1

Crossover design – definitions, notes, and limitations

A crossover or change-over design is a trial in which each eligible patient after obtaining his/her informed consent is randomly assigned to receive more than one treatment according to one of the predetermined treatment-receipt sequences. By contrast, a parallel groups design is a trial in which each eligible patient after obtaining his/her informed consent is randomly assigned to receive exactly one of the treatments under comparison. Similar to parallel groups design, the main goal of a crossover design is to study the difference between individual treatments (rather than the difference between treatment sequences). Because each patient serves as his/her own control, the crossover design is a useful alternative to the parallel groups design for increasing power or saving the number of patients needed via elimination of response variation between patients when treatments are compared (Hills and Armitage, 1979; Fleiss, 1986a; Senn, 2002). A review of the history of the crossover design can be found in Jones and Kenward (2014, pp. 5–7).

The simplest crossover design is the AB/BA design, in which patients are randomly assigned to either the AB group, in which patients receive treatment A first and then cross over to receive treatment B, or the BA group, in which patients receive treatment B first and then cross over to receive treatment A. The AB/BA design is also called the simple crossover or 2×2 design (Jones and Kenward, 1989). Because of its simplicity, the application of AB/BA design has accounted for a large proportion of crossover trials used in practice (Hills and Armitage, 1979; Senn, 2002, 2006; Mills *et al.*, 2009). As noted in a survey of 12 large pharmaceutical companies in the United States (Fava and Patel, 1986), more than half of the 72 cross-over trials conducted

Crossover Designs: Testing, Estimation, and Sample Size, First Edition. Kung-Jong Lui.

Companion website: www.wiley.com/go/lui/crossover

over a five-year period by these companies used the AB/BA design (Jones and Kenward, 1989, p. 6). Because of the characteristics of the study design features, there are limitations in use of a crossover design.

1.1 Unsuitability for acute or most infectious diseases

The crossover design is inapplicable for studying acute or most infectious diseases that are likely to be either fatal or curable, such as food poisoning, the common cold, or stomach flu. In treating these diseases, there will be nothing left to treat when the patient is scheduled to receive the second or later treatments. Thus, the crossover design should generally be reserved for non-curable chronic diseases for which the symptoms can be relieved only for a short time period after treatments and will reoccur shortly after the applied treatment is removed. These chronic diseases may commonly include asthma, epilepsy, angina pectoris, migraine, hypertension, and so forth (Fleiss, 1986a; Senn, 2002). The crossover design is also popular in studying the effects of antianxiety drugs on human performance (McNair, 1971).

1.2 Inappropriateness for treatments with long-lasting effects

The crossover design is appropriate for treatments that have a rapid, short, and reversible effect (e.g., bronchodilators in asthma) (Senn, 2002). The crossover design is not adequate for studying treatments that have a long-lasting effect, such as steroids or vaccine for rabies or rubella. The carry-over effect – that is, the persistent effect of a treatment applied in one period on patient responses in a subsequent period of treatments – is one of the major concerns for the use of a crossover trial. If there are carry-over effects in a crossover trial, the effect of a treatment observed at one time point can be confounded with the residual effects due to earlier treatments. Thus, our comparison between treatments and assessment of the relative treatment effect can be biased if we cannot appropriately adjust the carry-over effects due to earlier treatments. As noted by Hills and Armitage (1979), it can be difficult or even impossible to disentangle the treatment effect from the residual effects of earlier treatments. Although we may find intensive research in the development of models or testing strategies to account for the carry-over effect (Balaam, 1968; Kershner and Federer, 1981; Laska, Meisner, and Kushner, 1983; Ebbutt, 1984; Laska and Meisner, 1985; Willan and Pater, 1986; Lehmacher, 1991; Jones and Donev, 1996), most of these models are assumed for mathematical interests and convenience rather than practical utility (Fleiss, 1986a, 1986b, 1989; Senn, 1992, 2002; Senn, D'Angelo, and Potvin, 2004). Senn (1992) and Fleiss (1989) contended that the best strategy to deal with the carry-over effect is to employ an adequate washout period to assure that patients are weaned off these residual effects from earlier treatments.

