Supplemental Online Content

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods

As robustness checks, we provide alternative estimates of treatment effects. First, we used propensity score matching (nearest-neighbor and post-stratification). Second, we estimated treatment effects with linear regression models with covariate adjustment (akin to ANCOVA) with three-month outcome score as the response variable and fixed effects for study arm (surgery vs. control), baseline outcome score, age, race, ethnicity, insurance type, BMI, testosterone use duration, chest binding, parental support, and survey response time. Third, we conducted an analysis of gain scores, akin to a difference-in-difference analysis, using linear regression models with difference between baseline and three-month outcome score as the dependent variable. These models included main effects for study arm (surgery vs. control), age, race, ethnicity, insurance type, BMI, testosterone use duration, chest binding, parental support, and survey response time. Finally, we performed sensitivity analyses to evaluate the influence of unmeasured confounding. These analyses focused on estimation of the E-value, which characterizes the minimum strength of association between an unmeasured confounder and treatment selection and outcomes required to explain the entirety of the observed treatment-outcome association. A large E-value indicates that considerable unmeasured confounding would be needed to negate an effect estimate.

eResults

Supplemental eTable 1 reports results for nearest-neighbor propensity score matching (full cohort) and matching with post-stratification (under 18 years of age cohort). Nearest neighbor matching could not be performed in the under 18 years cohort due to significant loss of observations. Surgery was significantly associated with improvement in all three-month outcomes in both cohorts except for the BIS secondary sub-score.

Supplemental eTable 2 reports results for unweighted ANCOVA models with sensitivity analysis. Surgery was significantly associated with improvement in all three-month outcomes in both cohorts except for the BIS secondary sub-score. E-values and their lower confidence interval bounds for all models except the three-month BIS secondary model were large, indicating that significant confounding would need to occur to negate the observed association between surgery and outcome measure.

Supplemental eTable 3 reports results for gain score analysis. Surgery was significantly associated with improvement in all three-month outcomes in both cohorts except for the BIS secondary sub-score. E-values and their lower confidence interval bounds for all models except the three-month BIS secondary model were large, indicating that significant confounding would need to occur to negate the observed association between surgery and outcome measure.

eTable 1. Matching Propensity Score Model Results

Outcome	Entire Cohort (n = 44) (Surgery Estimate (95% CI))[a]	Under 18 Years Cohort (n = 29) (Surgery Estimate (95% CI))[b]
Primary Outcome		
Three-Month CDM	-24.67 (-29.57, -19.78)	-24.40 (-31.29, -17.50)
Secondary Outcomes		
Three-Month TCS	8.13 (5.14, 11.12)	6.45 (3.25, 9.65)
Three-Month TCS AC	7.66 (5.25, 10.08)	6.14 (3.53, 8.74)
Three-Month BIS	-5.30 (-13.35, 2.74)	-10.34 (-16.06, -4.61)
Three-Month BIS Secondary	-0.76 (-4.70, 3.18)	-1.31 (-4.64, 2.02)

Abbreviations: AC = appearance congruence; BIS = body image scale; CDM = chest dysphoria measure; CI = confidence interval; TCS = transgender congruence scale

[a]Nearest neighbor matching resulted in 22 patients in the control group and 22 patients in the surgical group.

[b]Due to loss of a significant number of observations with nearest neighbor matching, matching with post-stratification into five quintiles was performed.

eTable 2. Analysis of Covariance Model Results with Sensitivity Analysis

Outcome	Entire Cohort (n = 70) (Surgery Estimate (95% CI))	Sensitivity Analysis Surgery Estimate Range[a]	E-Value (95% Lower Bound)[b]	Under 18 Years Cohort (n = 29) (Surgery Estimate (95% CI))	Sensitivity Analysis Surgery Estimate Range[a]	E-Value (95% Lower Bound)[b]
Primary Outcome						
Three-Month CDM	-24.29 (-28.88, 19.70)	(-34.44, -14.14)	50.98 (27.79)	-26.16 (-31.36, -20.95)	(-28.48, -22.33)	78.22 (38.76)
Secondary Outcomes						
Three-Month TCS	7.27 (4.78, 9.76)	(4.20, 10.35)	11.43 (6.04)	7.15 (4.55, 9.74)	(5.59, 7.93)	14.80 (7.04)
Three-Month TCS AC	6.53 (4.14, 8.92)	(3.46, 9.60)	10.10 (5.30)	6.78 (4.27, 9.30)	(4.93, 7.63)	14.16 (6.73)
Three-Month BIS	-6.28 (-11.63, -0.92)	(-14.84, 2.29)	3.51 (1.51)	-7.26 (-11.93, -2.58)	(-10.42, -5.00)	5.38 (2.41)
Three-Month BIS Secondary	0.16 (-2.48, 2.80)	(-5.09, 5.41)	1.24 (1.0)	0.28 (-2.72, 3.27)	(-1.83, 1.93)	1.33 (1.0)

Abbreviations: AC = appearance congruence; BIS = body image scale; CDM = chest dysphoria measure; CI = confidence interval; TCS = transgender congruence scale

[a]Sensitivity analysis shows minimum and maximum treatment effect estimates assuming R^2 of unmeasured confounding greater than 0.5.

