

Table S1. Coefficient values for the population models used for the simulations

	<u>Optimal marginal improvement</u>	<u>Sample size per arm</u>	<u>Noise variables</u>	<u>Parametrization¹</u>
I. Baseline outcome prevalence : 30%				
A. Dense signal	0	100, 300, 500	0, 20, 50	Inverse Logit = -0.718 + 0.011*TRT - 0.27*SEX - 0.27*AGE - 0.27*EDU - 0.27*HDTOT
	0.05	100, 300, 500	0, 20, 50	Inverse Logit = -1.93 - 0.613*TRT+ 1*SEX + 1*AGE + 1* EDU + 1*HDTOT - 0.424809*TRT*(-SEX - AGE + EDU + HDTOT)
	0.1	100, 300, 500	0, 20, 50	Inverse Logit = -1.86 - 1.309*TRT + 1*SEX + 1*AGE + 1*EDU + 1*HDTOT - 0.907137*TRT*(-SEX - AGE + EDU + HDTOT)
	0.15	100, 300, 500	0, 20, 50	Inverse Logit = -1.4 - 1.982*TRT +1* SEX + 1*AGE + 1*EDU + 1*HDTOT - 1.373526*TRT*(-SEX - AGE + EDU + HDTOT)
	0.2	100, 300, 500	0, 20, 50	Inverse Logit = -1.97 - 5.15*TRT + 1*SEX + 1*AGE + 1*EDU + 1*HDTOT - 3.52044*TRT*(-SEX - AGE + EDU + HDTOT)
	0.25	100, 300, 500	0, 20, 50	Inverse Logit = -1.97 - 21.9*TRT + 1*SEX + 1*AGE + 1*EDU + 1*HDTOT - 13.49964*TRT*(-SEX - AGE + EDU + HDTOT)
B. Sparse signal	0	100, 300, 500	0, 20, 50	Inverse Logit = -0.89 + 0.011*TRT - 0.27*SEX
	0.05	100, 300, 500	0, 20, 50	Inverse Logit = -1.148 - 1.19*TRT - 0.87*SEX - 1.57*TRT*SEX
	0.1	100, 300, 500	0, 20, 50	Inverse Logit = -0.81 - 0.42*TRT + 0.1*SEX - 1.192*TRT*SEX
	0.15	100, 300, 500	0, 20, 50	Inverse Logit = -0.987 + 0.1*TRT + 1.2*SEX - 1.5774*TRT*SEX
	0.2	100, 300, 500	0, 20, 50	Inverse Logit = -0.987 - 0.18*TRT +1.2*SEX - 2.218*TRT*SEX
	0.25	100, 300, 500	0, 20, 50	Inverse Logit = -0.987 - 0.99*TRT + 1.2*SEX - 3.47*TRT*SEX
II. Baseline outcome prevalence: 50%				
A. Dense signal	0	100, 300, 500	0, 20, 50	Inverse Logit = 0.17 - 0.01*TRT - 0.27*SEX - 0.27*AGE - 0.27*EDU - 0.27*HDTOT
	0.05	100, 300, 500	0, 20, 50	Inverse Logit = -0.585 - 0.35*TRT+ 1*SEX + 1*AGE +1* EDU + 1*HDTOT - 0.394317*TRT*(-SEX- AGE + EDU + HDTOT)
	0.1	100, 300, 500	0, 20, 50	Inverse Logit = -0.62 - 0.707*TRT+ 1*SEX + 1*AGE +1* EDU + 1*HDTOT - 0.817047*TRT*(-SEX- AGE + EDU + HDTOT)
	0.15	100, 300, 500	0, 20, 50	Inverse Logit = -0.62 - 1.16*TRT+ 1*SEX + 1*AGE +1* EDU + 1*HDTOT - 1.325016*TRT*(-SEX- AGE + EDU + HDTOT)
	0.2	100, 300, 500	0, 20, 50	Inverse Logit = -0.62 - 1.848*TRT+ 1*SEX + 1*AGE + 1*EDU + 1*HDTOT - 2.10672*TRT*(-SEX- AGE + EDU + HDTOT)
	0.25	100, 300, 500	0, 20, 50	Inverse Logit = -0.62 - 3.34*TRT+ 1*SEX + 1*AGE + 1*EDU + 1*HDTOT - 3.871791*TRT*(-SEX- AGE + EDU + HDTOT)
B. Sparse signal	0	100, 300, 500	0, 20, 50	Inverse Logit = -0.03 + 0.012*TRT - 0.27*SEX
	0.05	100, 300, 500	0, 20, 50	Inverse Logit = -0.173 - 0.229*TRT - 1.2*SEX - 0.749*TRT*SEX
	0.1	100, 300, 500	0, 20, 50	Inverse Logit = -0.13 - 0.85*TRT - 1*SEX - 2.05*TRT*SEX
	0.15	100, 300, 500	0, 20, 50	Inverse Logit = -0.03 - 0.7*TRT - 0.4*SEX - 2.07*TRT*SEX
	0.2	100, 300, 500	0, 20, 50	Inverse Logit = -0.01 - 0.8*TRT + 0.01*SEX - 2.65*TRT*SEX
	0.25	100, 300, 500	0, 20, 50	Inverse Logit = -0.01 - 1.46*TRT + 0.19*SEX - 4.08*TRT*SEX

¹TRT = dichotomous treatment variable; AGE = continuous variable for respondent age, standardized to mean 0, std 1; SEX = dummy variable coded 1 for male and -1 for female; EDU = dummy variable coded 1 for more than high school education and -1 for high school or less education; HDTOT = continuous score on the 17-question Hamilton Depression Rating Scale, standardized to mean 0, std 1 (Hamilton, 1960). Trailing zeros are omitted for brevity.

References

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23, 56-62.