

OPEN

# Evaluating Glaucoma Treatment Effect on Intraocular Pressure Reduction Using Propensity Score Weighted Regression

Mengfei Wu<sup>1,2</sup>, Mengling Liu<sup>2,3\*</sup>, Joel S. Schuman<sup>1,4</sup>, Yuyan Wang<sup>2</sup>, Katie A. Lucy<sup>1</sup>, Hiroshi Ishikawa<sup>1</sup> & Gadi Wollstein<sup>1,4</sup>

Observational studies in glaucoma patients can provide important evidence on treatment effects, especially for combination therapies which are often used in reality. But the success relies on the reduction of selection bias through methods such as propensity score (PS) weighting. The objective of this study was to assess the effects of five glaucoma treatments (medication, laser, non-laser surgery (NLS), laser + medication, and NLS + medication) on 1-year intraocular pressure (IOP) change. Data were collected from 90 glaucoma subjects who underwent a single laser, or NLS intervention, and/or took the same medication for at least 6 months, and had IOP measures before the treatment and 12-months after. Baseline IOP was significantly different across groups ( $p = 0.007$ ) and this unbalance was successfully corrected by the PS weighting ( $p = 0.81$ ). All groups showed statistically significant PS-weighted IOP reductions, with the largest reduction in NLS group ( $-6.78$  mmHg). Baseline IOP significantly interacted with treatments ( $p = 0.03$ ), and at high baseline IOP medication was less effective than other treatments. Our findings showed that the 1-year IOP reduction differed across treatment groups and was dependent on baseline IOP. The use of PS-weighted methods reduced treatment selection bias at baseline and allowed valid assessment of the treatment effect in an observational study.

Glaucoma is a leading cause of visual impairment and blindness, both in the United States and worldwide. As a chronic and complex disease, its management often requires the use of medications and/or surgery, with the primary clinical goal of lowering intraocular pressure (IOP) and ultimately halting further damage. A variety of treatment options are now available including topical and oral medications<sup>1–5</sup>, laser<sup>6</sup>, and non-laser surgery (NLS)<sup>7</sup>. All are designed to either decrease aqueous humor production or increase aqueous outflow. Guidance for preferred treatment is typically derived from clinical trials that are mostly designed to evaluate the effect of a single medication or surgical procedure in highly controlled settings. However, many subjects require combined treatments to halt or slow damage over the course of their disease. Thus, results from clinical trials may not reflect the complex interplay among commonly used combination regimens or may not apply to subjects who do not fit stringent trial selection criteria.

Unlike randomized controlled trials, observational studies often have pre-treatment (baseline) characteristics that are associated with treatment assignment, thereby introducing confounding effects when evaluating treatment efficacy. The propensity score (PS) method is commonly used in observational studies to reduce bias caused by different patient characteristics at baseline aiming to mimic treatment randomization<sup>8</sup>. PS is defined as the probability of receiving treatment assignment conditional on observed baseline covariates<sup>9</sup>, and combining all baseline variables into a single score is particularly convenient when there are a large number of baseline covariates<sup>10</sup>. Several PS methods are commonly used: PS stratification, PS matching, and PS weighting. The PS weighting method is particularly useful because all participants can be included in the study, unlike PS matching

