REGULAR ARTICLE

Steady state trials: another valid substitution of counterfactual ideal to measure causal effects

Okujou Iwami · Masayuki Ikeda

Abstract

Objectives Many traditionally established medical interventions are not examined with randomized trials especially in emergency medicine. We researched what is the scientific basis of the measurement of the causal effect in these interventions and proposed another trial to measure causal effects.

Methods We deduced steady state trials from the counterfactual model and used Bayesian approaches to estimate causal effects statistically.

Results When the state of the observed person is fairly steady before an exposure, the ratio of the after-period to the before-period of the exposure is sufficiently small, and changes are obtained in relatively short time, it is possible to postulate that the state of the counterfactual person to be compared is almost equal to the state of the real person before the exposure. Bayesian approaches show that the causal effect of the exposure is estimated even in only one-person steady state trials, when large changes are observed.

Conclusions Steady state trials are valid methods to measure causal effects and can measure causal effects even in one-person trials. When we can measure the causal effect of interventions with steady state trials, these interventions should be regarded as scientific without use of randomized trials.

Keywords Cross-over trials · Counterfactual model · Steady state · Period ratio · Individual causal effect

Introduction

Evidence-based medicine (EBM) appeared as a handy tool kit for clinicians who had not understood the basic thinking of epidemiology [1]. After the advocates of EBM succeeded in nominating randomized trials to be paramount [2], the so-called "Hierarchy of Strength of Evidence" towered in medical practice and many clinical guidelines prostrated themselves in front of the pyramid [3, 4]. Many traditionally established medical interventions were stripped of their rank for reasons having to do with observational studies. Under these circumstances, Smith and Pell [5] asked a sarcastic question why protagonists of EBM did not participate in a randomized trial of parachute use.

In epidemiological studies, the counterfactual or potential-outcome model has become increasingly standard for causal inference [6–8]. However, the theoretical ideal to measure causal effects of exposure is impossible. To achieve a valid substitution for the counterfactual experience, we resort to various design methods that promote comparability. One approach is a cross-over study and another is a randomized trial. Other approaches might involve choosing unexposed study subjects who have the same or similar risk-factor profiles for disease as the exposed subjects [9]. Case-crossover design was introduced for estimating a short term, transient effect of intermittent exposures on acute-onset diseases [10, 11]. For
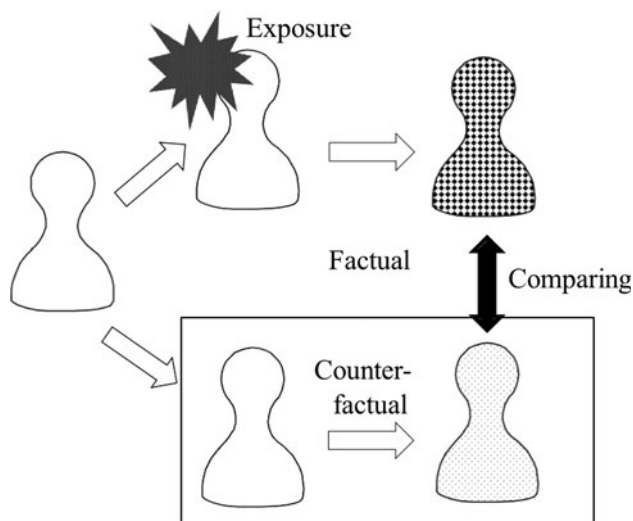
O. Iwami (✉)
Iwami Neurological Clinic, 1-5-2 Ohdohri,
Miyako 027-0083, Japan
e-mail: iwamineurol@mx5.et.tiki.ne.jp

M. Ikeda
Kyoto Industrial Health Association,
67 Nishinokyo-Kitatsuboicho, Nakagyo-ku,
Kyoto 604-8492, Japan

Fig. 1 Counterfactual model. We establish a hypothetical person in the counterfactual world in order to compare the outcome of the exposed person with the outcome of the unexposed person. After the exposure, both the conditions of the exposed person and the unexposed person are observed at the same time. As the only difference between the two settings is the exposure, it is possible to measure the effect of the exposure

each case, one or more predisease or postdisease time periods are selected as matched control periods for the case. The exposure status of the case at the time of the disease onset is compared with the distribution of exposure status for the same person in the control periods. The key feature of the case-crossover design is that each case serves as its own control. In this paper, we expand this key feature and propose another valid substitution of the counterfactual ideal to measure causal effects and show that parachute use and many interventions in emergency medicine have the scientific basis of the causal inference without randomized trials.