1.3 Loss of efficiency in the presence of carry-over effects

As mentioned, the main motivation of employing a crossover trial instead of a parallel groups design is to increase the power of test procedures or reduce the number of patients needed. Grizzle (1965) focused attention on the AB/BA design and proposed a random effects linear additive risk model. Grizzle (1965) showed that incorporating the carry-over effects into test procedures and estimators is equivalent to carrying out data analysis based on only the data at the first period as done in the parallel groups design. Thus, the gain of efficiency in use of an AB/BA crossover trial in the presence of carry-over effect under the random effects linear additive risk model will completely disappear. On the other hand, if we originally planned to carry out a parallel groups design, we would be able to take advantage of the resources that were used to take patient responses at period 2 in use of a crossover trial to recruit more patients for the parallel groups design. This implies that for a given fixed budget, we may even end up of losing the opportunity to detect what we are interested in if we employ an AB/BA crossover trial in the presence of carry-over effects. Thus, it is advisable that we put more efforts into prior consideration in the design stage of a crossover trial. If we cannot assure ourselves of elimination of the carry-over effect with an adequate washout period based on our subjective knowledge, we may not wish to consider using the crossover trial (Brown, 1980; Fleiss, 1986a, 1989; Senn and Hildebrand, 1991; Senn, 2002; Schouten and Kester, 2010).

1.4 Concerns of treatment-by-period interaction

The treatment-by-period interaction means that the effect of a treatment on patient responses varies between periods. The interaction between treatments and periods can result from (i) a treatment effect that may carry over from one period to the next period physically or psychologically; or (ii) a treatment effect that may vary according to the level of patient response (Hills and Armitage, 1979; Ebbutt, 1984; Jones and Kenward, 1989, p. 42). If there is an interaction between treatments and periods, as for studying the association between risk factors and diseases in epidemiology, a summary conclusion between treatments across periods can be misleading (Lui, 2004). Furthermore, interpretation of our findings about the treatment effect in the presence of treatment-by-period interaction can also be more difficult. Note that the carry-over effect is a special type of treatment-by-period interaction, except that for the carry-over effect, we cannot eliminate all other kinds of treatment-by-period interactions (Hills and Armitage, 1979; Jones and Kenward, 1989) by use of a washout period. When employing an AB/BA design, we often need to implicitly assume that the treatment-by-period interaction does not exist. The treatment-by-period interaction (or carry-over effect) is not really unique in a crossover trial and can exist in the parallel groups trial as well (Senn and Hildebrand, 1991; Senn, 2002).

1.5 Flaw of the commonly used two-stage test procedure

The carry-over effect and treatment-by-period interaction are not separately identifiable under an AB/BA design (Senn, 2002). Either of these effects can cause a bias in our inference of the treatment effect. Although we may employ an adequate washout period to attenuate the carry-over effects, we can never be certain that the washout period has worked. Grizzle (1965) proposed a two-stage test procedure as follows. We first test whether the carry-over effect exists. Because the test itself for the carry-over effect is subject to the response variation between patients, the power of this test is generally low and thereby a high nominal level of Type I error (10 or 15%) is usually chosen. If the test for the carry-over effect is nonsignificant, we will carry on analyses based on the difference between responses within patients as done for the crossover trial with assuming no carry-over effects. Otherwise, the test procedure using the data at the first period only is carried out as for the parallel groups design, and all data obtained at the second period are excluded from data analysis. Freeman (1989) carried out a thorough investigation of the two-stage test approach and concluded that this approach could be potentially misleading. This is because the statistic for testing the carry-over effect is highly correlated to that for testing equality based on the data at the first period. Thus, the test result based on the data at the first period is likely to be significant as well when the test result for the carry-over effect is significant. This leads to the actual Type I error for the two-stage test being higher than the nominal α-level. Due to this concern, we do not recommend use of the two-stage test procedure in practice (Senn, 1988, 1991, 1997, 2000, 2002, 2004, 2006). Other notes in use of the two-stage test in a crossover trial can be found elsewhere (Jones and Kenward, 2003, pp. 44–49; Cleophas, 1991; Freeman, 1991; Senn, 1996).

1.6 Higher risk of dropping out or being lost to follow-up

Because each patient is to receive more than one treatment, the duration of a crossover trial is expected to be longer than that of a parallel groups design. The longer the time length of a trial, the more difficult is to obtain patients' consent to participate in a trial and the higher is the risk for patients dropping out or being lost to follow-up after consent (Senn and Lee, 2004). Extra effort is also needed, as compared with a parallel groups design, to ensure that patients follow the study protocol closely for a crossover trial with a long duration. All the above concerns can be exacerbated if it is necessary for all patients to receive even more than two treatments. Thus, the crossover trial should generally be reserved for trials in which the number of treatments under comparison is not numerous to reduce the risk of dropping out in participating patients, while the acute effect of treatments can be quickly measured.

