APPENDIX I. BIAS MODELING APPROACH ONE

Step One: Record-level exclusion of individuals who are incorrectly included in the sample.

Inclusion of the responses from individuals who do not have the relevant anatomy for a given sex-specific question inflates the "no", "I don't know" or "refused" responses.[1] For transgender men and women, this record-level adjustment is straightforward and described in the main text.

For GNC individuals, we conduct a probabilistic record level adjustment using Monte Carlo sampling methods. We randomly sample 1,000 scenarios, where for each iteration, we remove a random subset of 30% of all GNC respondents, and only remove individuals if they responded "no", "I don't know" or "refused" under the assumption that they are individuals for whom the sex-specific questions we assume are not anatomically relevant. **Table A1** shows the percentage of GNC individuals who respond as "no", "I don't know" or "refused" that need to be omitted.

Table A1. Among GNC individuals, Proportion who are excluded by Outcome

Outcome	*Proportion of Total Responses that were "No", "I don't know" or "Refused"*	*Proportion of "No", "I don't know" or "Refused" Responses to Exclude*[1]
PSA test	0.59	0.50
Pap test	0.46	0.66
Hysterectomy	0.78	0.38
Currently Pregnant	0.97	0.31

[1] The proportion excluded is calculated as 0.3/ proportion *of total responses that were "no", "I don't know" or "refused"*

Step Two: Summary-level adjustment to account for individuals who are incorrectly excluded from the sample.

The prevalence ratio $\widehat{\mathrm{PR}}$ estimated from the record-level adjusted dataset can then be multiplied by a selection bias odds ratio determined by the selection probabilities shown in the contingency table of **Table A2**. We define the selection bias odds ratio as $\mathrm{OR}_{select} = (\mathrm{B} \times C)/(A \times D)$ and the prevalence ratio adjusted for selection bias as $\mathrm{PR}_{adj} = \widehat{\mathrm{PR}} \times \mathrm{OR}_{select}$.[2] The selection probability is the probability of being asked sex-specific questions across strata defined by gender identity and the outcome of interest.[3] Although we do not have disaggregated data on the values of A and C, **Table A3** shows the overall selection probabilities for transgender and gender nonconforming (TGNC) individuals for each outcome.

We can reasonably assume non-differential selection bias among cisgender respondents. Therefore, letting $B = D$ allows us to simplify $\mathrm{OR}_{select} = C/A$. When $A > C$, TGNC individuals with the outcome are preferentially included, or over selected. This implies that $\mathrm{OR}_{select} < 1$, so our bias adjusted prevalence ratio decreases. Conversely, when $A < C$, TGNC individuals with the outcome are preferentially excluded, or under selected. This implies that $\mathrm{OR}_{select} > 1$, so our bias adjusted prevalence ratio increases.

Table A2. Selection probability

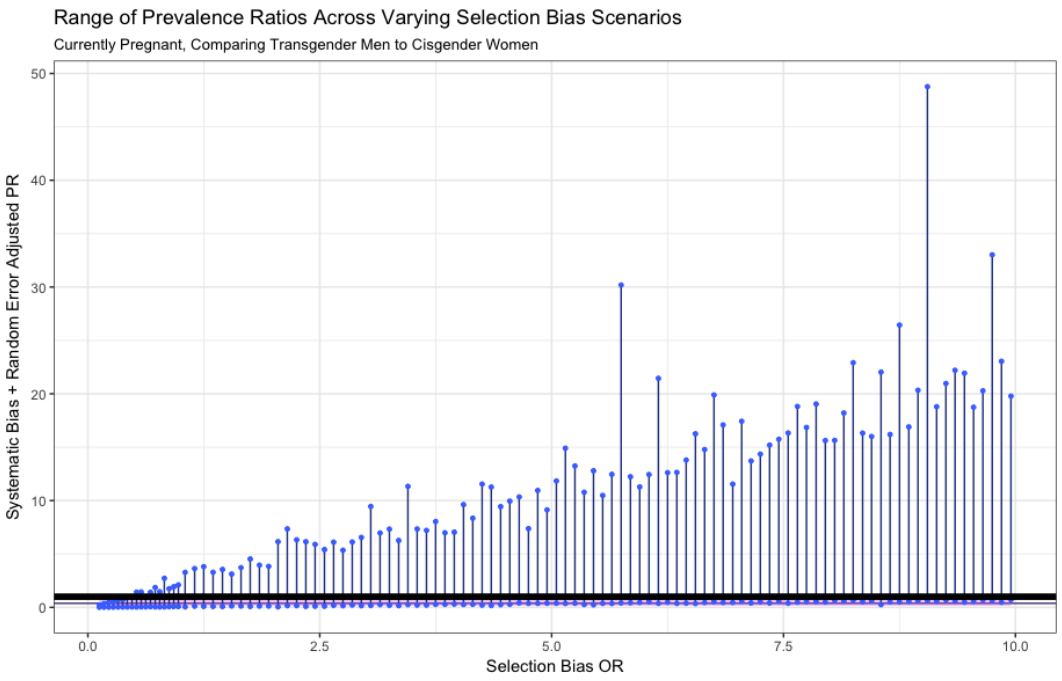
	TGNC	*Cisgender*
Outcome	A	B
No Outcome	C	D

