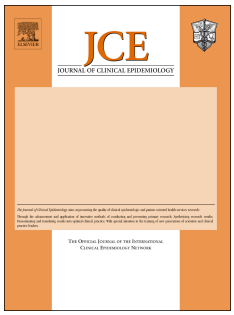
Identifying compliant participants through data matching improved estimation of intervention efficacy: randomized trials with opt-in/ opt-out

Accepted Manuscript



Identifying compliant participants through data-matching improved estimation of intervention efficacy: Randomised trials with opt-in/opt-out strategies

Shu Kay Ng, Joshua Byrnes, Paul Scuffham

PII: S0895-4356(19)30217-3

DOI: https://doi.org/10.1016/j.jclinepi.2019.07.013

Reference: JCE 9944

To appear in: *Journal of Clinical Epidemiology*

Received Date: 7 March 2019

Revised Date: 2 July 2019

Accepted Date: 17 July 2019

Please cite this article as: Ng SK, Byrnes J, Scuffham P, Identifying compliant participants through data-matching improved estimation of intervention efficacy: Randomised trials with opt-in/opt-out strategies, *Journal of Clinical Epidemiology* (2019), doi: https://doi.org/10.1016/j.jclinepi.2019.07.013.

Identification of compliant participants through data-matching improved estimation of intervention efficacy: Randomised trials with opt-in/opt-out strategies

Shu Kay Ng,[1*] Joshua Byrnes,[1] and Paul Scuffham[1]

[1] School of Medicine and Menzies Health Institute Queensland, Griffith University, Nathan, Q4111, Australia

***Corresponding author:** Shu Kay Ng, School of Medicine, Griffith University, Room 2.41, N78, 170 Kessels Road, Nathan Q4111, Australia; Phone: (617) 3735 9131; Email: s.ng@griffith.edu.au

Abstract

Obective: We propose a data-matching approach to estimate intervention efficacy for randomised controlled trials (RCTs) when there is non-compliance to the allocated treatment with induced selection bias.

Study Design and Settings: We considered a large RCT to compare healthcare costs and hospital length-of-stay 12 months post randomisation. Participants allocated to the intervention group were eligible to receive health-coaching and disease-management services. An opt-out approach was adopted for recruitment. Control-group participants received usual care but were allowed to opt-in to receive the intervention. Using "nearest-neighbour"-matched data, we identified compliant participants in both arms to estimate intervention efficacy. Results were compared with intention-to-treat (ITT), instrumental-variable (IV)-adjusted ITT, per-protocol (PP), and as-treated (AT) analyses.

Results: The ITT estimated an intervention effect of a 1.5% reduction in cost, but 56.7% of intervention-group participants did not receive health-coaching. The PP and AT found an increase in cost of 9.4% and 17.1%, respectively. The matching method estimated a 12.3%

reduction in cost. After adjustment for baseline covariates, the intervention group had lower same-day admission cost (13.6%; 95% CI: 7.3%-20.0%; p<0.001) and shorter hospital stay (11.2%; 95% CI: 2.6%-19.9%; p=0.021).

Conclusion: Opt-in/opt-out strategies in RCTs misled intervention comparisons and the matching approach improved estimation of intervention efficacy.

Keywords: Non-compliance, selection bias, randomised controlled trial, health coaching, nearest-neighbour matching

Running Title: Estimation of intervention efficacy in randomised trials

What is new?

- We showed that opt-in and opt-out recruitment strategies in RCTs inflated selection bias due to non-compliance (as shown in this study, participants with higher healthcare needs were more likely to engage).
- We considered a conceptual framework with three levels of "principal compliance" (always-takers, compliant participants, never-takers) and developed a new approach to identify compliant participants for comparison of intervention efficacy.
- Always-takers (opt-in) had the highest healthcare needs among the participants in the control group and never-takers (opt-out) had the lowest healthcare needs among those allocated to the intervention group.
- The ITT analysis actually measures intervention effectiveness (as there are opt-in and opt-out participants who do not comply with the assigned treatment); IV adjusts the ITT for non-compliance, but the assumption for IV adjustment is not always valid; PP and AT analyses measure treatment efficacy, but they are subject to selection bias.
- The new method identifies compliant participants in both treatment arms by data-matching on the basis of participants with known compliance information; this method is useful for analysing future RCTs that anticipate strong selection bias.

