

## Survey Methodology

# Author's response to comments on "Exchangeability assumption in propensity-score based adjustment methods for population mean estimation using non-probability samples"

by Yan Li

Release date: June 25, 2024



Statistics  
Canada

Statistique  
Canada

Canada

---

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, [www.statcan.gc.ca](http://www.statcan.gc.ca).

You can also contact us by

**Email at** [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Telephone**, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- |   |                |
|---|----------------|
| • Statistical Information Service                             | 1-800-263-1136 |
| • National telecommunications device for the hearing impaired | 1-800-363-7629 |
| • Fax line  | 1-514-283-9350 |

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on [www.statcan.gc.ca](http://www.statcan.gc.ca) under “Contact us” > “[Standards of service to the public](#).”

## Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2024

Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

**An [HTML version](#) is also available.**

*Cette publication est aussi disponible en français.*

---

# Author's response to comments on "Exchangeability assumption in propensity-score based adjustment methods for population mean estimation using non-probability samples"

Yan Li<sup>1</sup>

## Abstract

In this rejoinder, I address the comments from the discussants, Dr. Takumi Saegusa, Dr. Jae-Kwang Kim and Ms. Yonghyun Kwon. Dr. Saegusa's comments about the differences between the conditional exchangeability (CE) assumption for causal inferences versus the CE assumption for finite population inferences using nonprobability samples, and the distinction between design-based versus model-based approaches for finite population inference using nonprobability samples, are elaborated and clarified in the context of my paper. Subsequently, I respond to Dr. Kim and Ms. Kwon's comprehensive framework for categorizing existing approaches for estimating propensity scores (PS) into conditional and unconditional approaches. I expand their simulation studies to vary the sampling weights, allow for misspecified PS models, and include an additional estimator, i.e., scaled adjusted logistic propensity estimator (Wang, Valliant and Li (2021), denoted by sWBS). In my simulations, it is observed that the sWBS estimator consistently outperforms or is comparable to the other estimators under the misspecified PS model. The sWBS, as well as WBS or ABS described in my paper, do not assume that the overlapped units in both the nonprobability and probability reference samples are negligible, nor do they require the identification of overlap units as needed by the estimators proposed by Dr. Kim and Ms. Kwon.

**Key Words:** Conditional Exchangeability; Causal inferences; Propensity score; Randomized trials; Observational studies; SARS-CoV-2 seroprevalence study.

I want to thank the discussants for their insightful comments of my paper and for the excellent additional references they cite. I will begin by addressing Dr. Saegusa's discussion on two major points. The first contrasts the differences between the conditional exchangeability (CE) assumption for causal inferences versus the CE assumption for finite population inferences using nonprobability samples. The second point focuses on distinguishing between design-based versus model-based approaches for finite population inference using nonprobability samples.

## 1. Response to comments by Dr. Saegusa

### The CE assumption in causal inference and in finite population inference

Dr. Saegusa provided a thorough explanation of the CE assumption for estimating causal effects of treatments in randomized trials and observational studies. The key condition is that conditional exchangeability is satisfied in order to make causal inferences. In randomized trials, CE of potential outcomes is achieved through random assignment of treatments. Conversely, in the analysis of observational

---

1. Yan Li, Joint Program in Survey Methodology & Department of Epidemiology and Biostatistics, University of Maryland at College Park.  
E-mail: yli6@umd.edu.

studies, CE is assumed rather than guaranteed to draw causal conclusions. The CE assumption in observational studies asserts that the distribution of potential outcomes (for different treatments), given all *observed* covariates, are exchangeable across treatment groups. In this case, there are no *unobserved* covariates that influence both the treatment assignment and the outcome of interest; see Rubin (2007) for causal effect estimation using randomized trials and observational studies. Note that unlike for *randomized trials*, in the context of finite population (FP) inference using nonprobability samples, CE is presumed rather than guaranteed, which is similar to the CE assumption needed for causal inference in *observational* studies.

