Chapter 6
Methodological Considerations for N-of-1 Trials

Keumhee C. Carriere, Yin Li, Geoffrey Mitchell, and Hugh Senior

Abstract N-of-1 trials are extremely useful in subject-focused investigations, for example, medical experiments. As far as we are aware, no guidelines are available in the literature on how to plan such a trial optimally. In this chapter, we discuss the considerations when choosing a particular N-of-1 trial design. We assume that the outcome of interest is measured on a continuous scale. Our discussion will be limited to comparisons of two treatments, without implying that the designs constructed can apply to non-continuous or binary outcomes. We construct optimal N-of-1 trials under various models depending upon how we accommodate the carryover effects and the error structures for the repeated measurements. Overall, we conclude that alternating between AB and BA pairs in subsequent cycles will result in practically optimal N-of-1 trials for a single patient, under all the models we considered without the need to guess at the correlation structure or conduct a pilot study. Alternating between AB and BA pairs in a single trial is nearly robust to misspecification of the error structure of the repeated measurements.

Keywords N-of-1 trials • Optimal design • Clinical trials • Crossover design • Residual effects • Direct treatment effect • Error structure • Adaptive trial design

Introduction

Medicine is an ever-changing science. Clinical trials are employed to conduct biomedical studies on human subjects to obtain specific answers about the impact of drug therapy and related medical interventions or treatments, generating efficacy data. The majority of such studies employ randomized controlled trials (RCTs)

K.C. Carriere (✉) • Y. Li
Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Canada
e-mail: kccarrie@ualberta.ca; yin8@ualberta.ca

G. Mitchell
School of Medicine, The University of Queensland, Ipswich, QLD, Australia
e-mail: g.mitchell@uq.edu.au

H. Senior
School of Medicine, The University of Queensland, Brisbane, Australia
e-mail: h.senior@uq.edu.au

J. Nikles, G. Mitchell (eds.), *The Essential Guide to N-of-1 Trials in Health*,
DOI 10.1007/978-94-017-7200-6_6

(Armitage 1975; Kenword and Jones 1987; Wei and Durham 1978). One approach to designing an RCT is the use of optimal experimental designs. An optimal design is a technique designed to assist a decision maker in identifying a preferable choice among many possible alternatives. Among the many RCT designs available, the most useful and popular design is the crossover design. For example, in a survey done in 1980 of numerous studies on the effects of antianxiety drugs on human performance, 68 % of the studies used the crossover approach (Brown 1980). It is still certainly one of the most popular approaches being adopted in many epidemiologic and pharmaceutical trials (Figueiras et al. 2005).

To illustrate the logistics of choosing a particular design, we first note that there are a number of excellent articles on optimal designs in the RCT literature. See for example Cheng and Wu 1980; Carriere 1994; Carriere and Huang 2000; Liang and Carriere 2009; Laska and Meisner 1985; (Afsarinejed and Hedayat 2002; Kunert and Stufken 2002, 2008). However, most of these designs, if not all, focus on optimizing the treatment effect for an average patient. The average patient is a construct – a virtual person who responds to the intervention by the mean of the population's responses. Individuals enrolled in a trial will respond better or worse than, or simply differently from the average patient. The available optimal designs are not adequate when studying individual-based treatment effects is desired.

Multi-crossover single-patient trials, known as N-of-1 trials, are often employed when the focus is to make the best possible treatment decision for an individual patient. From a clinician's perspective, having clear evidence of the value of one treatment over another (or no treatment) is far more useful than knowing the average response. The average response gives the clinician the *probability* that a treatment will be effective, whereas N-of-1 trials give far more certainty about whether the treatment for the patient sitting in front of them will work or not.

The simplest two-treatment N-of-1 trial uses the AB (or BA) sequence for treatments A and B; this treatment sequence has one crossover pair over two treatment periods. Each period is chosen to be of sufficient length for the treatments being tested to show an effect. Two periods (such as AB or BA) constitute a single cycle in a N-of-1 trial. As the patient becomes his or her own control, N-of-1 trials provide individual-based clinical evidence for the treatment effect, free of between-patient variations. With the rising cost of patient care, N-of-1 trials have the potential to be extremely useful, as they can minimize clinic visits and time on suboptimal treatments (Greenfield et al. 2007; Guyatt et al. 1986; Kravitz et al. 2004; Larson 1990; Nikles et al. 2005). To obtain stable estimates of the treatment effect, we desire to replicate such evidence for each patient. The question is then how many such cycles are desirable and what is the optimal order of treatment administration. Quite naturally, designing an N-of-1 trial involves deciding on the number of cycles and proper sequencing of treatments in order to plan the study optimally to achieve the trial objectives. The literature is lacking in providing these guidelines for constructing optimal N-of-1 trials. A recent book on N-of-1 trials also leaves the choice of an ideal design to the clinician, while suggesting various possible designs to consider (Kravitz et al. 2014).

