Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Randomization schemes and Bayesian borrowing of survey-derived control information in a simulated cardiovascular prevention trial: a simulation study based on the 2022 Canadian Community Health Survey

Syed Ehsan Ar Rafi

srafi@ttu.edu

Texas Tech University

Research Article

Keywords: randomization, minimization, covariate adjustment, Bayesian clinical trials, external control information, robust mixture prior, simulation study, Canadian Community Health Survey

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11826760/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Randomization schemes and Bayesian borrowing of survey-derived control information in a simulated cardiovascular prevention trial: a simulation study based on the 2022 Canadian Community Health Survey

Syed Ehsan Ar Rafi

Department of Mathematics and Statistics, Texas Tech University

Keywords: randomization; minimization; covariate adjustment; Bayesian clinical trials; external control information; robust mixture prior; simulation study; Canadian Community Health Survey

Abstract

Background: Randomization strategies and external baseline-risk information can influence covariate balance and the operating characteristics of clinical trials. We evaluated these approaches using a population-based virtual patient population.

Methods: We used the 2022 Canadian Community Health Survey (CCHS) to construct a virtual population of adults aged 50 years or older for a hypothetical cardiovascular prevention trial. In Part A, four randomization schemes were compared across 2,000 simulated trials of 2,400 participants. In Part B, Bayesian 1:1 and 2:1 designs using a robust mixture prior informed by CCHS prevalence were evaluated across 10,000 simulations per scenario, with varying degrees of prior–data conflict.

Results: Minimization improved covariate balance, reducing the mean maximum standardized difference from 0.086 to 0.030, while differences in type I error and power across randomization schemes were small. When external information matched the true control rate, Bayesian borrowing increased power to 88.2% in the 2:1 design. However, prior–data conflict increased type I error to 6.1% and reduced power under certain scenarios.

Conclusions: Survey-derived information may improve trial efficiency, but its benefits depend on the accuracy of external baseline-risk assumptions. Prior–data conflict should be carefully evaluated when designing Bayesian trials.

1. Introduction

Randomization is intended to produce comparable treatment groups, but in a single trial chance imbalance in prognostic factors can occur, particularly in multicentre trials with several important baseline covariates. Restricted and covariate-adaptive schemes, including stratified permuted blocks and minimization, aim to improve balance on selected factors [1-3]. Their effect on inference depends on the analysis, because stratified randomization is usually accompanied by covariate adjustment to preserve nominal error rates and recover power [4,5].

A second design question is whether external information can reduce the number of participants assigned to control. Bayesian methods allow historical or external control data to enter through an informative prior, and robust mixture priors protect against prior-data conflict by combining an informative component with a vague one [6-9]. The benefit depends on whether the external control rate matches the control rate in the new trial, and borrowing can inflate type I error when it does not [6,7].

Evaluating these design choices by simulation requires a patient population with realistic covariate distributions and a plausible gradient in baseline risk. Population health surveys can provide both. The aim of this study was to use a national health survey to construct a virtual patient population and to compare (A) randomization schemes, with and without covariate adjustment, and (B) Bayesian designs that borrow survey-derived control information, with respect to balance, type I error, power and estimation accuracy.

2. Methods

2.1 Design and reporting

The study is a simulation study, planned and reported using the aims, data-generating mechanisms, estimands, methods and performance measures (ADEMP) framework [10]. The CCHS is a cross-sectional survey. It supplies the covariate distribution and a baseline-risk gradient for the simulations, but it contains no information on treatment effects or on incidence, so the trial outcome and the treatment effect were simulated, and the survey prevalence was used as a plausible baseline risk.

2.2 Virtual patient population

We used the 2022 Canadian Community Health Survey (CCHS) public use microdata file [11]. The target population for the hypothetical trial was adults aged 50 or older. Respondents with a valid cardiovascular condition status (heart disease, heart attack or the effects of stroke), sex and household income quintile were included (n = 37,753). Respondents were grouped into 1,792 cells defined by health region, age group (50 to 64, 65 or older), sex and income quintile, and the sampling weights were summed within each cell to give the cell population (14.26 million in total).

The probability of a cardiovascular condition in each cell was taken from a Bayesian multilevel logistic model fitted to the same survey data in a companion analysis (age by sex interaction, income quintile, and province and health region random effects, with survey weights entering through a pseudo-likelihood [12], fitted with brms [13]). The weighted mean of the cell probabilities was 12.6%. Virtual participants were drawn with replacement from the cells with probability proportional to the cell population. Each participant had five baseline variables: province (10 sites), age group, sex, income quintile and the cell-level baseline risk.

