

**Type of the Study:** Original Research Article

**Comparative Effectiveness of Traditional Chinese Medicine vs. Losartan on Blood Pressure:  
Real-World Insights from RCT-Eligible Populations**

**Short Running Title:** Generalize RCT Results to Real-World Trial-Eligible Population

Kuan Jiang<sup>1,2#</sup>, Xin-Xing Lai<sup>3,4#</sup>, Shu Yang<sup>5</sup>, Ying Gao<sup>3,4\*</sup>, Xiao-Hua Zhou<sup>1,2\*</sup>

<sup>1</sup> Department of Biostatistics, Beijing International Center for Mathematical Research, Peking University, Beijing, China PR

<sup>2</sup> Department of Biostatistics, School of Public Health, Peking University, Beijing, China PR

<sup>3</sup> Institute for Brain Disorders, Beijing University of Chinese Medicine, Beijing, China PR

<sup>4</sup> Department of Neurology, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China PR

<sup>5</sup> Department of Statistics, North Carolina State University, North Carolina, USA

# Kuan Jiang and Xin-Xing Lai contributed equally to this paper.

\* **Correspondence to:**

Ying Gao, M.D.

Institute for Brain Disorders, Beijing University of Chinese Medicine

No. 5 Haiyuncang, Beijing 100700, China PR

Email: [gaoying97@163.com](mailto:gaoying97@163.com)

Xiao-Hua Zhou, Ph.D.

Department of Biostatistics, School of Public Health, Peking University

Beijing 100871, China PR

Email: [azhou@math.pku.edu.cn](mailto:azhou@math.pku.edu.cn)

**Author contributions:** Concept and design: Kuan Jiang and Xin-xing Lai; data collection: Xin-xing Lai; data analysis: Kuan Jiang; critical revision of the article for important intellectual content: Yang Shu and Xiao-Hua Zhou; study supervision: Xiao-Hua Zhou and Ying Gao. All the authors approved the final article.

**Financial Support:** This work was supported by the Beijing Nova Program of Science and Technology (Z211100002121061) and the Young Elite Scientist Sponsorship Program by the China Association for Science and Technology (2021-QNRC1-04).

**Declarations of interest:** none

**Total word count :** 4561



## Comparative Effectiveness of Traditional Chinese Medicine vs. Losartan on Blood Pressure: Real-World Insights from RCT-Eligible Populations

### Abstract

When evaluating the effectiveness of a drug, a Randomized Controlled Trial (RCT) is often considered the gold standard due to its perfect randomization. While RCT assures strong internal validity, its restricted external validity poses challenges in extending treatment effects to the broader real-world population due to possible heterogeneity in covariates. In this study, we employed the augmented inverse probability of sampling weighted (AIPSW) estimator to generalize findings from an RCT comparing the efficacy of Songling Xuemaikang Capsule (SXC) — a traditional Chinese medicine (TCM) and Losartan on hypertension reduction to a real-world trial-eligible population. Additionally, we conducted sensitivity analyses to assess the robustness of the AIPSW estimation against unmeasured confounders. The generalization results indicated that although SXC is less effective in lowering blood pressure than Losartan on week 2, week 4, and week 6, there is no statistically significant difference among the trial-eligible population at week 8, and the generalization is robust against potential unmeasured confounders.

### Introduction

Hypertension is one of the most prevalent chronic diseases globally. Losartan, a recognized medication, is often prescribed for treating hypertension<sup>1-3</sup>. However, in China, Songling Xuemaikang Capsule (SXC) — a traditional Chinese herbal medicine, is recommended to treat mild hypertension clinically as an alternative in certain official guidelines<sup>4</sup>. Previous RCT comparing efficacy on lowering blood pressure between the two drugs demonstrated the non-inferiority of SXC to Losartan<sup>5</sup>. Although RCT serves as the golden standard for evaluating drug efficacy, it lacks representativeness, leading to debates when applying results to real-world situations. As evidence, there are substantive studies indicating the inconsistency between results obtained from RCT and a real-world study due to different population heterogeneity<sup>6-9</sup>. Compared with RCT, a registry study provides a sample more representative of the general population than RCT, but unmeasured confounders could diminish its reliability<sup>10</sup>. Since RCT and real-world study have their strengths and weaknesses, it is important to generalize RCT results to a real-world trial-eligible population to draw correct conclusions<sup>11</sup>.

There are some statistical methods available to solve the problem, which have been proven reasonable mathematically. Intuitively speaking, we can model outcomes in RCT directly and make predictions in real-world data by the fitted model (the outcome-model-based estimator)<sup>12</sup> or model the sampling score (probability of a unit being sampled into RCT population) and reweight RCT results by the inverse of sampling score (IPSW)<sup>13-15</sup>. However, these methods need a correctly specified model to yield consistent estimation. Augmented inverse probability of sampling weighted (AIPSW) method models outcome and sampling score simultaneously and can get consistent estimation if either of the two models is correctly specified, also called doubly robust estimation<sup>16,17</sup>. Recently, Lee et al. proposed a calibration weighting method to reweight

units in the RCT sample, and after calibration, the covariate distribution of the RCT sample empirically matches the real-world trial-eligible population<sup>18</sup>. These methods generalize RCT results to a real-world trial-eligible population by the following data structure: treatment assignment in RCT, the RCT outcomes, and observed common covariates in both RCT and real-world data. For reviews on existing methods, readers can refer to them for formulas and relative proofs<sup>17,19</sup>.