The Number of Cycles and Sequences

For two-treatment experiments, a general N-of-1 trial can have multiple AB or BA crossover pairs called a cycle, in a sequence of treatments for within-patient comparisons. Hence, the number of periods, p, is a multiple of 2 and it is desirable to have more than two of these pairs or cycles for a stable estimate of a treatment effect. Typically, possible sequences to consider rapidly increase with increasing p in multi-period designs. For example, N-of-1 trials with three cycles and therefore $p = 6$ would require considering $2^6 = 64$ treatment sequences before we could determine the optimal treatment sequence (s). Since it is only feasible to use a small set of cycles, we aim to determine the optimal sequences, while ensuring that each pair of periods consists of two distinct treatments. Therefore, such sequences as AAABBA or AAAABA are unlikely to be used in N-of-1 trials as AA are not two distinct treatments. In addition, we may consider block effects due to AB or BA, treating these as independent entities, which makes those sequences unsuitable to consider. Eliminating unsuitable sequences among 2^p for N-of-1 trials, we are left with $2^{p/2}$ distinct sequences to consider for p-period 2-treatment N-of-1 trials for $p/2$ cycles. For example, a study with 6 periods (3 cycles) will be left with 8 distinct sequences to consider for a 6-period 2-treatment N-of-1 trial. In this example, an individual patient would be randomized to only 1 of the 8 possible sequences of treatment order.

There are many optimal repeated measurement designs available in the literature (Cheng and Wu 1980; Hedayat and Yang 2003). However, due to the special conditions mentioned above, these N-of-1 trial designs cannot be optimally derived from the existing designs. It is known that the two-treatment design (*AB, AA*) and their duals (*BA, BB*) is found to be universally optimal for two-period experiments, with the duality defined as the sequence that switches *A* and *B* with the same effect. Similarly, it is known that the two-sequence design *ABB* and its dual *BAA* and the four-sequence design (*ABBA, AABB*) and their duals (*BAAB, BBAA*) are optimal for three- and four-period experiments, respectively (Carriere 1994; Laska and Meisner 1985).

Straight application of this two-treatment optimal design literature with *A* to AB and *B* to BA would suggest that optimal N-of-1 trials can use the 4 sequence design with ABBA, ABAB and their duals for two within-patient comparisons, the 2 sequence design with ABBABA and its dual for three within-patient comparisons, and the 4 sequence design with ABBABAAB, ABABBABA and their duals for four within-patient comparisons. It is not yet known whether all of these sequences are indeed optimally equivalent so that each sequence is optimal for each individual patient for 4, 6 and 8-period N-of-1 trials. Further, applying the results from the literature would require at least two patients to utilize these existing designs, as the optimal design uses at least two sequences and is unsuitable for N-of-1 trials. In this Chapter, we show that not all sequences in these repeated measurement designs are optimal for N-of-1 trials for estimating individual-based treatment effects.

Ideally, when aggregated, the series of N-of-1 trials that are optimal for individual patients can also provide an optimal estimate of the treatment effects for the average patient. For example, in a multi-clinic setting in three AB pair six-period N-of-1 studies, all eight possible sequences ($2^{6/2} = 8$) have been used, i.e., ABABAB, ABABBA, ABBAAB, ABBABA and their duals to estimate both individual- based and average treatment effects (Guyatt et al. 1990). However, it is not known whether each of these eight sequences is optimal for individual patients. Further, it is not known whether a collection of the optimal and not-so-optimal N-of-1 trials will lead to optimal designs for estimating the average treatment effects. In the next Sections, we discuss how these do not lead to optimal aggregated N-of-1 trials for estimating the treatment effects for the average patient. We first discuss issues arising due to the repeated nature of these experiments.