1. Introduction

Randomised controlled trials (RCTs) are considered the gold standard in clinical research for their potential to reduce bias through randomisation such that participants in the study arms have similar characteristics and any difference in outcomes between the study arms can be attributed to the intervention under study [1,2]. However, selection bias can occur when individuals are requested to give consent for participation in a trial and when individuals are assigned to a treatment arm once they have been accepted into a trial [3-8]. The former kind of selection bias may result in an over-representation of relatively advantaged participants (such as individuals with higher socio-economic status) who agree to participate in a trial; it can be reduced by randomisation but has impact on interpretation and generalisation of findings [9]. Conversely, the second kind of selection bias cannot be reduced by randomisation as it occurs after the assignment of treatment arms, inducing a serious impact on estimating intervention effects due to non-compliance with assigned treatments [10]. The situation may be complicated further by ethics approval requirements and new rulings in conducting clinical research [3]. An "opt-out" approach (contact was made unless individuals indicated unwillingness to participate) ensures a better response rate by following up non-response to an initial invitation [11]. Also, there may be an "opt-in" strategy for giving the participants assigned to the control group the opportunity to receive the same active treatment as those in the intervention group. When participants do not adhere to the protocol of the assigned treatment due to selection bias, the estimation of intervention effects and interpretation are not straightforward.

The most widely-used approach in estimating intervention effects is intention-to-treat (ITT), where treatment comparisons are based on the assigned treatment arms, regardless of whether the participants complied with the treatment. While ITT analysis preserves the benefits of randomisation, it provides valid measures of the intervention "effectiveness" rather

than the treatment "efficacy" (the effectiveness of an intervention when it is in fact taken) about which many researchers want to know. The ITT analysis can be adjusted for non-compliance to estimate intervention efficacy using the randomisation indicator as an instrumental variable (IV) [10,12-14]. Alternatively, per-protocol (PP) and as-treated (AT) analyses are simple approaches to estimate treatment efficacy, but they are subject to selection bias [10]. Many studies have identified strong selection bias in RCTs [3-4,7,9]. However, besides the proposal of allocation concealment and masking [15-17], few studies consider an appropriate way to correct for selection bias due to non-compliance in estimating intervention effects. How these methods (ITT with or without IV adjustment, PP, or AT) influence the comparison of treatments remains unclear, especially, when non-compliance of assigned treatment protocols and selection bias exist. In this paper, we propose a new approach with the use of "nearest-neighbour" matched data to identify compliant participants in both arms and hence to estimate intervention efficacy for RCTs with opt-in and/or opt-out recruitment strategies. Novelty lies in the idea of identifying compliant participants in both arms by data-matching on the basis of participants with known compliance information, not the matching method itself.

2. Methods

2.1. Setting

The Costs to Australian Private Insurance – Coaching Health (CAPICHe) is a parallel-group RCT of the relative impact of telephonic health-coaching support on healthcare cost and utilisation of participants in a disease-management program in Australia [18,19]. The trial enrolled participants sourced from Bupa Australia health-fund members and received ethical approval from Griffith University Human Research Ethics Committee (Ref.: MED/12/11/HREC). It was registered with the Australian New Zealand Clinical Trials

Registry (Ref.: ACTRN12611000580976). The published protocol provides information on pre-specified inclusion and exclusion criteria, sample-size calculation, data collection, and follow-up [18].

2.2. Randomisation, recruitment, and intervention program

Participants who met the inclusion criteria were randomly selected from Bupa Australia claims database. Independently of Bupa Australia, the samples were then randomised into the intervention or control groups in a 4:1 ratio stratified by five diagnosed chronic condition [18]. The effectiveness of randomisation for each batch of data was checked. The participants allocated to the intervention group were eligible to receive disease-management services provided by Bupa Health Dialog [19]. An opt-out approach (contact was made unless they contacted the intervention provider to signal unwillingness to participate) was adopted for recruitment.