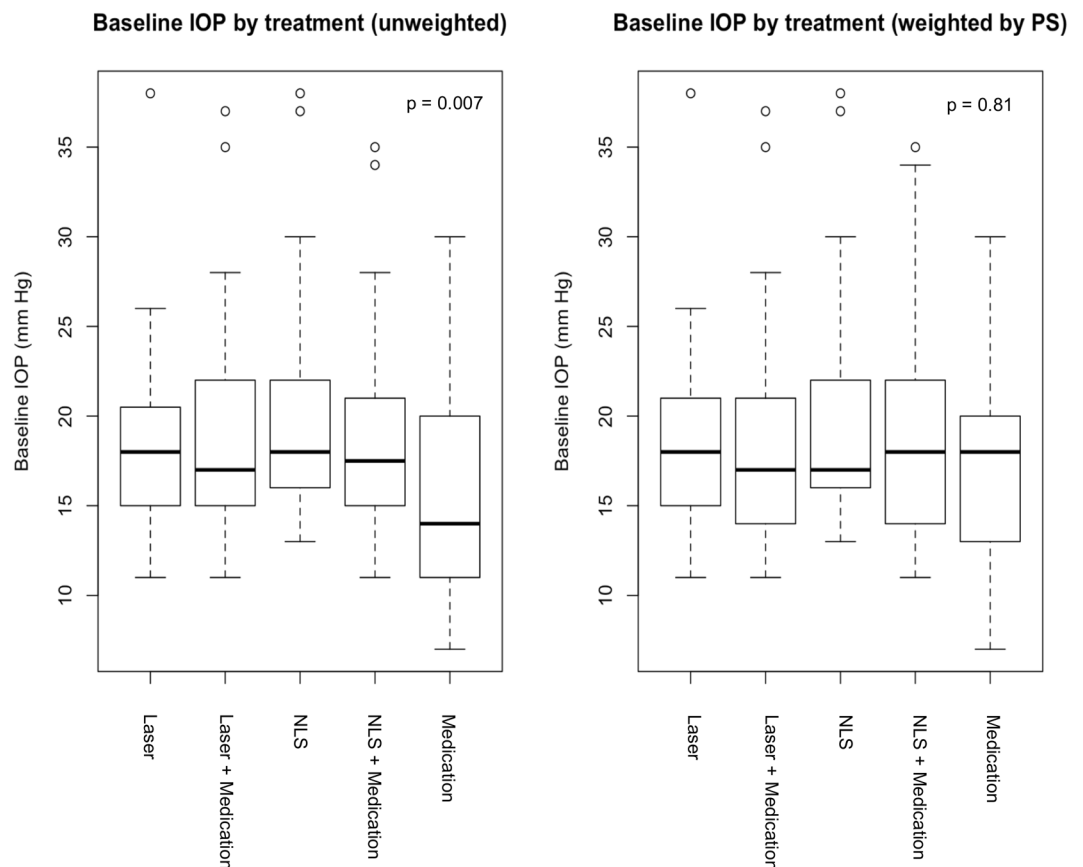
<sup>1</sup>Department of Ophthalmology, New York University (NYU) School of Medicine, NYU Langone Health, New York, NY, USA. <sup>2</sup>Department of Population Health, NYU School of Medicine, NYU Langone Health, New York, NY, USA.

<sup>3</sup>Department of Environmental Medicine, NYU School of Medicine, NYU Langone Health, New York, NY, USA.

<sup>4</sup>Center of Neural Science, NYU, New York, NY, USA. \*email: [mengling.liu@nyulangone.org](mailto:mengling.liu@nyulangone.org)

	Medication (n = 33)	Laser (n = 56)	NLS (n = 33)	Laser + Medication (n = 41)	NLS + Medication (n = 34)	p value
Age (years)	64.0 (11.4)	64.4 (8.6)	63.9 (8.4)	63.0 (13.9)	65.1 (9.9)	0.90
Caucasian*	30 (88)	43 (77)	25 (76)	36 (88)	29 (85)	0.29
Female*	17 (52)	34 (61)	21 (64)	28 (68)	20 (59)	0.28
Baseline IOP (mmHg)	15.2 (5.8)	18.2 (4.7)	20.3 (6.2)	19.0 (6.0)	18.4 (5.9)	0.007**

**Table 1.** Demographics by Treatment Group for All Segments. NLS = Non-Laser Surgery; IOP = Intraocular Pressure. \*Ethnicity and gender were reported in N (%), age and baseline IOP were reported in mean (standard deviation). \*\*Statistically significant (ANOVA test).



**Figure 1.** Baseline IOP by Treatment Group Before and After PS. PS = Propensity Score; NLS = Non-Laser Surgery; IOP = Intraocular Pressure.

where some subjects might be excluded due to matching criteria<sup>11</sup>. A successful application of PS weighting can balance baseline characteristics across treatment groups to improve the estimation of treatment effects.

The purpose of this study was to determine glaucoma treatment effect on IOP reduction in real-life settings. We applied the PS weighting method for a longitudinal cohort of subjects with glaucoma that were treated with a single treatment or a combination of treatments, including medication, laser, and NLS.

## Results

One hundred twenty-two eyes (90 subjects) were included to provide 197 treatment segments of 1-year duration. The total number of segments included 33 medication, 56 laser, 33 NLS, 41 laser + medication, and 34 NLS + medication segments. Demographics are presented in Table 1. Baseline IOP differed significantly across the five treatment groups ( $p = 0.007$ ), with the lowest baseline IOP in the medication group (mean  $\pm$  standard deviation:  $15.2 \pm 5.8$  mmHg), and the highest baseline IOP in the NLS group ( $20.3 \pm 6.2$  mmHg).