Residual Treatment Effect

The main attraction of crossover designs is that the subject provides their own control, as measurements are taken repeatedly from the same subject using different treatments. If the treatment effects lasting beyond the given period are equal, they can provide efficient within-subject estimators of direct short-term treatment effects by removing between-subject variations.

However, there is one critical issue plaguing these repeated measurement designs and preventing them being popular despite their practical appeal. It is because they suffer from a long-standing controversy regarding residual treatment effects that last beyond the given period. N-of-1 trials are no exception. Sometimes referred to as the carryover effect, the residual effect is the effect of a previous treatment that carries over into the subsequent treatment periods. Thus, the effect of a treatment in a given period can be carried over to influence the responses in a subsequent period. Often, the residual effect of a treatment may be ignorable after two periods. However, the residual effect between the responses over two consecutive treatment periods may not be assumed to be negligible. A "washout" period placed between treatment periods could reduce the carryover effects, but a long washout period may increase the risk of drop-outs. Also, there is no guarantee that it completely removes the residual effects. Therefore, careful planning is important, as the nature of the carryover effect may be such that the N-of-1 trial method is not feasible (Bose and Mukherjee 2003; Carriere and Reinsel 1992; Kunert and Stufken 2002).

Nevertheless, the presence of residual effects does not invalidate the use of crossover designs. Rather it is the inequality of the residual effects of each treatment that may be causing the controversies. If the residual effects are equal for the treatments, then this has an effect in a statistical sense as if the residual effects do not exist, because they cancel out mathematically.

Despite the concern over residual effects, ethicists apparently have less of a problem with self-controlled designs such as crossover designs than with completely randomized or parallel-group designs (Carriere 1994). For example, when a

trial involves treatments for patients with life-threatening conditions, it is almost beneficial to adopt these self-controlled designs. In parallel-group placebo-controlled studies, a group of sick patients is randomly allocated to a placebo or active group, with a 50 % chance of receiving a placebo or ineffective treatment for one or more periods of time, making them unethical for obvious reasons.

The way this is overcome in researching critical conditions depends on which of the following scenarios the researcher is addressing; either

1. The test treatment is given in addition to a proven or acceptable treatment in order to avoid a patient having to receive placebo alone, or
2. If the best known treatment has been shown to fail, only then considering a placebo-controlled design.

In the first scenario, the trial is: best treatment plus test intervention, compared with best treatment plus placebo. The second scenario is: test treatment vs placebo in a constrained population. Ethically either is acceptable. However, both alternatives raise interesting practical questions. The first is that if the proven treatment is effective, then the test treatment can only offer incremental effects. These are likely to be smaller, and therefore a larger sample size will be required to detect a clinically significant difference. Further, this design will not give information about the efficacy of the treatment on its own in the condition under consideration. In the alternative case of testing in the presence of failed proven treatments for a subset of patients, the potential participant pool will be far smaller, which may cause the proposed trial to be impractical to perform. Of course, test treatment vs placebo trials are ethically acceptable in critical conditions where there is currently no proven effective treatment.

In the literature, various models have been proposed to accommodate carryover effects. We will introduce and discuss the two most popular models in the next Section: (A) a model with a first-order residual effect and (B) a model with self and mixed carryover effects.

Models for N-of-1 Trials to Account for Carryover Effects

The most widely read statistical paper on the use of crossover experiments in clinical trials was published in 1965 by Grizzle, where the responses are modeled as:

Response = overall mean
+ period effects
+ sequence effects
+ direct treatment effects
+ residual treatment effects
+ measurement error.

Aside from the obvious overall mean effect and period effects, sequence effects may be present due to treatments given in a different order to patients, because some patients will be given AB or BA or some other order. While the primary objective is to study the

direct treatment effects, their effects may not be unique due to the unequal residual treatment effects. Crossover design models have typically assumed that the treatments assigned to subjects have lasting effects on their responses to treatments in subsequent periods. A two-step approach has been used quite extensively where the unequal residual effects are first estimated and tested for significance before proceeding to estimate the direct treatment effects (Carriere 1994; Kunert and Stufken 2008).

Model with a First-Order Residual Effect

When it is assumed that the carryover effects last for only one period, this is known as a first-order residual effect model. In such a model, no interaction is assumed between the treatment administered during the current period and the carryover effects from the previous period. This interaction gives rise to the second-order residual effect. Hence, the model under this assumption is basically equal to the Eq. (1), where the term for the residual treatment effects is just the first-order residual effects, which last for only one more period of the treatment administration.