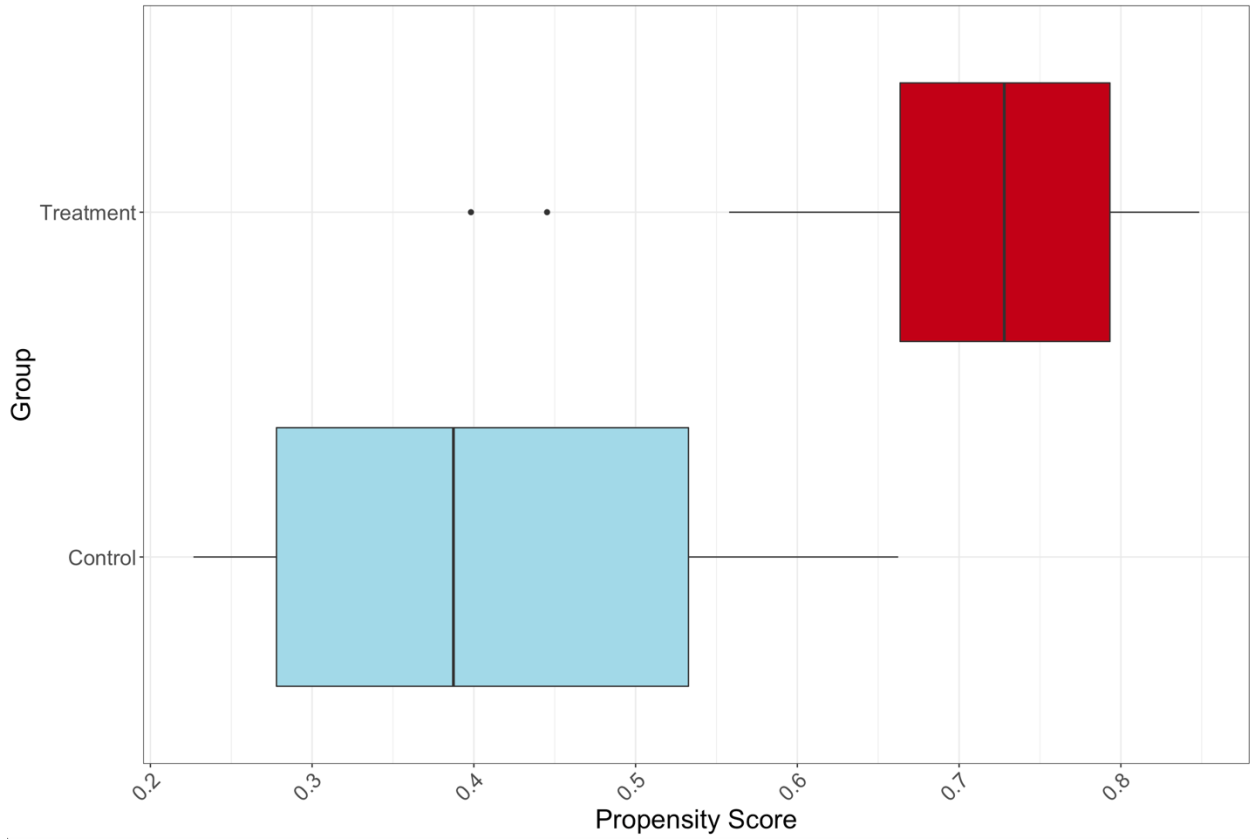[b]Represents the limit of the confidence interval closest to the null.

eTable 3. Analysis of Gains Model Results

Outcome	Entire Cohort (n = 70) (Surgery Estimate (95% CI))	E-Value (95% Lower Bound)[a]	Under 18 Years Cohort (n = 29) (Surgery Estimate (95% CI))	E-Value (95% Lower Bound)[a]
Primary Outcome				
Three-Month CDM	-26.27 (-31.34, -21.20)	43.87 (24.27)	-20.89 (-32.21, -9.57)	24.33 (6.68)
Secondary Outcomes				
Three-Month TCS	11.94 (7.86, 16.02)	10.09 (5.54)	12.25 (3.37, 21.13)	11.25 (3.08)
Three-Month TCS AC	10.73 (6.98, 14.48)	9.70 (5.31)	9.03 (0.89, 17.17)	7.78 (1.92)
Three-Month BIS	-5.58 (-10.93, -0.23)	3.15 (1.25)	-10.14 (-18.03, -2.24)	9.90 (2.64)
Three-Month BIS Secondary	0.30 (-2.79, 3.39)	1.31 (1.0)	-1.34 (-6.45, 3.76)	2.15 (1.0)

Abbreviations: AC = appearance congruence; BIS = body image scale; CDM = chest dysphoria measure; CI = confidence interval; TCS = transgender congruence scale.
[a]Represents the limit of the confidence interval closest to the null.

eFigure 1. Testosterone Use versus Age per Group

Testosterone Use versus Age per Group

Boxplots of Estimated Propensity Scores per Group

eFigure 3. Mean Propensity Scores by Group Across Quintiles

Mean Propensity Scores by Group Across Quintiles