Materials and methods

We deduce steady state trials from the counterfactual model. The scheme is presented in Fig. 1. Bayesian methods are used to estimate causal effects statistically [12, 13] (see appendix). Posterior distributions are computed with WinBUGS version 1.4.3, which reports two-sided equi-tail-area credible intervals [14]. We use these intervals for convenience, although highest posterior density intervals are more preferable.

Results

Steady state trials

For the purpose of discussion, letters are defined as follows;

t	time
T_0	the time when the observation starts
T_1	the time when the exposure is done
T_2	the time when the outcome is observed
$B = (T_1 - T_0)$	the period before the exposure
$A = (T_2 - T_1)$	the period after the exposure
n	the integer which gives the ratio of A to B, $A:B = 1:n$
S	the state of the observed person which is a function of time
X	the state S just before the time T_1
Y	the state S at the time T_2
Z	the state of the counterfactual ideal of the unexposed person which is a function of time
W	the state Z at the time T_2.

Steady state trials begin with the observation of the state of the object person (Fig. 2). Suppose the state is almost steady during the period B (Fig. 3). Namely, the derivative of the state with respect to time during the period B is
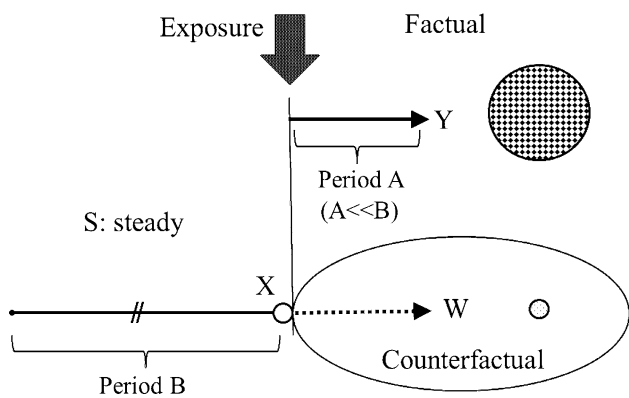
$$\frac{dS}{dt} = k + \delta,$$

where k is a constant and δ is noise which follows the normal distribution $N(0, \sigma^2)$. We observe the state S $(n + 1)$-times at the interval of the period A and obtain sample noises n-times ($\delta_i; i = 1, 2, \ldots, n$) during the period B. Just before the exposure, the state is recorded as X. When we observe Y at the end of the period A after the exposure, we get the mean value of $\frac{dS}{dt}$ during the period A:

$$\frac{dS}{dt} = (Y - X)/A.$$

When the ratio of the period A to the period B is sufficiently small, i.e., n is sufficiently large, we can postulate the derivative of the counterfactual unexposed state is k plus noise δz which follows the same distribution as δ_is:

$$\frac{dZ}{dt} = k + \delta z.$$

However, we cannot really observe $\frac{dZ}{dt}$, so the value of $(k + \delta z)$ is replaced with the observed value of $(k + \delta_{is})$. In order to estimate the difference between $(Y - X)/A$ and $(k + \delta_{is})$, we postulate that the distribution of $(Y - X)/A$ follows the normal distribution with the same variance as σ^2 which is estimated by the sample variance of $(k + \delta_{is})$. Then the difference between $(Y - X)/A$ and $(k + \delta_{is})$ can be statistically estimated with the t distribution. When the outcome Y has the quality different from the state X, the nominal scale is applied.

Fig. 2 Steady state trials. *S* the state, *X S* just before the exposure, *Y S* at the end of the period *A*, *W* the counterfactual state at the end of the period *A*. *S* is steady during the period *B*
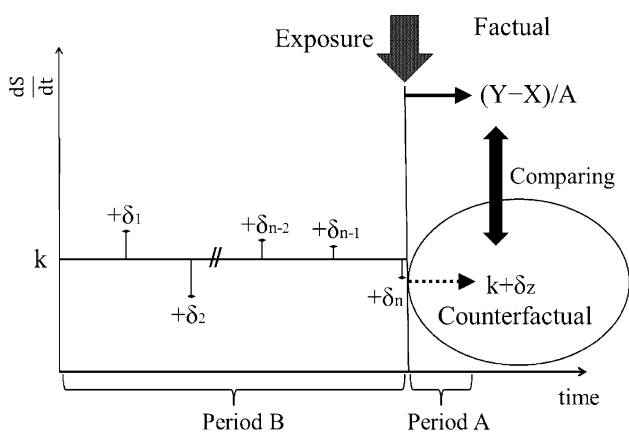
Fig. 3 Derivative of the state. *Period A:Period B* = 1:*n*, where *n* is a positive integer. $\frac{dS}{dt}$ is the derivative of *S* with respect to time; $\frac{dS}{dt} = k + \delta$, where *k* is constant and δ (δ_i; $i = 1, 2,..., n$) is noise which follows the normal distribution. We observe the state $(n + 1)$-times at the interval of the period *A* during the period *B*. The counterfactual derivative of the state is postulated as $(k + \delta z)$, where δz follows the same distribution as δ_{is}