Model with Self and Mixed Carryover Effect

Taking the treatment and period interactions into account, Kunert (Kunert and Stufken 2002, 2008) presented an alternate model with self and mixed carryover effects. The self carry-over effect occurs when the treatment administered in the current and the previous period is the same; alternatively, if two different treatments are administered between the periods, it is known as a mixed carryover effect. The model under this assumption is more elaborate. From the Eq. (1), the term for the residual treatment effects will be split and be replaced with the following two terms:

- +Self carryover effects (if a preceding treatment is the same as the current one)
- +Mixed carryover effects (if the preceding treatment is not the same as the current one).

Optimal designs are highly model dependent. These effects are assumed to exist unless proven insignificant and therefore a reasonable effort should be made to separate them for an unbiased estimation of the direct treatment effects. Sometimes, however, it is simply practically impossible to accommodate all effects in the model. With N-of-1 trials, not all of these effects can be included in the model. Because we are dealing with just one subject and p responses in total, the period effects cannot be accommodated.

Repeated responses from a subject can be correlated and also involve measurement errors. The most popular structures may be to consider the pair of measurements as being equally correlated. Such a simple structure may work in designs with small numbers of periods. In N-of-1 trials with at least 4 periods, we may need to consider an auto-regressive structure, as the correlations may diminish gradually as the pair of measurements comes farther apart in their treatment periods. Here, the

correlation is assumed large for a pair of measurements from two adjacent periods and decreases as the time between the two period increases. Some modification is possible by assuming them to be uncorrelated if they are more than two periods apart. See related discussion in Carriere (1994).

Design Parameters for N-of-1 Trials

In this section, we characterize N-of-1 trials in a small number of design parameters. Within each crossover pair AB and BA, the two treatments are to be different but, for two consecutive crossover pairs, the treatments assigned to the second period in the previous pair and the first period in the latter pair can be different or the same.

If an AB pair is followed by a BA pair, as in ABBA (or its dual, BAAB), we refer to the design as having an alternating pair in the sequence. Therefore, the sequence ABAB does not have an alternate cycle. The performance of an N-of-1 trial is related to how the pairs AB and BA alternate. By denoting s as the number of AA and BB and m as the number of AB and BA in a treatment sequence, we define $h = s - m$. There are $p - 1$ subsequences with a length of 2 in a p-period sequence. For example, for N-of-1 design with $p = 8$, ABABBABA, there are seven subsequences AB, BA, AB, BB, BA, AB, and BA. Here, $s = 1$ and $m = 6$, and $h = -5$. Table 6.1 shows the relationship between these design parameters for each of the eight possible eight-period sequences.

In the next section, we will see that sequences with the same h values have the same design properties. For instance, all three sequences ABABBAAB, ABBAABAB, ABBABAAB and their duals have $h = -3$, and contain the same amount of information about parameters of interest for this design. Hence, if this is the optimum value, the 8-period N-of-1 trial can use any one of these three sequences and their duals. Here, we immediately see from Table 6.1 that not all of the 8 sequence N-of-1 trials used elsewhere are optimal N-of-1 trials. Therefore, it is important to deliberate what the optimal N-of-1 trials are by examining how these cycles and pairs are organized.

Table 6.1 Sequences for $p = 8$ with corresponding design parameter values

h	Sequence	Alternation	s	m
−7	ABABABAB	0	0	7
−5	ABABABBA			
−5	ABABBABA	1	1	6
−5	ABBABABA			
−3	ABABBAAB			
−3	ABBAABAB	2	2	5
−3	ABBABAAB			
−1	ABBAABBA	3	3	4

Note: s is the number of AA and BB and m is the number of AB and BA in a treatment sequence with $h = s\text{-}m$

Optimal N-of-1 Trials

Typically, we are interested in estimating the direct and carryover treatment effects, while all others are treated as nuisance and secondary parameters. We build designs to this end. Due to the special nature of the design and the correlated data, we first consider how one could approach the data analysis. When data are obtained from an N-of-1 design, there are various ways to approach data analyses in order to gather any beneficial treatment evidence for the patient.

One is to observe effects between successive trial conditions, as shown in Fig. 6.1. Once the series of differences are observed, it would involve the usual repeated measures data analytic technique to plan and analyze them (Cantoni 2004; Davidian et al. 2009; Liang and Zeger 1986).