In the control group, the participants received a letter outlining the services from Bupa Health Dialog. These participants received usual care but were given the opportunity, via mail, to opt-in to receive health coaching. Opt-in participants received the same services as that provided to a participant assigned to the intervention group. Health coaches had the same quality of information available about the members and had all the same educational resources at their disposal as they did for the intervention group. Other than participants in the usual-care arm who opted-in to the intervention and received health coaching, the health coaches were blind to whether participants were in the CAPICHe intervention group or were receiving health coaching as part of the usual business for Bupa [19].

Outcome measures are non-maternity healthcare costs and hospital length-of-stay 12 months post randomisation, presented as a value and a percentage of difference.

2.3. Methodology and statistical methods

Our approach to addressing selection bias due to non-compliance of assigned treatment involved two data-matching on the basis of participants with known compliance information, as indicated by the hollow arrows in Figure 1. The correction of selection bias was implemented by considering three levels of "principal compliance" corresponding to "Always-takers", "Compliant participants", and "Never-takers" [10]. This approach is in contrast to the inverse-probability-weighting (IPW) method that has been used in RCTs in which data on compliance are missing [20]. As illustrated in Figure 1, always-takers in the control group (Cell A) are participants who opted in to receive coaching (n=153 out of 8,883 randomised to the control group), whereas the participants who did not engage (n=8,730) were either compliant participants (Cell B) or never-takers (Cell C). As the principal-compliance levels of these 8,730 participants were not observable, the actual numbers of participants in Cells B and C were unknown. In the intervention group, engaged participants (n=15,375) were either always-takers (Cell D) or compliant participants (Cell E), while never-takers (Cell F, n=20,160) are participants who did not engage in the program. With the same reason above, the actual numbers of participants in Cells D and E were unknown. For RCTs that adopt opt-in and/or opt-out recruitment strategies, it is anticipated that there would be selection bias in those participants who engaged in health coaching and likely to be associated with higher costs or longer hospital stay in the follow-up period.

(Figure 1 here)

The intervention efficacy is estimated on the basis of compliant participants who adhere to the protocol of the allocated treatment (i.e. Cell E versus Cell B). Let $\bar{y}_{rp}$ denote the sample means (some are not observable) for the corresponding randomised group (r=0: Control group; r=1: Intervention group) and compliance information (p=A: Always-takers; p=C: Compliant participants; p=N: Never-takers), the proposed data-matching approach provides an estimate of treatment effect as:

$$\hat{\delta}_{DM} = \bar{y}_{1C} - \bar{y}_{0C} \quad (1)$$

This new matching approach will provide a better estimation of intervention efficacy, as supported by that Cells B and E are the only groups of compliant participants for each treatment arm, respectively. The compliant participants were identified using nearest-neighbour data matching (see Supplementary material for Stata command [21] "teffects nnmatch"). Always-takers in the intervention group (Cell D) were matched with reference to the (observed) always-takers in the control group (Cell A), while never-takers in the control group (Cell C) were matched with the observed never-takers in the intervention group (Cell F); see Figure 1. The intervention-to-control ratio of 4:1 was assumed to remain in the matching, which is plausible when the randomisation is effective (effectiveness of randomisation can be verified from the data). For matching Cell D from Cell A, four nearest-neighbour matches for each participant in Cell A were obtained, based on the Mahalanobis distance measure [22] (the covariance-adjusted distance between two points) using age, randomisation batch, historical cost and admission rate, and the number of coexisting conditions, whereas exact match was performed for engagement level (3 categories: number of coaching sessions 1-2, 3-6, >6; only available for engaged participants), gender, state of residence, and clinical condition. That is, engaged participants in the intervention group (Cells D and E) were split into Cell D (participants matched from Cell A) and Cell E (non-matched remainders). For matching Cell C from Cell F, the nearest-neighbour match of each participant in Cell F was obtained using the same distance measure as above except with regards to the engagement level. To keep the 4:1 intervention to control ratio, the best n=5,040 (20,160 divided by 4) nearest neighbours based on the Mahalanobis distances were considered as never-takers in Cell C. A comparison study was conducted to compare the crude estimates of intervention effects (without adjustment for baseline covariates) using the

data-matching approach, the ITT (with or without adjustment using the randomisation indicator as an IV [10,12-14]), PP, and AT methods, where