2.3 Hypothetical trial and data-generating mechanism

The hypothetical trial compared an intervention with control in adults aged 50 or older, with provinces as sites and a binary endpoint. For participant i, the event probability under control was p_{0i}, the survey-derived cell probability. Under treatment, $logit(p_{1i}) = logit(p_{0i}) + log(OR)$. The true odds ratio was 1 (null) or 0.65 (alternative). The same participants, allocation and uniform random numbers were

used for the null and alternative outcomes within each simulated trial, so that the two scenarios were compared with common random numbers.

2.4 Estimands

In Part A, the estimand for the adjusted analysis was the conditional treatment odds ratio, and for the unadjusted analysis the marginal odds ratio. In Part B, the estimand was the marginal risk difference between treatment and control, and the success criterion was based on the posterior probability that the treatment event rate was lower than the control event rate.

2.5 Part A: randomization schemes

A trial of 2,400 participants was simulated 2,000 times under each of four schemes with 1:1 allocation: (1) simple randomization; (2) permuted blocks of size 4 within province; (3) permuted blocks of size 4 within province by age group by sex strata; and (4) Pocock-Simon minimization on province, age group, sex and income quintile using marginal imbalance and a biased coin with probability 0.8 [1,2]. Each simulated trial was analysed with an unadjusted logistic regression (treatment only) and an adjusted logistic regression (treatment, age group, sex, income quintile and province), using a one-sided Wald test at the 0.025 level, with benefit defined as a negative log odds ratio.

Performance measures were the maximum absolute standardized mean difference across 17 binary baseline indicators (sex, age group, five income categories and ten provinces), the absolute difference between arms in mean baseline risk (percentage points), the absolute difference in arm sizes, the rejection rate under the null (type I error) and under the alternative (power), and the bias of the estimated log odds ratio relative to the true value. The Monte Carlo standard error was 0.35 percentage points for a rejection rate of 2.5% and 0.9 percentage points for a rate of 80%.

2.6 Part B: Bayesian borrowing of survey-derived control information

A trial of 2,000 participants was simulated 10,000 times in each scenario. The number of events in the control and treatment arms was binomial with the true control rate and treatment rate. The external information was the weighted prevalence among respondents aged 50 or older (12.59%) and the Kish effective sample size of 11,910 [14]. The control event probability had a robust mixture prior [7]: weight 0.5 on a Beta distribution with mean 0.1259 and prior effective sample size 300 (about 2.5% of the survey effective sample size), and weight 0.5 on a vague Beta(1, 1) distribution. The treatment event probability had a Beta(1, 1) prior. Posterior mixture weights were updated using the beta-binomial marginal likelihoods, and the posterior probability that the treatment event rate was lower than the control event rate was computed by numerical integration. A trial was successful when this probability exceeded 0.975.

Five designs were compared: a frequentist 1:1 design with a one-sided pooled z test at the 0.025 level; Bayesian 1:1 and 2:1 (treatment to control) designs with vague priors; and Bayesian 1:1 and 2:1 designs with the survey-informed mixture prior. Prior-data conflict was created by multiplying the control odds by 0.75, 1, 1.25 or 1.5 relative to the external rate (true control rates 9.7%, 12.6%, 15.3% and 17.8%). Performance measures were the probability of success (type I error when the true odds ratio was 1, power when 0.65), the bias and root mean squared error (RMSE) of the posterior mean risk difference, the mean posterior weight on the informative component, and the number of control participants. The

Monte Carlo standard error was 0.16 percentage points for a success probability of 2.5% and 0.37 percentage points for 80%.

2.7 Implementation

All analyses were run in R version 4.5.1 with fixed random seeds. Part A used 16 minutes of computing time on a laptop. The R code is available at [srafi106/cchs-trial-simulation: Simulation of randomization schemes and Bayesian borrowing in a hypothetical cardiovascular trial, using CCHS 2022].

3. Results

3.1 Virtual patient population

The virtual population was based on 37,753 respondents in 1,792 cells, representing 14.26 million adults aged 50 or older. The weighted prevalence of a cardiovascular condition was 12.6%, and the Kish effective sample size was 11,910.