Statistical inference of causal effects

Suppose when we observe the outcome *Y* which belongs to the category different from the state *X* and the change is of practical importance, or when the difference between $(Y - X)/A$ and $(k + \delta_{is})$ is statistically significant and large enough to be of practical importance. We now discuss the causation of the incidence of such an important outcome *Y*, which we designate Y_{imp} in the following discussion.

The probability that Y_{imp} happens during the period *A* with the exposure is represented by the letter θe:

$$\theta e = P(Y_{imp} \,|\, E, C),$$

where *E* is the exposure, *C* is the condition that the state is steady during the period *B* and the vertical line represents conditioning. We postulate that Y_{imp} is a Bernoulli variable

during the period *A*. The probability that Y_{imp} happens during the period *A* without the exposure ($\neg E$) is represented by the letter θu:

$$\theta u = P(Y_{imp} \,|\, \neg E, C).$$

As the state is steady and the period *A* is small relative to the period *B*, we can postulate that the counterfactual condition of the unexposed state in the period *A* is equivalent to the real condition in the period *B*, and the probability that Y_{imp} happens within the time span of the period *A* during the period *B* is equal to θu. Then the period *B* has a sequence of *n*-times repetitions of a trial with constant probability θu.

Suppose that we observe Y_{imp} after the exposure and there is no incidence of Y_{imp} during the period *B* in one steady state trial. The components of the Bayesian model for steady state trials can be written as follows:

$$\begin{aligned} \text{prior distribution} \quad \theta e &\sim \text{Beta}(\alpha_1, \beta_1) \\ \theta u &\sim \text{Beta}(\alpha_2, \beta_2) \\ \text{likelihood} \quad p(ye|\theta e) &= \theta e \\ p(yu|\theta u) &= (1 - \theta u)^n \\ \text{posterior distribution} \quad \Delta &= \theta e|ye - \theta u|yu, \end{aligned}$$

,where *ye* is the success of Y_{imp} in one trial under the exposure, *yu* is the no success of Y_{imp} in *n* trials under the non-exposure, and Δ is the difference between θe and θu taking account of the trial evidence. Ideally, the likelihood of no success under the non-exposure should be computed with *n*-times trials in the real world and one-time trial in the counterfactual world. The trial in the counterfactual world cannot be observed. Thus, we approximately compute this likelihood with *n*-times trials in the real world. The posterior distribution Δ is computed with WinBUGS.

When we are uncertain about the prior distribution, we adopt Beta(0.5, 0.5) or Beta(1, 1) as the reference prior distribution for θe and θu. The posterior distribution of θu shifts to zero as *n* becomes larger. With the reference prior Beta(0.5, 0.5) for θe and θu, the lower limit of the 95 % credible interval of Δ is over zero when *n* is equal to or more than four. With the reference prior Beta(1, 1) for θe and θu, the lower limit of the 95 % credible interval of Δ is over zero when *n* is equal to or more than seven. Classical statistical approaches also show similar results [15, 16]. The larger the number of *n* is, the more credibility we can gain in the inference of the causal effect. However, the lower limit of the 95 % credible interval of Δ cannot be over 0.15 with the prior Beta(0.5, 0.5) and not over 0.16 with the prior Beta(1, 1), no matter how *n* may become large. This is the limitation of one-person trials. Population studies with many persons can show larger lower limits of the credible interval, if the success proportion is high.

Steady state implies that the previous observations of the same condition showed no incidence of Y_{imp} without the exposure. When we believe the previous evidence for the no-incidence of Y_{imp} under the non-exposure, we can adopt, for example, Beta(1, 1000000) as the prior distribution for θu. Adopting almost null prior distribution Beta(1, 1000000) for θu means that $p(\Delta)$ is practically equal to $p(\theta e | ye)$.

