The principal strata effect for post-randomization selection bias: application to a prostate cancer prevention trial

Yasutaka Chiba, Ikuko Banba, Kenichi Azuma

Department of Environmental Medicine and Behavioral Science,
Kinki University School of Medicine, Japan

Corresponding Author:
Yasutaka Chiba
Department of Environmental Medicine and Behavioral Science,
Kinki University School of Medicine,
377-2, Ohno-higashi, Osakasayama, Osaka 589-8511, Japan
Tel: +81-72-366-0221
Fax: +81-72-368-1192
email: chibay@med.kindai.ac.jp

Summary

Objectives. An approach that deals with post-randomization selection bias is to consider the principal strata effect (PSE). In this paper, the concept of the PSE is introduced, aimed at researchers. The practicality of a sensitivity analysis method that has recently been developed is also investigated. Although a formula for the method had been presented previously, its practicability with real data has not been discussed for general cases.

Methods. The PSE is introduced using a simple hypothetical example. Whether the sensitivity analysis method is practical is examined by applying the process to an actual prostate cancer prevention trial.

Results. The sensitivity analysis could be readily conducted, and the results could be displayed graphically using Monte Carlo sensitivity analysis.

Conclusions. The sensitivity analysis requires prior distributions of sensitivity parameters. Thus, it will be useful for disease-specific experts, who have sufficient clinical knowledge, and statisticians, who can conduct the sensitivity analysis, to work closely together for the sensitivity analysis.

KEY WORDS: *bounds, causal inference, Monte Carlo sensitivity analysis, potential outcome, principal stratification, randomized trial.*

Introduction

A multi-center, double-blind, randomized prostate cancer prevention trial (PCPT) studied the effects of finasteride on the prevalence of prostate cancer in healthy men screened for 7 years (1). The 18,882 men, aged 55 years or older, with no history or current indicator of prostate cancer, were randomized to receive either 5 mg of finasteride per day or placebo. At the annual follow-up examinations, participants with a prostate-specific antigen (PSA) level exceeding a threshold of 3 ng/mL or with an abnormal digital rectal examination were referred for a prostate biopsy. Additionally, all participants not diagnosed with prostate cancer during the trial were instructed to undergo an end-of-study prostate biopsy at their seventh and final visit.

Of the 8,935 men whose status was identified through biopsy, prostate cancer was detected in 757 (17.5%) of 4,322 men in the finasteride arm, compared with 1,068 (23.2%) of 4,613 men in the placebo arm, suggesting that finasteride lowered the risk for prostate cancer (see Table 1). However, the

Table 1. Participants and their Gleason score in the prostate cancer prevention trial [1].

		Finasteride (A = 0)	Placebo (A = 1)	Total
Randomized		9,377	9,378	18,755
Cancer status known by biopsy (R = 1)		4,322	4,613	8,935
Prostate cancer (S = 1)		757	1,068	1,825
Gleason score (Y)	2	4	9	13
	3	1	8	9
	4	15	38	53
	5	69	118	187
	6	388	658	1,046
	7	190	184	374
	8	45	25	70
	9	36	24	60
	10	9	4	13

mean Gleason score for the individuals with prostate cancer in the finasteride arm was 6.40, compared with 6.06 in the placebo arm. This makes interpretation of the results challenging, because the trial suggests that finasteride reduced the overall risk for prostate cancer, but accelerated the growth of high-grade tumors (2).

Although the finasteride prostate cancer cases had a higher grade on the Gleason score than the placebo prostate cancer cases, this may not be an appropriate measure of the effect of finasteride on cancer severity. The men diagnosed with prostate cancer were a subset of the individuals initially randomized in the trial. As this subset was selected after randomization, there is the possibility of selection bias (3). If the characteristics of the men diagnosed with prostate cancer differed between treatment arms, the apparent effect of finasteride on cancer severity may be explained by the correlation between the differing characteristics and cancer grade, rather than a causal effect of finasteride.

A relevant population for addressing the effect of finasteride on cancer severity is the latent subgroup of participants who would have developed prostate cancer regardless of treatment. The principal stratification approach (4, 5) can be used to define this population and the effect under this population, i.e., the principal strata effect (PSE) or equally the survivor average causal effect (6). Several authors have discussed methods for assessing treatment effects on outcomes defined by a post-randomization event (4–10), but many of these techniques are difficult to implement in practice or require special statistical programming. Recently, a simple sensitivity analysis method was presented (10), but its practicability was examined with real data only under the monotonicity assumption (11); there were no participants who would have developed prostate cancer had they received finasteride, but would not had they received the placebo. Nobody has examined whether the method is practical in a case without this assumption.