In addition to addressing observed covariates imbalance between RCT and real-world trial-eligible populations, exploring unobserved covariates is crucial and challenging for ensuring reliable generalization results by corresponding sensitivity analysis. Nguyen et al. proposed a set of methods to evaluate the bias of IPSW and model-based estimators when unobserved confounders exist. Still, the methods rely on a strict parametric setting<sup>20</sup>. Recently, based on Cinelli et al. and Douglas et al., Huang proposed several tools to assess the robustness of generalization results, including numeric statistics (robustness value, minimum relative confounder strength, etc.), bias counterplots, and a formal benchmarking approach. The advantage of Huang's method is that it does not require any assumptions on the data-generating process<sup>21-23</sup>.

In this article, we generalized the results in RCT comparing SXC with Losartan on the efficacy of treating hypertension in the real-world trial-eligible population through the AIPSW estimator. We found that conclusions in RCT and real-world populations were different. In RCT, there was no statistically significant difference in lowering systolic blood pressure (SBP) or diastolic blood pressure (DBP) between the SXC and Losartan group at all follow-up visits. However, in the real-world trial-eligible population, Losartan was more effective in lowering SBP in the early stages (weeks 2, 4, 6) but had a non-superior effect to SXC on lowering SBP and DBP at week 8. To address potential threats from unmeasured confounders, we performed a sensitivity analysis to assess the robustness of non-inferiority at week 8 by tools proposed in Huang et al. Corresponding sensitivity analysis on week 8 estimation indicated the robustness of generalization results against potential violation of assumptions. Therefore, SXC was not inferior to Losartan in lowering blood pressure (BP) in real-world trial-eligible populations in the long term.

## **Method**

### **Data Source**

We conducted two studies: an RCT and a registry study. The RCT was a multicenter, randomized, double-blind, non-inferiority trial that compared SXC with Losartan in terms of the efficacy of lowering blood pressure. It enrolled patients 18 to 65 years of age with mild essential hypertension (Grade I hypertension). In the treatment group, patients were assigned to receive SXC monotonously, and patients in the control group were assigned to receive Losartan. Details of the study design and the primary results were published previously<sup>5</sup>. The registry study collected patients in the real-world practice setting who used SXC monotonously or a combination of SXC and other medicines to lower hypertension. We gathered baseline covariates and conducted follow-up visits at weeks 2, 4, 6, and 8 in both studies. The study protocol was approved by the

Institutional Review Board of Dongzhimen Hospital affiliated to Beijing University of Chinese Medicine (approval number: ECSL-BDY-2011-19) and was registered on the Chinese Clinical Trial Registry Platform ([www.chictr.org.cn](http://www.chictr.org.cn); Unique identifier: ChiCTRONC-11001612)

### **Study Objective**

The primary objective of the study is to combine the RCT with real-world registry datasets to generalize RCT comparison results to a real-world trial-eligible population through the AIPSW estimator. Moreover, since the assumptions of generalization methods can be violated by unmeasured covariates, the robustness of the inference was assessed through corresponding sensitivity analysis.

### **Statistical Analysis**

#### **Common Variables Extraction, Missing Value Imputation and Trimming**

In the data preprocessing stage, we first extracted common baseline variables in RCT and real-world datasets to make the two datasets comparable. Because there are abnormal and missing values in variables 'weight' and 'marriage', we first deleted the abnormal values, treated them as missing values, and imputed them by multiple imputation. Because variables with missing values are continuous and binary, we chose the random forest method to impute the value due to its superior handling capabilities. Concerning the inclusion/exclusion criteria in RCT, we respectfully declined to include patients who didn't meet them in the real-world study.

### **Outcome**

In the RCT, the outcome was defined as an effect on lowering blood pressure, calculated by subtracting baseline systolic/diastolic blood pressure (BSBP/BDBP) from observed blood pressures at 2,4,6,8-week follow-up visits. Efficacy differences between treatment and control groups were defined as the differences in means between the two groups. If the difference is greater than zero, then the effectiveness of SXC on lowering BP was worse than that of Losartan.