Relation to cross-over trials

The simple cross-over design is outlined by Armitage et al. [17]. With two treatments, F and G, one randomly chosen group of patients receives treatments in the order FG, while the other group receives them in the order GF. The active response that is common to all subjects in a particular group and particular period with the treatment received is modeled as follows:

	Period 1	Period 2
Group I (FG)	$\mu + \tau_F + \pi_1$	$\mu + \tau_G + \pi_2 + \gamma_{FG}$
Group II (GF)	$\mu + \tau_G + \pi_1$	$\mu + \tau_F + \pi_2 + \gamma_{GF}$

Here, μ is a general mean, the τ terms represent treatment effects, the π terms represent period effects, and the γ terms represent the treatment × period interaction.

When F is no treatment, τ_F and γ_{FG} are null and the model of the group I is as follows:

	Period 1	Period 2
Group I (FG)	$\mu + \pi_1$	$\mu + \tau_G + \pi_2$

Suppose the ratio of the period 2 to the period 1 is 1:n. The constancy of $\frac{dS}{dt}$ means that the period effect π_1 is constant during the period 1. Under the condition of state steadiness which is confirmed by the $(n + 1)$ times observations during the period 1, when the period 2 follows successively the period 1 and n is sufficiently large, we can postulate that the π_2 is almost equal to $(\pi_1 + \pi_1/n)$ in the group I. The larger n is, the more we can believe the steadiness of the state and the approximation of π_2. Then the difference of the response between the two periods is $(\tau_G + \pi_1/n)$ and we can measure τ_G with the repeated observations of group I. Thus steady state trials are considered as variants of cross-over trials.

Prerequisite for steady state trials

How many figures should we adopt for n? In the above model, we postulate that $\frac{dZ}{dt}$ is equal to k, or π_2 is equal to $(\pi_1 + \pi_1/n)$. When n is infinitely large, this postulation is reasonable. However, when n is moderately large, the postulation receives criticism. There are many biological parameters which show cyclical or periodic variations, for example follicle-stimulating hormone or luteinizing hormone levels in female blood plasma. Another criticism is that the observed variable might reach the critical point after the steady state and change drastically without exposures. Before executing steady state trials in medicine, we have to examine biologically the trial condition for the possibility of cyclical or drastic state change. If some period ratio is thought to be critical, we have to avoid using such n for steady state trials.

Discussion

We have deduced steady state trials (SSTs) from the counterfactual model, from which randomized controlled trials (RCTs) were also deduced. Although RCTs are thought to be paramount trials in recent clinical research, STTs can also offer the valid method to measure causal effects, when the state before the exposure is steady and large changes are immediately observed. The smaller the ratio of the after-period to the before-period is, the better we can rely on the measurement of the causal effect. When the after-period is relatively long, the measurements of SSTs may be confounded and RCTs should be considered in such situations. RCTs are also necessary when outcomes long after the exposure are important, even if SSTs show causal effects immediately.

Individual causal effects are defined as a contrast of the counterfactual outcomes. Because only one of those values is observed, it has been proposed that individual causal effects cannot be identified in epidemiological research [18, 19]. The epidemiologic principle is that a person may be exposed to an agent and then develop disease without there being any causal connection between exposure and disease [9]. SSTs show that we can measure individual causal effects in the condition where repeated observations are performed, the state before the exposure is steady, and large changes are immediately obtained after the exposure. This approach could open the door to the individual causal inference and other conditions for the individual causation that should be researched in epidemiology.

One example of steady state trials is parachute use in skydiving [5]. At the height of 4000 m, we jump into the sky and we are falling at the terminal velocity of 55 m/s

after a few seconds. Within 3 s after opening parachutes, we usually fall at the next terminal velocity of 5 m/s. Now we record acceleration values at the interval of 3 s. Once we have the terminal velocity of 55 m/s, the acceleration value of 0 m/s² is observed about twenty times before opening parachutes and the deceleration value of 17 m/s² for 3 s is observed one time after opening parachutes. The Y_{imp} is the deceleration value of 17 m/s² for 3 s. After sampling one successful skydiving, the 95 % credible interval of the posterior distribution Δ with prior Beta(0.5, 0.5) is computed as 0.12–0.99. In 2010, 1308 members of the United States Parachute Association (USPA) reported skydiving injuries requiring medical attention [20]. During the same year, USPA members and first-time students made roughly 3 million jumps. These data may be translated into the following sample distribution.