The aim of this paper is to introduce the concept of the PSE to researchers other than statisticians, and to examine whether it is practical to use this simple method without the monotonicity assumption, by applying it to the PCPT data.

Methods

Notation and definitions

We use the following notation. A denotes the assigned treatment, where $A = 0$ if the participant was in the finasteride arm and $A = 1$ if in the placebo arm. R denotes the observed indicator that the status of prostate cancer was known; $R = 1$ if the participant had a biopsy and $R = 0$ if he did not. S denotes whether the biopsy detected prostate cancer; $S = 1$ if a participant had developed prostate cancer and $S = 0$ if not. S is missing when $R = 0$. Y denotes the Gleason score. Y is missing when $R = 0$ and is not defined when $S = 0$.

For each participant, it is also possible to consi-

der the potential outcomes (12), which correspond to the outcome of the participant if he had been in the other arm of the trial. S_1 and S_0 denote the prostate cancer status for each participant under $A = 1$ (and $R = 1$) and $A = 0$ (and $R = 1$), respectively. Also, Y_1 and Y_0 denote the outcomes for each participant under $A = 1$ (and $R = 1$) and $A = 0$ (and $R = 1$), respectively. The variables Y_1 and Y_0 are only defined if $S_1 = 1$ and $S_0 = 1$, respectively. Otherwise, the individual would not have developed prostate cancer, and the outcome Y would be undefined.

Crude and principal strata effects

To introduce the PSE, we will consider a hypothetical randomized trial for evaluating the effect of finasteride on Gleason scores. Note that this is different from the above PCPT. For simplicity, the outcome is dichotomized into high (Gleason score ≥ 7) and low (Gleason score ≤ 6), and we assume that all participants had a biopsy, i.e., $R = 1$ for all participants. Two thousand participants were randomized to finasteride and placebo arms (see Table 2). Of the 1,000 participants assigned to the finasteride arm, 400 participants developed prostate cancer, and 100 of these had high Gleason scores. For the placebo arm, 500 of the 1,000 participants developed prostate cancer, and 125 of these had high Gleason scores. A crude comparison of the proportion of participants with high Gleason scores between the two treatment arms is:

$E[Y \mid A = 0, R = 1, S = 1] - E[Y \mid A = 1, R = 1, S = 1] = 100/400 - 125/500 = 0.$

As noted in the introduction, this would not be a fair comparison, because those who developed prostate cancer without treatment may be a healthier group overall than those who developed prostate cancer despite treatment.

To make a fair comparison, a principal stratification approach (4, 5) could be used. This approach considers four types of participants that define four principal strata. (i) Always-developers: individuals who would develop prostate cancer regardless of the assigned treatment arm, i.e., $S_1 = S_0 = 1$. (ii) Never-developers: individuals who would not develop prostate cancer regardless of the assigned treatment arm, i.e., $S_1 = S_0 = 0$. (iii) Compliers: individuals who would develop prostate cancer if assigned to the finasteride arm, but would not if assigned to the placebo arm, i.e., $S_1 = 1$ and $S_0 = 0$. (iv) Defiers: individuals who would not develop prostate cancer if assigned to the finasteride arm, but would develop it if assigned to the placebo arm, i.e., $S_1 = 0$ and $S_0 = 1$. In the above hypothetical example, the number of these four principal strata might be as described in Table 3.

Comparisons of Gleason scores for each of these principal strata are fair, because the comparisons are made between two different treatment arms for the same populations. Of these four principal strata, we can make a comparison of Gleason scores only in the subpopulation of always-developers, because, for participants within the other principal strata, no developer exists and their Gleason scores cannot be defined in either the finasteride or placebo arm. This comparison between always-developers with $S_1 = S_0 = 1$ is the PSE:

$$PSE = E[Y_0 - Y_1 \mid S_1 = S_0 = 1].$$

Again, this comparison is fair, because it is made between the two different treatment arms for the same population. The data in Table 3 shows that the PSE estimate is:

$$PSE = 75/300 - 105/300 = -0.10.$$

Unfortunately, based on the observed data, we cannot know which participants are always-developers.

Table 2. Data from a hypothetical randomized trial.

