Insights into Different Results from Different Causal Contrasts in the Presence of Effect-Measure Modification

Til Stürmer[1,2], **Kenneth J Rothman**[1,3], and **Robert J Glynn**[1,2,4]

1 *Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, Boston, USA*

2 *Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, USA*

3 *Department of Epidemiology, Boston University School of Public Health, Boston, USA*

4 *Department of Biostatistics, Harvard School of Public Health, Boston, USA*

Abstract

Purpose—Both propensity score (PS) matching and inverse probability of treatment weighting (IPTW) allow causal contrasts, albeit different ones. In the presence of effect-measure modification, different analytic approaches produce different summary estimates.

Methods—We present a spreadsheet example that assumes a dichotomous exposure, covariate, and outcome. The covariate can be a confounder or not and a modifier of the relative risk (RR) or not. Based on expected cell counts, we calculate RR estimates using five summary estimators: Mantel-Haenszel (MH), maximum likelihood (ML), the standardized mortality ratio (SMR), PS matching, and a common implementation of IPTW.

Results—Without effect-measure modification, all approaches produce identical results. In the presence of effect-measure modification and regardless of the presence of confounding, results from the SMR and PS are identical, but IPTW can produce strikingly different results (e.g. RR=0.83 vs. RR=1.50). In such settings, MH and ML do not estimate a population parameter and results for those measures fall between PS and IPTW.

Conclusions—Discrepancies between PS and IPTW reflect different weighting of stratum specific effect estimates. SMR and PS matching assign weight according to the distribution of the effect-measure modifier in the exposed subpopulation, whereas IPTW assigns weights according to the distribution of the entire study population. In pharmacoepidemiology, contraindications to treatment that also modify the effect might be prevalent in the population, but would be rare among the exposed. In such settings, estimating the effect of exposure in the exposed rather than the whole population is preferable.

Keywords

Address for reprints: Til Stürmer, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont Street, Suite 3030, Boston, MA 02120, USA, Phone: +1 617 278 0627, Fax: +1 617 232 8602, Email: til.sturmer@post.harvard.edu.

INTRODUCTION

In the presence of effect-measure modification or interaction across strata. stratum-specific estimates need to be presented.[1] Nonetheless, some authors propose the use of standardized summaries, as opposed to summaries that assume uniformity, in such situations, since these retain valid interpretations and can be used to measure overall population impact.[1–3] Two analytic methods that are increasingly used to control for confounding, propensity scores (PS) and inverse probability of treatment weighting (IPTW), allow such summary estimates. In this paper we give a non-technical description of these methods and present numerical examples comparing PS, IPTW and other commonly used approaches to obtaining a summary estimate of effect across strata.

Propensity scores[4] (PS) are increasingly used in non-experimental medical research.[5] Propensity scores can be implemented in various ways, including matching. Under the assumption of no unmeasured confounding, unexposed observations individually matched to each exposed observation on the estimated PS can be conceptualized as counterfactuals, i.e. representing the experience of the exposed people if they had been unexposed. The crude relative risk obtained from a cohort matched on the PS can therefore be interpreted as the causal contrast between the outcome experience in the exposed (factual) and one counterfactual, i.e. what would have happened to the exposed if they had been unexposed. The target of this causal contrast thus is the population that is going to be getting the treatment.

Although alternative weighting strategies are available, a popular implementation of inverse probability of treatment weighting (IPTW) is based on the estimated probability of exposure in the exposed and the probability of non-exposure in the unexposed.[6,7] These analyses adjust for confounding by creating a re-weighted pseudo-population in which exposure is independent of the measured confounders. Under the assumption of no unmeasured confounding, this implementation of IPTW estimates the causal contrast between two counterfactuals: what would have happened if everyone in the population had been exposed vs. what would have happened if everyone had been unexposed. This population or marginal contrast is conceptually similar to the contrast that results from a randomized controlled trial including all subjects. The target of this causal contrast thus is the general population which might be justified in specific settings, e.g. for mass vaccinations or over-the-counter medications.

IPTW can be implemented in various ways leading to different causal contrasts. For the sake of clarity and of the argument, however, we will assume implementations of IPTW leading to a marginal contrast from now on. In situations in which there is confounding by a variable that is also affected by previous exposure, and information relating to repeated assessments of exposure and covariates, IPTW has been promoted as the basis to create marginal structural models, allowing the control of these confounding variables that cannot be controlled using traditional methods.[8,9] Here we use IPTW in a more basic situation that does not involve time-varying exposure or covariates, so that IPTW can be compared with alternative methods that would not be appropriate for time-varying exposures.

Recently, Kurth et al. reported striking differences of the association between thrombolytic therapy and mortality in patients with ischemic stroke comparing results from PS matching (OR = 1.2) and the marginal causal ratio obtained by IPTW (OR = 11).[10] In this paper we describe the differences between these analytic techniques in the presence of effect-measure modification and we discuss the choice among these analytic approaches in specific settings such as pharmacoepidemiology.