3.2 Part A: randomization schemes

Balance improved most with minimization (Table 1, Figure 1). The mean maximum standardized difference was 0.086 under simple randomization, 0.069 with province blocks, 0.062 with province by age by sex blocks and 0.030 with minimization, a 65% reduction relative to simple randomization; the 95th percentile was 0.121 for simple randomization and 0.058 for minimization. The mean absolute difference between arms in baseline risk was 0.28 percentage points for simple randomization and 0.27 for province blocks, compared with 0.10 for province by age by sex blocks and 0.08 for minimization. The mean absolute difference in arm sizes was 39.4 participants for simple randomization, compared with 2.2 to 4.6 for the blocked schemes and 1.4 for minimization.

Table 1. Balance, type I error and power of four randomization schemes (2,400 participants, 2,000 simulated trials)

Scheme	Mean max SMD	95th pct max SMD	Mean abs. difference in baseline risk (pp)	Mean abs. difference in arm size	Type I error, unadjusted (%)	Type I error, adjusted (%)	Power, unadjusted (%)	Power, adjusted (%)
Simple	0.086	0.121	0.28	39.4	2.6	3.1	86.1	87.7
Province blocks	0.069	0.108	0.27	2.2	2.3	2.7	87.5	89.0
Province x age x sex	0.062	0.103	0.10	4.6	2.0	2.7	86.2	87.6

Scheme	Mean max SMD	95th pct max SMD	Mean abs. difference in baseline risk (pp)	Mean abs. difference in arm size	Type I error, unadjusted (%)	Type I error, adjusted (%)	Power, unadjusted (%)	Power, adjusted (%)
blocks								
Minimization	0.030	0.058	0.08	1.4	2.5	2.6	86.2	87.7

SMD: standardized mean difference; pp: percentage points. One-sided test at the 0.025 level; true odds ratio 0.65 for power. Monte Carlo standard error: 0.35 pp (type I error), 0.9 pp (power).