	Finasteride (A = 0)	Placebo (A = 1)
Assignment	1,000	1,000
Prostate cancer (S = 1)	400	500
High Gleason score (Y = 1)	100	125

Sensitivity analysis formula

When treatment A is randomized, it is possible to conduct a sensitivity analysis using the following formula (10):

$$PSE = E_0 - E_1 - \frac{\pi_{01}}{p_0}\beta_0 + \frac{p_1 - p_0 + \pi_{01}}{p_1}\beta_1, \qquad [1]$$

where $E_a = E[Y \mid A = a, R = 1, S = 1]$ and $p_a = Pr(S = 1 \mid A = a, R = 1)$, and where β_0, β_1, and π_{01} are sensitivity parameters:

$\beta_0 = E[Y_0 \mid S_1 = 0, S_0 = 1] - E[Y_0 \mid S_1 = S_0 = 1]$,

$\beta_1 = E[Y_1 \mid S_1 = 1, S_0 = 0] - E[Y_1 \mid S_1 = S_0 = 1]$,

Table 3. Data from a hypothetical randomized trial under the principal stratification.

	Treatment group (A = 0)				Placebo group (A = 1)			
	Always	Never	Complier	Defier	Always	Never	Complier	Defier
Assignment	300	400	200	100	300	400	200	100
Prostate cancer (S = 1)	300	0	0	100	300	0	200	0
High Gleason score (Y = 1)	75	Undefined	Undefined	25	105	Undefined	20	Undefined

and $\pi_{jk} = \Pr(S_1 = j, S_0 = k)$. The derivation can be found elsewhere (10). The formula shows that it is possible to use the crude differences in Y, between the two different treatment arms between those who developed prostate cancer, $E_0 - E_1$. p_a can be estimated from the data. It is possible to conduct a sensitivity analysis by setting some plausible values of β_0, β_1, and π_{01}.

β_0 and β_1 are the differences in the outcome that would have been observed under different treatment conditions comparing two populations. β_0 contrasts the average outcomes under the placebo arm between defiers and always-developers. β_1 contrasts the average outcomes under the finasteride arm between compliers and always-developers. $\pi_{01} = \Pr(S_1 = 0, S_0 = 1)$ is the proportion of defiers. Note that $\pi_{01} = 0$ under the monotonicity assumption.

Formula [1] can be applied with plausible values of β_0, β_1, and π_{01} under the assumption that the missing-data mechanism for the biopsy was "missing completely at random" (13). However, this assumption may be unlikely for the PCPT data, because one of the criteria for interim biopsies was a PSA level exceeding 4.0 ng/mL or an abnormal digital rectal examination (14). However, finasteride reduces the PSA level by half, because it shrinks the volume of the prostate. Thus, the actual PSA criterion for referral to biopsy in the finasteride arm was a PSA value multiplied by a factor of approximately 2.0 that exceeded 4.0 ng/mL (1). Then, the missing-data mechanism is likely to be "missing at random" (13). To take the missing biopsy data into account, the following formula will be used rather than formula [1]:

$$\mathrm{PSE} = E_0 - E_1 - \frac{\pi_{01}}{p_0}\beta_0 + \frac{p_1 - p_0 + \pi_{01}}{p_1}\beta_1 + \theta, \quad [2]$$

where θ is a sensitivity parameter indicating bias due to the missing biopsy data.

Monte Carlo sensitivity analysis

There are several approaches that can be used in a sensitivity analysis. The simplest is to obtain plausible ranges of values for sensitivity parameters, and select several of the values within that range for analysis. This approach may be useful if the number of sensitivity parameters is one, because it is easy to display how the PSE estimate changes as a sensitivity parameter changes. However, it is difficult to conduct this analysis when there are multiple parameters, as in formulas [1] and [2]. Thus, a more sophisticated approach is employed: the Monte Carlo sensitivity analysis (MCSA) (15–19). In this approach, investigators assume the prior distributions of sensitivity parameters, and generate a large number (L) of PSE estimates by drawing L sets of random values from their distributions. Then, frequency distributions are generated using the L PSE estimates. To account for random errors in the PSE estimates, L sets of random values for $E_0 - E_1$ and p_a are drawn using the mean and variance estimated from the observed data. An illustration of an application of the MCSA to the PCPT data is described in the next section.

Results

Prior distributions of sensitivity parameters

As we did not have information about β_0, β_1 and π_{01}, the possible ranges of these values were assumed to be uniform distributions. π_{01} is the probability of defiers, and the principal stratum with $S_1 = 0$ and $S_0 = 1$ might be the smallest: $\min\{\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}\} = \pi_{01}$. This assumption derives the range of $0 \le \pi_{01} \le \min\{p_0/2, (1 - p_1)/2\}$. The derivation is found elsewhere (7, 10).