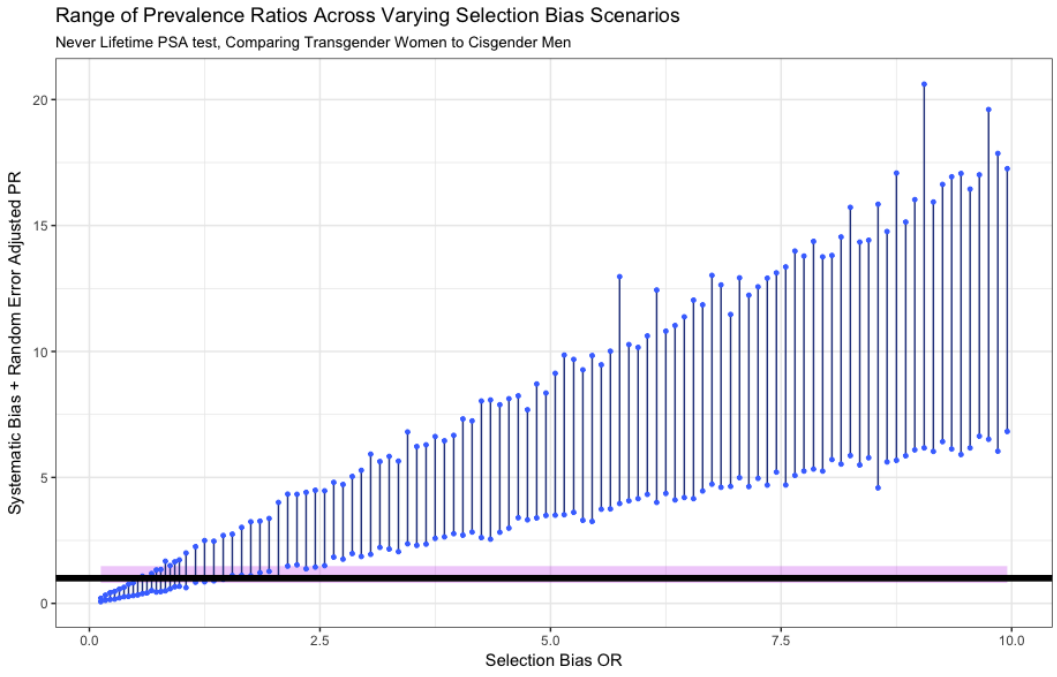
Now we are able to conduct a summary level selection bias adjustment using Monte Carlo simulation methods.[3] We do not have external data to inform the direction or degree of selection bias, so we consider a wide range of selection bias ratios under a uniform distribution where the parameter OR_{select} ranges between 0.10 and 10.0. For each outcome, we sample from the uniform distribution of parameter OR_{select} 100,000 times. For each iteration, we estimated two bias-adjusted prevalence ratios, one modeling systematic bias alone, and a second combining systematic bias and random errors. Random error is incorporated at each step in the simulation: letting $\mathrm{se}_{regression}$ be the robust standard error estimated from the complete cases Poisson regression, we calculate:

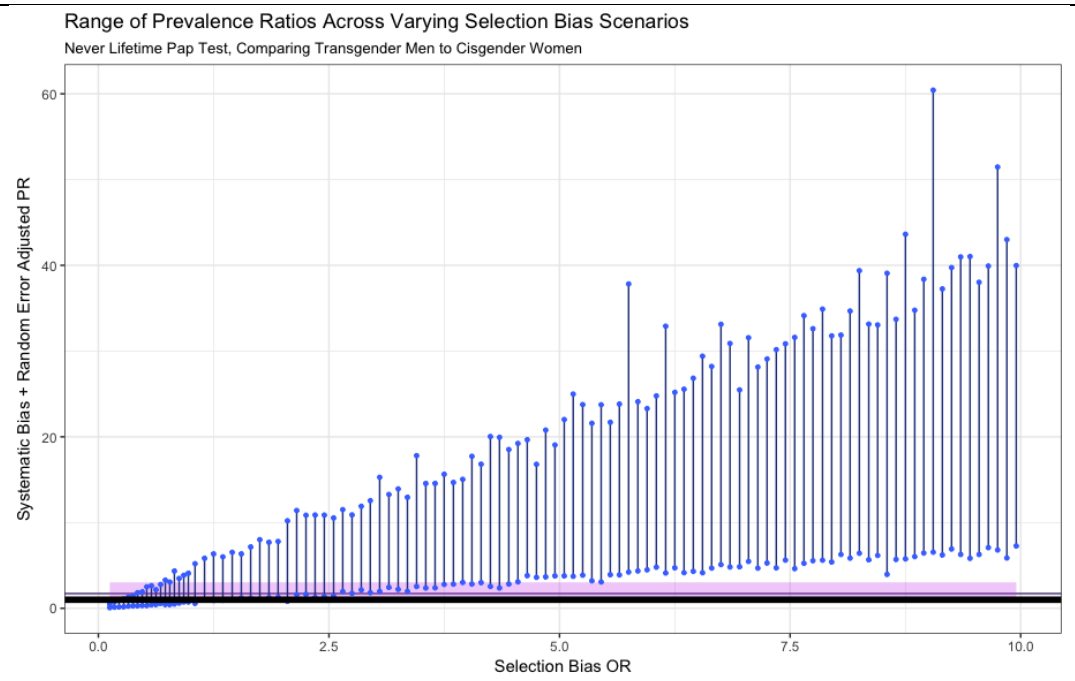
$$\mathrm{PR}_{adj} = \exp\left(ln\left(\widehat{PR} \times OR_{select}\right) + N(0,1) \times se_{regression}\right)$$