METHODS

To illustrate the analytic techniques using numerical examples, we chose the simplest possible scenario, i.e. a dichotomous exposure E (E=1 or 0 for exposure present or absent) that has a prevalence P_E, a dichotomous covariate X (X=1 or 0 for the covariate being present or absent) with a prevalence P_X, and a dichotomous outcome Y (Y=1 or 0 according to whether the outcome occurred or did not occur) with an incidence proportion IP. The covariate X can be either confounding or not and it can either be an effect-measure modifier of the relative risk for the exposure-outcome association or not.

We first calculated the expected number of observations with and without the outcome within each stratum of X=i from the following parameters: the total size of the study (N), the prevalence of the covariate (P_X), the prevalence of the exposure within each stratum of the covariate (P_{Ei}), and the risks for the outcome IP_{1i} and IP_{0i} among exposed and unexposed in each stratum of X, respectively.

The Mantel-Haenszel (MH) summary estimator was calculated as a weighted average of the stratum-specific relative risks with weights based on the standard MH weights

$$w_i = N_{0i} {}^* IP_{0i} {}^* N_{1i} \big/ N_{+i},\qquad(1)$$

where N_{1i}, N_{0i} and N_{+i} are the number of exposed, unexposed, and total in the i-th stratum of X (i=0,1), respectively.

We also calculated the maximum likelihood (ML) estimator of the RR, an iterative calculation involving the solution of a set of i + 1 equations with i + 1 unknowns, where i is the number of strata, as described by Rothman.[11] In the presence of effect-measure modification, the MH estimator and the ML estimator, both of which are intended to estimate uniform effects, do not consistently estimate any meaningful population parameter.[12] We present these estimates here for comparison with the other estimates, without implying that they are valid or that they should be used in the presence of effect-measure modification of relative risk. In many realistic settings, the two estimators have been shown to approximate relative risks standardized to the total population.[13]

We then calculated the standardized mortality or morbidity ratio (SMR), which has weights based on the distribution of the covariate in the exposed. The SMR is calculated as the ratio of the number of events among those exposed divided by the number of events expected for the exposed based on the morbidity experience of the unexposed within each stratum. The expected number is thus the sum of the stratum-specific risks in the unexposed (IP_{0i}) multiplied by the size of the exposed group within each stratum of the covariate:

$$\text{Expected} = \Sigma_i N_{1i} {}^* IP_{0i}.\qquad(2)$$

The propensity score in this simple setting is the probability of being exposed within each level of the covariate and thus equals the prevalence of the exposure within each stratum of X:

$$e(E)_i = P_{Ei}\qquad(3)$$

We matched on this PS by choosing a number of unexposed equal to the number of exposed in each stratum of X and multiplying this number by IP_{0i} to obtain the expected number of events within the matched unexposed group. Once this matching is completed for both strata of X, the two strata of X are collapsed into a single 2x2 table and the RR_{EY} is calculated by

taking the ratio of the observed risk among the exposed by the observed risk among the matched unexposed.

Another way to use PS is to stratify on the score. In our setting with one dichotomous covariate, the estimate obtained after stratifying on the PS and combining these strata using ML corresponds to the ML-estimate presented above. In this simple setting, the PS stratified estimate therefore corresponds to the estimate obtained by "conventional" control for a covariate in a traditional outcome model. It should be pointed out, however, that estimates stratified on the PS can be combined by other methods, including the SMR.

One commonly used way to implement an IPTW estimator is to use weights that are the inverse of the PS in those exposed and the inverse of 1-PS (i.e. the probability of "non-exposure") in those unexposed. To obtain a weighted population of equal size to compare with the original population, we used stabilized weights that include the marginal prevalence of the observed exposure status, i.e. P_E in the exposed and $1-P_E$ in the unexposed, in the numerator.[9] The following weights were calculated accordingly:

$$w_{1i} = (P_E) \big/ [e(E)_i] = (P_E) / [P_{Ei}] \text{ and} \tag{4}$$

$$w_{0i} = (1 - P_E) \big/ [1 - e(E)_i] = (1 - P_E) / [1 - P_{Ei}] \tag{5}$$

in the exposed and unexposed, respectively.

Multiplying the number of observations in each of the 4 groups defined by exposure and covariate by the corresponding weights, we obtained a new population defined by the number of observations within these four strata. The number of outcomes within each of these four strata corresponds to the expected value, i.e. the number of observations multiplied by the observed stratum-specific risks for the outcome IP_{1i} and IP_{0i} in exposed and unexposed, respectively. The two strata of X are collapsed into a single 2x2 table and the RR_{EY} is calculated by dividing IP_1 by IP_0.

All the above formulas were entered into an Excel-spreadsheet to allow easy calculation with any input data. The spreadsheet, named 'CausalComp.V1.xls' is offered as open code under the conditions of the GNU-GPL license,[14] and can be downloaded at www.members.aol.com/epidemiol/CausalComp.V1.xls. The analytic strategies assessed are summarized and compared with emphasis on their theoretical behavior in the setting of nonuniform effects in table 1.