$$\hat{\delta}_{ITT} = \bar{y}_{1(A+C+N)} - \bar{y}_{0(A+C+N)} \quad (2)$$

$$\hat{\delta}_{IV} = \hat{\delta}_{ITT}/(1 - \hat{\pi}_{0A} - \hat{\pi}_{1N}) \quad (3)$$

$$\hat{\delta}_{PP} = \bar{y}_{1(A+C)} - \bar{y}_{0(C+N)} \quad (4)$$

$$\hat{\delta}_{AT} = \hat{\delta}_{PP} + \bar{y}_{0A} - \bar{y}_{1N}, \quad (5)$$

and where $\hat{\pi}_{rp}$ denotes the proportion of samples with principal compliance (p) in the randomised group (r). From (1) and (4), it can be seen that the data-matching approach estimates treatment efficacy by restricting analysis to participants who comply with the assigned allocation (like PP analysis) and simultaneously correcting for selection bias from always- and never-takers due to non-compliance (Cell D and Cell C in Figure 1).

To adjust for baseline covariates, zero-inflated regression models [19,23,24] were adopted to estimate the intervention efficacy on healthcare costs and hospital length-of-stay between compliant participants (Cells B and E) identified using the data-matching approach. Regression covariates were historical costs and admission counts, age, gender, state of residence, randomisation batch, as well as the diagnosed chronic condition, the count of chronic conditions, and the proportion of admissions due to surgical treatment within the follow-up period. Outcome measures with incomplete follow-up were adjusted in the analyses using the observed days of follow-up as exposure risk [19]. Covariates with a p-value greater than 0.05 were removed from the models. Analyses were undertaken in Stata IC 13.1; StataCorp, College Station, TX.

3. Results

The study enrolled a total of 44,418 participants, of whom 35,535 participants (80%) were allocated to the CAPICHe intervention group and 8,883 participants (20%) were allocated to

the usual-care control group. The randomisation of treatment groups was effective, as there were no major differences in baseline characteristics (including age group, gender, state of residence, historical healthcare cost and frequency of admissions for the prior 12 months to randomisation) between the two groups; detailed baseline characteristics of participants are reported elsewhere [19].

Table 1 presents the dose-response relationship. Strong selection bias was observed in that participants with higher engagement levels in health coaching were also associated with significantly higher costs or longer hospital length-of-stay within 12 months post randomisation.

(Table 1 here)

The average healthcare costs 12 months post randomisation for the three levels of principal compliance after data-matching were presented in Figure 2. The trend of increasing costs for never-takers, compliant participants, and always-takers confirmed the existence of strong selection bias. It can be observed that always-takers in the control group had higher costs compared to those in the intervention group, while never-takers in the intervention group had higher costs compared to those in the control group.

The crude estimation of intervention efficacy on the healthcare cost was summarised in Table 2. The ITT compares the outcomes of participants between the intervention group ($4,840) versus the control group ($4,914), regardless of whether they complied with the treatment allocation (a cost saving of $74, or 1.5%). With adjustment for non-compliance using the IV method [10,12], the adjusted ITT estimate for intervention efficacy is a saving of $178 (i.e. $74/(1-153/8883-20160/35535), or 3.6%). The PP analysis restricts the comparison to participants who comply with the assigned treatment (Cells D and E in the intervention group ($5,338) versus Cells B and C in the control group ($4,881); an additional cost of $457, or 9.4%), whereas the AT analysis compares participant's outcomes according to the

treatment actually received (participant received the intervention (Cells A, D, E; $5,351) versus those who did not (Cells B, C, F; $4,571); an additional cost of $780, or 17.1%). In the presence of selection bias, biased estimation of intervention efficacy was obtained with the PP and AT analyses because it involved comparisons across the three levels of principal compliance (see shaded cells in Figure 1). Comparing compliant participants between Cells B and E, the intervention efficacy was $5,332 versus $6,080 (a saving of $748, or 12.3%).