Dr. Saegusa adeptly linked self-selection into nonprobability samples to treatment assignment in causal inference, and defined that being self-selected into the non-probability sample  $C$  versus in the rest in the finite population (i.e.,  $U \setminus C$ ) as the two treatments. However, in FP inference using nonprobability samples, we are interested in estimating the FP mean for a single outcome, rather than the treatment effects, i.e., the difference between two potential outcomes under the treatments of  $C$  and  $U \setminus C$ . There are no multiple treatments applied to “like” groups to obtain multiple potential outcomes. Instead, only one single potential outcome is realized. As a result, the CE assumption in FP inference differs from CE in observational studies, where it asserts that the distribution of the (*single*) outcome is exchangeable between the nonprobability sample  $C$  and the finite population  $U$ , given all *observed* covariates. Under this CE assumption, the FP mean can then be inferred using  $C$  without observing the outcome in  $U \setminus C$ .

In summary, CE in FP inference is similar to CE in causal inference using observational studies. However, unlike in causal inference, the exchangeability pertains to a single outcome between  $C$  and  $U$  in FP inference.

### **Model-based vs. design-based methods for FP inferences**

Dr. Saegusa effectively outlined the fundamental differences of model-based vs. design-based methods for FP inference. Model-based methods treat the outcome as the random variable while the design-based methods consider the selection (into the sample) indicator as random (with the outcome constant). In this paper, a set of design-based pseudoweights were constructed for the nonprobability sample to estimate FP mean under the CE assumption of  $E\{y|b(\mathbf{x}),C\} = E\{y|b(\mathbf{x}),U\}$ , where the expectation  $E(\cdot)$  is with respect to two levels of randomness of 1) random realization of the FP from a superpopulation, and 2) random self-selection into  $C$  from the finite population  $U$ . Dr. Saegusa further indicated that “a clear definition of  $y$  in  $C$  and/or  $U$  is desired”. The results, however, for obtaining unbiased estimation of the FP mean, apply as long as the FP is a random realization of a superpopulation. Only the existence of a distribution function with appropriate finite moments of variables is needed for the superpopulation. There is no need to specify a specific form of the parametric model; see for example Graubard and Korn (2002) for further details.

## 2. Response to comments by Dr. Jae-Kwang Kim and Ms. Yonghyun Kwon

In the following, I address the thoughtful discussion by Dr. Kim and Ms. Kwon in which they first presented a comprehensive framework established for categorizing existing approaches for estimating propensity scores (PS) into conditional and unconditional approaches. They further conducted simulation studies comparing different estimators, and I am glad to see our proposed ABS estimator worked well in their simulations.

In the conditional approaches two phases are involved with the first phase of sampling  $S_1 = A \cup B$  from  $U$  and the second phase of sampling  $B$  from  $S_1$ . The model parameters  $\phi$  of the conditional inclusion probability for the second phase  $P(i \in S_2 | i \in S_1) = p(x_i; \phi)$  is estimated by

$$\hat{\phi} = \arg \max_{\phi} \sum_{i \in S_1} [\delta_i \log(p(x_i; \phi)) + (1 - \delta_i) \log(1 - p(x_i; \phi))],$$

where  $\delta_i = I(i \in B)$  is the indicator function of the event  $i \in B$ . Under a statistical model, say logistic regression model  $\text{logit}\{p(x_i; \phi)\} = x_i' \phi$ , the  $\hat{\phi}$  can be obtained by solving for  $\phi$

$$\sum_{i \in B} (1 - p(x_i; \phi)) x_i - \sum_{i \in A \text{ and } i \notin B} p(x_i; \phi) x_i = 0.$$

Note the overlapped units that are selected into both samples of A and B need to be identified and removed from the sample A for the second summation above. There was an accidental omission of “and  $i \notin B$ ” under the second summation in the discussion. Based on the estimate  $\hat{\phi}$ , Dr. Kim and Ms. Kwon proposed the CPS pseudoweight for the  $i^{\text{th}}$  unit in B, given by

$$\hat{w}_i^{(B)} = 1 + w_i^{(A)} \left( \frac{1}{p(x_i; \hat{\phi})} - 1 \right),$$

where  $w_i^{(A)}$  is often unknown and estimated under a parametric model in practice.

