effects
Version 1: $Y_i = \beta_0 + \beta_1 T_i + S_i + \varepsilon$
Version 2 (main): $Y_i = \beta_0 + \beta_1 T_i + D_i + X_i + \varepsilon$

Research Question 2: Does additional reminder messaging to patients that (1) emphasizes the value of tracking blood glucose data to the patient OR (2) emphasizes the value of tracking blood glucose data to the doctor OR (3) informs patient of their selection for a chance to receive an award conditional on tracking increase adoption relative to no reminder messaging?

Outcome Measure: Comparison of individuals across reminder messaging assignment groups (within treatment practices only)-- Outcomes 1, 3, and 4.
Specification: OLS with Lin covariate adjustment, HC2 standard errors, Y=outcome, T=treatment indicator, D= doctor fixed effects , X= covariates
Version 1: $Y_i = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + \beta_3 T_{3i} + \varepsilon$
Version 2: $Y_i = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + \beta_3 T_{3i} + D_i + X_i + \varepsilon$
Version 3 (main): Same as version 2, but limited to observations with a flowsheet order (outcome 5==1)

Research Question 3: Does promotion of adoption of electronic blood glucose tracking through the means described above result in the following intent-to-treat effects:
(a) reduction in most recent patient HbA1c (test prior to study begin compared to most recent test after intervention begins)
Outcome Measures: Intent to treat comparison of individuals between treatment and control practices of the following measures at the end of the intervention period and follow-up period-- Outcomes 2, 6, 7, 8, and 9.
(b) increase in frequency of doctor-patient interaction
Outcome Measures: Intent to treat comparison of individuals between treatment and control practices of the following measures during the intervention period and follow-up period-- Outcomes 10, 11, 12, 13 and 14
For all of these outcomes, the specification that includes controls/covariates will include as a covariate a baseline measure of the outcome that is calculated of over the same length of time as the outcome period.
(c) changes to treatment plan path
Outcome Measure: Intent to treat comparison of individuals between treatment and control practices during the intervention period and follow-up period-- Outcomes 15-16
For all of these outcomes, the specification that includes controls/covariates will include as a covariate a baseline measure of the outcome that is calculated of over the same length of time as the outcome period.

Specification: OLS with Lin covariate adjustment, CR2 standard errors clustered at practice level, Y=outcome, T=treatment indicator, D= doctor fixed effects , X= covariates S= strata fixed effects
Version 1: $Y_i = \beta_0 + \beta_1 T_i + S_i + \varepsilon$
Version 2 (main): $Y_i = \beta_0 + \beta_1 T_i + D_i + X_i + \varepsilon$

Research Question 4: Do reminder messaging treatments that induce more intensive use of flowsheets impact the outcomes described under research question 3 (a)-(c) above?

Outcome Measure: Intent to treat comparison of individuals across reminder messaging assignment groups (within treatment practices). Outcomes same as RQ3: 2, 6-16
Specification: OLS with Lin covariate adjustment, HC2 standard errors, Y=outcome, T=treatment indicator, D= doctor fixed effects , X= covariates
Version 1: $Y_i = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + \beta_3 T_{3i} + \varepsilon$
Version 2: $Y_i = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + \beta_3 T_{3i} + D_i + X_i + \varepsilon$
Version 3 (main): Same as version 2, but limited to observations with a flowsheet order (outcome 5==1)

Research Question 5: Entries of blood glucose data will be predictive of HbA1c and will lower over the study period.
Outcome Measure: Descriptive analysis of flowsheet entry values in the treatment group during the implementation and follow-up period. Outcome=17
Specification: Non-parametric/Summary statistics

Follow-Up Analyses:
For outcomes with significant treatment effects, I will examine heterogeneous treatment effects for patients below/above A1c=7 at baseline, by sex, and by age below/above median.

Inference Criteria, Including Any Adjustments for Multiple Comparisons:
I will be using 2-tailed tests with the following cutoff p-values: 0.10, 0.05, 0.01 to infer statistical significance of treatment effects. I will not correct for multiple inferences as outcomes are expected to be highly correlated/interdependent. See: Rothman, Kenneth J. "No adjustments are needed for multiple comparisons." Epidemiology (1990): 43-46.

Limitations:
Reminder messaging groups will be pseudo-randomly assigned based on first letter of last name (due to logistical infeasibility of random assignment). Thus, for this portion of the experiment, causal interpretation will require the assumption that grouped last name spelling is not independently related to likelihood of flowsheet adoption and other outcomes, controlling for documented ethnicity.

Additionally, low take-up of the practice level intervention (bulk ordering of flowsheets) would significantly hamper power to look at other downstream outcomes.

Exploratory Analysis:
TBD

Link to an Analysis Code/Script:
N/A