#### **Generalization of RCT results in the real-world population**

We built two logistic regression models to estimate the weights: one is to regress the sampling variable (binary variable, 1 if in RCT dataset, 0 otherwise) on covariates in the combined dataset, and the other is to regress treatment assign variable (binary variable, 1 if in SXC group, 0 Losartan group) on covariates in the RCT dataset. We use the random forest to model the potential outcomes of two groups in the RCT dataset. Then we calculated the point estimation of RCT generalization results using the AIPSW estimator. Details about the formula can be found in section 5.3 in Dahabreh et al.<sup>17</sup>. Meanwhile, as a statistical method, the estimator depends on some necessary assumptions to get correct generalization results, which are based on potential outcome framework in causal inference<sup>24</sup>. Some of the assumptions (Consistency assumption, Mean exchangeability assumption in RCT, Positivity of treatment probability in the trial, Positivity of

trial participation probability) make sense obviously in our setting, while a crucial assumption—the mean generalizability assumption is hard to check. Further illustrations of the assumptions are relegated in Appendix S2.

To estimate the standard error and 95% confidence interval (CI) of the above estimators, we used the bootstrap method, which repeats resampling 1000 times. If 95% CI contains 0, then it is not statistically significant at 0.05. Statistical analysis was performed using R version 4.2.2 for Windows.

### **Sensitivity analysis**

To ensure robust estimation against unmeasured variables that may violate the mean generalizability assumption, we conducted a sensitivity analysis on the AIPSW estimator. We assessed the robustness in the following aspects. First, we used robustness value (RV) to measure the strength a confounder needs to cause the bias equal to estimation using observed variables. Then, we examined the plausibility for such an unmeasured variable to exist by using observed covariates as a benchmark. We used MRCS,  $k_{\sigma}^{min}$  and  $k_{\rho}^{min}$  to evaluate the relative strength a potential confounder needs to disturb the estimation. Intuitively, MRCS measures how much the relative confounding strength an omitted variable must have to result in a killer confounder, and  $k_{\sigma}^{min}$  and  $k_{\rho}^{min}$  can be similarly interpreted. We described the benchmarking process on a bias contour plot, where the benchmark variables were annotated by their confounder strength. For simplicity, we relegated the detailed descriptions of the above measurements to Appendix S2, readers can refer to them for further explanation.

## **Results**

### **Descriptive Analysis**

We initially recruited 602 patients with grade I hypertension into the RCT, 300 of whom were assigned to the treatment group randomly. Meanwhile, 3,000 patients were enrolled in a registry study, and they only used SXC (N = 1567) or a mixture of SXC and other Western medicine (N = 1433) to lower blood pressure. After trimming the real-world data, 804 patients were left in the real-world cohort as a trial-eligible population.

### **Baseline covariates at visit**

We extracted seven common variables in both datasets: Age, Sex, Body Mass Index (BMI), Marriage, Smoking, BSBP, and BDBP. Summary statistics of the variables are presented in Table 1. Of the baseline covariates, the mean of age was significantly greater in the real-world dataset (RWD) population than in the RCT group but was balanced between the treatment group and the control group within the RCT population. The standard deviation of BDBP in the RWD population is greater than the RCT population. Other baseline covariates were similarly distributed. Kolmogorov-Smirnov test (K-S test) was performed on continuous variables, and p-values were less than 0.05, indicating that the distribution of variables between the two populations was significantly different at the 0.05 level. Similarly, binary variables were tested by a two-sample Z-test. Box plots of continuous covariates were provided in Figure 2. In the real world, people using SXC were older than those in RCT and had lower BDBP than in the real-world trial-eligible population.

## **RCT results**

We presented the difference between SBP at follow-up visits and BSBP (DSBP) in the treatment and the control group in Figure 2(A) and DDBP in Figure 2(B) similarly. Compared to Losartan (control group), SXC was less effective in lowering SBP/DBP in the short term (weeks 2, 4, 6) but had a similar effect in the long term (week 8). The difference between the effect and 95% CI was provided in Table 2, and all CIs contain 0, implying that the difference between the two groups was not statistically significant at 0.05. Thus, we concluded that SXC was not inferior to Losartan in curing hypertension in the RCT population.

## **RCT generalization to RWD**

Comparison of the two drugs on lowering SBP and DBP in RCT and its generalization to real-world trial-eligible populations are in Figures 3(A) and 3(B). On lowering SBP, although no statistically significant result was reported in RCT when generalized to the real world, the difference was significant except at week 8. In contrast, RCT and AIPSW estimations were similar in lowering DBP except at week 2, where AIPSW estimation was slightly significant. The comparison showed a discrepancy in conclusion between the two populations. In RCT, the efficacy of SXC on lowering SBP was comparable to Losartan, while in real-world trial-eligible populations, the conclusion held only in the long term (8w). As for DBP, conclusions from RCT and its generalization were similar, where SXC was not inferior to Losartan.

## **Sensitivity analysis**

We performed sensitivity analysis on the upper and lower bounds of the AIPSW estimation in week 8. Table 3 reported the minimum relative strength over benchmark variables for an unmeasured variable to be a killer confounder for the lower bound, which means that the confounder can overturn the negative value to zero or even positive if observed. Once the lower bound becomes positive, we can conclude SXC's significant inferiority to Losartan in lowering blood pressure, then the generalization result on week 8 will be unreliable. Take DSBP as an example; the required relative strength (i.e. the ratio of bias caused by omitted confounders over benchmarking variables) for an unobserved confounder is no less than 58.16 times gender, 24.77 times age, and 23.2 times marriage, etc., to overturn the negative lower bound. Since we had collected all key confounders according to our clinical practice, we were confident in denying the existence of such variables. A similar conclusion was for DDBP at week 8.