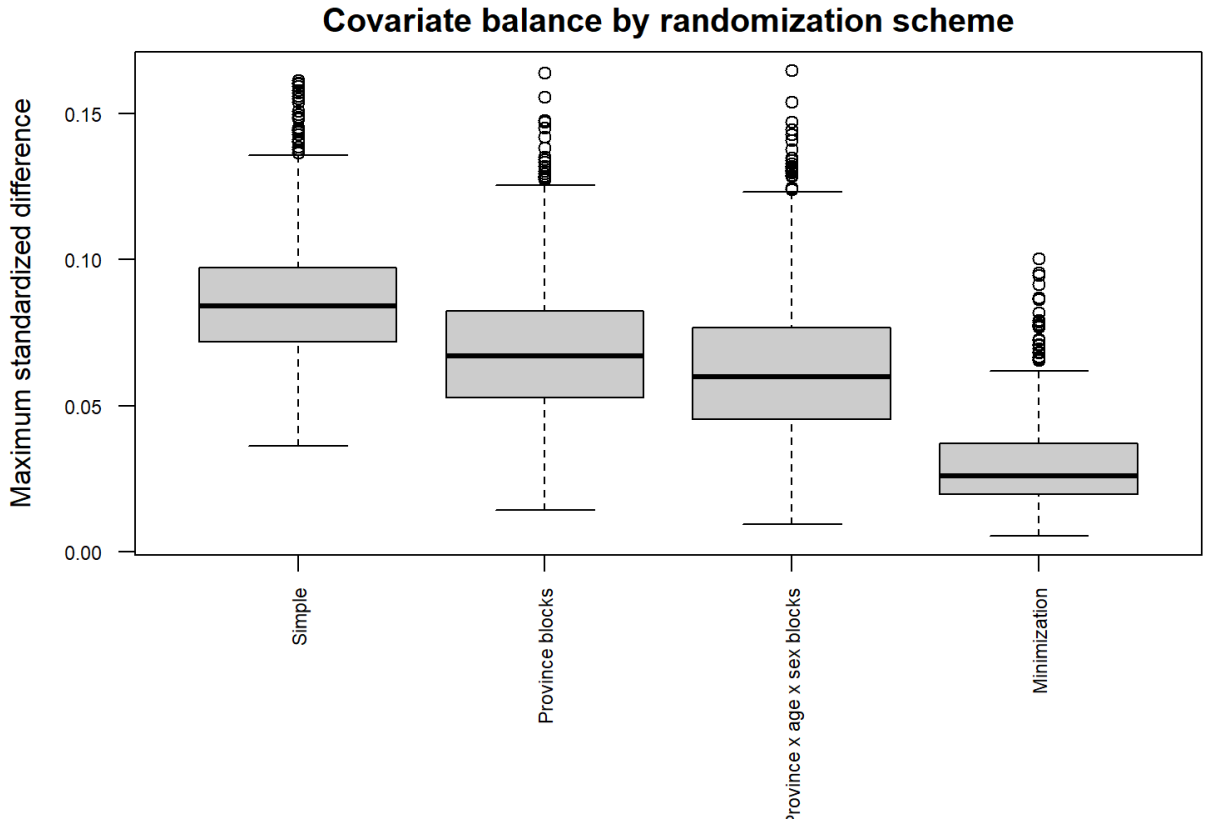

Figure 1. Maximum standardized difference across 17 baseline indicators by randomization scheme (2,000 simulated trials per scheme).

Type I error ranged from 2.0% to 2.6% for the unadjusted analysis and from 2.6% to 3.1% for the adjusted analysis, the latter slightly above the nominal 2.5% and highest under simple randomization. Power ranged from 86.1% to 87.5% for the unadjusted analysis and from 87.6% to 89.0% for the adjusted analysis. Covariate adjustment increased power by about 1.5 percentage points under every scheme, and the largest difference in power between schemes (1.4 percentage points, province blocks

versus simple randomization) was within about 1.5 Monte Carlo standard errors. The mean unadjusted log odds ratio was 0.021 to 0.024 above the true conditional value of log(0.65), whereas the adjusted estimate was within 0.006 of it.

3.3 Part B: Bayesian borrowing of control information

When the true control rate equalled the external rate (odds multiplier 1.00), the designs that borrowed information were conservative under the null and more powerful under the alternative (Table 2, Figure 2). The 2:1 design with the survey-informed prior, with 667 control participants, had 88.2% power and 1.4% type I error, compared with 83.9% power and 2.8% type I error for the standard 1:1 design with 1,000 controls. The same 2:1 design with vague priors had 80.2% power. The frequentist and Bayesian vague-prior 1:1 designs gave practically identical results in every scenario.

Table 2. Probability of success (%) by design and control-rate drift (2,000 participants, 10,000 simulated trials per scenario)

True OR	Odds multiplier	True control rate (%)	Frequentist 1:1	Bayesian 1:1, vague	Bayesian 1:1, survey-informed	Bayesian 2:1, vague	Bayesian 2:1, survey-informed
1.00	0.75	9.7	2.7	2.7	5.2	2.5	6.1
1.00	1.00	12.6	2.8	2.8	2.1	2.5	1.4
1.00	1.25	15.3	2.3	2.3	0.8	2.6	0.5
1.00	1.50	17.8	2.5	2.5	1.4	2.5	1.2
0.65	0.75	9.7	74.3	74.3	86.8	71.0	87.3
0.65	1.00	12.6	83.9	83.9	87.8	80.2	88.2
0.65	1.25	15.3	89.4	89.4	85.7	85.9	82.5
0.65	1.50	17.8	92.8	92.7	84.5	90.3	78.2

OR: odds ratio. Rows with true OR 1.00 show type I error; rows with true OR 0.65 show power. The 2:1 designs had 667 control and 1,333 treatment participants. Monte Carlo standard error: 0.16 pp (2.5%), 0.37 pp (80%).

Prior-data conflict affected the borrowing designs in both directions. When the true control rate was lower than the external rate (multiplier 0.75), type I error rose to 5.2% (1:1) and 6.1% (2:1), about double the nominal level. When the true control rate was higher (multipliers 1.25 and 1.50), the borrowing designs were conservative, with type I error between 0.5% and 1.4%, and power fell to 84.5% (1:1) and 78.2% (2:1) at multiplier 1.50, compared with 92.8% and 90.3% for the corresponding vague-prior designs.

The mean posterior weight on the informative component was 0.94 when the external rate was correct, 0.86 at multipliers 0.75 and 1.25, and 0.55 to 0.58 at multiplier 1.50 (Table 3). The robust prior therefore reduced borrowing under larger conflict, but not enough to prevent type I error inflation at multiplier 0.75. When the external rate was correct, the survey-informed designs had lower RMSE of the risk difference than the vague-prior designs (1.21 versus 1.36 percentage points for 1:1, and 1.20 versus 1.50 for 2:1), whereas under conflict they were biased in the direction of the external rate (-0.47 to +0.87 percentage points) and had larger RMSE at multiplier 1.50.

