Evaluation Strategy

Mississippi Medicaid Demonstration Project

University of Memphis Center for Community Research and Evaluation

November 11, 2021

This document describes the evaluation strategy for the December 2021 evaluation of the Mississippi Medicaid Demonstration Project, in particular the Prediabetes Program and the Healthy Pregnancy Program. The CCRE conducts evaluations of the two programs at least annually to provide Delta Health Alliance (DHA) information on program fidelity and program impact, particularly with respect to quality-of-life outcomes, clinical outcomes, and costs paid by Mississippi Medicaid.

Evaluation procedures are expected to be limited in scope due to the termination of Delta Health Alliance's contract with Cerner Corporation. Because technical issues relating to cost savings were not resolved at the time of contract termination, we do not anticipate evaluating cost savings at this time. We also do not have access to updated data from electronic health records networks other than that of Delta Health Alliance. We anticipate combining legacy Cerner data with an updated data pull from Delta Health Alliance's electronic health records system to make inferences where possible.

We emphasize that the development of a preregistered research plan is extremely difficult given the nature of the data collection and the evolving nature of the Project. Because of the loss of Cerner data, as well as changes in the patient pool due to the reduction in funding for the Project, it is difficult to assess *a priori* exactly what parameters will be available for evaluation over what time periods. Nevertheless, this document states what we expect to analyze.

Prediabetes Program – Clinical Outcomes

We will assess whether there are statistically significant changes between baseline and follow-up clinical values for treatment groups.

Clinical values. Body weight; hemoglobin A1c; blood glucose; diastolic blood pressure; systolic blood pressure. These values will be collected by pooling historical Cerner data with a fresh pull of Delta Health Alliance's electronic health records system for more recent records (with the limitation that DHA's system has limited coverage of the partnering clinics; this analysis also is conditional on our ability to match identifiers in both data files). We do not expect to be able to credibly evaluate diabetes incidence or death without Cerner data.

Treatment group. Individuals enrolled in the intervention and having both a baseline and follow-up value for the clinical value under consideration.

Control group. Individuals randomized to the control group and having both a baseline and follow-up value for the clinical value under consideration.

Baseline and follow-up values. For body mass index and blood pressure, the baseline equals the valid value closest to and within 90 days of the randomization date; follow-up values equal the valid value closest to and within 90 days of the randomization date plus 180, 360, or 540 days. For hemoglobin A1c and blood glucose, as these are collected more sporadically, the baseline has a window of 180 days, and the follow-up value corresponds to the valid value closest to and within 180 days of the randomization date plus 360 days.

Statistical test. Paired t-tests will be used to assess whether there are statistically significant intra-group differences between baseline and follow-up. Difference-in-differences regression with controls for race, sex, age, and fixed effects for year of randomization will be used to assess whether the treatment group difference is distinguishable from that of the control group. Statistical tests will be conducted for each pairing of follow-up period and clinical value; to adjust for the multiple comparisons problem, conclusions about specific clinical outcomes will be made by examining whether the p-value for any statistical test exceeds the Bonferroni-corrected p-value with family-wise error rate of 5%. If race, sex, or age is missing, the median value within the subgroup under consideration will be imputed.

Limitations. We anticipate that this analysis will allow for an evaluation of the impacts of the intervention sufficient to inform project decisions. However, there are known limitations that hinder the ability to make causal inferences, including but not limited to the following.

- **Selection bias** attributable to individuals in the treatment group being systematically different from the control group. Control group individuals are set aside without being contacted; treatment group individuals are both recruited and elect to join the intervention. The decision to select into intervention is likely to occur for reasons that correlate with health outcomes, biasing estimated treatment effects. We can attempt to mitigate this by controlling for observable differences such as race, sex, and age, but we cannot control for selection bias from remaining unobserved factors. Due to a relatively low recruitment rate, we do not commit to conduct an intent-to-treat analysis that would leverage all individuals ever randomized to the recruitment group.
- **Exclusion bias** attributable to a reliance on health outcomes incidental to clinical services. Unlike a standard clinical trial, we are not able to estimate clinical outcomes for those patients who do not seek medical care within the clinics covered in this intervention. This problem is aggravated during the COVID-19 pandemic, as healthcare shifted to a telehealth model which necessarily involves less collection of clinical data points. The loss of Cerner's system also is expected to limit our coverage of EHR data in 2021.

Prediabetes and Healthy Pregnancy Programs – Longitudinal Self-Report Outcomes

For self-reported measures collected during intervention for the treatment group, we will assess whether there are statistically significant changes between baseline and follow-up values.

Self-reported measures. Depression, anxiety, and stress as measured by the DASS-21 screener; physical activity, nutrition, and health measures as measured by the RAND-36 and REAP screeners.

Baseline and follow-up values. The first measure collected for the individual shall constitute the baseline; the follow-up values shall equal the valid value closest to and within 180 days of the randomization date plus 365 days.

Statistical test. McNemar's test will be used to compare dichotomous measures; all other measures will be converted to interval measures and paired t-tests will be used to assess whether there are statistically significant differences between baseline and follow-up.

Limitations. We anticipate that this analysis will allow for an evaluation of the impacts of the intervention sufficient to inform project decisions. However, there are known limitations that hinder the ability to make causal inferences, including but not limited to the following.

- **Attrition bias** attributable to a lack of data for individuals who withdraw from the intervention.
- **The lack of a control group** will limit our ability to know what changes would have been occurred absent intervention.

Prediabetes and Healthy Pregnancy Programs – Other Outcomes

- We will use program data to evaluate rates of preterm birth, low birthweight, and very low birthweight, and will compare this to the most recent statewide rates for Black women available from the Mississippi Department of Health.
- We will report descriptively on number of program enrollments and number of program activities occurring.
- We will disaggregate selected analyses by year to report on most recent patients.