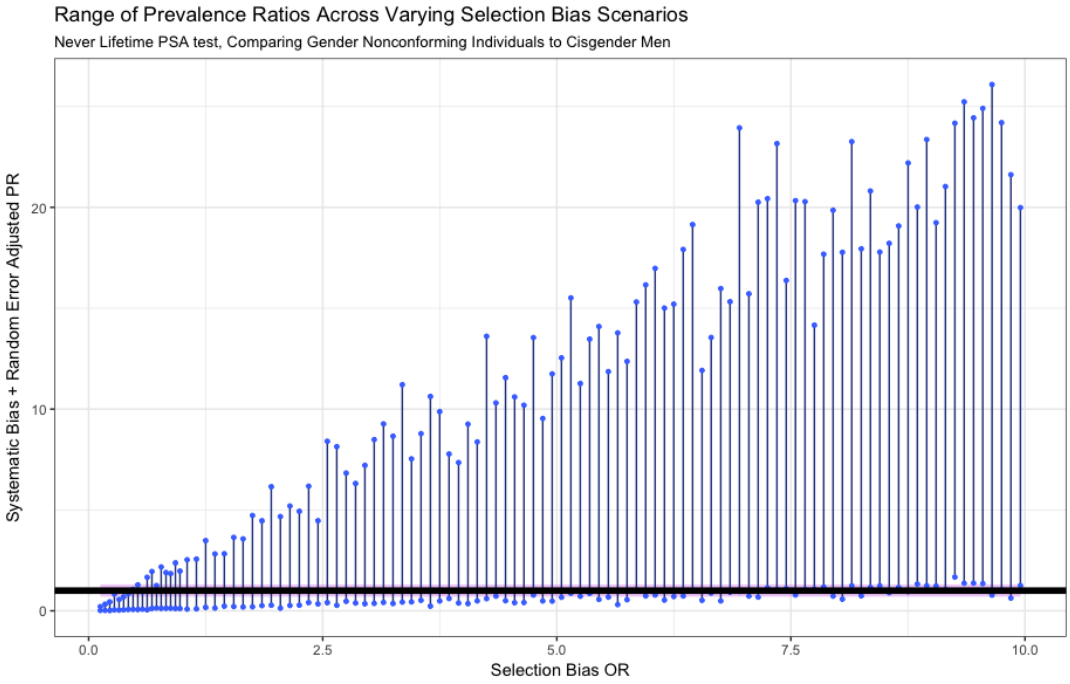
where $N(0,1)$ is a random number drawn from the standard normal distribution.[3] **Figure A1** displays the resulting relationship between OR_{select} and the systematic bias and random error adjusted prevalence ratios.

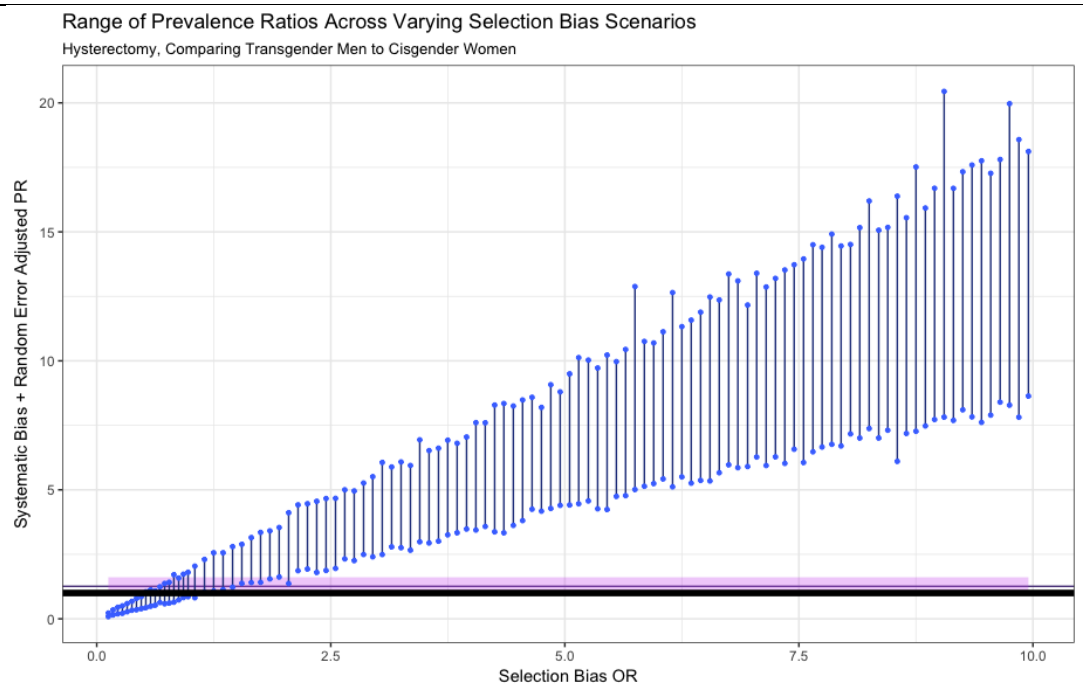
Figure A1. Range of PR adjusted for Systematic Bias + Random Error by Varying Selection Bias Odds Ratios Ranging from 0.10 to 10.0 for each Outcome and Gender Identity