The bias contour plot for the AIPSW estimation bounds and confounding strength of benchmarking variables for the lower bound at week 8 were provided in Figure 4(A) and Figure 4(B). In Figure 4(A), if unmeasured confounders exist in the green zone, then the upper bound can be reversed, indicating SXC's superiority over Losartan. In contrast, the lower bound can be reversed by unmeasured confounders in the red area, meaning that SXC is inferior to Losartan. In Figure 4, all the benchmark variables were not in the red area, and we concluded that our estimation was robust against potentially unmeasured confounders. This was reasonable because when designing clinical trials, we attempted to gather all major confounding variables and believed there were no stronger confounders than what we have included based on clinical practices.



## Discussion

In this article, we compared the efficacy of lowering blood pressure using SXC and Losartan in RCT. We used the AIPSW estimator to generalize the results to a real-world trial-eligible population. We found inconsistency between the RCT and its generalization results. Although the difference between the two drugs in treating hypertension was statistically insignificant in RCT, the efficacy of lowering SBP at week 2, week 4, and week 6 of SXC was inferior to Losartan significantly when it was generalized to the real-world trial-eligible population. As for lowering diastolic blood pressure, SXC was inferior to Losartan at week 4, but similar at week 6 and week 8. In the long term, we can conclude that SXC is not inferior to Losartan in treating hypertension in the real-world trial-eligible population. The difference in SXC efficacy in two populations can be attributed to covariate distribution shifts. Take age as an example. From Figure 1 and Table 1, the real-world population has more older patients than the RCT population. This may explain the difference in the study conclusions to some extent, perhaps because SXC works to lower blood pressure for seniors in the long term compared to Losartan. However, there was a higher proportion of older patients in the real-world trial-eligible population, thus SXC behaved less effectively than RCT in the short term. In terms of generalization results, our analysis generated important hypotheses for further exploring the mechanisms of how SXC works and served as a guideline for using SXC to treat hypertension in the real-world population. In the real world, using a mixture of SXC and Losartan to treat hypertension may be a better solution to balance efficacy and side effects, which had been proven in existing studies<sup>25</sup>.

The discrepancy between RCT and the real-world study was mainly due to covariates heterogeneity between RCT and real-world populations, including those that cannot be measured or observed directly. This has led to the inability of RCT populations to accurately represent trial-eligible populations in real-world contexts. There are numerous studies revealing the heterogeneity, many of which used matching methods to compare results obtained from RCT and real-world studies<sup>6-8</sup>. For instance, a previous study utilized the propensity score matching (PSM) method to match SXC users in both RCT and real-world cohorts and concluded the SXC's non-inferiority in treating hypertension in real-world population to RCT population. In their framework, after balancing observed outcomes, the heterogeneity comes from external unmeasured covariates. However, to generalize RCT results to the real world, the imbalance between the observed variables of the RCT population and the real-world trial-eligible population must be considered. Therefore, in our study, we sought to address heterogeneity from both observed and unobserved variables simultaneously. We first used the AIPSW estimator to make a generalization, which addressed the covariates distribution shift between the two populations under study. However, the statistical method depends on the mean generalizability assumption, which is potentially threatened by unobserved variables. Therefore, we made a sensitivity analysis to illustrate that our generalization results were robust against heterogeneity from unobserved variables.

In addition to the AIPSW estimator, we applied other methods introduced in the previous section to generalize the RCT results. The corresponding estimation tables and figures can be found in supplemental materials. As shown in Figures S3.1 and S3.2, despite the similarity of point estimation, the three methods exhibited different standard errors. The difference in standard errors was explained by Dahabreh et.al., that when the probability model and outcome model are correctly specified, the large-sample variance of AIPSW estimators will be larger or equal to that



of outcome model-based method and no larger than that of IPSW estimators<sup>17</sup>. The explanation was in accordance with our practical results.

There are still some limitations in our study. Firstly, we relied on statistical generalization rather than evidence-based generalization, meaning we could not access real-world trial-eligible population data to verify our findings. Obtaining such evidence would be challenging, as designing RCT on the target population would be difficult, and conducting real-world studies may be affected by unmeasured confounders, which could introduce bias. However, the mathematical property of the estimator ensured that our results were reliable even in the absence of real-world evidence. Secondly, in the study, we dropped outpatients in the real-world cohort who did not meet the inclusion/exclusion criteria of the RCT to make a more accurate estimation. Therefore, our generalization conclusion was limited to the trial-eligible population. Recently, Paul et al. proposed a synthesis parametric model to deal with the non-positivity generalization problem, but this method requires subjective information, making it challenging to apply in the practice field<sup>26,27</sup>.