In the unconditional approaches, only one step was involved. The conditional maximum likelihood estimator of  $\phi$  was estimated from the combined sample  $S_1 = A \cup B$ . Same as the CPS estimator, the proposed unconditional propensity score (UCPS) approach is also based on the assumption that the units that belong to the intersection of A and B can be identified.

The proposed CPS and UCPS were evaluated by simulation studies, considering varying sample sizes in sample B selected using stratified simple random sampling (SSRS) with one categorical stratification variable. This design, although simple, is clever. It aligns with the true underlying PS model for all methods considered, ensuring a fair comparison. To further evaluate the performance of the proposed estimates, we expanded the simulation studies by including an additional estimator under the same SSRS sampling design but with varying sampling weights. Recall the population size is  $N = 5,000$ , sample A size  $n_A = 250$ , and varying sample B sizes  $n_B = 250$  and 2,500. We consider the three estimators that have the smallest root

mean squared errors (RMSE) in Table 4.1 of their discussion: WBS, ABS and CPS. The UCPS performs similarly to CPS, and therefore not considered. Recall that WBS refers to the adjusted logistic propensity estimator, proposed by Wang et al. (2021). In the same paper, the authors also proposed the scaled WBS estimator, denoted by sWBS, where the scaled weights are the value of one for sample B units and  $n_s w_i^{(A)} / \sum_{i \in S_A} w_i^{(A)}$  for unit  $i$  in sample A. Sampling fractions vary within sampling strata. In stratum 1,  $n_{B1} = f_1 n_B$  samples are selected by simple random sampling. In stratum 2,  $n_{B2} = (1 - f_1) n_B$  samples are selected by simple random sampling. The value of  $f_1$  is varied as 0.7, 0.8, and 0.9 to produce different values of the coefficient of variation of the SSRS sampling weights (CVWT). In the PS analysis, we consider two models: M1) the main effects of  $(x_1, x_2, x_3)$  and their pairwise interaction effects; M2) the main effects  $(x_{1c}, x_2, x_3)$  and their pairwise interaction effects, where  $x_{1c} = I\{x_1 < 0\}$ , the indicator function of the event  $x_1 < 0$  where  $x_1$  are generated from a  $N(0,1)$ . Note that M2 aligns with the SSRS design while M1 is misspecified by including the continuous variable  $x_1$  in the PS analysis.

Four observations are made: 1) All the four estimators are approximately unbiased under the true PS model. 2) ABS and CPS perform similarly for a small sample size of  $n_B = 250$  under both models. 3) When the sample size is large  $n_B = 2,500$ , CPS consistently has smallest SE and RMSE under the true model. These results are as expected, given that there is a large percentage of overlapped units in both samples. Therefore, efficiency is gained by the CPS method, which assumes that the overlapped units can be identified. 4) Under the misspecified PS model, sWBS consistently has the smallest bias, especially when CVWT is large. In contrast, CPS has the largest bias and SE when CVWT and  $n_B$  is large. The biasness and the loss of efficiency from CPS can be attributed to the misspecified modeling of  $P(i \in B | i \in A \cup B)$ , the limited sample size by removing overlapped units (~50%) from sample A, and the variable sampling weights. CPS is sensitive to model misspecification, especially when  $n_B$  and CVWT are large.