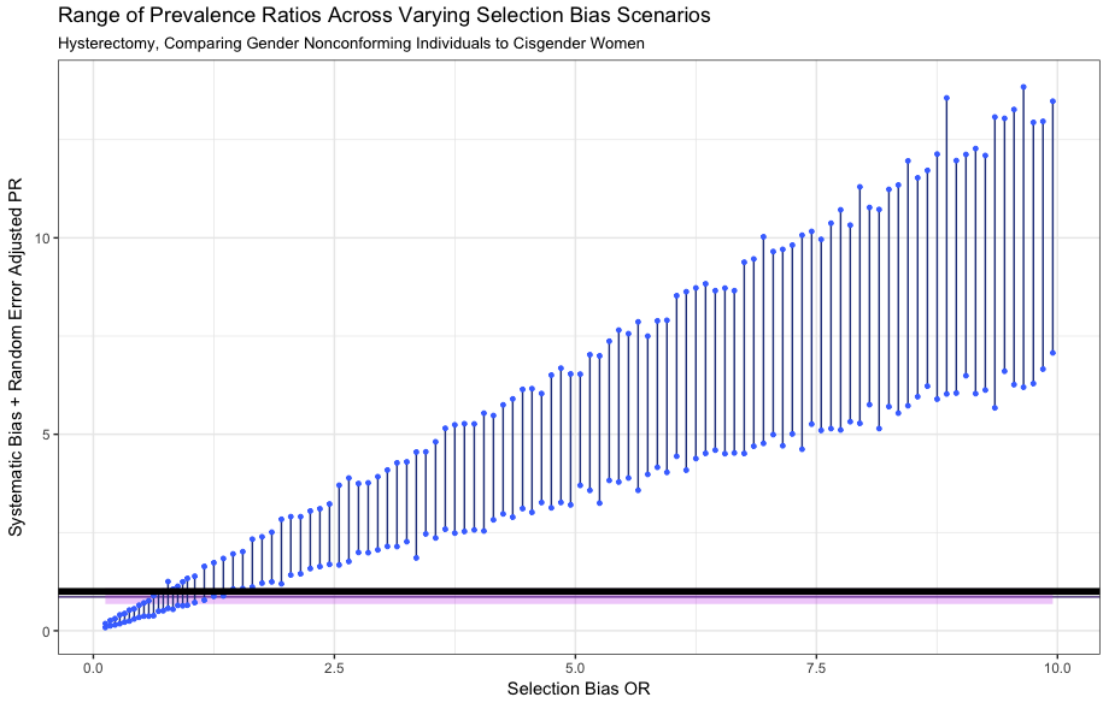


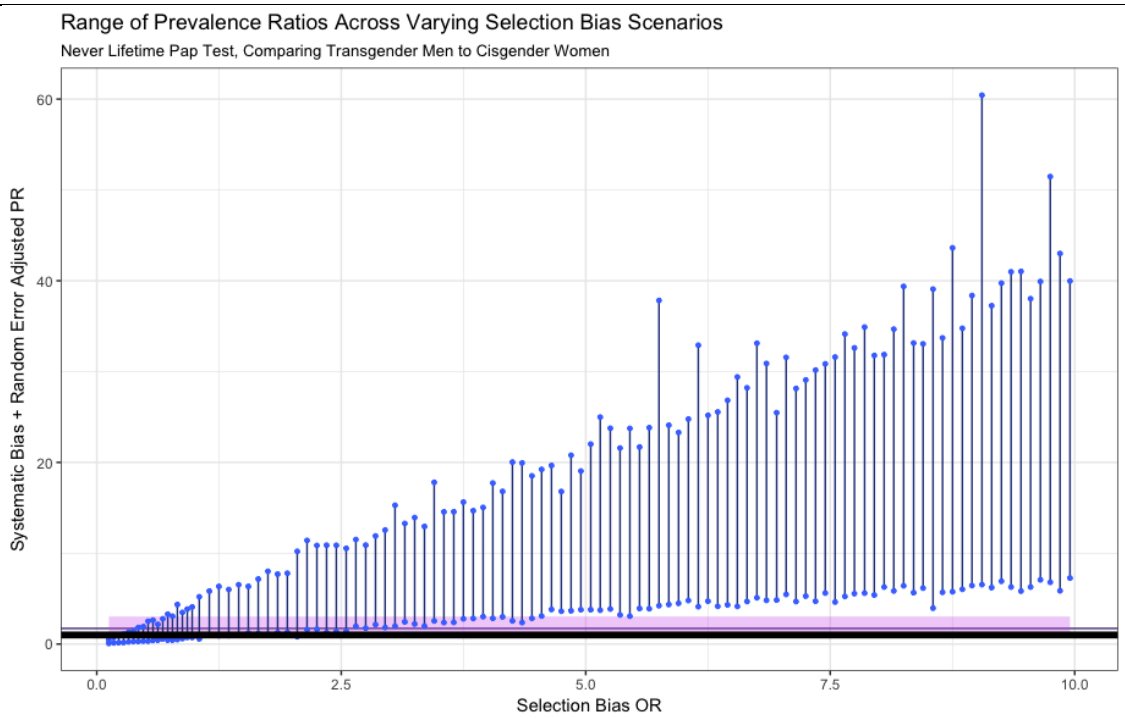


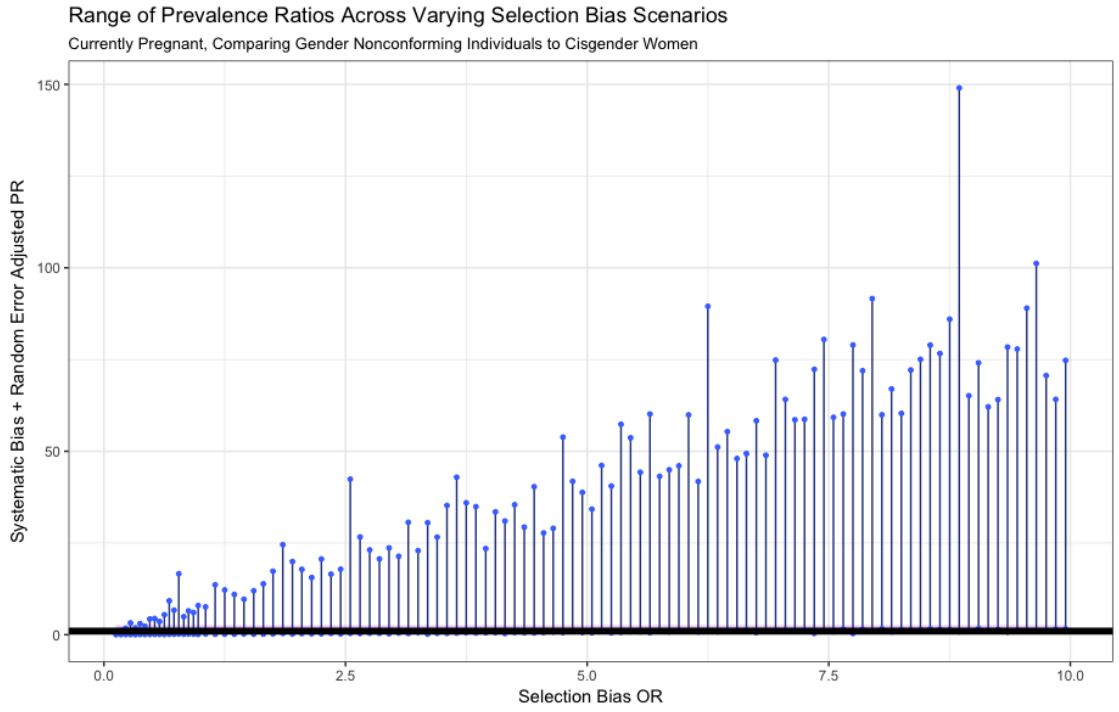

Range of Prevalence Ratios Across Varying Selection Bias Scenarios
Never Lifetime Pap Test, Comparing Transgender Men to Cisgender Women
Systematic Bias + Random Error Adjusted PR
60
40
20
0
0.0
2.5
5.0
7.5
10.0
Selection Bias OR


Range of Prevalence Ratios Across Varying Selection Bias Scenarios
Never Lifetime PSA test, Comparing Gender Nonconforming Individuals to Cisgender Men
Systematic Bias + Random Error Adjusted PR
20
10
0
0.0
2.5
5.0
7.5
10.0
Selection Bias OR


Range of Prevalence Ratios Across Varying Selection Bias Scenarios
Hysterectomy, Comparing Transgender Men to Cisgender Women
Systematic Bias + Random Error Adjusted PR
Selection Bias OR
0
5
10
15
20
0.0
2.5
5.0
7.5
10.0


Range of Prevalence Ratios Across Varying Selection Bias Scenarios
Hysterectomy, Comparing Gender Nonconforming Individuals to Cisgender Women
Systematic Bias + Random Error Adjusted PR
Selection Bias OR
0
5
10
0.0
2.5
5.0
7.5
10.0


Range of Prevalence Ratios Across Varying Selection Bias Scenarios
Never Lifetime Pap Test, Comparing Transgender Men to Cisgender Women
Systematic Bias + Random Error Adjusted PR
60
40
20
0
0.0
2.5
5.0
7.5
10.0
Selection Bias OR


Range of Prevalence Ratios Across Varying Selection Bias Scenarios
Currently Pregnant, Comparing Gender Nonconforming Individuals to Cisgender Women
Systematic Bias + Random Error Adjusted PR
150
100
50
0
0.0
2.5
5.0
7.5
10.0
Selection Bias OR

APPENDIX II. BIAS MODELING APPROACH TWO

Step One: Multiple Imputation by Chain Equations

Multiple imputation by chained equations (MICE) was executed in Stata version 15.1 using the standard multiple imputation package.[4] The largest among of missingness was observed in the sex-specific outcomes, with up to 30% of data missing. Therefore, we create 30 imputed datasets, each with a 10 iteration burn in.[5–7]