Alternately, one can observe and analyze treatment differences from each pair, as shown in Fig. 6.2. The same strategy of a longitudinal and repeated measures data analysis method as the first approach applies here, as the N-of-1 trial data were replaced by differences from successive pairs. The paired differences are simply regarded as repeated measurements, and more data simply means a better precision for the treatment effect (Carriere and Huang 2000). Such repeated measurements can be analyzed using various parametric and nonparametric methods (Liang and Zeger 1986; Senn 2002).

The third way to look at these N-of-1 trials data is similar to the first approach, but differs in that it holds the judgment or decision of a beneficial treatment effect for the given patient till the end of the trial, by using likelihood methods (Kravitz et al. 2014). As it analyzes the entire data set based on the employed model, the Type I error probability is minimized while possibly making multiple interim analyses and decisions. This is another consideration, discussed in this chapter in the development of optimal N-of-1 trial design. Such a model-based general approach could be more efficient than the first two approaches analyzing each cycle separately within each patient (Carriere 1994; Kravitz et al. 2014).

In this chapter, we recognize that N-of-1 designs deal with small samples, and thus we discuss an optimal strategy to collect the data based on the model by constructing an optimal design rather than specific data analysis strategies. To do so, we first need to define the information matrix under a particular model, which will contain information about all parameters of interest under the assumed model. More details

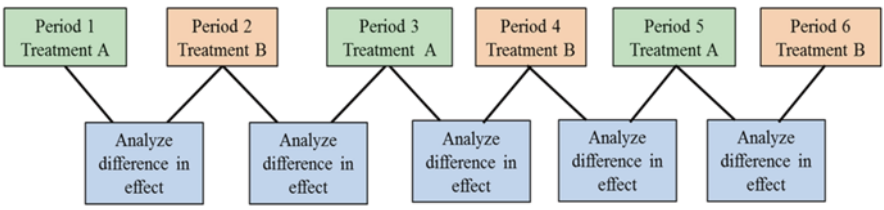

Fig. 6.1 Analysis of successive trial conditions in a six period N-of-1-design

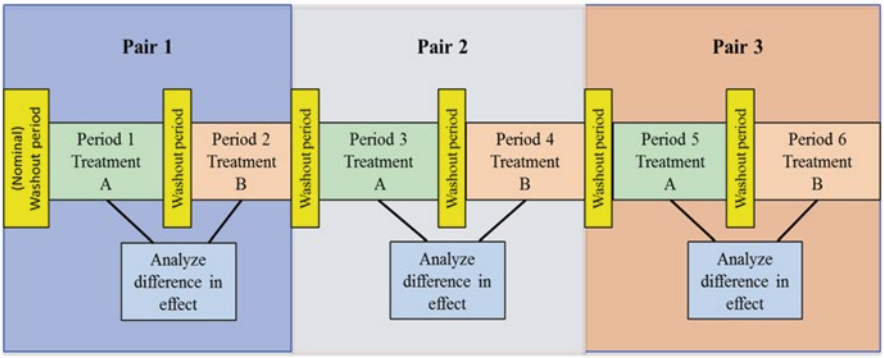

Fig. 6.2 Analysis of results from each successive pair in a six period of N-of-1 design

about an information matrix are found in much of the design literature, for example, Carriere and Reinsel (1993). Then, the optimal design for a set of parameters of interest is constructed using various optimality criterion. We use the D-optimality, which finds the optimal design by maximizing the determinant of the information matrix. See also Cheng and Wu (1980) and Kiefer (1975). Carriere (1994) and Carriere and Huang (2000) describe a practical approach to find optimal designs.

For N-of-1 trials, the optimal sequence is completely determined by h, as noted in the previous section and therefore, is much simpler to construct than previously. One could also find the optimal design that simultaneously optimizes all parameters of interest or the one that optimizes the carryover effects under the constraint that it optimizes the direct treatment effects. We note that the approach can also apply to find optimal designs for estimating some linear combinations of the parameters of interest. See also Carriere and Reinsel (1993). Since we are primarily interested in the optimal estimation of the direct treatment effects, we do not consider these cases. We summarize important results based on a D-optimality criterion in the next subsections (Carriere and Reinsel 1992).

