PRIME Care (PRecision medicine In MEntal health Care)
Statistical Analysis Plan
Derived 7/2020 with the publication of the Methods paper (Oslin et al 2021)
Reviewed and finalized prior to Analysis by Executive Committee on 8/16/21

The following statistical analysis plan was specified in the methods paper submitted in July 2020 for publication [34]). The methods paper was an expansion of the original grant application, which had limited specificity in the approach to analysis. In the paper, the original hypotheses were unaltered from the grant application, although edited for clarity. Prior to the conclusion of the study and prior to initiation of analyses, the Statistical Analytic Plan was reviewed by the Executive Committee and approved as amended below. We highlight only substantive changes from the methods paper as noted.

The primary hypotheses to be tested in the clinical trial component are that:

1) During the first six months of treatment, Veterans with MDD whose care is guided by the results of the PGx battery (the intervention group) will have a higher rate of remission of depression than the delayed results group. Remission is defined as a Patient Health Questionnaire (PHQ-9) score of 5 or less.
2) Over the first 30 days after randomization, provider/patient dyads in the intervention group will use fewer antidepressant medications that have potential gene-drug interactions based on commercial PGx test results than dyads in the delayed results group.
3)

Rationale for amendments made– Neither the grant nor the methods paper specified the time frame of Hypothesis #2. This was discussed and approved by the Executive Committee on 8/16/21, and 30 days was agreed upon to evaluate the time most proximal to receiving the PGx results. A secondary outcome will evaluate the proportion of days on which patients were prescribed antidepressants with a significant drug-gene interaction over the 24 weeks with the expectation of small changes in prescribing over time.

The secondary hypotheses are that:

4) Veterans in the intervention group will have better secondary outcomes than the delayed results group over the 24 weeks post-randomization, including rates of depression symptom response (as indicated by a 50% reduction in PHQ-9 score from baseline) and changes in PHQ-9 score.
    - Other secondary outcomes will include side effect rates, treatment adherence rate, and functional improvement.
5) Three months after receiving the PGx battery results, the delayed results group will show time-dependent changes in antidepressant prescribing similar in direction and magnitude to those seen in the intervention group.

Rationale for amendments made– To clarify the intent of Hypothesis #4, it was agreed that this analysis would not be included in the primary paper.

Sample Size and Power Considerations.
We have two primary hypotheses, so we will test each at a 2.5% significance level, using two-sided tests.

The primary outcome for Hypothesis #1 is a set of repeated binary indicators of remission from depression. We assume a 20% loss to followup, distributed evenly over time, and across the two groups. The anticipated remission rate in usual care is based on results from STAR*D [2]. For the base rate, we consider rates of 20-25-30-35%, and for the treatment effect, we look at effects of about 5%. To account for correlations between the results for two patients of the same provider, we assume an intra-class correlation (ICC) of 0.05 or lower for providers, which should also address the nesting within sites [21]. We assume our design will have approximately 200 providers, with an average of ten patients per provider, so the design effect for an ICC of 0.05 is 1.45. For a base rate of 30% (based on results from STAR*D [2]), and an assumed within-patient correlation of 0.4 or lower in a compound symmetry structure, we have 82% power for a 5% treatment effect and 92% power for the same effect in an autoregressive (ar(1)) structure, with higher power for the same 5% effect with lower base rates. For a provider ICC of 0.01, where we have an effective sample size of 1835, we have power in excess of 90% for 5% effects, under either covariance structure.

For Hypothesis #2, the primary outcome is an ordinal response indicating the gene-drug interaction potential (coded as green, yellow, and red based on the PGx test for non, moderate and significant drug- gene interactions) of the initial prescription made during the intervention period. Assuming that the overall proportions in the sample are approximately 40%/40%/20% for green/yellow/red, respectively, then our sample of 2000 provides 80% power to detect a common odds ratio of 1.29 under a proportional odds model, at a significance level of 2.5%.

Data Analysis Plan for the Completed Sample.
Prior to performing analyses, we will apply standard data-cleaning procedures. Specifically, we (1) screen the data for data-entry errors, (2) check for outliers, (3) assess the extent and pattern of missing data, and (4) check that appropriate assumptions of normality for data distributions are met whenever necessary.