RESULTS

The first scenario (table 2) has a relative risk between exposure E and outcome Y that is confounded by the covariate X while it is identical within both strata of X (no effect-measure modification of the risk ratio). We assumed a protective effect of an exposure, which can be conceptualized as a drug intervention that prevents outcome Y, such as anti-hypertensive medication to prevent a specified level of cognitive decline in an elderly population of hypertensives.[15] The confounding is due to the exposure being more prevalent in individuals with X ($P_{E1} = 0.5$) compared with individuals without X ($P_{E0} = 0.1$) and X being a risk factor for the outcome in the unexposed ($IP_{01} = 0.4$ vs. $IP_{00} = 0.2$). X might be conceptualized as the severity of hypertension.

Since we are using relative risks (as opposed to odds ratios), the presence of confounding is demonstrated by the crude estimate differing from stratum-specific estimates.[16] The absence of effect-measure modification is demonstrated by the stratum-specific relative risk estimates

being identical. With no effect-measure modification, all analytic techniques control effectively for confounding and result in identical relative risk estimates of 0.5.

In table 3, we examine an example without confounding but with effect-measure modification of the relative risk. There is no confounding because X is not associated with the outcome in the unexposed ($IP_{0i} = 0.2$). The exposure of interest is still protective in individuals that have the covariate X, but the exposure is now a risk factor for the outcome in those individuals without the covariate X (i.e. where X=0). Continuing with the example of hypertension, the latter group might be conceptualized as individuals with essentially normal blood pressure in whom antihypertensive treatment increases rather than decreases the risk of cognitive decline, e.g. by an unwanted drop in blood pressure.[15] Whereas the crude relative risk and the relative risks obtained by SMR and propensity score matching are identical (RR=0.83), the RR obtained by IPTW is 1.5. The MH and ML estimators are 1.03 and 1.09, respectively, slightly closer to the results obtained by SMR and PS than to the IPTW estimator.

The reason for this difference is apparent upon comparing the number of individuals in each stratum of X with PS matching and IPTW. With PS matching, there are far fewer individuals in the stratum with X=0 than in the stratum with X=1. This distribution reflects the prevalence of X among exposed people. Both the PS and the SMR weight the summary according to the distribution of X among exposed people and are based on the same counterfactual number of expected events for the exposed had they been unexposed (N=60).

The difference between SMR or PS-matching and IPTW can also be understood as a difference in the causal contrasts: whereas SMR and PS-matched analysis base their weighting on the distribution of the covariate X in the exposed, IPTW is based on the distribution of the covariate X in the entire study population.

In table 4, we present an example when there is both confounding and effect-measure modification. Confounding is introduced by creating an association between X and the outcome in unexposed ($IP_{01} = 0.2$ and $IP_{00} = 0.1$). This example can again be conceptualized in the setting of antihypertensive treatment as the combination of severity of disease and lack of indication. The pattern is essentially the same as in table 3 with the exception that the crude RR is now biased owing to the presence of confounding by X. Quantitatively, the MH and ML estimators are again closer to the relative risk estimates obtained by SMR and PS than to the one from IPTW.

All of the numerical examples presented so far assume a prevalence of the covariate X of 0.5. The results from varying this proportion from 0.1 to 0.9 while keeping the other parameters constant are presented in table 5. The differences between the estimates obtained from IPTW and the estimates obtained by SMR or PS matching are most pronounced when the prevalence of X is 0.5. This result is not surprising given that differences in weighting based on the exposed and those based on the entire study population will tend to vanish as the prevalence of X in the entire study population approaches either 0 or 1.

We present a simplified, hypothetical numerical example of a study on the association between antithrombotic therapy and risk of stroke in an elderly population (say over 80 years of age) in table 6. The effect measure modifier in this setting is atrial fibrillation: antithrombotic therapy (e.g. warfarin) reduces the risk of embolic stroke in individuals with atrial fibrillation[17] but increases the risk of hemorrhagic stroke in individuals without atrial fibrillation[18]. We assume a prevalence of atrial fibrillation of 10% in this elderly population. [16] Without antithrombotic therapy, the risk of stroke is 5 times higher in individuals with atrial fibrillation (15%) than in individuals without atrial fibrillation (3%).[16] The prevalence of antithrombotic therapy is 50% in those with atrial fibrillation compared with 1% in those individuals without atrial fibrillation and the relative risks of stroke associated with

antithrombotic therapy are 0.33 and 2.7 in individuals with[17] and without[18] atrial fibrillation, respectively. In a hypothetical study of N=100,000, these parameters lead to the expected numbers of individuals in each of the 8 cells defined by atrial fibrillation, antithrombotic therapy, and stroke presented in table 6. The relative risk obtained by SMR and PS matching is 0.42 compared with 1.86 from IPTW. Using this implementation of IPTW leading to a marginal risk ratio, patients that do not have atrial fibrillation (and therefore do not profit from antithrombotic therapy with respect to risk of stroke) but receive antithrombotic therapy are up-weighted (weight = 5.9) whereas those individuals with atrial fibrillation who receive antithrombotic therapy as indicated to reduce the risk of stroke are down-weighted (weight = 0.118). This weighting is the basis for the marginal causal risk ratio (RR=1.86), providing an estimate of the effect of treating everyone compared with treating no one.