(Figure 2 here)

(Table 2 here)

The baseline characteristics of compliant participants in Cells B and E are different, providing further support of the hypothesis that there would be selection bias in those participants who choose to opt-in or opt-out (see Supplementary Table 1). Table 3 presents comparisons of treatment efficacy for the data-matching method on the basis of compliant participants with adjustment for baseline covariates. Adjusted estimates of healthcare costs were lower for the intervention group compared to the control group, especially the cost due to same-day admissions (intervention group: $516; 95% CI: $493-$539; control group: $597; 95% CI: $554-$641; reduced cost: $81; 95% CI: $38-$125; $p<0.001$). The intervention group had a shorter adjusted estimate of hospital stay (4.164 days; 95% CI: 3.967-4.361) compared to the control group (4.691 days; 95% CI: 4.275-5.107) with reduced hospital stay of 0.527 days (95% CI: 0.081-0.974; $p=0.021$).

(Table 3 here)

4. Discussion

For RCTs, the widely-used ITT analysis preserves the benefits of randomisation by comparing intervention effects according to the assigned treatments and provides a valid measure of treatment effectiveness, such as a reduction in cost of 1.5% ($74) in this health-

coaching trial. This is important information to policy makers and health planners, by knowing the benefit from offering the intervention program to individuals (who may not adhere to treatment protocols). In this trial, 56.7% of participants assigned to the intervention group did not engage in any health coaching. Many researchers thus also want to know the efficacy of the treatment effect when it is in fact taken. While the ITT analysis can be adjusted for non-compliance using the IV method to estimate intervention efficacy, the IV method requires the "exclusion restriction" assumption that mean effects (healthcare costs in this example) are the same between the intervention and control groups for always-takers (Cells A and D) as well as never-takers (Cells C and F) [10,12]. According to our findings displayed in Figure 2, these assumptions, however, are not valid. The PP and AT analyses were also inappropriate as they suffered from selection bias by comparing participants across different compliance levels. Also, it is expected that the PP analysis tends to give, on average, higher estimates of intervention effect than the ITT analysis, especially in RCTs with dropouts [25-28]. But the findings obtained in this trial indicated the opposite because of the strong selection bias due to non-compliance (see Table 2). We proposed the matching approach to correct for selection bias due to non-compliance by identifying compliant participants in the intervention and control groups, with whom the intervention efficacy was estimated. We found that, after adjustment of baseline covariates, the total healthcare costs in the intervention group were $107 (95% CI: -$241 to $455) lower than the control group, representing a 2.0% ($107 of $5440) reduction in cost (p=0.546). Moreover, the total cost due to same-day admissions was lower ($81 or 13.6%; 95% CI: $38-$125 or 7.3%-20.0%; p<0.001) and the hospital length-of-stay was shorter (0.527 days or 11.2%; 95% CI: 0.081-0.974 or 2.6%-19.9%; p=0.021) in the intervention group compared to the control group.

Estimation of intervention efficacy in RCTs by correcting for selection bias due to non-compliance is vital for the scientific rigour of medical and health research to establish

evidence-based practices. While well-planned trial designs with proper allocation concealment and masking certainly help to minimise selection bias [15-17], it is still required in the analysis to correct for selection bias due to different compliance levels of participants, as selection bias may not be fully removed by adjustment for covariates [10]. Our findings demonstrated strong selection bias in compliance with reference to assigned treatment arms when opt-in and opt-out methods were in place. As illustrated by this health-coaching trial, there is a tendency for participants with higher engagement levels to be associated with higher healthcare costs in the follow-up period (differences across: engagement levels, Table 1; compliance levels, Figure 2). Always-takers in the control group who opted in to receive coaching had the highest healthcare costs among the participants in the control group, whereas never-takers in the intervention group who opted out of coaching engagement had the lowest healthcare costs relative to other participants in the intervention group (Figure 2). Differences within compliance levels were also found: Always-takers in the control group had higher healthcare costs compared to those in the intervention group, whereas never-takers in the intervention group had higher healthcare costs relative to those in the control group. These findings will have great impact on the choice of analysis methods for estimating intervention efficacy in RCTs with strong selection bias due to non-compliance.