To determine the ranges of β_0 and β_1, "large sample bounds" were used, as presented by Zhang and Rubin (7). When $\min\{p_0/2, (1 - p_1)/2\} = p_0/2$, using

the bounds, the ranges of β_0 and β_1 are determined as:

$$2(E_0 - B_{0U}) \le \beta_0 \le 2(E_0 - B_{0L}), \quad [3]$$

$$\frac{2p_1}{2p_1 - p_0}(E_1 - B_{1U}) \le \beta_1 \le \frac{2p_1}{2p_1 - p_0}(E_1 - B_{1L}), \quad [4]$$

where B_{0L} and B_{0U} are the lower and upper large sample bounds on $E[Y_0 \mid S_1 = S_0 = 1]$, respectively, and B_{1L} and B_{1U} are those on $E[Y_1 \mid S_1 = S_0 = 1]$, respectively. The derivations of the large sample bounds and inequalities [3] and [4] are given in the Appendix. Furthermore, it is considered that, of participants assigned to placebo arm, Gleason scores for those who would not develop prostate cancer had they received finasteride ($S_1 = 1$ and $S_0 = 0$) might be lower overall than those who would develop prostate cancer even if they had received finasteride ($S_1 = S_0 = 1$). This observation indicates that $E[Y_1 \mid S_1 = 0, S_0 = 1] \le E[Y_1 \mid S_1 = S_0 = 1]$, i.e., $\beta_1 \le 0$. Thus, we improve the range of β_1 as:

$$\frac{2p_1}{2p_1 - p_0}(E_1 - B_{1U}) \le \beta_1 \le 0.$$

Note that the large sample bounds on the PSE are $B_{0L} - B_{1U} \le PSE \le B_{0U} - B_{1L}$.

For the prior distribution of θ indicating the missing biopsy data, the inverse-probability-weighting approach (20) was used. Although we did not have the data for the covariates, a quasi-population representing a number of participants with the same status as participant i ($1 / \Pr(R = 1 \mid A = a_i)$) was developed. The different random values among participants were generated following the binomial distribution with the probability of $\Pr(R = 1 \mid A = a)$. For each set of random values, θ was calculated by the difference between the crude estimate under this quasi-population and $E_0 - E_1$.

Sensitivity analysis

The observed data yielded the crude estimate of $E_0 - E_1 = 0.34$ (95% confidence interval: 0.24, 0.44; $p < 0.0001$), where standard error = 0.049. p_a and $\Pr(R = 1 \mid A = a)$ were computed as $p_0 = 757/4322 = 0.18$, $p_1 = 1068/4613 = 0.23$, $\Pr(R = 1 \mid A = 0) = 4322/9377 = 0.46$, and $\Pr(R = 1 \mid A = 1) = 4613/9378 = 0.49$. The above assumptions yield the ranges of $0 \le \pi_{01} \le 0.09$, $-1.43 \le \beta_0 \le 1.43$, and $-1.18 \le \beta_1 \le 0$.

From these ranges, 100,000 sets of random values were generated (the normal distribution for $E_0 - E_1$, binomial distributions for p_a and $\Pr(R = 1 \mid A = a)$, and uniform distributions for π_{01}, β_0,