### **Conclusion**

We concluded that although in the RCT population, the SXC was non-inferior to Losartan in lowering SBP and DBP, in real-world trial-eligible population, the SXC is inferior to Losartan in lowering SBP in the short term but non-inferior in the long term. Besides, our conclusion in week 8 (the non-inferiority conclusion) was robust, against potentially unmeasured confounders.

### **Acknowledgments**

We thank all patients and their families for participating in the registry or the SXC-BP trial and all the centers and investigators who devoted their effort and time to the two studies.

### **Funding**

This work was supported by the Beijing Nova Program of Science and Technology (Z211100002121061) and the Young Elite Scientist Sponsorship Program by the China Association for Science and Technology (2021-QNRC1-04).

### **Disclosures**

No potential conflict of interest was reported by the author(s).

## Reference

1. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. *The Lancet*. 2005;365(9455):217-223. doi:10.1016/S0140-6736(05)17741-1
2. Mills KT, Bundy JD, Kelly TN, et al. Global Disparities of Hypertension Prevalence and Control. *Circulation*. 2016;134(6):441-450. doi:10.1161/CIRCULATIONAHA.115.018912
3. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. *Nat Rev Nephrol*. 2020;16(4):223-237. doi:10.1038/s41581-019-0244-2
4. Dongzhimen Hospital, Beijing University of Chinese Medicine, Editing Group of Guidelines for Prevention and Treatment of Hypertension. Guidelines for prevention and treatment of hypertension. *Journal of Beijing University of Traditional Chinese Medicine*. 2023;46(8):1063-1075. doi:110. 3969 / j. issn. 1006-2157. 2023. 08. 003
5. Lai X, Dong Z, Wu S, et al. Efficacy and Safety of Chinese Herbal Medicine Compared With Losartan for Mild Essential Hypertension: A Randomized, Multicenter, Double-Blind, Noninferiority Trial. *Circ: Cardiovascular Quality and Outcomes*. 2022;15(3):e007923. doi:10.1161/CIRCOUTCOMES.121.007923
6. Munk NE, Knudsen JS, Pottegård A, Witte DR, Thomsen RW. Differences Between Randomized Clinical Trial Participants and Real-World Empagliflozin Users and the Changes in Their Glycated Hemoglobin Levels. *JAMA Netw Open*. 2020;3(2):e1920949. doi:10.1001/jamanetworkopen.2019.20949
7. Hong YD, Jansen JP, Guerino J, et al. Comparative effectiveness and safety of pharmaceuticals assessed in observational studies compared with randomized controlled trials. *BMC Med*. 2021;19:307. doi:10.1186/s12916-021-02176-1
8. Lai X, Fang Z, Dong Z, Wu S, Zhou X, Gao Y. A propensity score matched comparison of blood pressure lowering in essential hypertension patients treated with antihypertensive Chinese herbal Medicine: comparing the real-world registry data vs. randomized controlled trial. *Clinical and Experimental Hypertension*. 2023;45(1):2249269. doi:10.1080/10641963.2023.2249269
9. Rivera-Caravaca JM, Esteve-Pastor MA, Marín F, et al. A Propensity Score Matched Comparison of Clinical Outcomes in Atrial Fibrillation Patients Taking Vitamin K Antagonists: Comparing the “Real-World” vs Clinical Trials. *Mayo Clinic Proceedings*. 2018;93(8):1065-1073. doi:10.1016/j.mayocp.2018.01.028
10. Victora CG, Habicht JP, Bryce J. Evidence-Based Public Health: Moving Beyond Randomized Trials. *Am J Public Health*. 2004;94(3):400-405. doi:10.2105/AJPH.94.3.400
11. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based

- medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71-72. doi:10.1136/bmj.312.7023.71
- 12.Nie L, Soon G. A covariate-adjustment regression model approach to noninferiority margin definition. *Statistics in Medicine*. 2010;29(10):1107-1113. doi:10.1002/sim.3871
- 13.Cole SR, Stuart EA. Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *American Journal of Epidemiology*. 2010;172(1):107-115. doi:10.1093/aje/kwq084
- 14.Tipton E. Improving Generalizations From Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts. *Journal of Educational and Behavioral Statistics*. 2013;38(3):239-266. doi:10.3102/1076998612441947
- 15.Buchanan AL, Hudgens MG, Cole SR, et al. Generalizing Evidence from Randomized Trials Using Inverse Probability of Sampling Weights. *Journal of the Royal Statistical Society Series A: Statistics in Society*. 2018;181(4):1193-1209. doi:10.1111/rssa.12357
- 16.Zhang Z, Nie L, Soon G, Hu Z. New methods for treatment effect calibration, with applications to non-inferiority trials. *Biometrics*. 2016;72(1):20-29. doi:10.1111/biom.12388
- 17.Dahabreh IJ, Robertson SE, Tchetgen EJ, Stuart EA, Hernán MA. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*. 2019;75(2):685-694. doi:10.1111/biom.13009
- 18.Lee D, Yang S, Dong L, Wang X, Zeng D, Cai J. Improving trial generalizability using observational studies. *Biometrics*. 2023;79(2):1213-1225. doi:10.1111/biom.13609
- 19.Colnet B, Mayer I, Chen G, et al. Causal inference methods for combining randomized trials and observational studies: a review. Published online January 10, 2023. Accessed September 18, 2023. <http://arxiv.org/abs/2011.08047>
- 20.Nguyen TQ, Ebnesajjad C, Cole SR, Stuart EA. Sensitivity Analysis for an Unobserved Moderator in Rct-to-Target-Population Generalization of Treatment Effects. *The Annals of Applied Statistics*. 2017;11(1):225-247.
- 21.Cinelli C, Hazlett C. Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2020;82(1):39-67. doi:10.1111/rssb.12348
- 22.Faries D, Gao C, Zhang X, et al. *Real Effect or Bias? Best Practices for Evaluating the Robustness of Real-World Evidence through Quantitative Sensitivity Analysis for Unmeasured Confounding*. Preprints; 2023. doi:10.22541/au.168793390.00676676/v1
- 23.Huang M. Sensitivity Analysis in the Generalization of Experimental Results. Published online February 7, 2022. Accessed November 8, 2023. <http://arxiv.org/abs/2202.03408>