DISCUSSION

Our comparison of results of summary estimates in the presence of effect-measure modification using MH, ML, SMR, PS matching, and IPTW should clarify and help to understand not widely appreciated similarities between SMR and PS matching and striking discrepancies between these analytic approaches and a popular implementation of IPTW. The results from MH and ML lie between SMR and PS matching on the one hand and IPTW on the other. MH and ML, however, do not estimate a meaningful population parameter in the presence of pronounced effect-measure modification.[12] In the presence of less pronounced and more realistic effect-measure modification, MH weighting and ML have been shown to produce results closer to a population-standardized estimate (e.g. IPTW as implemented here).[13]

Although discrepant results due to differences in weighting of unequal stratum-specific estimates from SMR and IPTW have already been noted,[10,19] the similarities between SMR and PS with weights based on the distribution of the effect-estimate modifier in the exposed and the discrepancy of IPTW with weights based on the distribution of the effect-estimate modifier in the entire study population may not have been widely appreciated. Weighting according to the distribution of effect-measure modifiers observed in the whole study population is ideal in studies with well defined inclusion and exclusion criteria (i.e. everyone has the indication under study, and no one has a contraindication). In such studies the operating theme is that anyone entering the study could be either exposed or unexposed, similar to a randomized controlled trial where the exposure only depends on the outcome of the random assignment. The target of this causal contrast thus is the general population which might be justified in specific pharmacoepidemiologic settings, e.g. for mass vaccinations or over-the-counter medications.

Outside of such an idealized setting, however, treatment is less likely to be given to individuals with borderline indication, no indication, or even contraindications. In such a setting, weighting according to the distribution of the effect-measure modifier in only those exposed is easier to interpret, since the target of this causal contrast is the population that is going to be getting the treatment. The counterfactual question, "what would have happened if everyone had been treated", might be moot in many pharmacoepidemiologic contexts, including our example used to conceptualize the results presented, i.e. antithrombotic therapy and risk for stroke in a population containing many people who do not have atrial fibrillation and who would therefore not likely be treated.

Kurth et al. recently reported striking differences of the association between thrombolytic therapy and mortality in patients with ischemic stroke between PS matching (OR = 1.2) and the marginal causal ratio obtained by IPTW (OR = 11)[10] This discrepancy between approaches might be explained in part by a large proportion of the study population having a contraindication to thrombolysis, e.g., an onset of symptoms of more than 3 hours before

hospital admission.[20] Owing to an increased risk of intracerebral bleeding, thrombolysis is clearly associated with an increased mortality in patients treated more than 3 hours after symptom onset, a widely accepted contraindication.[21] Thrombolysis does not increase mortality in patients treated within 3 hours of symptom onset, the period during which this treatment might be indicated (because it can reduce stroke sequelae). The prevalence of the contraindication is low among those who are actually treated, but it is high (around 50%) in the entire study population.

We have presented a simple hypothetical numerical example of a study of the association between antithrombotic therapy and risk of stroke in an elderly population. The parameter values we used were derived from published literature. It is straightforward to make the discrepancies more extreme. The calculations clearly show how patients that do not have atrial fibrillation (and therefore do not profit from antithrombotic therapy with respect to risk of stroke) but receive antithrombotic therapy (for whatever reason) are up-weighted in this implementation of IPTW. Conversely, those individuals with atrial fibrillation who receive antithrombotic therapy as indicated to reduce the risk of stroke are down-weighted. These weights are used for the calculation of the marginal causal risk ratio (RR=1.86), providing an estimate of the effect of treating everyone compared with treating no one in a population in which only 50 percent have the indication (with respect to risk of stroke).

Rather than obtaining any summary estimate in such data, it might be preferable to exclude patients without atrial fibrillation from the analysis and present results only for those who have the indication (i.e. conditional on the indication; RR=0.33). Such a restriction makes the study setting closer to what would prevail in a randomized trial where treatment might be assigned to any participant who meets inclusion criteria. Residual confounding and effect-measure modification within each stratum might again be addressed using the methods presented, taking the differences between these methods in the presence of effect-measure modification into account. Alternatively, if the intent is to assess the population impact, including off-label use, it might be preferable to weight subjects using standardization based on the observed off-label use (e.g., SMR or PS matching; RR=0.42) or any reasonable explicit set of weights. Such summary estimates have a clear advantage of providing information on the usually reduced effectiveness of the treatment under realistic conditions that cannot be obtained from randomized trials or with an analysis conditional on having the indication for treatment.