Given the imbalance in baseline IOP across groups, we applied the generalized boosted model (GBM) for PS weighting and found that the weighted baseline IOPs were no longer different across groups ( $p = 0.81$ ; Fig. 1).

We estimated the average 1-year IOP changes for all treatment groups with 95% confidence intervals (CIs) either unadjusted or adjusted by the PS weights from GBM (Table 2). Overall, the treatment groups showed statistically significant difference in their effects on IOP reduction with ( $p = 0.02$ ) or without ( $p = 0.003$ ) PS weighting.

	1-year IOP change (mmHg)	
	Mean (95% CI)	
	Unweighted	PS-weighted
Medication (n = 33)	-1.88 (-3.56, -0.20)	-3.05 (-4.71, -1.39)
Laser (n = 56)	-2.81 (-4.24, -1.38)	-2.79 (-4.20, -1.37)
NLS (n = 33)	-7.30 (-9.76, -4.84)	-6.78 (-9.02, -4.54)
Laser + Medication (n = 41)	-3.42 (-5.54, -1.30)	-2.82 (-4.86, -0.79)
NLS + Medication (n = 34)	-4.53 (-6.64, -2.42)	-4.97 (-7.31, -2.62)
p value	0.003*	0.02*

**Table 2.** Estimated Treatment Effect by Treatment Group with 95% CI. CI = Confidence Interval; PS = Propensity Score; NLS = Non-Laser Surgery; IOP = Intraocular Pressure. \*Statistically significant (ANOVA test).

Parameters (Laser as Reference)	Estimate (SE)	p value
Intercept	12.73 (1.58)	<0.001*
Medication	-6.69 (2.35)	0.006*
NLS	-1.45 (2.06)	0.48
Laser + Medication	3.57 (2.96)	0.23
NLS + Medication	-1.25 (4.05)	0.76
Baseline IOP	-0.86 (0.06)	<0.001*
Baseline IOP × Medication	0.40 (0.12)	0.001*
Baseline IOP × NLS	-0.06 (0.11)	0.57
Baseline IOP × (Laser + Medication)	-0.14 (0.16)	0.38
Baseline IOP × (NLS + Medication)	-0.01 (0.23)	0.89

**Table 3.** PS-weighted Multiple Regression on 1-year IOP Change. PS = Propensity Score; NLS = Non-Laser Surgery; IOP = Intraocular Pressure; SE = Standard Error. \*Statistically significant.

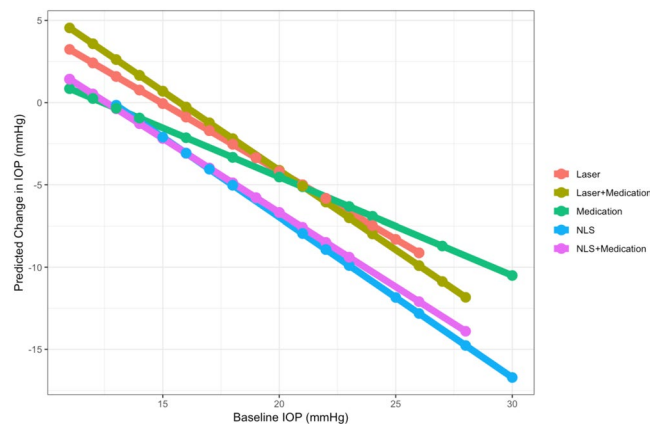
The NLS and NLS + medication groups showed the largest IOP decreases, and their PS-weighted estimates were -6.78 mmHg (95% CI: [-9.02, -4.54]) and -4.97 mmHg (95% CI: [-7.31, -2.62]), respectively. The medication, laser, and laser + medication groups had smaller IOP reductions. Notably, using PS weighting revealed a larger treatment effect in the medication group after correcting for the IOP imbalance at baseline. For example, the unweighted effect of medication was -1.88 mmHg (95% CI: [-3.56, -0.20]), while the PS-weighted estimate was -3.05 mmHg (95% CI: [-4.71, -1.39]). This demonstrated that the medication effect was potentially larger than it appeared to be once the baseline selection bias was adjusted.

