Statistical Advances in Clinical Trials and Clinical Research

Guogen Shan
University of Nevada, Las Vegas, guogen.shan@unlv.edu

Sarah Banks
Cleveland Clinic Lou Ruvo Center for Brain Health

Justin B. Miller
Cleveland Clinic Lou Ruvo Center for Brain Health

Aaron Ritter
Cleveland Clinic Lou Ruvo Center for Brain Health

Charles Bernick
Cleveland Clinic Lou Ruvo Center for Brain Health

Repository Citation

Authors

Guogen Shan, Sarah Banks, Justin B. Miller, Aaron Ritter, Charles Bernick, Joseph Lombardo, and Jeffrey
L. Cummings

Featured Article

Statistical advances in clinical trials and clinical research

Guogen Shan[a,*], Sarah Banks[b], Justin B. Miller[b], Aaron Ritter[b], Charles Bernick[b],
Joseph Lombardo[c], Jeffrey L. Cummings[b]

[a]Epidemiology and Biostatistics Program, Department of Environmental and Occupational Health School of Community Health Sciences, University of Nevada Las Vegas, Las Vegas, NV, USA
[b]Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, NV, USA
[c]National Supercomputing Institute, University of Nevada Las Vegas, Las Vegas, NV, USA

Abstract

Introduction: New treatments for neurodegenerative disease are urgently needed, and clinical trial methods are an essential component of new drug development. Although a parallel-group study design for neurological disorder clinical trials is commonly used to test the effectiveness of a new treatment as compared to placebo, it does not efficiently use information from the on-going study to increase the success rate of a trial or to stop a trial earlier when the new treatment is indeed ineffective.

Methods: We review some recent advances in designs for clinical trials, including futility designs and adaptive designs.

Results: Futility designs and noninferiority designs are used to test the nonsuperiority and the non-inferiority of a new treatment, respectively. We provide some guidance on using these two designs and analyzing data from these studies properly. Adaptive designs are increasingly used in clinical trials to improve the flexibility and efficiency of trials with the potential to reduce resources, time, and costs. We review some typical adaptive designs and new statistical methods to handle the statistical challenges from adaptive designs.

Discussion: Statistical advances in clinical trial designs may be helpful to shorten study length and benefit more patients being treated with a better treatment during the discovery of new therapies for neurological disorders. Advancing statistical underpinnings of neuroscience research is a critical aspect of the core activities supported by the Center of Biomedical Research Excellence award supporting the Center for Neurodegeneration and Translational Neuroscience.

© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Adaptive designs; Clinical trials; Futility design; Neurological disorders; Proper statistical inference

1. Introduction

In clinical trials for neurological disorders, a parallel group study is commonly used to assess the effectiveness of a new treatment as compared to the placebo group [1–4]. Patients are randomized to either the treatment arm(s) or the placebo arm following a prespecified randomization schedule. At the end of the study, the change of the primary outcome from the end to the baseline, calculated from the treatment arm, is compared with that from the placebo arm to make a conclusion whether the new treatment has sufficient activity to move to the next phase for further investigation. The primary outcome to assess the cognitive performance can be measured by established assessment tools, such as the Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog), the Unified Parkinson's Disease Rating Scale (UPDRS), Clinical Dementia Rating, and the Amyotrophic Lateral Sclerosis Functional Rating Scale-revised (ALSFRSr). The commonly used parallel-group design is able to study the

*Corresponding author. Tel.: 702-895-4413; Fax: 702-895-5573.
E-mail address: guogen.shan@unlv.edu

effectiveness of the new treatment with influential covariates being balanced during the randomization procedure; however, it may not be efficient enough for the purpose of rapidly screening out nonpromising treatments or identifying the most promising treatments [1,5–10].

Futility designs are widely used in early phase neurological disorder trials to screen out new treatments that are highly unlikely to produce successful results [11–15]. Futility designs can be used in a single-arm study with the threshold estimated from historical controls or in a parallel-group study with a nonsuperiority alternative hypothesis [16–18]. The purpose of the futility design is to screen out an unpromising treatment with fewer patients and a much shorter study time period. As compared to the futility design, the commonly used parallel-group study is often used to test the superiority or noninferiority of the new treatment over the placebo. In this article, we review the difference between the futility design and the noninferiority design which is also widely used in clinical trials to test the noninferiority of a new treatment. We also provide some guidance on the proper usage of such designs [19–24].