Under the exposure   2998692 Y_{imps} out of 3000000 trials
Under the non-exposure   No Y_{imp} out of 3000000 × 20 trials

We adopt Beta(0.5, 0.5) as the prior distributions for θe and θu. However, the huge sample size fixes practically the same posterior distribution as when we believe the prior probability of θu is almost null. The 95 % credible interval of Δ is computed as 0.9995–0.9996 with WinBUGS. Classical statistics show that the 95 % confidence interval of the safe skydiving proportion is 0.99954–0.99959.

SSTs are practicable in the situation where immediate clinical responses are important, such as in the emergency room, where confounders are under the control of practitioners. Many treatments in emergency medicine have a long good history of SSTs in innumerable persons and can be regarded as scientific medical interventions without RCTs, such as intravenous injection of glucose for patients in hypoglycemic coma, injection of adrenalin (epinephrine) for patients with anaphylactic shock, a tourniquet for bleeding patients, and so on.

For example, one-person SST is presented in the use of adrenalin injection for an imaginary adult patient with anaphylactic shock. The data of the patient is shown in the Table 1. The systolic blood pressures (SBP) were recorded at the interval of 1 min. The patient had an intravenous injection of 0.1 mg adrenalin at time 10 min. We first check whether $\frac{dS}{dt}$ in the period B follows normal distribution. If there are any outliers, SSTs is not the choice for this trial. The data of the table can be considered as following normal distribution. The mean of $(k + \delta_{is})$ is one. The estimated variance of $(k + \delta_{is})$ is
$[(-1)^2 + (-2)^2 + 0^2 + 2^2 +$
$(-1)^2 + 0^2 + 1^2 + 0^2 + 0^2 + 1^2]/(10 - 1) = 1.33.$
The standard error of $[(Y - X)/A - (k + \delta_{is})]$ is
$\sqrt{[1.33(1/1 + 1/10)]} = 1.21.$
Two-sample t statistic is $(32 - 1)/1.21 = 25.6,$

which follows the t distribution on $(1 + 10 - 2) = 9$ degrees of freedom. The P value is computed as $<10^{-8}$. In this statistical estimation, we postulate that the distribution of $(Y - X)/A$ follows the normal distribution with the same variance as σ^2 which is estimated by the sample variance of $(k + \delta_{is})$. However, we do not really have data for the estimation of the variance of $(Y - X)/A$. We propose the above P value as an informal index to be used as a measure of discrepancy between $(Y - X)/A$ and $(k + \delta_{is})$. We recommend that this P value should be smaller than 0.001 to show discrepancy. SBP over 90 mmHg is a clinically important value in shock, and we can consider that $\frac{dS}{dt}$ of 32 mmHg/min for 1 min after the exposure is Y_{imp}. With prior Beta(0.5, 0.5) for θe and θu, the 95 % credible interval of the posterior distribution Δ is computed as 0.09–0.99. When we have the prior information that Y_{imp} have not been observed in the past unexposed states for 1000 min in total, we can adopt Beta(0.5, 1000.5) as the prior distribution for θu and n can be smaller to infer causal effects, if we confirm that unexposed state in the period B is in the same condition as the previously reported unexposed states. After observing one success of Y_{imp} after the exposure, with prior Beta(0.5, 0.5) for θe, prior Beta(0.5, 1000.5) for θu and $n = 1$, the 95 % credible interval of the posterior distribution Δ is computed as 0.15–1.00. Practically, n will be more than two in order to confirm the equal conditions, even when we adopt almost null prior for θu. When we execute SSTs in population studies, if some period ratio is thought to be critical, the participants are divided into two or three groups to which different values of n are allocated. RCTs between the groups are also possible in this setting.

Table 1 Systolic blood pressure in an imaginary adult patient with anaphylactic shock

Time (min)	S (mmHg)	dS/dt (mmHg/min)
0	50	
1	50	0
2	49	−1
3	50	1
4	53	3
5	53	0
6	54	1
7	56	2
8	57	1
9	58	1
10	60	2
Exposure		
11	92	32

Systolic blood pressure of the patient was recovering slowly without treatments. The patient had an intravenous injection of 0.1 mg adrenalin at time 10 min

S Systolic blood pressure

SSTs are also possible in treatment trials of neurodegenerative diseases whose patients almost always show progressive deteriorations. For example, although RCTs has not been performed in the levodopa therapy of Parkinson's disease [21], neurologists will admit that symptoms lasting for several months of early Parkinson's patients almost always improve within a few days after receiving levodopa.