Primary hypotheses.
Patient response to treatment (Hypothesis #1). The primary outcome measures will be a set of binary indicators of remission (PHQ-9 score ≤ 5) across the intervention phase time points (weeks 4, 8, 12, 18, and 24). We will use a mixed-effects logistic regression model to test the hypothesis that patients whose treatment is informed by the PGx results are more likely to be in remission than those receiving usual care. The model will have a binary indicator of intervention group as the main explanatory variable, together with variables for time and group-by-time effects, and a fixed effect for site. To accommodate nesting of repeated measures within patients, and nesting of patients within providers, the models will include a random intercept for provider and correlated residuals within patient; we will use information-criteria comparisons to compare the fit of different covariance structures to ensure adequate fit. As there is particular interest in the remission rate of patients at the end of treatment, we will report on a comparison between the groups at the endpoint of the intervention phase (week 24), based on our final model.

Patient and provider use of PGx results (Hypothesis #2). For each provider-patient dyad in the trial, we will use the results of the PGx battery to classify each antidepressant medication prescribed by the provider to that patient during the study period as green, yellow, or red, thus creating an ordinal measure for each time the provider prescribes a medication. The primary outcome for Hypothesis #2 will be the ordinal variable corresponding to prescriptions written in the first 30 days after randomization. In the case of Veterans having simultaneous prescriptions, we will use a "worst-case" approach, assigning the highest level of the green, yellow and red scale. We will compare the groups on their probabilities of prescriptions of green, yellow, or red using a mixed-effects ordinal regression model, including random intercepts for providers and fixed effects for sites.

Rationale for amendments made– The study design has limited control on the number of providers recruited per site, or the number of Veterans treated per provider. In settings where these numbers vary considerably over sites and providers, there can be considerable challenges in fitting mixed effects models that take detailed account of the correlations within Veterans, providers, and sites. If necessary, we will adjust some aspects of the model specification above in order to fit the models.

In supplementary analyses, we will include some covariates in the analyses described above, to examine the effects of certain patient-level and provider-level characteristics. These comprise of age, race, gender, presence and severity of PTSD, and provider-level location of care (primary care vs mental health outpatient). To examine heterogeneity of effect, we will also include interactions between these variables and the intervention variable.

Rationale for amendments made– Based on prior literature showing that the PGx test results may only apply to 10-20% of the randomized population, we will use the PGx test information to define a "genetic susceptibility" variable for each participant. The susceptibility variable will correspond to the proportion of antidepressant medications that are "red" for each Veteran. This is in essence a marker of genetic risk for having gene-drug interactions. We will use this variable as a Veteran-level covariate in reruns of the primary analyses. This change was added as more thought was given to results from studies released after the start of the trial and in consideration of genotype prevalence.

<u>Dropout and Missing Data.</u> We anticipate that there will be dropout from the study as well as intermittent missing data, in particular for the patient outcome responses of Hypothesis #1, while some Veterans may not receive an antidepressant prescription in the 30-days post-randomization for Hypothesis #2. The mixed-effects models described above for Hypothesis #1 make use of all available responses, can accommodate the heterogeneity of response times caused by intermittent missing data, and provides valid estimates and hypothesis tests under an assumption of ignorable missingness. To assess the sensitivity of our conclusions to this assumption, we will extend the mixed-effects models to consider inverse probability-weighted selection models and pattern mixture models [22]. We will perform these analyses under a range of different prediction models and assess the sensitivity of the results. For Hypothesis #2 we will use a multiple imputation approach. [32-33]

Secondary Hypotheses.

Secondary Responses (Hypothesis #3). We will compare the intervention groups on their repeated depression response outcomes using mixed effects logistic regression models, as described for primary Hypothesis #1. For the PHQ-9 summary score of depression severity, we will use mixed-effects linear models.

We will use generalized linear regression models to compare the groups on side effect rate, treatment adherence rate, and functional improvement. Regression models will include random effects to accommodate correlations due to nesting of patients within providers, and repeated measures (where necessary).

Rationale for amendments made– Number of outpatient visits was removed from the secondary outcomes as there are no standards for an acceptable frequency of outpatient visits, particularly when comparing specialty mental health care and primary care.

We will also compare the groups on the rates of side effects that are moderate or severe. Side effects may be more plausibly related to serum antidepressant levels and thus PGx results than depressive symptoms. The data will include ordinal ratings of headache, nausea or vomiting, diarrhea, constipation, and sexual dysfunction at weeks 4, 8, 12, 18, and 24. We will use a mixed-effects ordinal logistic regression model to test whether there are differences in the probabilities of moderate or severe side effects between the two intervention groups. We will compare the groups on functional improvement, using the mental health component of the VR-12 score, with analogous generalized linear mixed-effects models appropriate to the response distribution.