We use conditional imputation to allow us to restrict the impute of sex-specific outcomes only to the relevant subpopulations. We restricted the imputation of the hysterectomy and Pap test variables to cisgender women and transgender men. We restricted the imputation of the pregnancy variable to cisgender women and transgender men age 44 and younger. Lastly, we restricted the imputation of the PSA test variable to cisgender men and transgender women age 40 and older. These age restrictions mirror BRFSS skip patterns based on self-reported age.

Perfect prediction and non-convergence occurs often in imputation models with a large number of categorical variables.[8] Therefore, we employed three strategies to achieve model convergence. First, we model all ordinal categorical variables as grouped linear variables to reduce the number of categorical variables in the imputation model. Second, we use an augmented regression approach to prevent perfect prediction.[9] Lastly, we estimate a separate imputation model for each of the four sex-specific variables, since the inclusion of multiple conditional variables markedly increased computation time and likelihood of non-convergence. Imputed values were not rounded, since rounding can introduce additional bias.[10]

We assessed convergence and goodness of fit by visually inspecting the trace plots for each imputed variable (**Figure A2**). Due to the size of the dataset (518,982 observations), a high degree of missingness, and reliance on a large number of categorical and conditional variables, this multiple imputation task was computationally intensive, with a total computation time over 60 hours.