The SMR has often been suggested and is often used as the analysis of choice to summarize data in the presence of effect-measure modification.[1,3] The SMR offers the advantage of a causal contrast between the experience of the exposed and their counterfactual experience if they had been unexposed. The SMR is also one of the components of the crude relative risk (together with the confounding risk ratio, i.e. the net effect of differences in the confounder distributions in exposed and unexposed)[22,23] and is closely linked to the concept of etiologic fraction[3,24]. The SMR uses as a standard the distribution of the effect-measure modifier in the exposed sub-population. Standardization to the distribution of the effect-measure modifier in the entire population (which results in a standardized effect measure, but not an SMR) would yield identical results to IPTW in this simple setting.[3,25] Reciprocally, Sato & Matsuyama proposed IPTW weights leading to a causal contrast based on the distribution of the effect-measure modifier in the exposed sub-population that would give the SMR.[19] They derived these weights algebraically from the SMR and propose using IPTW based on these "SMR" weights in marginal structural models.[19]

Given that in this setting the SMR and the PS matched analysis are identical, they both offer an interpretation as a causal contrast based on the counterfactual described above. The MH estimate and the ML estimate (and its multivariable counterpart), based on the assumption of uniform effects, do not represent a causal contrast. MH and SMR were developed before the

widespread availability and use of multivariable analysis techniques and have limitations when multiple confounders need to be controlled for. If the PS is estimated from several covariates (e.g. using multivariable logistic regression) rather than calculated non-parametrically, SMR and PS matched analyses will not be identical, as in this example, but should be similar. In those settings, the PS matched analysis or an IPTW analysis using "SMR" weights might be preferred to the SMR, since these analytic strategies can be seen as multivariable extensions of the SMR, allowing the inclusion of many covariates.

The interpretation of results based on stratification of the PS will depend on the method used to summarize stratum-specific estimates. A summary of stratum-specific estimates by the method of MH or ML assumes a uniform effect and thus does not provide the advantage of PS matching noted above. If stratum-specific estimates are summarized using the SMR, however, the result will again allow a causal contrast without the disadvantage of PS matching, which results in the loss of information from exclusion of unexposed individuals that are not selected as matches to exposed individuals.

Here we used IPTW in a setting with a single, non time-varying exposure and covariate assessment. IPTW is mostly used as the basis for marginal structural models (MSM) in longitudinal studies with repeated measures of the exposure and covariates. For repeated time-varying exposures in longitudinal studies, the analytic approaches presented are not useful alternatives to IPTW, insofar as MSMs allow one to obtain valid estimates of treatment effects in the presence of time-varying confounders that are affected by prior treatment.[8,9] G-estimation would be an alternative strategy in such a situation, but is more complex to implement.[26] Since the basic issue presented here remains, when applying MSMs it might be prudent to consider ways to implement IPTW leading to non-marginal risk ratios[19] and inspect individuals that get extreme weights to understand the extent to which their inclusion influences the results.

In our examples, we always found unexposed matches for all exposed individuals when matching on the PS, allowing a causal contrast. In reality, this will not always be the case. Not being able to include all exposed individuals will affect the results and furthermore will lead to a contrast that is not a causal contrast for the entire set of exposed persons. Matching avoids comparisons outside a range of the covariates common to both exposed and unexposed, thus avoiding a potential bias in multivariable outcome models due to inadequacies of the model.[27] Differences in the estimated PS within matched pairs, i.e. incomplete matching, might lead to residual confounding.[28]

We did not look at precision of the estimates, as this issue appeared secondary given the striking differences in point estimates. In less extreme situations and in situations without effect-measure modification, the choice of analytic technique will depend on statistical efficiency and investigator's knowledge. Since the PS can be estimated using logistic regression and IPTW is based on the PS, availability of software is not a problem, albeit getting correct standard errors for some of these estimates may involve resampling[8]. Also, we only present results based on risk ratio measures.[11,12] Modification of the spreadsheet and the parameter values to assess confounding and effect-measure modification of the risk difference is straightforward, however, and the patterns of similarities and differences are exactly the same.

We conclude that differences in the interpretation of causal contrasts in the presence of effect-measure modification need to be understood and kept in mind when interpreting results from these analytic strategies. In pharmacoepidemiology, the marginal risk ratio based on a popular implementation of IPTW, while valid under the built-in assumptions, might lead to misleading causal contrasts when summarizing the effects of treatment in the presence of a substantial proportion of individuals with heterogeneity of risk associated with treatment, e.g.

contraindications. Causal contrasts based on weighting according to the distribution of the effect modifier observed in the exposed rather than in the whole population, i.e. SMR, matching on the PS, stratification on the PS summarized by SMR, or implementations of IPTW providing non-marginal risk ratios, are easier to conceptualize and interpret in such settings.