4.1. Strengths and limitations

This is a large RCT for health coaching in chronic-disease management, with 44,418 participants in five preselected diagnosed chronic conditions. Previous trials either had relatively smaller sample sizes or focussed only on one specific condition [29-31]. With large sample sizes, our trial had sufficient power to detect true differences between the treatment arms if present. Detailed evaluation of the effectiveness of this disease-management program on healthcare cost and utilisation of participants using the ITT analysis with adjustment for baseline covariates was reported elsewhere [19], which provides important information to

policy makers and health planners regarding the intervention effectiveness of CAPICHe. The present trial had a number of limitations including a small number of contacts conducted per participant (median: 2; inter-quarter range: 2-4) and that clinical, surrogate or summary health outcomes were not measured.

The capability of this new approach for correcting selection bias due to non-compliance has been illustrated in Table 2. Unlike the exclusion restriction assumption for the IV method regarding zero intervention effects for always- and never-takers (which are invalid in this health-coaching trial), the proposed data-matching approach needs only effective randomisation of participants and the availability of key determinants of the outcome under study (as covariates) for matching participants with known compliance information. The former requirement is a typical assumption for treatment comparisons in RCTs. Other distance measures may be used in data matching, such as propensity scores [32]. An advantage of the nearest-neighbour approach is that it allows performing exact matches for selected covariates, such as engagement level, gender, state of residence, and clinical condition in the present trial. This feature offers a more robust matching of participants. It is worth mentioning that the two data matching did not have equal “precision”. The matching of Cell D from Cell A may not be sufficiently close because four nearest-neighbour matches for each participant in Cell A were considered as always-takers in Cell D. For RCTs without an opt-in strategy for the control group (empty Cell A), matching of Cell C from Cell F is still possible to identify complinant participants in the control group (Cell B). Subsequently, through the matching from Cell B, compliant participants in the intervention group (Cell E) can be found. However, further sensitivity analyses are required to study the impact of the uncertainty in compliance information within this latter matching process on treatment comparisons for these more general RCTs.

4.2. Implications

In conducting RCTs, the selection bias arising from the opt-in and/or opt-out recruitement strategies cannot be reduced by the means of randomisation and must be corrected for in the analysis in order to obtain a bias-corrected estimate of intervention effects. Our method is able to identify (nearest-neighbour) participants from participants in the other treatment arm with observed compliance levels, thus enables the estimation of intervention efficacy by comparing outcomes of participants who adhere to the protocol of the allocated treatment in both arms. Our findings are useful for the analysis of future RCTs, where strong selection bias due to non-compliance is anticipated.

Funding

This work was supported by the Bupa Health Foundation, Australia, which is an independent Foundation with charity status under the Australian Taxation Office rules. The funders had no role in the trial design, data collection, data interpretation, data analysis, or writing of the report.

Acknowledgment

We thank Christine Pollicino, Stan Goldstein and Raimundo Gomes (Bupa Australia) and David Cross (Bupa Health Dialog) for enrolling participants and collecting (extracting) data for the randomised controlled trial.

Conflict of interest: The authors have declared no conflict of interest.

References

[1] D'Agostino RB Jr, D'Agostino RB Sr. Estimating treatment effects using observational data. J Amer Med Assoc 2007;297:314-6.

[2] Hewitt CE, Kumaravel B, Dumville JC, Torgerson DJ. Trial attrition study group. Assessing the impact of attrition in randomized controlled trials. J Clin Epidemiol 2010;63:1264-70.

[3] Woolf SH, Rothemich SF, Johnson RE, Marsland DW. Selection bias from requiring patients to give consent to examine data for health services research. Arch Fam Med 2000;9:1111-8.

[4] Montoy JCC, Dow WH, Kaplan BC. Patient choice in opt-in, active choice, and opt-out HIV screening: randomized clinical trial. Brit Med J 2016;352:h6895.

[5] Chalmers I. Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. Int J Epidemiol 2001;30:1156-64.

[6] Braunholtz DA, Edwards SJL, Lifford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". J Clin Epid 2001;54:217-24.

[7] Tabar L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 2001;91:1724-31.

[8] Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. Brit Med J 2001;323:42-6.

[9] Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schunemann H, et al. Randomisation to protect against selection bias in healthcare trials. Cochrane Database Syst Rev 2011;13:MR000012.

[10] Little RJ, Long Q, Lin X. A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. Biometrics 2009;65:640-9.

[11] Junghans C, Feder G, Hemingway H, Timmis A, Jones M. Recruiting patients to medical research: double blind randomised trial of "opt-in" versus "opt-out" strategies. Brit Med J 2005;331:940.