Step Two: Summary-level adjustment to account for unknown confounding

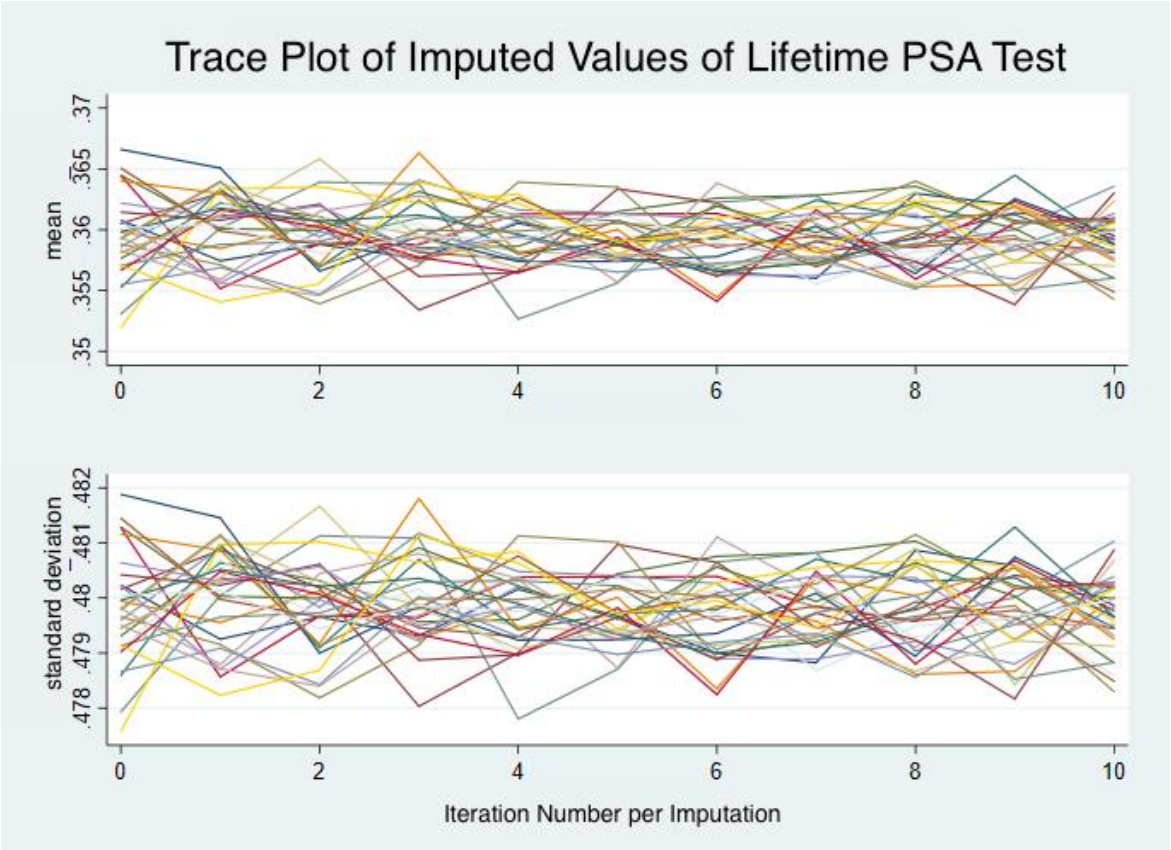
Since multiple imputation methods assume missingness at random (MAR), we subsequently conduct probabilistic adjustment for unknown confounding. We define the risk ratio due to confounding RR_{conf} so that prevalence ratio adjusted for confounding as $\mathrm{PR}_{adj} = \frac{\widehat{\mathrm{PR}}}{\mathrm{RR}_{conf}}$.[3]

There is no existing literature to inform the direction or degree of confounding, nor is there information to bound our estimates of the risk ratio due to confounding RR_{conf}.[3] Therefore, we use Monte Carlo methods to randomly sample 100,000 confounding scenarios. we consider a wide range of selection bias ratios under a uniform distribution where the parameter RR_{conf} ranges between 0.10 and 10.0. For each outcome, we sample from the uniform distribution of parameter RR_{conf} 100,000 times. For each iteration, we estimated two bias-adjusted prevalence ratios, one modeling systematic bias alone, and a second combining systematic bias and random errors. Random error is incorporated at each step in the simulation: letting $se_{imputed}$ be the robust standard error estimated from the Poisson regression on the imputed datasets, we calculate:

$$\mathrm{PR}_{adj} = \exp\left(ln\left(\frac{\widehat{\mathrm{PR}}}{\mathrm{RR}_{conf}}\right) + N(0,1) \times se_{inputed} \right)$$

where $N(0,1)$ is a random number drawn from the standard normal distribution.[3]