We will compare differential outcomes by randomized group on adherence rates. For Hypothesis #1, there are two types of adherence to be considered. First, a provider/patient dyad in the intervention group may not use a recommended medication (this is the focus of Hypothesis #2 with the analytic plan described above). The second form of adherence is whether the patient is actually taking their medication as prescribed. We assess medication adherence using electronic health record data on prescriptions filled. From the prescriptions filled data, we will calculate an antidepressant medication possession ratio. For each Veteran, we will create a binary variable, indicating whether they filled medication prescriptions for at least 80% of days through the 24-weeks treatment period, and compare the groups on this outcome using logistic regression models, as described above.

Rationale for amendments made– The above section on adherence was amended from the methods paper to separate the analysis for Hypothesis #2 and to further clarify medication adherence as a secondary outcome. Originally proposed was a self-reported variable on adherence. However, during the trial is was apparent that the validity of this assessment was suspect as patients often didn't remember their medications and had difficulty assessing adherence to specific medications. Prescription fill rates is considered a more reliable method for ascertaining availability of medication though it does not assess whether the patient has taken the medication. Below we also clarify that the proposed adherence by treatment interaction analysis was to be explored in secondary manuscripts.

In secondary manuscripts, we will assess the sensitivity of the results of the analyses described above to non-adherence, by extending the models described above for primary and secondary outcomes using instrumental variable approaches. Here, the instrumental variable is randomization that, under some conditions, will control for unmeasured bias (due to self-selection to adhere or not) when estimating the relative effects of the medications in those who adhere to their assigned medication. Results will use data from all patients who show various degrees of adherence as measured by the availability of medication or adherence to the PGx recommendation and will yield estimates of the intervention effects that would have been seen in a study with full adherence. For a single overall measure of adherence, we will use the methods of Nagelkerke et al. [25]; if there is sufficient within-patient variation across time in adherence, we will use the longitudinal methods of Small et al. [26].

Change in prescribing in delayed results group post-intervention (Hypothesis #4). At the conclusion of the six-month intervention phase of the trial, the PGx battery results are returned to the providers (and patients) in the delayed group. We hypothesize there will be an increase in use of antidepressant medications categorized as green after the results are returned. Thus, within the delayed results group, we will use a mixed-effects ordinal response model to estimate and test a time-dependent change in the probability of receiving antidepressant medications of the three types (red, yellow, green) after the intervention. In addition, we will perform a bivariate analysis, combining the ordinal model with a linear model for the PHQ-9 score, to examine whether change in prescribing practice (i.e., increased prescribing of green-type medications) is related to change (i.e., an expected decline) in PHQ-9 scores.

Further descriptive analyses:
Rationale for amendments made– Not specified in the methods paper are descriptions of the impact of the study on routine clinical care. To describe potential delays in care associated with the time of collecting and releasing PGx results, we will present data on the time from randomization to the release of the PGx results in the intervention arm. Descriptive data of participating providers and randomized participants will also be provided in the primary outcomes.

References:

2. Trivedi, M.H., et al., *Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice.* Am J Psychiatry, 2006. **163**(1): p. 28-40.
21. Ten Have, T.R., A.R. Kunselman, and L. Tran, *A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering.* Stat Med, 1999. **18**(8): p. 947-60.
22. Fitzmaurice, G., N. Laird, and J. Ware, *Applied Longitudinal Analysis*. Vol. 2nd Edition. 2011, New York: John Wiley and Sons.
25. Nagelkerke, N., et al., *Estimating treatment effects in randomized clinical trials in the presence of non-compliance.* Statistics in Medicine, 2000. **19**: p. 1849 - 1864.
26. Small, D.S., et al., *Random effects logistic models for analysing efficacy of a longitudinal randomized treatment with non-adherence.* Stat Med, 2006. **25**(12): p. 1981-2007.

32. Van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research*, 2007. 16, 219–242,
33. Van Buuren S., Groothuis-Oudshoon K. (2011) MICE: *Multivariate Imputation by Chained Equations in R. Journal of Statistical Software*. http://lib.stat.cmu.edu/R/CRAN/web/packages/mice/index.htm
34. Oslin DW, Chapman S, Duvall SL, Gelernter J, Ingram EP, Kranzler HR, Lehmann LS, Lynch JA, Lynch KG, Pyne JM, Shih MC, Stone A, Thase ME, Wray LO. *Study design and implementation of the PRecision Medicine In MEntal health Care (PRIME Care) Trial*. Contemp Clin Trials. 2021 Feb;101:106247. doi: 10.1016/j.cct.2020.106247. Epub 2020 Dec 11. PMID: 33316457.