Take home messages

- Effect-measure modification is likely in pharmacoepidemiologic studies owing to inclusion of individuals in whom treatment is not indicated or even contraindicated due to potential lack of response or risk for adverse outcomes

- In the presence of effect-measure modification, reporting stratum-specific results is preferred, but a summary estimate might still be valuable

- In contrast to standard analytic techniques, propensity scores (PS) and inverse probability of treatment weighting (IPTW) allow summary estimates that pertain to defined populations in the presence of effect-measure modification

- PS, standardized mortality ratio (SMR), and IPTW using SMR-weights estimate the effect of exposure in those who were exposed (target: those who are going to get the treatment) whereas IPTW using "marginal" weights estimates the effect of exposure if the whole population were to be exposed (target: general population)

- In populations with a substantial proportion of individuals in whom treatment is not indicated, estimating the effect of treatment in the exposed rather than the whole population is preferable

References

1. Rothman KJ, Greenland S, eds. *Modern Epidemiology.* 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998.

2. Miettinen OS. Standardization of risk ratios. Am J Epidemiol 1972;96:383–388. [PubMed: 4643670]

3. Greenland S. Interpretation and estimation of summary ratios under heterogeneity. Stat Med 1982;1:217–227. [PubMed: 7187095]

4. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.

5. Stürmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of applications of propensity score methods showed increased use but infrequently different estimates compared with other methods. J Clin Epidemiol (in press).

6. Hirano K, Imbens GW. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Services Outcomes Res Method 2001;2:259–278.

7. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statist Med 2004;23:2937–2960.

8. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–560. [PubMed: 10955408]

9. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561–570. [PubMed: 10955409]

10. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect . *Am J Epidemiol.* 2006; (in press online).

11. Rothman KJ. *Modern Epidemiology.* 1st ed. Boston/Toronto: Little, Brown & Company; 1986: page 193.

12. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol 1987;125:761–768. [PubMed: 3551588]

13. Greenland S, Maldonado G. The interpretation of multiplicative-model parameters as standardized parameters. Stat Med 1994;13:989–999. [PubMed: 8073203]

14. Free Software Foundation. The GNU General Public License, accessed June 8, 2005. (www.gnu.org/licenses/licenses.html).

15. Waldstein SR, Giggey PP, Thayer JF, Zonderman AB. Nonlinear relations of blood pressure to cognitive function. The Baltimore longitudinal study of aging. Hypertension 2005;45:374–379. [PubMed: 15699446]

16. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci 1999;14:29–46.

17. Stinger DE, Albers GW, Dalen JE, Go AS, Halperin JL, Manning WJ. Antithrombotic therapy in atrial fibrillation. The seventh ACCP conference on antithrombotic and thrombolytic therapy. Chest 2004;126(Suppl):429S–456S. [PubMed: 15383480]

18. Torn M, Bollen WLEM, van der Meer FJM, van der Wall EE, Rosendaal FR. Risk of oral anticoagulant therapy with increasing age. Arch Intern Med 2005;165:1527–1532. [PubMed: 16009869]

19. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. Epidemiology 2003;14:680–686. [PubMed: 14569183]

20. Heuschmann PU, Berger K, Misselwitz B, Hermanek P, Leffmann C, Adelmann M, Buecker-Nott HJ, Rother J, Neundoerfer B, Kolominsky-Rabas PL. German Stroke Registers Study Group; Competence Net Stroke. Frequency of thrombolytic therapy in patients with acute ischemic stroke and the risk of in-hospital mortality: the German Stroke Registers Study Group. Stroke 2003;34:1106–1113. [PubMed: 12663875]

21. Lindley RI, Wardlaw JM, Sandercock PA. Alteplase and ischaemic stroke: have new reviews of old data helped? Lancet Neurol 2005;4:249–53. [PubMed: 15778104]

22. Miettinen OS. Components of the crude risk ratio. Am J Epidemiol 1972;96:168–72. [PubMed: 4261806]

23. Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol 1981;114:593–603. [PubMed: 7304589]

24. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol 1974;99:325–332. [PubMed: 4825599]

25. Hernan MA, Robins JM. Estimating causal effects from epidemiologic data. *J Epidemiol Commun Health.* (in press)

26. Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology 1992;3:319–36. [PubMed: 1637895]

27. Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic Strategies to adjust confounding using Exposure Propensity Scores and Disease Risk Scores: Nonsteroidal Antiinflammatory Drugs (NSAID) and Short-Term Mortality in the Elderly. Am J Epidemiol 2005;161:891–898. [PubMed: 15840622]

28. Rosenbaum PR, Rubin DB. The bias due to incomplete matching. Biometrics 1985;41:103–116. [PubMed: 4005368]

Table 1

Summary of analytic strategies assessed and theoretical comparison with emphasis on non-uniform effects

Analytic strategy	Target population	Causal contrast	Uniformity assumption	Calculation	Multiple covariates
Mantel- Haenszel	None (approximates whole population)	No	Yes	Weighted average of stratum-specific estimates	Practicality limited by number of strata
Maximum likelihood	None (approximates whole population)	No	Yes	Usually iterative procedure	Modeling
Standardized morbidity ratio	Exposed (treated)	Yes	No	Weighted average of stratum-specific estimates	"SMR" weighted analysis[19]
Propensity score matching	Exposed (treated)	Yes	No	Matching unexposed to exposed on estimated probability of exposure	Yes
Propensity score stratification	Depending on method used to combine strata	If strata combined with SMR	MH, ML: yes SMR: no	Stratification on estimated probability of exposure	Yes
Inverse probability of treatment weighting	Whole population	Yes	No	Weighting by inverse of estimated probability of actual exposure (PS in exposed, (1-PS) in unexposed)	Yes
SMR- weighting	Exposed (treated)[19]	Yes	No	Weighting of exposed by 1, unexposed by PS/(1- PS)	Yes