Considering that baseline IOP in the medication group was significantly lower than the other treatment groups, we further investigated the interaction between treatments and baseline IOP (Table 3) using a PS-weighted multiple linear regression model. The F-test showed that the overall interaction effect was significant ( $p = 0.03$ ), mainly driven by the significant interaction between the medication group and baseline IOP ( $p = 0.001$ ). Figure 2 demonstrates the results of the multiple regression model. The medication group had a significantly flatter slope than other groups due to the positive interaction with baseline IOP, indicating that medication alone became less effective than other treatments as the baseline IOP increased and was the least effective when baseline IOP >21 mmHg. The NLS and NLS + medication groups consistently showed the largest 1-year IOP reduction for patients regardless of their baseline IOPs.

## Discussion

Our goal was to assess the effect of medication, laser, NLS, and their combination on 1-year IOP reduction in a retrospective analysis of a longitudinal cohort of glaucoma subjects. In practice, when a clinician decides what treatment to offer, the choice is heavily influenced by the subject's baseline characteristics and the expected outcome of treatment. Subjects with persistently high IOP at baseline might be assigned to more invasive interventions, especially if they previously failed to respond to medication. This could be a reason why baseline IOP was highest in the NLS treatment group and lowest in the medication only group. The imbalance of baseline IOP can also introduce bias in evaluating the effect of different treatments, as a bigger IOP reduction might be the result of a higher baseline IOP, not necessarily the use of a more effective treatment. Consequently, removing the imbalance in baseline characteristics is critical to achieve unbiased treatment comparison.

Our analysis successfully demonstrated that the application of PS weighting effectively reduces treatment selection bias at baseline in an observational cohort and allowed a more accurate assessment of treatment effect. Our PS-weighted results showed that treatment effects differed across five treatment groups, with NLS demonstrating a larger 1-year IOP reduction than medication and laser groups, similar to what was published by a



**Figure 2.** PS-weighted Predicted 1-year IOP change by Treatment Group. NLS = Non-Laser Surgery; IOP = Intraocular Pressure.

previous meta-analysis<sup>12</sup>. We also demonstrated that treatment effect was dependent on baseline IOP. For example, subjects with a lower baseline IOP appeared to benefit more from medication than laser or laser + medication, while subjects with a higher baseline IOP benefitted more from treatments that included laser and NLS.

By evaluating various treatment options and aggregating interventions into medication, laser, NLS, and their combinations, we were able to estimate treatment effect and obtain results in line with current clinical knowledge and published research<sup>13–16</sup>. This method can be extended for further analysis considering additional treatment options, more baseline characteristics, and by accounting for factors that affect the treatment outcome such as side effects and complications. This method also has the advantage of providing strong clinical implication as it reveals the clinical reality of a longitudinal cohort, and can be used to guide treatment selection and prospective clinical studies in the future.

A post-hoc sample size calculation showed that our study was sufficiently powered, using either unweighted or PS-weighted effect estimates. Based on the unweighted estimates for treatment effects and an average between-group variance of 40 from Table 2, a sample of 145 (29 per group) would provide the study with 80% power to show difference in the mean treatment effects across groups with type I error of 5% using an ANOVA test. A more conservative power analysis using the PS-weighted estimates from Table 2 and assuming the same between-group variance showed that a sample of 200 (40 per group) would be needed to achieve 80% power.

There are a number of limitations to our study. Even though PS weighting minimized the imbalance of baseline IOP across groups, it is not a perfect substitute for a randomized clinical trial as it cannot completely remove selection bias. Also, baseline IOP was not very high in our cohort, indicating that many subjects had likely been on treatment before baseline, potentially affecting the treatment effect estimations in this study. One possible explanation to why patients with higher baseline IOPs benefited less from medication could be that within the medication group, patients with a low baseline IOP might only need a single medication to manage glaucoma; while patients with a high baseline IOP might have previously failed to respond well to a single medication and need double or triple medications. Also, some patients with higher baseline IOPs in our cohort could be referrals to our clinics after being treated elsewhere, consequently presented a greater need for surgical interventions. Furthermore, the impact of other concomitant eye surgeries, such as cataract removal, was not addressed. Prior meta-analysis has indicated that there is weak evidence that combined cataract and glaucoma surgery may result in better IOP control 1-year after surgery compared with cataract surgery alone<sup>17</sup>. Lastly, VF and OCT measurements, which are highly relevant in the context of treatment selection and outcome, were not considered in this study. Nevertheless, these limitations should be guarded in the context of our proof of principle study.