In recent years, adaptive designs have been introduced and used in trials for neurological disorders to reduce resource use and study length [25–28]. There are a few definitions for an adaptive design. In 2010, the Food and Drug Administration published a draft guidance document on adaptive designs and defined an adaptive design as "a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study" [29].

Adaptive designs provide opportunities to modify or change the trial during the study while maintaining the validity and integrity of the trial. These opportunities are prespecified when certain conditions are met. In 2008, Chow and Chang [27] reviewed 10 adaptive designs used in clinical trials, including an adaptive randomization design that allows modification of randomization schedules; a group sequential design that allows early stopping due to futility, efficacy, or both; a sample size re-estimation design allowing sample size adjustment; a pick-the-winner design; an adaptive dose-finding design; a biomarker-adaptive design; an adaptive treatment-switching design; an adaptive seamless design; a hypothesis-adaptive design; and a multiple adaptive design. In this article, we review the following two commonly used adaptive designs in neurological disorder trials. The response-adaptive randomization design uses the patients' responses from the current on-going study to modify the assignment probabilities to each treatment arm, with more patients being treated in the better arms. The response-adaptive randomization design belongs to the adaptive randomization design that also includes treatment-adaptive randomization and covariate-adaptive randomization [27]. The other adaptive design discussed in this article is the adaptive dose-finding design that in-

creases the accuracy of the estimation for the maximum tolerated dose or minimum effective dose [30].

Studies designed by an adaptive method may introduce new challenges in data analysis. It is important that intended statistical analysis should guide the study design [23,31,32]. For this reason, new statistical analysis approaches to analyze the data from adaptive designs properly are also discussed. Review of novel, efficient, and proper statistical approaches in neuroscience research is an important service of the Data Management and Statistics Core of the Center for Neurodegeneration and Translational Neuroscience supported by the Center of Biomedical Research Excellence award from the National Institute of General Medical Sciences.

2. Futility designs

The futility design, also known as the nonsuperiority design, can be used to screen out a new treatment candidate who is not promising for further investigation. It can be implemented in a single-arm study or a parallel group study to investigate the effectiveness of a new experimental treatment. Suppose μ_e and μ_c are the mean of the primary outcome in a new experimental treatment group and in the control group, respectively, in a parallel group study. For a single-arm study, we may use the same notation μ_c to represent the estimated value from historical data. Let $\Delta = \mu_e - \mu_c$ be the difference between the two groups.

For clinical trials in neurology, the primary outcome of interest to measure disease symptoms is often computed from some well-established assessment tools, for example, ADAS-Cog, UPDRS, and ALSFRSr. The change of these measurements from the end to the baseline (post–pre) is often used as the primary outcome, for example, $\mu_e = \mu_{e1} - \mu_{e0}$, where μ_{e1} and μ_{e0} are the outcome of patients from the treatment group at the end and at baseline, respectively. It should be noted that a treatment with a smaller increase (slowing disease progression) or a larger decrease (improving the disease symptoms) in the outcome is considered as a better treatment in some assessment tools (e.g., ADAS-Cog, UPDRS), whereas it is reversed when others are used (e.g., ALSFRSr).

When ADAS-Cog or UPDRS is used to measure the disease symptom, suppose δ_0 is the maximum allowable progression threshold, the statistical hypotheses for the futility design are presented as

$$H_0 : \Delta \leq \delta_0 \text{ against } H_a : \Delta > \delta_0, \tag{1}$$

where δ_0 is a clinically meaningful threshold to measure the disease symptom [33–35]. For example, a clinical trial to assess the effectiveness of coenzyme Q10 and GPI-1485 in Parkinson's Disease (PD) patients [36] was designed as a futility study with $\delta_0 = -3.19$, which is 30% of the total UPDRS change of participants in the placebo group from the Deprenyl and Tocopherol Antioxidant Therapy of

Parkinsonism trial (DATATOP), $\mu_c = 10.65$. This trial is designed as a single-arm futility study with the hypotheses:

$$H_0 : \mu_e - \mu_c \leq -3.19 \text{ against } H_a : \mu_e - \mu_c > -3.19.$$

If the null hypothesis is rejected, we can conclude that the new experimental treatment is not promising for further investigation. The sample size for each arm in this study [36] was calculated as 58 participants per arm to attain 85% power at the significance level of 0.1, with $\mu_e = 7.46$ under the null hypothesis and $\mu_e = 10.65$ under the alternative hypothesis.