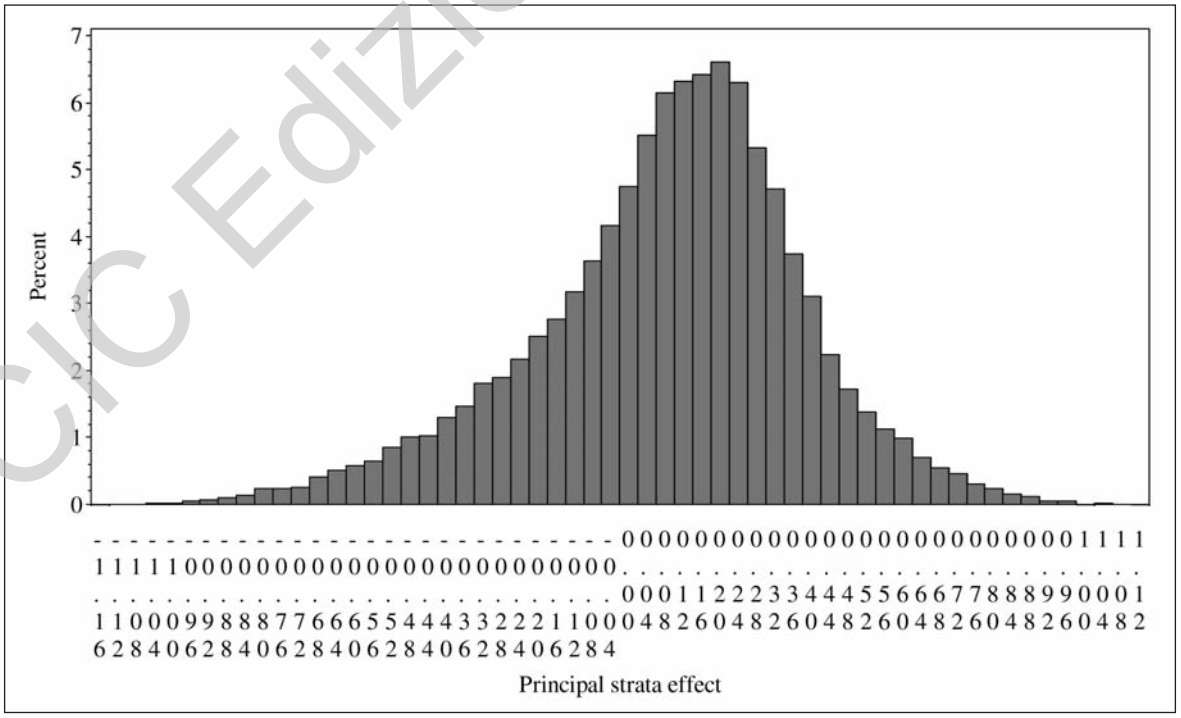

Figure 1. The result of the Monte-Carlo sensitivity analysis of the principal strata effect for the PCPT data.

and β_1) with frequency distributions for 100,000 PSEs. The results are shown in Figure 1. The 50th percentiles of the resulting PSE distribution were 0.12 (2.5th percentile: –0.59, 97.5th percentile: 0.63). Based on the findings of the distributions in this analysis, the difference in Gleason scores between the two treatment arms shifted to 0 but it was not significant, although the crude estimate yielded a converse effect of finasteride on cancer severity. Note that the large sample bounds yielded bounds on the PSE of $-1.11 \leq \text{PSE} \leq 1.75$.

Discussion

In this paper, we have introduced the concept of PSE for researchers other than statisticians, and have examined a simple sensitivity analysis for the PSE in a case without the monotonicity assumption (i.e., that no defier exists). Although the formula for sensitivity analysis has been presented previously (10), it has not been examined in such a case. In this paper, through the PCPT data, we show that the method can be used readily and that the result can be graphically displayed by MCSA.

We note that Egleston et al. (9) introduced a further method for the PSE in the clinical trial literature, but they also assumed monotonicity. We also note that the PCPT data have been analyzed by Shepherd et al. (8), however, they dichotomized the outcome into high (Gleason score ≥ 7) and low (Gleason score ≤ 6). Furthermore, their method is difficult to implement in practice and requires special statistical programming.

When the MCSA technique is applied for a sensitivity analysis, the result is affected by the prior distributions of sensitivity parameters. Thus, it is important to obtain plausible information about the distributions. If investigators can assume plausible prior distributions with single peaks, the resulting median PSE will yield a reliable PSE estimate. If investigators have information to develop plausible ranges for sensitivity parameters, the resulting 2.5th and 97.5th percentiles will be significant. It is then possible for investigators to conclude that the PSE has a positive effect, for example, if the 2.5th percentile is larger than 0, and a preventive effect if the 97.5th percentile is smaller than 0.

In many situations, it may be difficult for experts on a specific disease to conduct a sensitivity analysis, and statisticians, who can conduct one, may not have sufficient clinical knowledge. The sensitivity analysis introduced here requires prior distributions of sensitivity parameters. Thus, it will be useful for disease-specific experts and statisticians to work closely together for the sensitivity analysis.