Table 2

Comparison of analytic strategies using a numerical example with confounding without effect-measure modification

Data

X=0

	E=0	E=1	
Y=0	360	45	
Y=1	90	5	
	450	50	500

RR=0.5

X=1

	E=0	E=1	
Y=0	150	200	
Y=1	100	50	
	250	250	500

RR=0.5

Crude

	E=0	E=1	
Y=0	510	245	
Y=1	190	55	
	700	300	1000

RR=0.68

MH   **RR=0.50**
ML   **RR=0.50**
SMR   **RR=0.50**

w(X=0)=9   w(X=1)=50
O=55   E=110

Propensity Score Matching (expected)

pE(X=1)=0.5
PE(X=0)= 0.1

X=0

	E=0	E=1	
Y=0	40	45	
Y=1	10	5	
	50	50	100

X=1

	E=0	E=1	
Y=0	150	200	
Y=1	100	50	
	250	250	500

PS-matched

	E=0	E=1	
Y=0	190	245	
Y=1	110	55	
	300	300	600

RR=0.50

IPTW

w(E=0,X=0)= 0.7778   w(E=1,X=0)= 3
w(E=0,X=1)= 1.4   w(E=1,X=1)= 0.6

X=0

	E=0	E=1	
Y=0	280	135	
Y=1	70	15	
	350	150	500

X=1

	E=0	E=1	
Y=0	210	120	
Y=1	140	30	
	350	150	500

IPTW

	E=0	E=1	
Y=0	490	255	
Y=1	210	45	
	700	300	1000

RR=0.50

Parameters: $(IP_{E=e,X=x})$: $IP_{00} = 0.2$; $IP_{10} = 0.1$; $IP_{01} = 0.4$; $IP_{11} = 0.2$; $(P_{EX=x})$; $P_{E0} = 0.1$; $P_{E1} = 0.5$; $P_X = 0.5$;

N = 1000; RR: relative risk, MH: Mantel-Haenszel, ML: maximum likelihood, SMR: standardized

mortality ratio, IPTW: inverse probability of treatment weighting

Table 3

Comparison of analytic strategies using a numerical example without confounding and with effect-measure modification

Data

X=0
	E=0	E=1	
Y=0	360	25	
Y=1	90	25	
	450	50	500

RR=2.5

X=1
	E=0	E=1	
Y=0	200	225	
Y=1	50	25	
	250	250	500

RR=0.5

Crude
	E=0	E=1	
Y=0	560	250	
Y=1	140	50	
	700	300	1000

RR=0.83

MH   RR=1.03
ML   RR=1.09
SMR RR=0.83

$w(X=0)=9$ $w(X=1)=25$
O=50   E=60

Propensity Score Matching (expected)

$pE(X=1)=0.5$
$pE(X=0)=0.1$

X=1
	E=0	E=1	
Y=0	200	225	
Y=1	50	25	
	250	250	500

PS-matched
	E=0	E=1	
Y=0	240	250	
Y=1	60	50	
	300	300	600

RR=0.83

IPTW

$w(E=0,X=0)=0.7778$ $w(E=1,X=0)=3$ $w(E=0,X=1)=1.4$ $w(E=1,X=1)=0.6$

X=0
	E=0	E=1	
Y=0	280	75	
Y=1	70	75	
	350	150	500

X=1
	E=0	E=1	
Y=0	280	135	
Y=1	70	15	
	350	150	500

IPTW
	E=0	E=1	
Y=0	560	210	
Y=1	140	90	
	700	300	1000

RR=1.50

Parameters: ($IP_{E=e,X=x}$): $IP_{00} = 0.2$; $IP_{10} = 0.5$; $IP_{01} = 0.2$; $IP_{11} = 0.1$; ($P_{EX=x}$): $P_{E0} = 0.1$; $P_{E1} = 0.5$; $P_X = 0.5$;

N = 1000; RR: relative risk, MH: Mantel-Haenszel, ML: maximum likelihood, SMR: standardized

mortality ratio, IPTW: inverse probability of treatment weighting

Table 4

Comparison of analytic strategies using a numerical example with confounding and with effect-measure modification

Parameters: $(IP_{E=e,X=x})$: $IP_{00} = 0.1$; $IP_{10} = 0.5$; $IP_{01} = 0.2$; $IP_{11} = 0.1$; $(P_{E,X=x})$; $P_{E0} = 0.1$; $P_{E1} = 0.5$; $P_X = 0.5$;

