Addressee:
Toni VanPelt
President, National Organization for Women

From:
Armin Hamp
Juan Castro
Aaron Morales

Summary

We believe that obesity's effect on the Gender Pay Gap warrants NOW's attention, as obesity amongst females has risen considerably in the past decades, and obesity was found to impact real earnings by (Bhattacharya & Bundorf, 2009). Taking a step further to analyse the data used by (Bhattacharya & Bundorf, 2009) and its replication paper (Turabi & Saynisch, 2014), we use Genetic Matching to remove the baseline differences among the groups of obese women, sorted into three racial categories (white, black, and other). We find that the dataset employed by both studies is highly unreliably for this purpose, as we were only able to confidently remove imbalance on the subgroup of white women. Calculating a causal estimate for this group, we found a significant effect (-$2.61) of obesity on hourly wages, and so we suggest NOW to flag this subgroup for lawmakers as a group who are economically endangered by the effects of obesity, and consequently threaten our goal of closing the Gender Pay Gap.

Background

The Gender Pay Gap between men and women has received considerable attention in the United States for the past few decades, where systematic legislation had an equalising effect. While the rate of convergence was highly promising in the 70s and 80s, this effect kept slowing down since the 90s up until today (Blau & Kahn, 2007). While the legislation in place should guarantee equal pay, numerous gender-specific factors have been identified as slowing down this process. Amongst others, obesity has been identified to have significant effects on the real earnings of workers (Blau & Kahn, 2017). Since 2002, there has been a significant increase in obesity among working-age women; according to the latest nationwide-estimates, the rate of obesity for 2009-2010 has been 32.2% for men and 35.5% for women (Ogden et al., 2014).

In their 2009 article, Jay Bhattacharya and Kate Bundorf investigate the causal relationship between obesity and hourly wage and find that obesity has a significant negative impact on workers who were insured by their employers (Bhattacharya & Bundorf, 2009)[1]. The appeal of their investigation is the identification of the causal mechanism of obesity's effect on real earnings. We believe that by expanding upon their findings, we can identify the groups most endangered by obesity negative effects on income.

Given the reputation and authority of the National Organization for Women, we believe that we could press lawmakers to pass the appropriate health measures and help the endangered groups in fighting obesity, and thus prevent a further slowing down of the convergence of the Gender Pay Gap.

Replication Paper & Objective of Expansion

B&B work from a nationally representative sample, the data of the National Longitudinal Survey of Youth or NLSY between 1989 and 2002. It is a large dataset that they have coupled with the Medical Expenditure Panel Survey (MEPS) in order to get up-to-date data on insurance status. B&B stratified by insurance status of employers and conducted a difference-in-difference test to calculate the effect of obesity on hourly wages. We constructed a simple linear regression model to replicate these results.

	Difference in Hourly Wage	
	B&B	**Regression Model**
Insured	-1.42 (0.40)	-1.45 (0.39)
Uninsured	0.25 (0.50)	0.57 (0.51)
Unadjusted Difference-in Difference Estimate	-1.68 (0.63)	-2.02*

Table 1. The effect of obesity on Hourly Wages for insured and uninsured individuals, as calculated in B&B and our linear regression model. In brackets are SE calculated for our estimates, except for *.

[1] Henceforth B&B.

Given the representative nature of the dataset, B&B did not undertake any steps of preprocessing beyond the integration of the NLSY and MEPS datasets. Their final dataset was of 31,176 observations of over 200 variables, of which they have identified 33 variables to use as controls. In a replication study, (Turabi & Saynisch, 2014)[2], conducted Mahalanobis matching on the original dataset (with the intention to achieve balance on the same 24 covariates as B&B) and stratified for race and gender (only distinguishing between black and white) to find that the only significant offset of wages was present amongst white women. However, we find the matching results of T&S to be poor, with multiple covariate distributions having <0.001 p-values post-matching.

Therefore, we decided to further investigate if achieving appropriate covariate balance on the original dataset was possible and if our results with potentially better balance measures would show the same trends as expected from T&S. Furthermore, we extend our analysis to the racial category of "others", so as to be able to give appropriate guidance on endangered groups if significant results are found. Also, we would like to further scrutinise any significant results with a sensitivity analysis.

Analysis and Results

Table 2 shows the comparison between the balances among the covariates considered in both the original and replication paper.

Category	**Original**	**Mahalanobis**	**Genetic Matching**
Other	2.22e-16	9.1122e-08	0.064
White	2.22e-16	2.22e-16	**0.10078**
Black	0.00019897	0.024363	0.055833

Table 2. Lexical p-values for the balance on the 33 covariates used in the original and replication papers on women for white, black and other races.

[2] Henceforth T&S.

As a rule of thumb, we considered good balance when the p-value was higher than 0.10. As we can see on the table, Genetic Matching for White is the only matching that achieved a p-value higher than 0.10. This implies that only the estimates calculated for white women can be considered to fulfill the assumptions of the model of causal inference (one of the most important of which is the lack of baseline differences between control and treated units), and could then be considered significant if the p-value for such estimate suggests so. In Table 3, we included the estimates and p-values for all three of the racial categories (black, white and other), which serves as a nice comparison between the Mahalanobis results of T&S and our Genetic Matching results. However, given the balance results from Table 2, we are mostly interested in the estimate and p-value for White women after genetic matching.