Figure A2. Trace Plots for Sex-Specific Outcome Variables



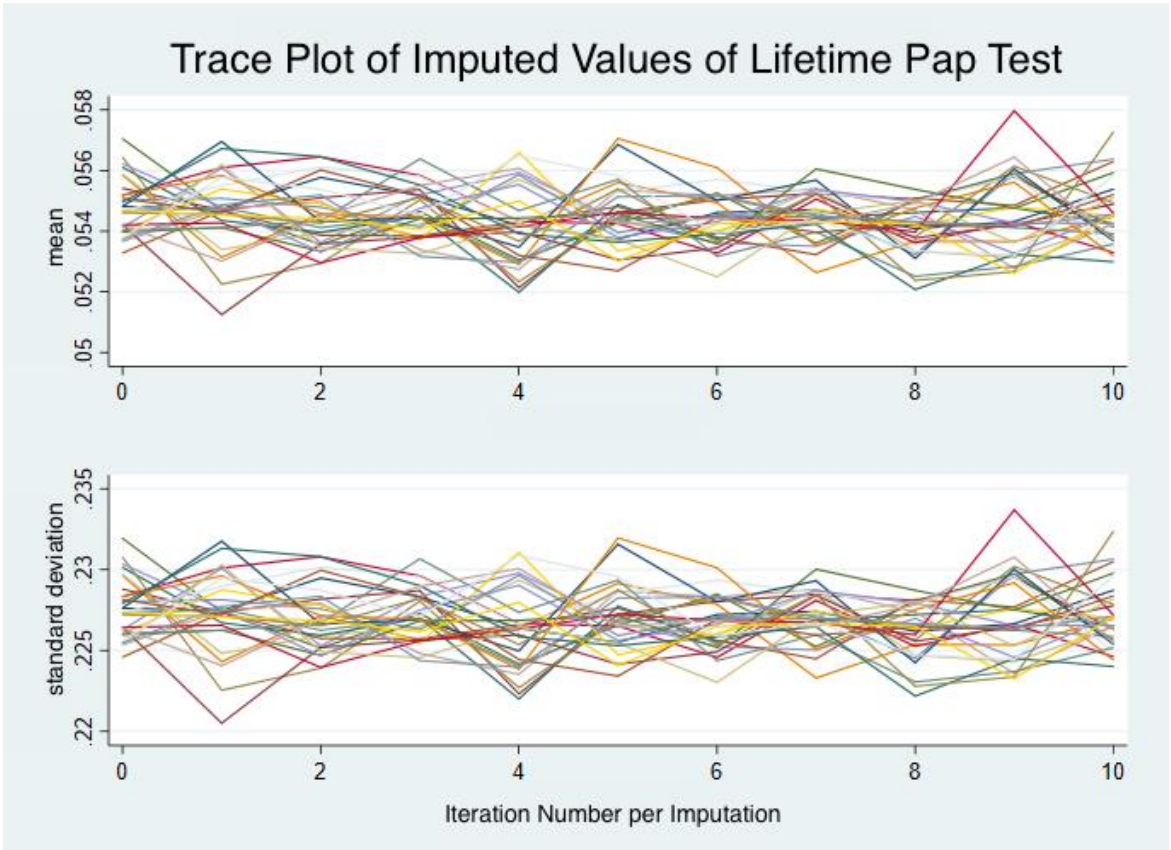
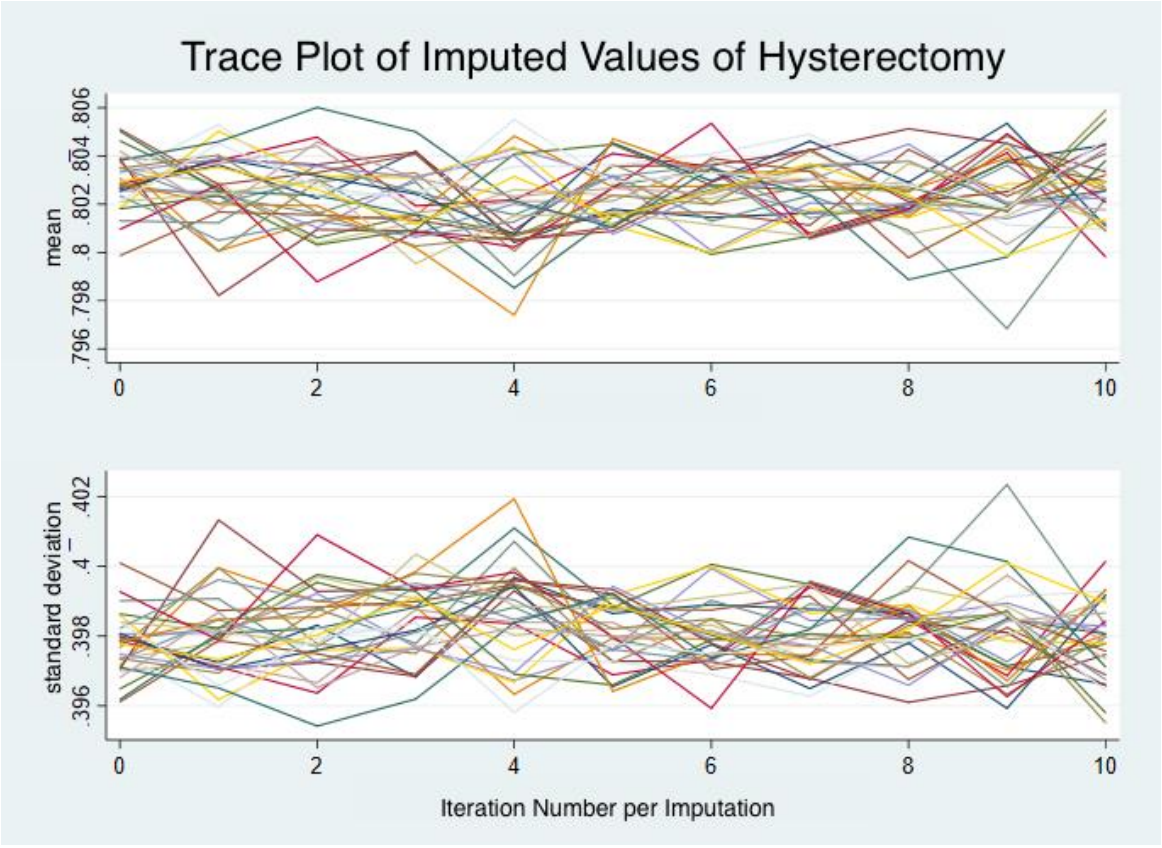
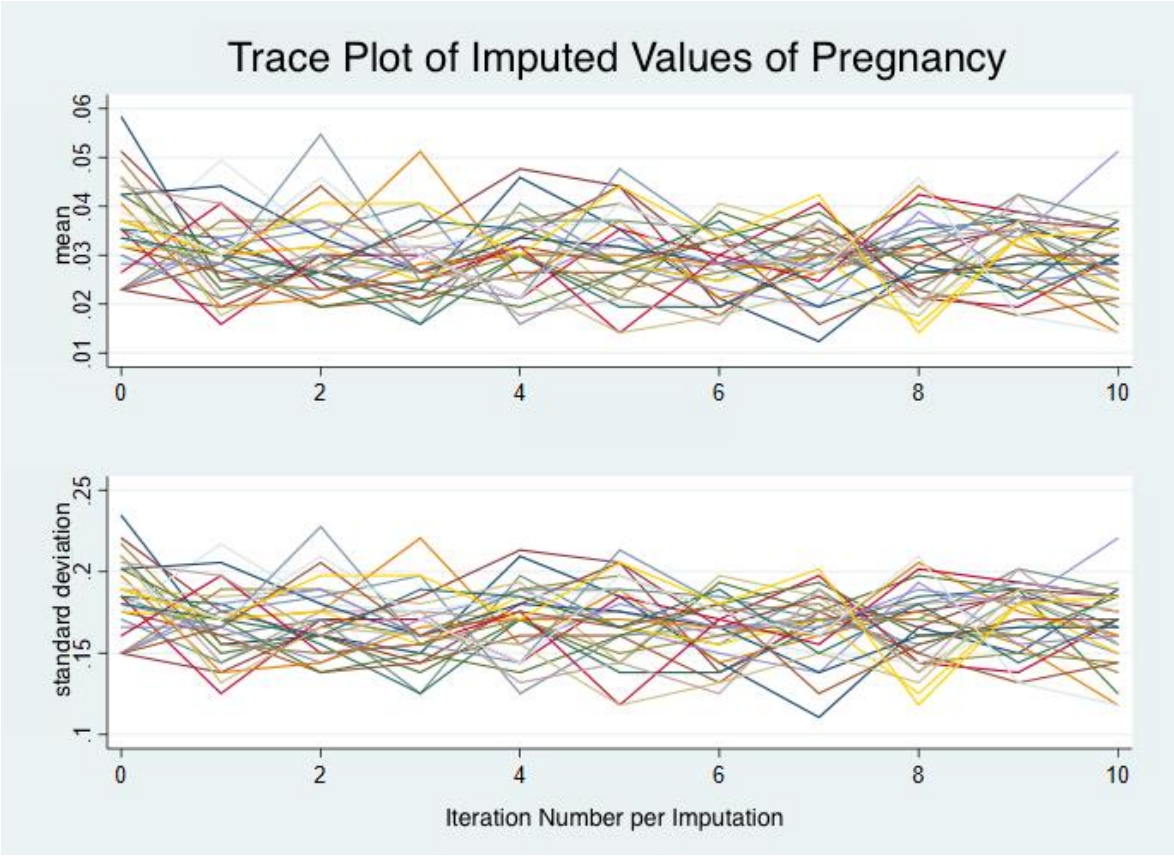