Under the Traditional Model

Under the traditional model, we consider the error structure to be equally correlated between two measurements within the patient. We find the optimal design to consist of pairs of AB and BA appearing alternatively throughout the trial. Therefore, we find the following result.

Result 1 The optimal N-of-1 trial for estimating both the direct and residual effects is the one sequence design that consists of pairs of AB and BA appearing alternately.

For example, the optimal designs for N-of-1 trials with 4, 6, and 8 periods are the one sequence designs, ABBA, ABBAAB, and ABBAABBA, respectively. One could switch A and B to obtain a dual sequence with the same effect.

Repeated measures may be correlated highly when they are close together in time, while they may be negligible when they occur at a distance in time. One possible model for such a situation is to consider auto-regressive errors, accounting for correlation of measurements between two adjacent periods to be stronger than those far apart. It turns out that to discuss designs with auto-regressive errors, we need to consider whether the treatment given in the first and last periods are the same or not. However, the optimal designs are still determined by the value of h, and in general, the optimal design is to alternate AB and BA cycles as above.

Under the Model with Self and Mixed Carryover Effects

We also considered the model with self and mixed carryover effects. The optimal design was constructed by obtaining information matrices for the relevant parameters. Unlike the previous case, the optimal designs for estimating the direct treatment effect are not the same as those for residual effects.

The optimal design for estimating the direct treatment effect is the sequence with only AB pairs, such as ABABABAB. Although it may be of less interest, the optimal sequence for estimating the self carryover effect is to alternate between AB and BA pairs, while the optimal sequence for estimating the mixed carryover effect is to repeat the AB pair with no alternation. We summarize our findings in Result 2.

Result 2 The optimal N-of-1 trial for estimating the direct treatment and mixed carryover effect is the sequence with only AB pairs with no alternation, such as ABABABAB, while the optimal N-of-1 trial for estimating the self carryover effect is the sequence with AB and BA alternating throughout the trial.

Numerical Comparison

Although we constructed the optimal designs, it would be of interest to determine the practical benefit of having followed specific guidelines in adopting the optimal clinical trial design. To appreciate the practical performance of the optimal N-of-1 trials we constructed, we compare the efficiencies of some selected designs in estimating the treatment and carryover effects under the two models. We limit the comparison to the cases with independent and equi-correlated errors.

Recall that the optimal N-of-1 trials are either to alternate between AB and BA pairs or simply to repeat AB pair in a sequence. Under the traditional model, the optimal N-of-1 trial uses ABBAAB and ABBAABBA for 6 and 8 period experiments, respectively. We refer to them as S63 and S83. Under the self and mixed effects model, the optimal N-of-1 trial is to use ABABAB and ABABABAB for 6 and 8 period experiments, respectively, which we refer to as S61 and S81. Some other sequences are also considered, as defined in Table 6.2.

Table 6.2 Numerical illustration of various 6- and 8-period N-of-1 trials

		Traditional		Self/mixed		
Sequence	*h*	var(t)	var(g)	var(t)	var(s)	var(m)
S61: ABABAB	−5	1.208	1.500	1.208	NE	1.500
S62: ABABBA	−3	0.242	0.300	1.250	3.000	1.500
S63: ABBAAB	−1	0.173	0.214	1.214	1.714	1.714
S64: ABBABA	−3	0.242	0.300	1.250	3.000	1.500
S81: ABABABAB	−7	1.146	1.333	1.146	NE	1.333
S82: ABABBABA	−5	0.229	0.267	1.167	2.667	1.333
S83: ABBAABBA	−1	0.127	0.148	1.150	1.600	1.400
S84: ABBABAAB	−3	0.150	0.174	1.147	1.647	1.412

Note: Entries are relative variances for a direct treatment effect *t*, a first-order carryover effect *g*, a self carryover effect *s*, and a mixed carryover effect *m,* under the traditional and self/mixed carryover effects models. S61-S64 are 6-period N-of-1 trials and S81-S84 are 8-period N-of-1 trials with the order of cycles as given

Table 6.2 tabulates the variances of the treatment effects for 8 period N-of-1 trials under the two models. Table 6.2 reveals that the effects of choosing a specific sequence under the self and mixed carryover effect model are rather minimal. Careful examination also reveals that the optimal sequence for all three (direct and residual carryover) treatment effects simultaneously is the same as that under the traditional model.