Appendix

The large sample bounds are derived as follows. As shown in Table 3, participants who developed prostate cancer in the finasteride arm of the trial are either always-developers or defiers. Then, the proportion of always-developers in the finasteride arm is $\pi_{11} / (\pi_{11} + \pi_{01}) = (p_0 - \pi_{01}) / p_0 \geq (p_0 - p_0/2) / p_0 = 1/2$ under $\min\{\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}\} = \pi_{01}$ and $\min\{p_0/2, (1 - p_1)/2\} = p_0/2$. Thus, of 757 participants who developed prostate cancer in the finasteride arm, half (379) must be always-developers. When it is assumed that they have the largest values of Y, the mean is B_{0U}, and when it is assumed that they have the smallest values of Y, the mean is B_{0L}. Likewise, because participants who developed prostate cancer in the placebo arm are either always-developers or compliers, the proportion of always-developers in the placebo arm is $\pi_{11} / (\pi_{11} + \pi_{10}) = (p_0 - \pi_{01}) / p_1 \geq (p_0 - p_0/2) / p_1 = p_0 / (2p_1)$. Thus, at least $1068 \times p_0 / (2p_1) \leq 404$ participants in placebo arm must be always-developers. When it is assumed that they have the largest values of Y, the mean is B_{1U}, and when it is assumed that they have the smallest values of Y, the mean is B_{1L}. See Zhang and Rubin (7) for details.

Using β_0 and β_1, $E[Y_a \mid S_1 = S_0 = 1]$ is expressed as follows (10):

$$E[Y_0 \mid S_1 = S_0 = 1] = E[Y_0 \mid A = 0, R = 1, S = 1] - \frac{\pi_{01}}{p_0}\beta_0,$$

$$E[Y_1 \mid S_1 = S_0 = 1] = E[Y_1 \mid A = 1, R = 1, S = 1] - \frac{p_1 - p_0 + \pi_{01}}{p_1}\beta_1.$$

Thus, substituting the large sample bounds into these equations yield

$$E_0 - \mathrm{B}_{0U} \le \frac{\pi_{01}}{p_0}\beta_0 \le E_0 - \mathrm{B}_{0L},$$

$$E_1 - \mathrm{B}_{1U} \le \frac{p_1 - p_0 + \pi_{01}}{p_1}\beta_1 \le E_1 - \mathrm{B}_{1L}.$$

Because these inequalities hold for all values within the range of $0 \le \pi_{01} \le p_0/2$, inequalities [3] and [4] are obtained by substituting $\pi_{01} = p_0/2$ into these inequalities.

References

1. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. New England Journal of Medicine 2003; 349: 215–224.
2. Scardino PT. The prevention of prostate cancer – the dilemma continuous. New England Journal of Medicine 2003; 349: 295–297.
3. Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. Journal of the Royal Statistical Society, Series A 1984; 147: 656–666.
4. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics 2002; 58: 21–29.
5. Rubin DB. Direct and indirect causal effects via potential outcomes. Scandinavian Journal of Statistics 2004; 31: 161–170.
6. Hayden D, Pauler DK, Schoenfeld D. An estimator for treatment comparisons amongst survivors in randomized trials. Biometrics 2005; 61: 305–310.
7. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death." Journal of Educational and Behavioral Statistics 2003; 28: 353–368.
8. Shepherd BE, Redman MW, Ankerst DP. Does finasteride affect the severity of prostate cancer? A causal severity analysis. Journal of the American Statistical Association 2008; 103: 1392–1404.
9. Egleston BL, Cropsey KL, Lazev AB, Heckman CJ. A tutorial on principal stratification-based sensitivity analysis: application to smoking cessation studies. Clinical Trials, 2010; 7: 286–298.
10. Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. American Journal of Epidemiology (in press).
11. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American Statistical Association 1996; 91: 444–455.
12. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 1974; 66: 688–701.
13. Little RJA, Rubin DB. Statistical Analysis with Missing Data. New York, Wiley, 2002.
14. Baker SG. Analyzing a randomized cancer prevention trial with a missing binary outcome, an auxiliary variable, and all-or-non compliance. Journal of the American Statistical Association 2000; 95: 43–50.
15. Greenland S. The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. Journal of the American Statistical Association 2003; 98: 47–54.
16. Steenland K, Greenland S. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. American Journal of Epidemiology 2004; 160: 384–392.
17. Greenland S. Multiple-bias modeling for analysis of observational data. Journal of the Royal Statistical Society, Series A 2005; 168: 267–291.
18. Chiba Y. Bias analysis of the instrumental variable estimator as an estimator of the average causal effect. Contemporary Clinical Trials 2010; 31: 12–17.
19. Chiba Y. Bias analysis for the principal stratum direct effect in the presence of confounded intermediate variables. Journal of Biometrics and Biostatistics 2010; 1: 101.
20. Robins JM, Rotnitzky A, Zhang LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association 1995; 90: 106–121.