24. Imbens GW, Rubin DB. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press; 2015. doi:10.1017/CBO9781139025751
25. Yang X chen, Xiong X jiang, Yang G yan, Wang H ran, Wang J. Songling Xuemaikang Capsule () for primary hypertension: A systematic review of randomized controlled trials. *Chin J Integr Med*. 2015;21(4):312-320. doi:10.1007/s11655-014-1709-6
26. Zivich PN, Edwards JK, Shook-Sa BE, Lofgren ET, Lessler J, Cole SR. Synthesis estimators for positivity violations with a continuous covariate. Published online November 15, 2023. Accessed November 24, 2023. <http://arxiv.org/abs/2311.09388>
27. Zivich PN, Edwards JK, Lofgren ET, Cole SR, Shook-Sa BE, Lessler J. Transportability without positivity: a synthesis of statistical and simulation modeling. Published online August 16, 2023. Accessed November 16, 2023. <http://arxiv.org/abs/2303.01572>

## Tables

Characteristics	treatment group(N = 300)	RCT(N = 602)		RWD(N = 840)	p-value
		control group(N = 302)	overall(N = 602)		
Age	50.3(8.83)	50.3(9.39)	50.3(9.11)	53.1(8.31)	<0.001
Sex	0.537(0.499)	0.517(0.501)	0.527(0.50)	0.556(0.497)	0.29
BMI	25.1(3.38)	24.9(3.02)	25.0(3.20)	24.3(3.20)	0.005
Smoking	0.233(0.424)	0.219(0.414)	0.226(0.419)	0.231(0.421)	0.91
Marriage	0.977(0.151)	0.964(0.188)	0.970(0.170)	0.989(0.103)	0.01
BSBP	145.17(8.50)	144.11(7.84)	145.63(8.19)	144.45(8.98)	<0.001
BDBP	92.01(5.03)	92.01(5.55)	92.0(5.29)	86.88(7.58)	<0.001

Table 1 characteristic of baseline covariates and test between RCT and real-world data (RWD) population

Week	DSBP		DDBP	
	point estimation	95% CI	point estimation	95% CI
2w	1.230(0.811)	(-0.370,2.822)	0.662(0.593)	(-0.591,1.736)
4w	1.080(0.773)	(-0.400,2.627)	0.911(0.644)	(-0.318,2.079)
6w	0.920(0.781)	(-0.537,2.466 )	0.104(0.643)	(-1.251,1.358)
8w	0.073(0.858)	(-1.674,1.759)	0.239(0.670)	(-1.103,1.394)

Table 2 Difference of efficacy in lowering BP between the treatment (SXC) and the control group (Losartan) in RCT

Outcome	Variable	$k_{\sigma}^{min}$	$k_{\rho}^{min}$	MRCS	RV
8W DSBP	gender	15.11	-2.96	58.16	
	age	0.66	-6.17	24.77	
	marriage	7.41	1.69	-23.2	
	BMI	1.67	1.04	-6.77	
	smoke	6.04	-0.97	12.08	
	BSBP	6.25	-0.6	7.57	
	BDBP	0.08	-2.88	2.9	0.01
8W DDBP	gender	7.23	-1.74	23.5	
	age	0.32	1.45	-3.99	
	marriage	3.55	2.66	-25.1	
	BMI	0.8	0.32	-1.42	
	smoke	2.89	1.58	-13.46	
	BSBP	2.99	1.07	-9.29	
	BDBP	0.04	0.49	-0.34	0.03

