**Plan of Analysis for Sub-Patent Malarial Data:**

**Further Data Extraction:**

1. Indicators for PCR and Microscopy Quality.
   1. Jane to inform on this front.
   2. Data extraction to be carried out following Jane’s input.
2. Sourcing of MAP/Other Estimates for Transmission History in Non-African Settings.
   1. Awaiting email from Sam about this.
   2. Could possibly source these from other places, e.g. the World Malaria Reports etc.

**Preliminary Data Analysis:**

1. Investigate a range of different transformations in order to assess which best produces normally and symmetrically distributed datasets.
   1. Kolmogorv-Smirnov Test, Shapiro-Wilk Test, Jargue-Bera Test, Anderson Darling Test.
   2. Q-Q Plot.
   3. gvlma() function from the gvlma package allows automated checking of the assumptions of linear regression.

**Exploratory Data Analysis:**

1. **Basic Model Fitting- Across All Transmission Settings.**
   1. Old data.
   2. New data.
   3. Combined.
2. **Basic Model Fitting- Stratified By Global Region.** 
   1. Further divide up some regions, namely East Africa into two separate regions (for high and low transmission intensities).
   2. Fit the full model and look to see whether any differences are apparent between global regions.
3. **Basic Model Fitting- Stratified By Transmission Setting.**
   1. Do this for both PCR and Microscopy transmission setting stratification and evaluate the comparative sensitivities in this context.
4. **Transmission Setting and History Model Fitting.**
   1. Fit the basic model to both Transmission History and Current Transmission Setting independently to see whether they influence anything.
      1. Initially fit separate models for high and low current and historical transmission setting settings and evaluate them separate to see whether any differences are apparent.
   2. Then create covariates to incorporate each of history and current separately into the model, but allowing joint fitting of high and low together.
   3. Update model to allow joint fitting of high and low AND current and historical transmission intensity, all simultaneously.
5. **Focussing In On Low Transmission Settings.**
   1. Do the below analyses for a variety of different transmission intensity cutoffs, including 0.1, 0.15 and 0.2.
   2. Basic model fitting to see whether any differences manifest between regions- compare to the model fitting to the whole dataset in 2 to see whether there are any discrepancies, and if so, why this might be.
6. **PCR and Microscopy Quality Model- Investigating Diagnostic Quality.**
   1. Following data extraction and Jane’s input, modify basic model to contain quality indicators for either PCR, microscopy, or both.
   2. Stratify data according to this, and investigate whether the impact of varying quality is significant or not.

**Note:** Investigate the lm function some more in R, particularly by adding extra 0s in to stratified datasets to see if they change anything (they should) to make sure it’s working properly. Also investigate the Bayesian logit() function and how it works with the fact that a lot of the data points have 0 prevalence by LM or PCR.