Specifically, Table 6.2 shows that the optimal individual-based N-of-1 trials are S63 and S81 for estimating the direct treatment effects under respective models, as expected. However, there are no real practical differences among various N-of-1 trials under the self and mixed model. Under the self and mixed carryover model, although S81 repeating AB pairs in each cycle is optimal for estimating the direct treatment effect, it does not allow estimation of self carryover effects, making S63 and S83 preferable. Therefore, for robust and optimal N-of-1 trials it is recommended to use a sequence alternating between AB and BA pairs, such as S63 and S83, under all models.

In summary, it appears that there is no discernable or sizable advantage to distinguish among the two models and possibly various error structures. The above comparison remains true under both independent and equi-correlated error structures.

Overall, S63 and S83 for single N-of-1 trials seem to be the best under both models. They are optimal for estimating direct treatment and mixed carryover effects. Further, they are optimal for estimating both the treatment and carryover effects under the traditional model.

Table 6.2 also shows that increasing the number of periods from 6 to 8 will result in an efficiency gain of 0.173/0.127 = 36 % under the traditional model, while the gain is not as substantial under a complex model.

In general, we suggest that alternating AB and BA pairs in sequence will result in a nearly optimal design, if not the optimal one, under all models we considered, for estimating individual effects in N-of-1 trials.

Adaptive Trial Design

Adaptive designs are gaining popularity in recent years. Liang and Carriere (2009) outline how one could plan response adaptive designs utilizing outcomes in a given experiment while achieving multiple objectives. For example, clinicians may wish to achieve good estimation precision, effective treatment of patients, or cost effectiveness (Carriere and Huang 2000). Recently, Liang, Li, Wang, and Carriere (Liang and Carriere 2009) extended their approach to binary responses. For N-of-1 trials, designs can be found by updating AB or BA pairs successively as the trial progresses. Such objectives as maintaining a balance or counterbalancing between AB and BA pairs as suggested by Kravitz et al. (2014) can also be considered. A Bayesian framework may be the most natural for adaptive design and decision-making. However, not much attention has been given to finding optimal designs for binary data in the literature, and further research is needed.

Discussion

N-of-1 trials are extremely useful in subject-focused investigations, for example, medical experiments. As far as we are aware, no guidelines are available in the literature on how to plan such a trial optimally. In this Chapter, we construct optimal N-of-1 trials under various models depending upon how we accommodate the carryover effects and the error structures for the repeated measurements. We also suggest constructing optimal aggregated N-of-1 trials for both the individual and average patients.

A straight application of the two-treatment optimal design results in the literature with A to AB and B to BA can result in an inferior design unsuitable for individual patient treatment care. We showed that not all of the suggested sequences are optimal to be used in N-of-1 trials. Further, they may not be optimal for estimating effects at the average level. For example, when $p=4$, the literature gives *ABBA* and *AABB* and their duals as the optimal design. Applying this result to 8-period N-of-1 trials, we would need to consider at least four sequences, ABBABAAB, ABABBABA and their duals. However, we showed that none of these four sequences are optimal for N-of-1 trials for $p=8$.

For the traditional first-order residual effects model with uncorrelated and equal-correlated errors, the optimal N-of-1 trial design is to use the sequence consisting of alternating AB and BA pairs. We can use its dual sequence with the same effect. For example, the optimal 8 period N-of-1 trial design is to use ABBAABBA or BAABBAAB under equal or uncorrelated errors. For the self and mixed effect model, the optimal N-of-1 trial uses the sequence consisting of only AB pairs. If self and mixed carryover effects are a concern, the optimal 8 period N-of-1 trial is ABABABAB for estimating the direct treatment effects. Hence, once again we find that optimal designs are strongly model dependent. It would depend on how we accommodate the residual effects and what type of errors are practically feasible.

Surprisingly, however, the results do not change by very much in practice with adopting not so optimal sequences, as we examined in Section "Optimal N-of-1 Trials". A numerical calculation of the estimation precision using several 6- and 8-period designs revealed the actual performance of a particular design, giving us practical guidelines. Overall, we conclude that alternating between AB and BA pairs in subsequent cycles will result in practically optimal N-of-1 trials for a single patient, if not the optimal, under all the models we considered without the need to guess at the correlation structure or conduct a pilot study. Alternating between AB and BA pairs in a single trial is nearly robust to misspecification of the error structure of the repeated measurements.