N = 1000; RR: relative risk. MH: Mantel-Haenszel. ML: maximum likelihood. SMR: standardized

mortality ratio, IPTW: inverse probability of treatment weighting

Table 5

Summary of relative risks obtained by Mantel-Haenszel (MH), maximum likelihood (ML), standardized mortality ratio (SMR), propensity score matching (PS), and inverse of probability of treatment weighting (IPTW) when assessing the association between an exposure and an outcome according to the prevalence and the relation of a potential confounder and effect-measure modifier with exposure and outcome using numerical examples

Parameters								RR_{EY}					
P_X	P_{E1}	P_{E0}	RR_{EY1}	RR_{EY0}	RR_{X0}	Confounding	Effect-measure modification	Crude	MH	ML	SMR	PS	IPTW
0.1	0.5	0.1	0.5	0.5	2	Yes	No	0.64	0.50	0.50	0.50	0.50	0.50
0.5	0.5	0.1	0.5	0.5	2	Yes	No	0.68	0.50	0.50	0.50	0.50	0.50
0.9	0.5	0.1	0.5	0.5	2	Yes	No	0.54	0.50	0.50	0.50	0.50	0.50
0.1	0.5	0.1	0.5	2.5	1	No	Yes	1.79	2.03	2.25	1.79	1.79	2.30
0.5	0.5	0.1	0.5	2.5	1	No	Yes	0.83	1.03	1.09	0.83	0.83	1.50
0.9	0.5	0.1	0.5	2.5	1	No	Yes	0.54	0.58	0.56	0.54	0.54	0.70
0.1	0.5	0.1	0.5	5	2	Yes	Yes	3.38	3.28	4.17	2.63	2.63	4.18
0.5	0.5	0.1	0.5	5	2	Yes	Yes	1.23	1.19	1.26	0.91	0.91	2.00
0.9	0.5	0.1	0.5	5	2	Yes	Yes	0.59	0.59	0.57	0.55	0.55	0.74

Parameters: P_X = prevalence of covariate; P_{E1} and P_{E0} = prevalence of exposure in those with or without the covariate, respectively; RR_{XY0} = relative risk of the covariate on disease in those without the exposure, respectively; RR_{EY1} and RR_{EY0} = relative risk of exposure on disease in those with or without the covariate, respectively; RR_{EY1} and RR_{EY0} = relative risk of exposure on disease in those with or without the exposure

Table 6

Hypothetical example of application of analytic strategies in a study on the association between antithrombotic therapy and risk of stroke in a fictitious population of 100,000 elderly; antithrombotic therapy reduces the risk of (embolic) stroke in individuals with atrial fibrillation but increases the risk of (hemorrhagic) stroke in individuals without atrial fibrillation

Data

Atrial fibrillation

No				Yes				Crude		
Antithrombotic therapy				Antithrombotic therapy				Antithrombotic therapy		
Stroke	No	Yes		Stroke	No	Yes		Stroke	No	Yes
No	86,427	827		No	4,250	4,750		No	90,677	5,577
Yes	2,673	73		Yes	750	250		Yes	3,423	323
N	89,100	900	90,000	N	5,000	5,000	10,000	N	94,100	5,900	100,000
	RR=2.70				RR=0.33				**RR=1.50**	

MH	w(X=0)= 26.73	w(X=1)= 375	RR=0.49
ML			RR=0.44
SMR	O= 323	E= 777	RR=0.42

Propensity Score Matching (expected)

Atrial fibrillation

No				Yes				PS-matched		
pTherapy = 0.01				pTherapy = 0.5						
Antithrombotic therapy				Antithrombotic therapy				Antithrombotic therapy		
Stroke	No	Yes		Stroke	No	Yes		Stroke	No	Yes
No	873	827		No	4,250	4,750		No	5,123	5,577
Yes	27	73		Yes	750	250		Yes	777	323
N	900	900	1,800	N	5,000	5,000	10,000	N	5,900	5,900	11,800
									RR=0.42	

IPTW

Atrial fibrillation

No				Yes				IPTW		
Antithrombotic therapy				Antithrombotic therapy				Antithrombotic therapy		
Stroke	w=0.9505	w=5.9		Stroke	w=1.882	w=0.118		Stroke	No	Yes
No	82,149	4,879.3		No	7,998.5	560.5		No	90,147.8	5,439.8
Yes	2,540.7	430.7		Yes	1,411.5	29.5		Yes	3,952.2	460.2
	84,690	5,310	90,000		9,410	590	10,000		94,100	5,900	100,000
									RR=1.86	

Assumptions (parameters derived from references 14 and 15):

- Individuals without atrial fibrillation:

  o Risk of stroke (no antithrombotic therapy) = 3%

  o Relative risk of stroke associated with antithrombotic therapy = 2.70

  o Prevalence of antithrombotic therapy = 1%

- Individuals with atrial firbrillation:

  o Risk of stroke (no antithrombotic therapy) = 15%

  o Relative risk of stroke associated with antithrombotic therapy = 0.33

  o Prevalence of antithrombotic therapy = 50%

- Prevalence of atrial fibrillation = 10%

RR: relative risk, MH: Mantel-Haenszel, ML: maximum likelihood, SMR: standardized mortality

ratio, IPTW: inverse probability of treatment weighting