When a larger observed outcome represents a better treatment (e.g., the ALSFRSr score), the hypotheses for a two-arm futility design are presented as

$$H_0 : \Delta \geq \delta_0 \text{ against } \Delta < \delta_0, \qquad (2)$$

where δ_0 is the minimum worthwhile efficacy, and it is often a positive value. Although the hypotheses can be presented in two different formulae as in Equations (1) and (2) which depend on the direction of the assessment tool, they are statistically identical by reversing one of the assessment tools. Both the alternative hypotheses suggest nonsuperiority of the new experimental treatment as compared to placebo.

3. Noninferiority designs

The aforementioned futility design (also known as the nonsuperiority design) should not be confused with other designs, such as a superiority design or a noninferiority design. We compare these designs to the futility design with the hypotheses in Equation (2) by assuming that a higher score represents a better treatment. The hypotheses of a superiority design or a noninferiority design are expressed as

$$H_0 : \Delta \leq \delta_1 \text{ against } \Delta > \delta_1, \qquad (3)$$

where δ_1 is the margin. Let the clinically meaningful estimate of Δ be Δ_{CME}. This estimate is very important in clinical trials to show clinically meaningful improvement by a new treatment. When $\delta_1 > \Delta_{CME}$, Equation (3) represents a superiority design. It becomes a noninferiority design when $\delta_1 < \Delta_{CME}$. In a noninferiority trial, the aim is to show that a new treatment is not much worse than the standard care or not clinically inferior to the standard care. Lesaffre [37] compared the difference between superiority trials and noninferiority trials with two real noninferiority trial examples along with the discussion of the noninferiority margin.

In a randomized clinical trial to investigate the validity and reliability of online delivery of the Lee Silverman Voice Treatment for PD patients with speech and voice disorder [38], the change in sound pressure level (dB-C) after the treatment was the primary outcome. The clinically relevant improvement was estimated as $\Delta_{CME} = 4.5$ dB, with an esti-mated standard deviation of 2.48 dB. This study was designed as a noninferiority trial to compare the performance with the Lee Silverman Voice Treatment between online and face-to-face administration, with the noninferiority marginal of 2.25 which is half of the estimated clinically relevant improvement, $\delta_1 = 2.25 < \Delta_{CME}$. A sample size of 15 per arm was required to attain 90% power at the significance level of 0.025 for this noninferiority trial.

In another noninferiority study reported by Winblad et al. [39], the rivastigmine capsule was compared with placebo for AD patients by using the ADAS-Cog change from baseline as the primary outcome. The noninferiority margin was set as 1.25 points decrease on the ADAS-Cog, which is half of the estimated treatment difference from other existing studies. This noninferiority margin is considered as the minimum clinically meaningful difference.

In the aforementioned hypotheses for either a futility design or a noninferiority design, the primary outcome is computed as the change from the end to baseline, for example, $\mu_e = \mu_{e1} - \mu_{e0}$. When the primary outcome is measured as the change from baseline to the end (pre–post, e.g., $\mu_e = \mu_{e0} - \mu_{e1}$), the aforementioned hypotheses can still be applied. For example, in a study to confirm the noninferiority of rotigotine to ropinirole for PD patients on concomitant levodopa therapy, the primary outcome was the change of the UPDRS Part III (ON state) sum score from baseline to the end [40]. In this study, a larger observed value (the difference of change between rotigotine and ropinirole) represents a better treatment. For this reason, the hypotheses presented in Equation (3) should be used in this study to assess the noninferiority of rotigotine to ropinirole.