Lastly, we suggest that when an experiment has been carried out with the optimal N-of-1 trial and additional patients are accrued in the trial, we can plan and aggregate these N-of-1 trials optimally by allocating the same number of patients to its dual sequence by reversing the treatment order, thereby optimizing the trial for both the individual and average patients.

Acknowledgments This research was made possible due to the funding from the Natural Sciences and Engineering Research Council of Canada.

References

Afsarinejed K, Hedayat A (2002) Repeated measurement designs for a model with self and mixed carryover effects. J Stat Plan Inference 106:449–459

Armitage P (1975) Sequential medical trials. Blackwell, Oxford

Bose M, Mukherjee B (2003) Optimal crossover designs under a general model. Stat Probab Lett 62:413–418

Brown BWJ (1980) The crossover experiment for clinical trials. Biometrics 36:69–79

Cantoni E (2004) A robust approach to longitudinal data analysis. Can J Stat 32:169–180

Carriere KC (1994) Cross-over designs for clinical trials. Stat Med 13:1063–1069

Carriere KC, Huang R (2000) Crossover designs for two-treatment cancer clinical trials. J Stat Plan Inference 87:125–134

Carriere KC, Reinsel GC (1992) Investigation of dual-balanced crossover designs for two treatments. Biometrics 48:1157–1164

Carriere KC, Reinsel GC (1993) Optimal two-period repeated measurements designs with two or more treatments. Biometrika 80:924–929

Cheng CS, Wu CF (1980) Balanced repeated measurements designs. Ann Stat 8:1272–1283

Davidian M, Verbeke G, Molenberghs G (2009) Longitudinal data analysis. Int Stat Rev 77:1857–1936

Figueiras A, Carracedo-Martinez E, Saez M, Taracido M (2005) Analysis of case-crossover designs using longitudinal approaches: a simulation study. Epidemiology 16:239–246

Greenfield S, Kravits R, Duan N, Kaplan SH (2007) Heterogeneity of treatment effects: implications for guidelines, payment, and quality assessment. Am J Med 120:S3–S9

Grizzle JE (1965) The two-period change over design and its use in clinical trials. Biometrics 21:461–480

Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S (1986) Determining optimal therapy–randomized trials in individual patients. N Engl J Med 314:889–892

Guyatt GH, Heyting A, Jaeschke R, Keller J, Adachi JD, Roberts RS (1990) N of 1 randomized trials for investigating new drugs. Control Clin Trials 11:88–100

Hedayat AS, Yang M (2003) Universal optimality of balanced uniform crossover designs. Ann Stat 31:978–983

Kenword M, Jones B (1987) A log-linear model for binary crossover data. Appl Stat 36:192–204

Kiefer J (1975) Construction and optimality of generalized Youden designs. In: Srivastava JN (ed) A survey of statistical designs and linear models. North-Halland, Amsterdam

Kravitz RL, Duan N, Breslow J (2004) Evidence-based Medicine, heterogeneity of treatment effects, and the trouble with averages. Milbank Q 82:661–687

Kravitz RL, Duan N (eds) and The Decide Methods Center N-of-1 Guidance Panel (Duan N, EI, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S) (2014) Design and implementation of N-of-1 trials: a user's guide. Agency for Healthcare Research and Quality, Rockville

Kunert J, Stufken J (2002) Optimal crossover designs in a model with self and mixed carryover effects. J Am Stat Assoc 97:896–906

Kunert J, Stufken J (2008) Optimal crossover designs for two treatments in the presence of mixed and self-carryover effects. J Am Stat Assoc 103:1641–1647 (correction 1999, 86, 234)

Larson EB (1990) N-of-1 clinical trials. West J Med 152:52–56

Laska E, Meisner M (1985) A variational approach to optimal two-treatment crossover designs: application to carryover-effect models. J Am Stat Assoc 80:704–710

Liang YY, Carriere KC (2009) Multiple-objective response-adaptive repeated measurement designs for clinical trials. J Stat Plan Inference 139:1134–1145

Liang KY, Zeger S (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22

Nikles CJ, Clavarino AM, Del Mar CB (2005) Using N-of-1 trials as a clinical tool to improve prescribing. Br J Gen Pract 55:175–180

Senn S (2002) Crossover trials in clinical research. Wiley, Hoboken

Wei LJ, Durham S (1978) The randomized play the winner rule in medical trials. J Am Stat Assoc 73:840–843