Figure A3 continued.





Appendix References

1. Riley NC, Blosnich JR, Bear TM, Reisner SL. Vocal Timbre and the Classification of Respondent Sex in US Phone-Based Surveys. *Am J Public Health*. 2017;107(8):1290-1294. doi:10.2105/AJPH.2017.303834.
2. Nohr EA, Liew Z. How to investigate and adjust for selection bias in cohort studies. *Acta Obstet Gynecol Scand*. 2018;97(4):407-416. doi:10.1111/aogs.13319.
3. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Springer; 2009.
4. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. doi:10.1002/sim.4067.
5. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1987. doi:10.1002/9780470316696.
6. Graham JW. Missing Data Analysis: Making It Work in the Real World. *Annu Rev Psychol*. 2009;60(1):549-576. doi:10.1146/annurev.psych.58.110405.085530.
7. Graham JW, Olchowski AE, Gilreath TD. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prev Sci*. 2007;8(3):206-213. doi:10.1007/s11121-007-0070-9.
8. Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*. 1984;71(1):1-10. doi:10.1093/biomet/71.1.1.
9. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Comput Stat Data Anal*. 2010;54(10):2267-2275. doi:10.1016/j.csda.2010.04.005.
10. Horton NJ, Lipsitz SR, Parzen M. A Potential for Bias When Rounding in Multiple Imputation. *Am Stat*. 2003;57(4):229-232. doi:10.1198/0003130032314.