4. Sample size determination and statistical inference

Sample size calculation plays a very important role in clinical trials to ensure a prespecified level of power when type I error rate (α) is controlled. Type I error rate and power are generally computed by using the estimated Δ values under the null and alternative hypotheses, δ_0 and δ_a. Accurate estimates of δ_0 and δ_a would increase the success rate of a trial with the computed sample size adequate to detect the difference between the treatment arms.

The hypotheses discussed in this article are all one-sided; therefore, z_α, instead of $z_{\alpha/2}$, is used in the sample size determination (see Levin [41] for the detailed sample size calculation formula). It should be noted that the sample size calculation provided by Levin [41] is based on asymptotic approaches, which should be used with caution for a study with sample size that is small to medium. For a study with binary outcome (more than 50% decrease in the Inventory of Depressive Symptomatology-Clinician score from baseline in PD [42]), the proper type I error rate should be computed under the null hypothesis ($H_0: \Delta \leq \delta_0$), not just at the boundary of the associated hypothesis space ($\Delta = \delta_0$) [43–45]. In the trial to compare response rates between atomoxetine and placebo [42], we suppose the

null hypothesis is presented as H_0: $p_a \leq p_c$, where p_a and p_c are the response rate for the atomoxetine arm and the placebo arm, respectively. The response rate for the placebo arm is estimated to be 10% from historical data. An exact unconditional approach [46] may be used to calculate the error rate, then the actual type I error rate is computed as

$$TIE(p_a, p_c) = \max_{p_a \leq p_c \leq 10\%} \sum_{T(X_a, X_c) \in \Omega} f(X_a, N_a, p_a) f(X_c, N_c, p_c),$$

(4)

where Ω is the rejection region and $f(.)$ is the probability density function of a binomial distribution. The type I error rate should be properly computed over the null space $p_a \leq p_c \leq 10\%$ as in Equation (4). Often, its error rate is computed on the boundary with $p_a = p_c = 10\%$, which is not proper for use without a theoretical proof to show that the error rate occurs at the boundary of the null space [43].

When two proportions are compared, efficient sample size calculation approaches are recommended for use to provide valid sample sizes, such as simulation-based approaches and exact approaches [3,23,47–50]. In addition, the actual type I error rate could be highly affected by the estimated δ_0 [15], and we would encourage researchers to compute and provide sample sizes under multiple possible δ_0 and δ_a scenarios.

At the end of a trial, observed data are analyzed to make statistical inference. Statistical analysis should be intentionally consistent with the design. For a randomized placebo-controlled futility study with the hypotheses given in Equation (1), the null hypothesis is rejected when a large Δ value is observed. The progression threshold δ_0 is the boundary of the one-sided lower confidence interval computed from the observed data. When δ_0 is outside of the interval (δ_0 is less than the lower limit), we have enough evidence to reject the null, and the new experimental treatment is not promising for further investigation. Otherwise, if the computed lower limit is less than δ_0, we fail to reject the null hypothesis. Similarly, the $1 - \alpha$ lower limit of Δ is used in testing the hypotheses in Equation (3), whereas the $1 - \alpha$ upper limit is used for the hypotheses in Equation (2) to make valid statistical inference. The lower or upper limit of δ_0 should be computed properly for different types of data. For a matched-pairs study, a study design that accounts for the matching information should be considered.

5. Adaptive designs

Adaptive designs have been increasingly used in clinical trials to increase the flexibility of a trial by allowing a trial to be stopped earlier for futility when a new treatment is not promising and/or allowing more patients being assigned to a better treatment and so on. The adaptations in a trial have to be prospectively planned to guarantee the validity and integrity of the trial.

An adaptive design provides opportunities for a trial to be modified during the course of the trial; however, it has to be prospectively planned. In other words, any modification of a trial (e.g., adding or dropping a treatment arm) is specified during the planning stage when certain conditions are met. These conditions include the comparison of results based on the observed data from the on-going study. In general, it takes more effort for the research team to prepare an adaptive design than a traditional nonadaptive design. A significant number of simulation studies have to be conducted to investigate all possible outcomes during the planning stage.

