

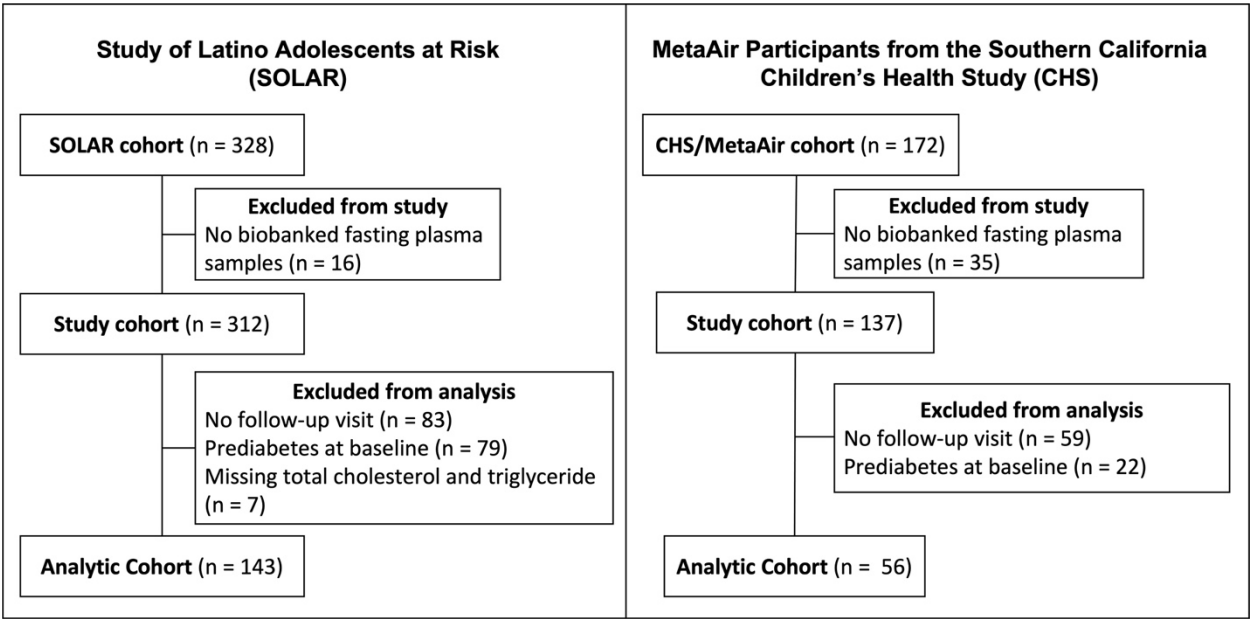
## Online Only Supplemental Figures

---

### Table of Contents

<b>Figure S1.</b> Population flowchart of SOLAR and CHS cohorts. ....	2
<b>Figure S2.</b> Empirical distribution of AUC values in the CHS (validation) cohort under the null hypothesis that no metabolites predicted prediabetes .....	3

**Figure S1.** Population flowchart of SOLAR and CHS cohorts.



**Figure S2.** Empirical distribution of AUC values in the CHS (validation) cohort under the null hypothesis that no metabolites predicted prediabetes. The empirical AUC distribution was calculated by repeatedly randomly selecting two metabolites and performing a logistic regression with the general risk factors, specific risk factors, and randomly selected metabolites in the SOLAR (discovery) cohort, and then using these coefficients to predict the outcomes in the CHS (validation) cohort. The AUC of 0.70 represents the AUC for the final prediction model using the coefficients from SOLAR to predict prediabetes in the CHS cohort. The final model included the LASSO-selected metabolites, general risk factors (sex and BMI), and specific risk factors. The shaded area of the empirical AUC distribution represents the upper 5% of the distribution; AUC values for the final prediction model which fall in this region were considered significant at a Type I error rate of 0.05. Since the AUC of the final model in the validation cohort falls in the shaded region, this indicates that the model with the LASSO-selected features performs significantly better than randomly selected features. See main text for additional details.