Category	**Mahalanobis**		**Genetic Matching**	
	estimate	p-value	estimate	p-value
Other	-0.014234	0.98372	-1.0295	0.14834
White	-2.5068	0.0025748	**-2.6067**	**0.0010156**
Black	0.27703	0.59333	-0.035119	0.94816

Table 3. Estimate a p-value of the treatment effect between insured and obese and non-obese women for white, black, and other races.

From the table, we can now say the treatment effect of -2.6067 for white women has a significant p-value of 0.001. One more analysis will now be made to determine the reliability of this result. Table 4 shows the Rosenbaum Sensitivity Test on the found significant treatment effect.

Gamma Γ	Lower bound	Upper bound
1.0	0	0.0000
1.1	0	0.0001
1.2	0	0.0005
1.3	0	0.0014
1.4	0	0.0034
1.5	0	0.0071
1.6	0	0.0134
1.7	0	0.0233
1.8	0	0.0373
1.9	0	0.0562
2.0	0	0.0803

Table 4. Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-value on white women.

Rosenbaum's Sensitivity Test measures the amount of bias that can be present in the pre-treatment differences among our units. The Gamma value is the factor ≥ 1 for which the odds of treatment for one unit can be greater than for another. The sensitivity analysis will tell us how large the Gamma value can be before the conclusions start to differ from our original findings. The greater this value is, the best our conclusion is as it allows for more hidden bias in our units that would still not change our qualitative conclusions. (Rosenbaum, 2005). For our white women, the upper bound surpasses the 0.05 value between 1.8 and 1.9, meaning that our last significant Gamma value is 1.8. This value means that anyone to refute the significance of this result must suggest that there exist an unobserved covariate, which is 1.8 times more common amongst obese people than non-obese. Which may be true, but is probably unlikely.

Data

The core of the dataset, that was used to calculate the results in this report come from the national longitudinal survey NLSY, which were adopted by B&B, who aggregated the data and averaged the covariates of interest for the period 1989-2002. The results obtained by B&B (such as the treatment effects calculated in Table 1) were based on the assumptions that the dataset is nationally representative and is large enough sample (>30,000 observations). However, these assumptions remain implicit, and are irrelevant for establishing a reliable causal estimate for the question of interest, the effect of obesity on real earnings.

As pointed out by T&S as well, in order to obtain reliable estimates of the causal effect of obesity a covariate balance between obese and non-obese groups would be necessary. However, as our results in Table 2 suggest, there is a terrible balance among the covariates (with values much lower than 0.10) that both the original and replication paper use. Therefore, one cannot rely on the estimates proposed by the former papers even when significant p-values for such have been found.

The Genetic Matching method as outlined in (Diamond & Sekhon, 2013) is a tool that so far has proven to give us the highest chance to achieve balance among covariates, especially when these are not few. After the Genetic Matching, we found that only an estimate for white women fulfills the lack of baseline differences (removed by our Genetic Matching), is significant, and has a moderately low sensitivity towards bias.

Thus, our only reliable estimate concerns the group of white women, who when obese earn on average $2.61 less than if not. These are shocking results, and could reasonably impact the livelihood of these women, as well as the status of the Gender Pay Gap. We then suggest NOW to flag this demographic for lawmakers as endangered. The estimates for women other than black or white also seems impactful (-$1.03), but due to the unreliability of the data, even after the preprocessing with Genetic Matching, we cannot say anything for certain. However, it is a question important enough to warrant further investigation through the construction of a new and more balanced dataset either in the framework of an observational study or experiment.

Conclusion

Following the failure of preprocessing techniques in T&S and motivated by the goal to understand which demographic of women are most impacted by obesity in terms of real earnings, we performed Genetic Matching. We find that out of our three strata, black, white and other women covariate balance could only be achieved on one of them. This group, white women, were also the only whose treatment estimates were significant. We further sophisticated our analysis by performing a Rosenbaum Sensitivity Test and found that the estimate becomes insignificant at $\Gamma > 1.8$, which suggests that this estimate is rather insensitive to hidden bias.

The causal estimate for obesity on hourly wages for white women is staggering. We advise NOW, to pay particular attention to this subgroup and take an active part in advocating for institutional programmes designed to reduce obesity in this subgroup. We hope that such measures will have an impact on our ultimate goal for the eradication of the Gender Pay Gap. As we have been unable to obtain reliable estimates for the other subgroups, we suggest further studies to be conducted to establish whether they are warranted NOW's attention.

Appendix

Code for replication

https://gist.github.com/hamparmin/2491c0ec051b41d55809801bd7c679fb

Data for replication

https://drive.google.com/file/d/1dRj_Mfey5NJQzq8EGkNmjUdYYfReNqvV/view?usp=sharing

References

Blau, F. D., & Kahn, L. M. (2007, 02). The Gender Pay Gap. *Academy of Management Perspectives, 21*(1), 7-23. doi:10.5465/amp.2007.24286161

Blau, F., & Kahn, L. (2016, 01). The Gender Wage Gap: Extent, Trends, and Explanations. doi:10.3386/w21913

Diamond, A., & Sekhon, J. S. (2013, 07). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics, 95*(3), 932-945. doi:10.1162/rest_a_00318

Rosenbaum, P. R. (2005, 10). Sensitivity Analysis in Observational Studies. *Encyclopedia of Statistics in Behavioral Science*. doi:10.1002/0470013192.bsa606

Rosenbaum, P. R. (2014, 09). Sensitivity Analysis in Observational Studies. *Wiley StatsRef: Statistics Reference Online*. doi:10.1002/9781118445112.stat06358