[12] Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722-9.

[13] Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. Brit Med J 2010;340:c2073.

[14] van der Velden JM, Verkooijen HM, Young-Afat DA, Burbach JP, van Vulpen M, Relton C, et al. The cohort multiple randomized controlled trial design: a valid and efficient alternative to pragmatic trials? Int J Epidemiol 2017;46:96-102.

[15] Norris SL, Engelgau MM, Venkat Narayan KM. Effectiveness of self-management training in Type 2 diabetes. Diabetes Care 2001;24:561-87.

[16] Pildal J, Hrobjartsson A, Jorgensen KJ, Hilden J, Altman DG, Gotzsche PC. Impact of allocation concealment on conclusions drawn from meta-analyses of randomized trials. Int J Epidemiol 2007;36:847-57.

[17] Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: Importance of allocation concealment and masking. Fam Med 2007;39:132-7.

[18] Byrnes JM, Goldstein S, Venator B, Pollicino C, Ng SK, Veroff D, et al. The impact of population-based disease management services for selected chronic conditions: the Costs to Australian Private Insurance - Coaching Health (CAPICHe) study protocol. BMC Public Health 2012;12:114. Available at:

http://www.bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-114.

[19] Scuffham PA, Byrnes JM, Pollicino C, Cross D, Goldstein S, Ng SK. The impact of population-based disease management services on health care utilisation and costs: Results of the CAPICHe trial. J Gen Intern Med 2019;34:41-48.

[20] Sheikn K. Investigation of selection bias using inverse probability weighting. Eur J Epidemiol 2007;22:349-50.

[21] StataCorp. Stata 13 Base Reference Manual. College Station, TX: Stata Press; 2013. Available at: http://www.stata.com/manuals13/teteffectsnnmatch.pdf

[22] Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci 2010;25:1-21.

[23] Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. Health Econ 2011;20:897-916.

[24] Ng SK, Xiang L, Yau KKW. Mixture Modelling for Medical and Health Sciences. Boca Raton, FL: Chapman & Hall/CRC; 2019, Chapter 4.

[25] Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. J Clin Epid 2007;60:663-9.

[26] Lachin JM. Statistical considerations in the intent-to-treat principle. Control Clin Trials 2000;21:167-89.

[27] Sedgwick P. What is per protocol analysis? Brit Med J 2013;346:f3748.

[28] Sheiner LB, Rubin DB. Intention-to-treat analysis and the goals of clinical trials. Clin Pharmacol Ther 1995;57:6-15.

[29] Gordon LG, Bird D, Oldenburg B, Friedman RH, Russell AW, Scuffham PA. A cost-effectiveness analysis of a telephone-linked care intervention for individuals with Type 2 diabetes. Diabetes Res Clin Pract 2014;104:103-11.

[30] Wolever RQ, Dreusicke M, Fikkan J, Hawkins TV, Yeung S, Wakefield J, et al. Integrative health coaching for participants with type 2 diabetes: a randomized clinical trial. Diabetes Educ 2010;36:629-39.

[31] Sangster J, Church J, Haas M, Furber S, Bauman A. A Comparison of the Cost-effectiveness of Two Pedometer-based Telephone Coaching Programs for People with Cardiac Disease. Heart Lung Circ 2015;24:471-9 .

[32] Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surveys 2008;22:31-72.

<table>
<tr><td rowspan="2">Randomised group</td><td colspan="3">Principal compliance</td><td rowspan="2">Overall</td></tr>
<tr><td>Always-takers</td><td>Compliant participants</td><td>Never-takers</td></tr>
<tr><td>Control group</td><td rowspan="2">Cell A
(opted in to receive coaching as Always takers; n=153)</td><td>Cell B
(not engaged due to Compliance; n=unknown)</td><td>Cell C
(not engaged as Never takers; n=unknown)</td><td>n=8883</td></tr>
<tr><td colspan="2">n=8730 not engaged
(either Compliance or Never takers)</td><td rowspan="2">n=35,535</td></tr>
<tr><td>Intervention group</td><td>Cell D
(engaged as Always takers; n=unknown)</td><td>Cell E
(engaged due to Compliance; n=unknown)</td><td>Cell F
(not engaged as Never takers; n=20,160)</td></tr>
<tr><td></td><td colspan="2">n=15,375 engaged
(either Always takers or Compliance)</td><td></td></tr>
</table>