Table 3. Borrowing and estimation performance, true OR 0.65

Odds multiplier	Posterior weight on informative component (1:1 / 2:1)	Bias of risk difference, survey-informed (pp, 1:1 / 2:1)	Bias, vague (pp, 1:1 / 2:1)	RMSE, survey-informed (pp, 1:1 / 2:1)	RMSE, vague (pp, 1:1 / 2:1)
0.75	0.86 / 0.86	-0.47 / -0.68	+0.01 / -0.05	1.21 / 1.33	1.21 / 1.31
1.00	0.94 / 0.94	+0.08 / +0.04	+0.01 / -0.07	1.21 / 1.20	1.36 / 1.50
1.25	0.86 / 0.86	+0.58 / +0.71	+0.01 / -0.06	1.49 / 1.59	1.49 / 1.63
1.50	0.55 / 0.58	+0.66 / +0.87	+0.03 / -0.02	1.81 / 2.05	1.58 / 1.72

pp: percentage points. Risk difference is treatment minus control; a negative bias means the treatment appeared more beneficial than it was.

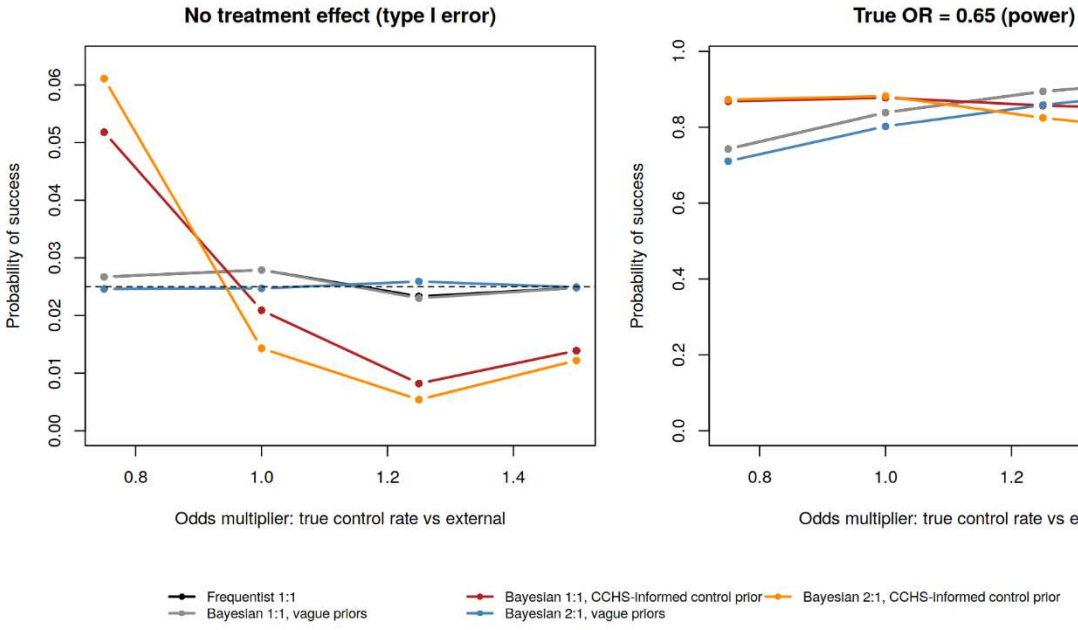

Figure 2. Probability of success by design and control-rate drift: type I error (left) and power (right). The dashed line is the nominal 2.5% level.

4. Discussion

4.1 Principal findings

In a trial of 2,400 participants, all four randomization schemes produced similar type I error and power, but minimization gave the best covariate balance, and the largest reduction in the maximum standardized difference, and stratified blocks gave much closer arm sizes than simple randomization. Covariate adjustment increased power by about 1.5 percentage points under every scheme. For a design question that involves control information, borrowing from a survey-based external rate reduced the number of control participants needed and increased power when the external rate was accurate, but it inflated type I error when the true control rate was lower than the external rate, and it lost power when the true control rate was higher.

4.2 Interpretation

The small differences in power and type I error between randomization schemes are consistent with the expectation that balance on a few measured covariates has limited effect on inference when the analysis adjusts for those covariates and the trial is large. The adjusted analysis had a type I error slightly above nominal (2.6% to 3.1%), as is expected for Wald tests in logistic regression with many covariates, and the effect was largest under simple randomization. The difference between the unadjusted and adjusted log odds ratios reflects non-collapsibility of the odds ratio, which makes the marginal and conditional estimands differ, and is not bias of the estimator [15].