Table 3. Summary statistics and relative strength to benchmarking variables for AIPSW lower

bounds at week 8

## Figures

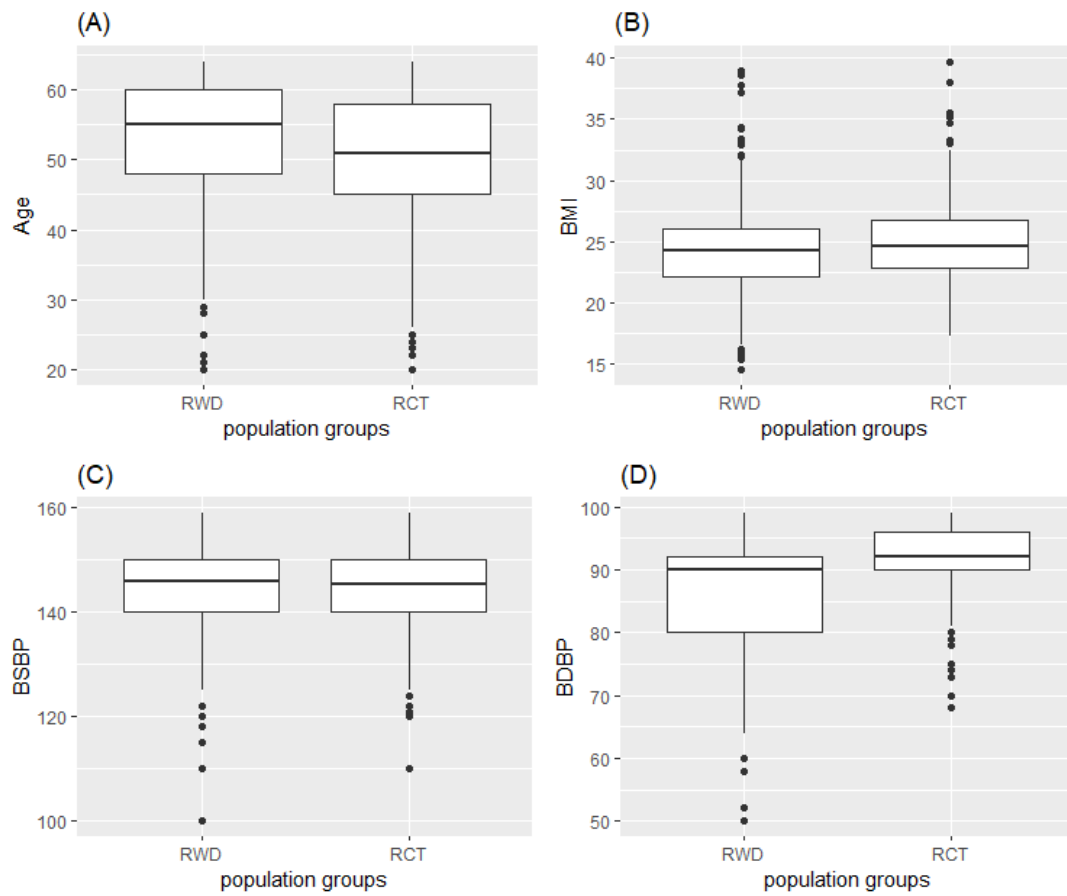


Fig 1. Box plots of continuous baseline variables in RWD and RCT sample: (A).Age, (B).BMI, (C). BSBP, and (D).BDBP

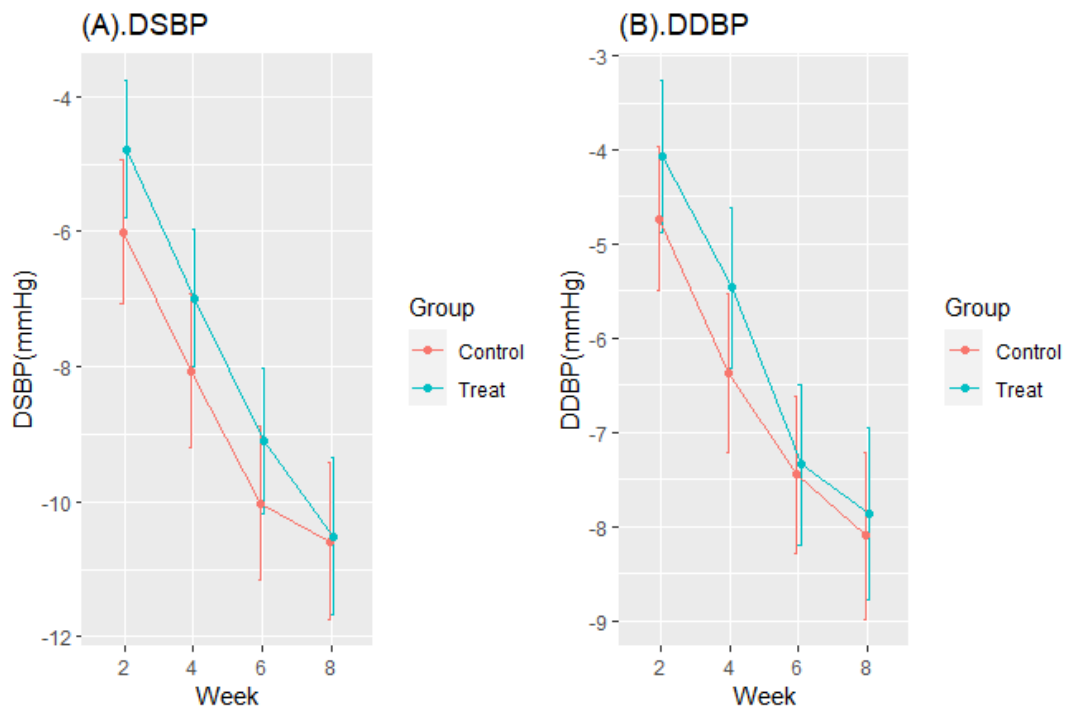


Fig 2 Difference of blood pressure at following-up visits in RCT population. (A). DSBP (B).DDBP