In summary, under true PS model, ABS and CPS perform similarly when  $n_B$  is small; when  $n_B$  is large, CPS estimator is more efficient due to increasing number of units that are selected and identified in both samples A and B. Under the misspecified PS model, sWBS (Wang et al., 2021) overperformed or was comparable to the other estimators. Effects of various misspecified PS models or scalars on the performance of sWBS require further investigation. Secondly, the estimators ABS, WBS, sWBS, as well as CPS, are developed without assuming that the overlapped units in both samples are negligible. For large sample size  $n_B = 2,500$ , the sampling rate for sample B,  $n_B/N = 50\%$ , is non-negligible. All estimators, as shown in Table 1, are approximately unbiased under the true PS model, which empirically proves that all the four methods do not require the assumption that the overlapped units in both samples are negligible. Finally, it is of practical importance for the reader to be aware that the CPS estimator requires the identification of overlap units. This may not be feasible in many situations. For example, in the NIH SARS-CoV-2 Seropositivity Study discussed in my paper, this identifying information was not collected.

**Table 1**

**Bias, standard error, and root mean square error ( $\times 100$ ) under SSRS with varying CV of sample weights (CVWT) after 5,000 repetitions.**

	Correctly Specified PS model ( $x_{1c}$ )						Misspecified PS model ( $x_1$ )					
	$n_B = 250$			$n_B = 2,500$			$n_B = 250$			$n_B = 2,500$		
	BIAS	SE	RMSE	BIAS	SE	RMSE	BIAS	SE	RMSE	BIAS	SE	RMSE
CVWT = 0.44												
Mean C	-3.25	2.98	4.41	-3.33	0.94	3.46	4.84	2.87	5.63	4.74	0.93	4.83
WBS	0.04	3.60	3.60	-0.04	1.42	1.42	0.51	3.06	3.10	0.23	1.54	1.56
sWBS	0.02	3.56	3.56	-0.05	1.41	1.41	0.33	3.05	3.07	0.18	1.62	1.63
ABS	0.02	3.54	3.54	-0.04	1.38	1.38	0.56	2.98	3.04	0.56	1.45	1.55
CPS	0.01	3.53	3.53	0.01	1.20	1.20	0.55	2.99	3.04	0.12	1.36	1.36
CVWT = 0.75												
Mean C	-4.97	2.89	5.76	-4.99	0.92	5.07	7.10	2.97	7.70	7.10	0.95	7.16
WBS	-0.03	4.16	4.16	0	1.56	1.56	1.03	3.18	3.34	0.47	1.55	1.62
sWBS	-0.08	4.09	4.09	-0.03	1.55	1.55	0.39	3.23	3.25	0.18	1.64	1.65
ABS	-0.08	4.08	4.08	-0.02	1.52	1.52	1.13	3.15	3.34	1.11	1.52	1.88
CPS	-0.08	4.07	4.07	0.05	1.34	1.34	1.10	3.16	3.34	-0.41	1.64	1.69
CVWT = 1.33												
Mean C	-6.58	2.88	7.19	-6.66	0.90	6.72	9.49	3.07	9.98	9.45	0.97	9.50
WBS	0.11	5.65	5.65	0.00	1.91	1.91	2.74	3.49	4.44	1.39	1.67	2.17
sWBS	0.06	5.54	5.54	-0.03	1.89	1.89	1.07	3.78	3.93	-0.05	1.85	1.85
ABS	0.03	5.49	5.49	-0.04	1.86	1.86	2.60	3.50	4.36	1.94	1.69	2.58
CPS	0.03	5.49	5.49	0.03	1.71	1.71	2.54	3.54	4.36	-3.00	3.10	4.31

## References

- Graubard, B.I., and Korn, E.L. (2002). Inference for superpopulation parameters using sample surveys. *Statistical Science*, 17(1), 73-96.
- Rubin, D.B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Stat Med.*, 26(1), 20-36. Doi: 10.1002/sim.2739. PMID: 17072897.
- Wang, L., Valliant, R. and Li, Y. (2021). Adjusted logistic propensity weighting methods for population inference using nonprobability volunteer-based epidemiologic cohorts. *Stat Med.*, 40(24), 5237-5250. Doi: 10.1002/sim.9122. Epub 2021 Jul 5. PMID: 34219260; PMCID: PMC8526388.