In preventive medicine, oral rehydration therapy is effective against diarrhea [22]. RCTs have compared oral rehydration with intravenous hydration [23]. SSTs can offer the measurement of the effect difference between the treated and the nontreated in acute stage diarrhea. Because of the necessity of controlling confounders, SSTs may be restricted within a narrow set of research topics in preventive medicine. However, if we have sufficient past databases of disease incidence and the derivative of the incidence rate with respect to time is constant, we may use SSTs for the measurement of effects of exposures which have a latent period of several years.

Appendix

The essence of Bayes theorem is as follows. Suppose θ is some quantity that is currently unknown, and let $p(\theta)$ denote the prior distribution of θ. Suppose we have some observed evidence y, whose probability of occurrence is assumed to depend on θ. This dependence is formalized by $p(y|\theta)$, known as the likelihood. The posterior distribution taking account of the evidence y is denoted $p(\theta|y)$. Bayes theorem is

$$p(\theta|y) \propto p(y|\theta) \times p(\theta).$$

The beta distribution is a conjugate family for the binominal likelihood. This means that a beta prior distribution and a binominal sample distribution provide a beta posterior distribution by Bayes theorem.

Prior distribution	Beta(α, β)
Observed sample proportion	r successes out of n trials
Posterior distribution	Beta($\alpha + r$, $\beta + n - r$)

The mean of Beta(α, β) is $\frac{\alpha}{\alpha+\beta}$.

References

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268:2420–5.
2. Sackett DL. Chapter 6: the principles behind the tactics of performing therapeutic trials. In: Haynes RB, Sackett DL, Guyatt GH, Tugwell P, editors. Clinical epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2006. p. 173–243.
3. Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-based medicine. How to practice and teach it. 4th ed. London: Elsevier Churchill Livingstone; 2011.
4. Centre for evidence based medicine. Oxford centre for evidence-based medicine 2011 levels of evidence. http://www.cebm.net/mod_product/design/files/CEBM-Levels-of-Evidence-2.1.pdf.
5. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ. 2003;327:1459–61.
6. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health. 2000;21:121–45.
7. Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol. 2002;31:422–9.
8. Höfler M. Causal inference based on counterfactuals. BMC Med Res Methodol. 2005;5:28.
9. Rothman KJ. Chapter 3: measuring disease occurrence and causal effects. In: Rothman KJ, editor. Epidemiology: an introduction. New York: Oxford University Press; 2002. p. 24–56.
10. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133:144–53.
11. Maclure M, Mittleman MA. Should we use a case-crossover design? Annu Rev Public Health. 2000;21:193–221.
12. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester: Wiley; 2004.
13. Gelman A, Carlen JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. New York: Chapman & Hall/CRC; 2004.
14. Medical Research Council and Imperial College of Science, Technology and Medicine. WinBUGS PACKAGE. http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.
15. Miettinen O, Nurminen M. Comparative analysis of two rates. Stat Med. 1985;4:213–26.
16. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med. 1998;17:873–90.
17. Armitage P, Berry G, Matthews JNS. Cross-over trials in Chapter 18. In: Armitage P, Berry G, Matthews JNS, editors. Statistical methods in medical research. 4th ed. Oxford: Blackwell Science; 2002. p. 627–36.
18. Hernán MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health. 2004;58:265–71.
19. Suzuki E, Komatsu H, Yorifuji T, Yamamoto E, Doi H, Tsuda T. Causal inference in medicine part I—counterfactual models—an approach to clarifying discussions in research and applied public health. Nihon Eiseigaku Zasshi. 2009;64:786–95. (in Japanese).
20. United State Parachute Association. Skydiving safety. http://www.uspa.org/AboutSkydiving/SkydivingSafety.

21. Ad Hoc Committee on the Guidelines for the Treatment of Parkinson's Disease, Japanese Neurological Society. Guidelines for the Treatment of Parkinson's Disease. Clin Neurol. 2002;42:421–94. (in Japanese).

22. Victora CG, Bryce J, Fontaine O, Monasch R. Reducing deaths from diarrhea through oral rehydration therapy. Bull World Health Organ. 2000;78:1246–55.

23. Bellemare S, Hartling L, Wiebe N, Russell K, Craig WR, McConnell D, et al. Oral rehydration versus intravenous therapy for treating dehydration due to gastroenteritis in children: a meta-analysis of randomised controlled trials. BMC Med. 2004;2:11.