Supplementary Figure 1: Pre-specified regressions and variables utilized in model development and prediction

A. Regressions	B. Variables
• Sample mean	• Age
Generalized Linear Models:	• Biological sex
• Including the infection status for each additional worm	• Recruitment site
• Including total number of worm infection types	• Hemoglobin
• With splines for age, viral load, and hemoglobin	• Pregnant
• Step Forward	• Malaria infection
• Step Forward + two-way interactions	• Any helminth infection
• Elastic-net regularized + two-way interactions stratified by HIV status	• Ascaris infection
Nonparametric:	• Trichuris infection
• Ranger: Mtry (2, 4, 6)	• Hookworm infection
• Xgboost: Max tree depths (2, 4, 6)	• Number of helminth species
• rpartPrune	• HIV status
• Cross Validated Earth (5 fold)	• Viral load

A. Regressions. Sample mean serves as a simple benchmark algorithm for prediction. Generalized linear models were user-generated as indicated. Nonparametric regressions were available within the SuperLearner package and were implemented with the hyperparameter values indicated in parenthesis. **B. Variables.** Variables which had the potential to confound the results and were available for all participants were included in model development and prediction.

Supplementary Figure 2: CD4 Count Sensitivity Analysis of Estimated Distribution of Mtb Infection Outcomes
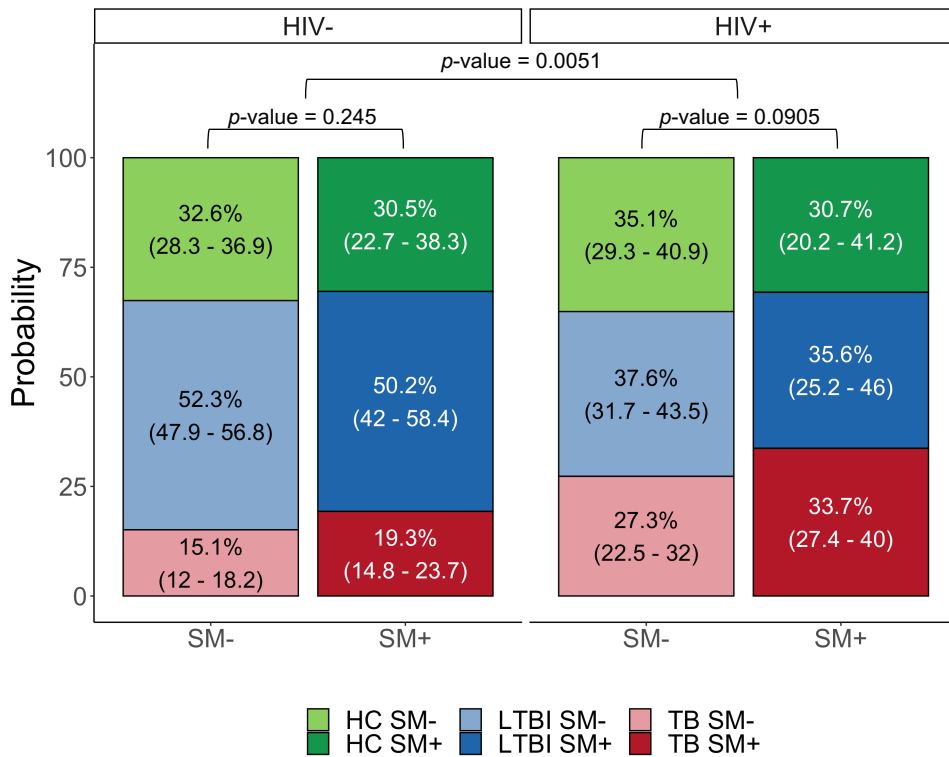
Probabilities are derived from TMLE estimation and are represented as stacked bar plots $\pm$ 95% confidence interval. Standard error estimates were obtained using estimated influence functions and the delta method. Significance level was calculated using Wald tests. The distribution of Mtb infection outcomes differs between SM$^-$ and SM$^+$ HIV$^+$ accounting for CD4 counts in HIV$^+$ individuals.