The findings for Bayesian borrowing illustrate the trade-off described for historical control designs [6,7]. When the external control rate was correct, the survey-informed designs improved power and precision, and the 2:1 design with fewer controls had higher power than the standard design. The same designs were exposed to inflation of type I error under one direction of drift, and the asymmetry (inflation when the true rate is below the external rate, power loss when it is above) matters for planning, because the direction of drift is rarely known in advance. Because the survey measures prevalence in the general population, a trial population with a different case mix or follow-up would be expected to have a different control rate. Operating characteristics at the external rate alone are therefore not sufficient to justify borrowing.

4.3 Strengths and limitations

Strengths include the use of a real national survey to supply covariate distributions and a baseline-risk gradient, common random numbers across schemes, and a Monte Carlo error small relative to the effects of interest.

The study has limitations. The survey is cross-sectional, so prevalence was used as a proxy for the event probability and the treatment effect was assumed. The results apply to the chosen sample sizes, event rates and effect size, and were not examined across a wider grid. The covariate structure of the virtual participants comes from cells and does not reflect correlation among covariates within cells. The borrowing analysis used a single prior effective sample size (300) and mixture weight (0.5), a binary endpoint and a final analysis only, with no interim analyses, and did not include dynamic borrowing methods such as commensurate or power priors with data-driven weights [8]. The success threshold of 0.975 was not calibrated to control type I error across the drift scenarios, so the type I error values are

those of a fixed rule, not of a calibrated design. Part A used simple covariate adjustment and one block size.

4.4 Implications and future research

Trial designers who consider borrowing survey-derived information should examine type I error and power over a plausible range of control rates, and consider calibrating the success threshold to limit type I error across that range, or using dynamic borrowing. Further work could add interim analyses and adaptive allocation [9], time-to-event endpoints, a grid of prior effective sample sizes and mixture weights, and trial populations that differ from the survey population in measured covariates.

5. Conclusions

In a simulated cardiovascular prevention trial based on a national health survey, randomization scheme affected covariate balance but changed type I error and power very little, and covariate adjustment provided a small, consistent power gain. Borrowing survey-derived control information reduced the number of control participants needed when the external rate was accurate but was sensitive to drift, which supports pre-specified drift scenarios and dynamic borrowing in designs that use external control information.

Declarations

Ethics approval. This study used publicly available, de-identified survey data (Statistics Canada Open Licence) and simulated trial data.

Data and code availability. The CCHS 2022 public use microdata file is available from Statistics Canada. The simulation code is available at [Canadian Community Health Survey: Public Use Microdata File].

Competing interests. None

Funding. None

Declaration of Generative AI. During the preparation of this manuscript, the author used ChatGPT (OpenAI) and Claude (Anthropic) for limited assistance with language refinement and manuscript organization. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of this article.

References

1. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics. 1975;31(1):103-15.
2. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. Clin Pharmacol Ther. 1974;15(5):443-53.
3. Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. Control Clin Trials. 2002;23(6):662-74.
4. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. BMJ. 2012;345:e5840.

5. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. Trials. 2014;15:139.
6. Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, et al. Use of historical control data for assessing treatment effects in clinical trials. Pharm Stat. 2014;13(1):41-54.
7. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics. 2014;70(4):1023-32.
8. Ibrahim JG, Chen MH. Power prior distributions for regression models. Stat Sci. 2000;15(1):46-60.
9. Berry SM, Carlin BP, Lee JJ, Müller P. Bayesian adaptive methods for clinical trials. Boca Raton (FL): CRC Press; 2010.
10. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074-102.
11. Statistics Canada. Canadian Community Health Survey (CCHS) annual component: complement to the user guide, 2022 public use microdata file. Ottawa (ON): Statistics Canada; 2025 Jun.
12. Savitsky TD, Toth D. Bayesian estimation under informative sampling. Electron J Stat. 2016;10(1):1677-708.
13. Bürkner PC. brms: an R package for Bayesian multilevel models using Stan. J Stat Softw. 2017;80(1):1-28.
14. Kish L. Survey sampling. New York (NY): John Wiley and Sons; 1965.
15. Greenland S, Pearl J, Robins JM. Confounding and collapsibility in causal inference. Stat Sci. 1999;14(1):29-46.