1.7 More assumptions needed in use of a crossover design

In a crossover trial, responses taken from the same patients are likely correlated and the period effect on patient responses often exists. Furthermore, in an AB/BA trial the carry-over effect, the group effect, and the treatment-by-period interaction are all aliased and not separately estimatible (Jones and Kenward, 1989, 2003; Senn, 2002). Thus, we need to make more model assumptions to account for the above factors, and statistical analyses for a crossover design are generally more complicated than those for a parallel groups design. Although many subtle and sophisticated models have been developed recently for a crossover design, the need to make more model assumptions for the crossover design than that for the parallel groups trial is undesirable. This is because our inferences can be vulnerable to more model assumptions, which are often difficult to be completely justified in practice. To alleviate this concern, it is logically appealing that we should put more effort into designing a trial to simplify the underlying assumed model structure instead of constructing complicated models and developing sophisticated analyses.

1.8 General principle and conditional approach used in the book

To avoid the possible loss of substantial efficiency in adjusting carry-over effects and the bias of our estimators derived under models unable to appropriately account for the complicated structure of carry-over effects (Senn, D'Angelo, and Potvin, 2004), we follow Fleiss (1986a, 1989), Senn (1992, 2002) and Schouten and Kester (2010) and recommend use of the study design to deal with the carry-over effect. We focus our discussions on situations in which the carry-over effect is ignorable by assuming that practitioners will use their best knowledge to cautiously apply an adequate washout period. In case we should unexpectedly encounter carry-over effects (which are found and supported by internal evidence rather than by statistical tests exclusively), however, we discuss hypothesis tests and estimation related to the carry-over effect, as well as assess the relative treatment effect in the presence of carry-over effect at the end of some chapters.

To account for the intraclass correlation between responses taken within patients, we assume that the effects due to individual patients are random rather than fixed (Gart, 1969). Except for a section (discussing the test for carry-over effects) in Chapter 2 (for continuous data), we do not assume the random effects due to patients to follow any specified parametric distribution (such as a normal distribution as commonly done in random effects models). To eliminate these random effects, we adopt the conditional approach rather than the likelihood-based approach. The former may be less efficient than the latter if the random effects of patients do follow the assumed normal distribution. However, it is difficult or impossible to justify any specified distribution for these random effects in practice. From the practical point of view, using a

simple valid method that is easy to understand and calculate, as well as requires fewer assumptions, can be more useful and desirable than using a possibly more efficient approach derived under more restricted situations. Furthermore, using the conditional arguments can easily lead to deriving the exact tests and exact interval estimators. These exact methods and results will be of use for investigators who have the concerns of employing asymptotic test procedures and interval estimators when the size of their trials is actually small.

To consider a period or simple trend effect, we focus discussions on models including the period effect throughout the book. When there is a period effect, test procedures (or estimators) without accounting for the period effect can lose accuracy (or be biased). Because one of the main reasons for using more than one treatment-receipt sequence in a crossover design is to allow for a possible period effect (Jones and Kenward, 2014, p. 99), an analysis with assuming no period effect seems to internally contradict the design itself. Note that a period effect may also be called an "order effect" (Gart, 1969; Prescott, 1981).

The scope of the book is modest. We concentrate our attention under the most commonly encountered basic crossover designs on various statistical aspects, including tests of non-equality, non-inferiority and equivalence, interval estimation, sample size determination, estimation and adjustment of unexpected carry-over effects, as well as tests of the period effect and treatment-by-period interaction. We consider the AB/BA design for continuous data (Chapter 2), dichotomous data (Chapter 3), ordinal data (Chapter 4), and frequency data (Chapter 5). To reduce the number of patients receiving an inert placebo in a random clinical trial, it can sometimes be appealing to compare more than one experimental treatment with a placebo in a single trial instead of separate trials, each having its own an experimental and a placebo arms. Thus, we extend the results for the AB/BA design to accommodate a three-treatment three-period crossover design for continuous data (Chapter 6), dichotomous data (Chapter 7), ordinal data (Chapter 8), and frequency data (Chapter 9). The extension of methods and ideas presented here to accommodate the cases for more than three-treatment three-period crossover trials (Senn, 2002) is simply straightforward. To decrease the number of groups with different treatment-receipt sequences in a crossover trial, we include some discussions on the use of Latin squares in exercises when comparing three or more treatments. To reduce the lengthy duration of a crossover trial comparing more than two treatments, we also include a chapter (Chapter 10) discussing hypothesis testing and estimation under an incomplete block crossover design for both continuous and dichotomous data. Using similar ideas, we can easily extend these results to accommodate the ordinal and frequency data under the incomplete block crossover design. For other specific topics, such as high-order designs, fixed effects models, the search for optimal designs, detailed discussions of carry-over effects, and the Bayesian approach, we refer readers to excellent books (Jones and Kenward, 1989, 2003, 2014; Senn, 2002), all of which were truly inspiring during the preparation of this book.