In conclusion, applying PS weighting to real-life clinical data reduces selection bias and allows evaluation of the IOP reduction effect of the commonly used glaucoma treatment options.

## Methods

**Subject inclusion and treatment groups.** Subjects in this study came from a large, ongoing, National Institute of Health (NIH)-funded, longitudinal cohort of healthy subjects, glaucoma suspects, and subjects with glaucoma (R01 EY13178). All subjects in this cohort underwent comprehensive ophthalmic evaluation including medical history, IOP measurement, clinical examination of anterior and posterior segments, visual field (VF), and optical coherence tomography (OCT) testing. The institutional review boards and ethics committees at the New York University School of Medicine and the University of Pittsburgh approved the study. The study followed the tenets of the Declaration of Helsinki and was conducted in compliance with the Health Insurance Portability and Accountability Act (HIPAA). Informed consent was obtained from all subjects. Subjects who were diagnosed with glaucoma by their ophthalmologists and did not have confounding systemic or ocular comorbidities were qualified for the current study. Glaucoma diagnosis was based on the presence of typical structural and functional glaucomatous damage. Medication and surgical records were collected starting from the subjects' first visit to our clinics.

Throughout follow-up, glaucoma patients could be treated with medication, laser, NLS, or a combination of medication and surgery. The following five treatment groups were considered: medication, laser, NLS,

laser + medication, and NLS + medication. Medication included the use of prostaglandin analogues, beta blockers, carbonic anhydrase inhibitors, and/or alpha-adrenergic agonists. NLS included trabeculectomy, shunts/tubes, and minimally invasive glaucoma surgery. Subjects, who received a single laser or NLS intervention and/or took the same medication during a period of at least 6 months and had IOP measures before the treatment and 12-months after, were included in our study.

**12-Month treatment segments for analysis.** A treatment segment was defined as the period of time when a subject was maintained on the same treatment without adding, subtracting, or switching to a different medication or undergoing a second surgery. The first 12 months of any qualified segment were selected for analysis.

**Study outcome.** IOP information was collected at each clinical visit. The last IOP measured within 3 months prior to treatment initiation was used as the baseline IOP of a treatment segment. One-year change in IOP was the outcome. If a subject did not have IOP measured at exactly the end of 12 months, two IOP measurements immediately before and after the end of one year were used to impute the value by linear interpolation.

**Statistical analysis.** For continuous variables, the mean and standard deviation (SD) were reported and the analysis of variance (ANOVA) test was used to compare baseline covariates across the five treatment groups. For categorical variables, counts and proportions were reported and the chi-squared test was used to compare covariates across groups.

Generalized boosted model (GBM)<sup>18</sup>, a machine learning method, was used to calculate PS weights. Previous studies have shown that machine learning methods generally outperform traditional multinomial logistic models by achieving better balance among treatment groups with regards to baseline characteristics and producing more stable PS weights<sup>19</sup>. This is especially the case when there are more than two treatment groups, as in our study. A detailed description of the PS and GBM methods is provided in the supplementary information.

GBM analysis was performed using the Toolkit for Weighting and Analysis of Nonequivalent Groups (TWANG) R-package<sup>20</sup>. PS weights for treatment assignment were calculated based on each individual's baseline IOP, age, gender, and ethnicity. To assess the performance of PS weighting in removing selection bias, a balance analysis was then conducted for baseline characteristics before and after the implementation of PS weighting<sup>21</sup>. Treatment effect on the 1-year IOP change before and after PS weighting was estimated with 95% CIs and compared across treatment groups using the ANOVA test.

Given that baseline IOP was unbalanced across treatment groups, a PS-weighted multiple regression model was further built to evaluate the interaction effect of baseline IOP and treatments on the 1-year IOP change. The laser group served as the reference for treatment comparison. The F-test was used to assess whether the interaction effect was significant. The estimated 1-year IOP change was plotted to demonstrate the treatment effect along the range of baseline IOPs. For all analyses,  $p < 0.05$  was considered statistically significant.

## Data availability

The data analyzed in this study are available from the corresponding author on reasonable request.