Strong selection bias anticipated (higher costs for participants with higher engagement levels; represented by the semi-transparent and dashed arrow across the three levels of principal compliance)

Figure 1. Classification of participants by treatment group and principal compliance (unobservable principal compliance information was presented in italic; shaded cells correspond to participants who received coaching. Nearest-neighbour data matching for Cells C and D with reference to known observed compliance from Cells F and A were indicated by the hollow arrows). Through data-matching, we have n=5040 in Cell C and hence n=3690 in Cell B, whereas n=599 in Cell D due to overlapping of 13 matched participants and hence n=14,776 in Cell E.

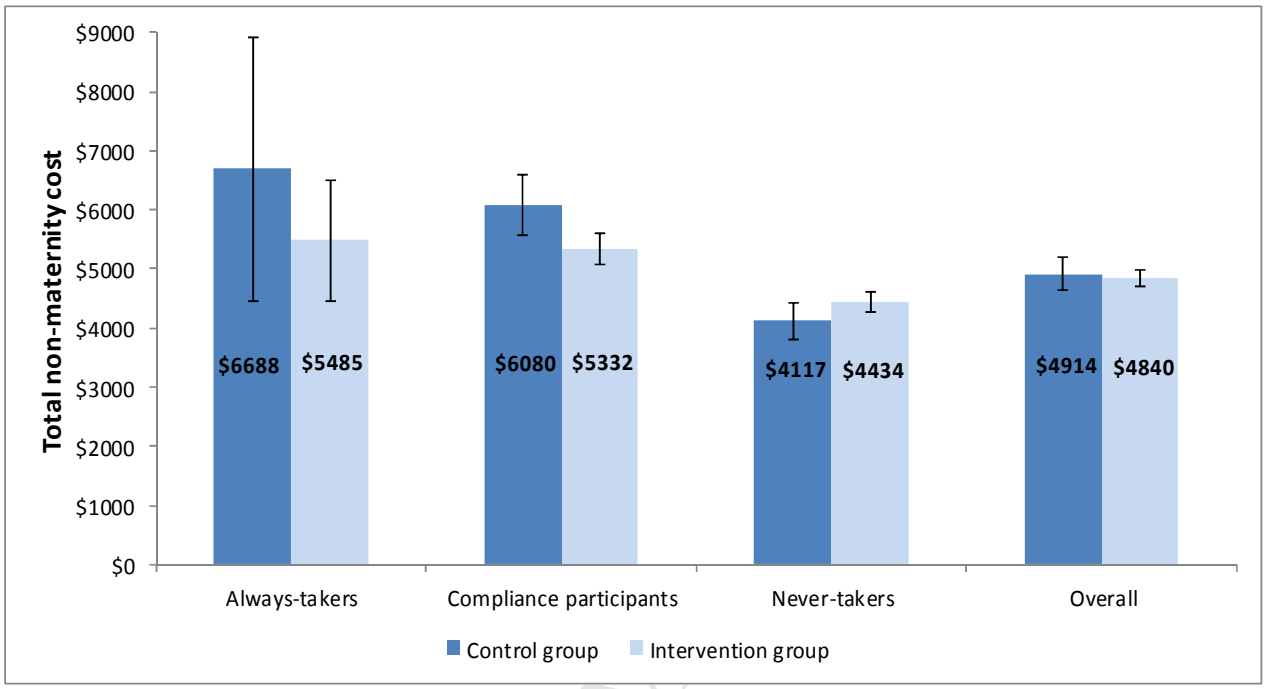


Figure 2. Mean non-maternity costs 12 months post randomisation by treatment group and principal compliance (the numbers of participants in the control group, n_c, and the intervention group, n_i are: Always-takers, n_c=153, n_i=599; Compliance participants, n_c=3690, n_i=14,776; Never-takers, n_c=5040, n_i=20,160; Overall, n_c=8883, n_i=35,535). The percentage of opt-in in the control group=1.7% (153/8883) and the percentage of opt-out in the intervention group=56.7% (20160/35535).