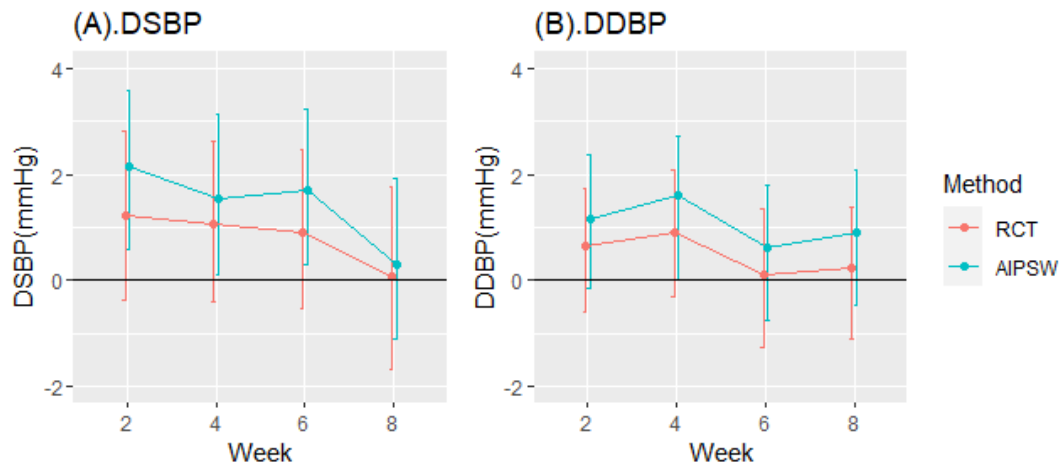


Fig 3. Difference of efficacy on lowering BP in RCT and its generalization to the real-world trial-eligible population by AIPSW estimator



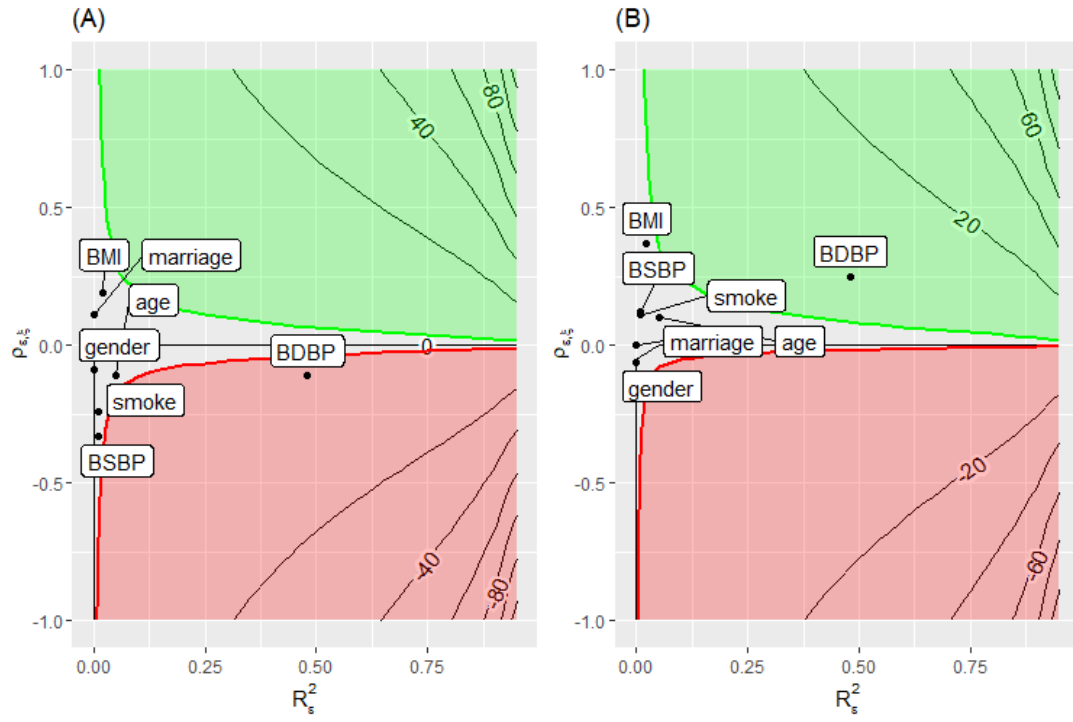


Fig 4. Bias plots for AIPSW bounds at week 8. Benchmarking variables are for the lower bound.