Received: 8 August 2019; Accepted: 12 October 2019;

Published online: 29 October 2019

## References

- Weinreb, R. N., Aung, T. & Medeiros, F. A. The Pathophysiology and Treatment of Glaucoma. *JAMA* **311**, 1901–1911 (2014).
- Cantor, L. B. Brimonidine in the treatment of glaucoma and ocular hypertension. *Ther Clin Risk Manag* **2**, 337–346 (2006).
- Brooks, A. M. & Gillies, W. E. Ocular beta-blockers in glaucoma management. Clinical pharmacological aspects. *Drugs Aging* **2**, 208–221 (1992).
- Mincione, F., Scozzafava, A. & Supuran, C. T. The development of topically acting carbonic anhydrase inhibitors as anti-glaucoma agents. *Curr Top Med Chem* **7**, 849–854 (2007).
- Lindén, C. & Alm, A. Prostaglandin Analogues in the Treatment of Glaucoma. *Drugs & Aging* **14**, 387–398 (1999).
- Wong, M. O. M., Lee, J. W. Y., Choy, B. N. K., Chan, J. C. H. & Lai, J. S. M. Systematic review and meta-analysis on the efficacy of selective laser trabeculectomy in open-angle glaucoma. *Survey of Ophthalmology* **60**, 36–50 (2015).
- Razeghinejad, M. R., Fudenberg, S. J. & Spaeth, G. L. The Changing Conceptual Basis of Trabeculectomy: A Review of Past and Current Surgical Techniques. *Survey of Ophthalmology* **57**, 1–25 (2012).
- Austin, P. C. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* **46**, 399–424 (2011).
- Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
- Deb, S. *et al.* A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. *Canadian Journal of Cardiology* **32**, 259–265 (2016).
- Olmos, A. & Govindasamy, P. A Practical Guide for Using Propensity Score Weighting in R. *Practical Assessment, Research & Evaluation* **20**, 13 (2015).
- Boland, M. V. *et al.* Comparative Effectiveness of Treatments for Open-Angle Glaucoma: A Systematic Review for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* **158**, 271 (2013).
- Glaucoma Laser Trial Research Group. The Glaucoma Laser Trial (GLT) and Glaucoma Laser Trial Follow-up Study: 7. Results. *American Journal of Ophthalmology* **120**, 718–731 (1995).
- Samples, J. R. *et al.* Laser trabeculectomy for open-angle glaucoma: A report by the American academy of ophthalmology. *Ophthalmology* **118**, 2296–2302 (2011).
- Juzych, M. S. *et al.* Comparison of long-term outcomes of selective laser trabeculectomy versus argon laser trabeculectomy in open-angle glaucoma. *Ophthalmology* **111**, 1853–1859 (2004).

16. Netland, P. A. *et al.* Travoprost compared with latanoprost and timolol in patients with open-angle glaucoma or ocular hypertension. *Am. J. Ophthalmol.* **132**, 472–484 (2001).
17. Zhang, M., Hirunyachote, P. & Jampel, H. Combined surgery versus cataract surgery alone for eyes with cataract and glaucoma. *Cochrane Database of Systematic Reviews* 1465–1858, <https://doi.org/10.1002/14651858.CD008671.pub3> (2015).
18. McCaffrey, D. F. *et al.* A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine* **32**, 3388–3414 (2013).
19. Westreich, D., Lessler, J. & Funk, M. J. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology* **63**, 826–833 (2010).
20. Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L. & Griffin, B. A. Toolkit for Weighting and Analysis of Nonequivalent Groups: A tutorial for the twang package, <https://CRAN.R-project.org/package=twang> (2017).
21. Parast, L. *et al.* Optimizing Variance-Bias Trade-off in the TWANG Package for Estimation of Propensity Scores. *Health Serv Outcomes Res Methodol* **17**, 175–197 (2017).

## Acknowledgements

This work was funded by National Institutes of Health (NIH) through grant R01-EY013178 (Bethesda, MD).

## Author contributions

M.W. and M.L. conceived the original idea of this work. M.W. and K.A.L. cleaned the data. M.W. performed the analyses and wrote the paper. M.L., Y.W. and G.W. contributed to the discussion of the results and suggested improvements. All authors reviewed the manuscript.

## Competing interests

Dr. Schuman receives royalties for intellectual property licensed by Massachusetts Institute of Technology and Massachusetts Eye and Ear Infirmary to Zeiss.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-52052-5>.

**Correspondence** and requests for materials should be addressed to M.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019