In a very recent phase II trial to evaluate the BAN2401 (a monoclonal antibody targeting amyloid protofibrils) for the treatment of AD patients, the response-adaptive randomization model was used in the study design. The probability of the next patient being assigned to one of the treatment arms or the control arm is determined by the probability of that treatment arm being the most effective treatment arm among all arms. The cumulative data of patients from this on-going study are used for calculating these probabilities. An adaptive randomization design allows a trial to assign more patients to better treatment arms, which may lead to imbalances in the sample-size allocation and the distribution of influential covariates across treatment arms. Recently, Saville and Berry [51] proposed using odds ratios to modify the probability in the response-adaptive randomization for each arm to improve the covariate balance. When the primary outcome is binary, a new patient allocation scheme to adjust the covariate imbalance issue during the adaptive randomization procedure has been proposed [52]. New and proper statistical methods are needed to overcome the emergent statistical challenges from the response-adaptive randomization. Alternatively, a study can be designed as a covariate-adaptive randomization to balance the allocation of multiple arms across a set of influential covariates without compromising randomness. That would help reducing the complexity of the final data analysis.

In neurology, adaptive designs are often used in early phases to learn the safety of a new treatment and select the dose for the following trials. In a dose-finding study, there are often a few arms with different dose levels and a placebo arm to estimate the placebo effect. Adaptive methods can be used to stop a dose earlier due to futility, accept a dose due to efficacy, or add a new dose to the study. The aforementioned response-adaptive randomization could be used in conjunction with the adaptive dose-finding design. For example, a phase II trial to evaluate the safety and efficacy of ABT-089 in AD patients was designed by using a Bayesian response-adaptive randomization method to allocate patients to one of the seven arms (six arms of ABT-089 with different doses and placebo) after having at least five patients in each arm [53]. This study was also designed to allow stopping for efficacy or futility based on the conditional power calculated from the on-going study. The objective of this adaptive dose-finding study was to identify the minimum effective dose resulting in at least an average of 1.75-point

ADAS-Cog improvement over placebo. In other studies, the maximal tolerated dose may be the target dose [54].

In addition to the aforementioned adaptive designs, Chow and Chang [27] reviewed other commonly used adaptive design methods in clinical trials: an adaptive seamless phase II/III design, a biomarker-adaptive design, an adaptive treatment-switching design, and so on. More details about these adaptive designs may be found in the literature [1,9,22,24,25,27,55–57].

6. Discussion

The futility design is used in early phase clinical trials to screen out unpromising treatments and save resources for other treatment candidates. By contrast, a futility-stopping boundary is used to drop a treatment arm from a trial or stop a trial earlier if the treatment will not show efficacy based on the observed results. The futility boundary could be defined as the prespecified conditional power or the pre-specified confidence limit. They are two different concepts, that is, the futility design is a study design, whereas a futility boundary is a threshold in the study design [41,58].

Adaptive designs are attractive to increase the flexibility of a trial, but they also introduce new statistical challenges to analyze the final observed data properly. The aforementioned imbalance issue on the inferential covariates from the response-adaptive randomization is one of them. When a study is designed by an adaptive approach, the data analysis should align with the study design. For an adaptive two-stage design in which the second-stage sample size depends on the results from the first stage, the data analysis that uses only the final observed data without considering the nature of a two-stage adaptive design is not appropriate [10,59]. For example, a single-arm two-stage adaptive design was used to assess the effectiveness of a new treatment for PD patients with the primary outcome as a binary endpoint. The required sample size for the first stage is $n_1 = 22$ participants, and the second-stage sample size, $n_2(X_1)$, is a function of the number of the responses from the first stage (X1), (e.g., $n_2(X_1) = 35$ when $X_1 = 11$) [59]. At the beginning of the study, the number of responses from the first stage X_1 is unknown. Therefore, the sample space having all possible X_1 from 0 to n_1 should be used in statistical inference, such as P-value calculation and confidence interval calculation.

In practice, it is possible that the final observed sample size is different from that is planned. Then, a statistical approach that incorporates adaptive elements and the observed sample size is valid for data analysis [10,45].

Acknowledgments

RESEARCH IN CONTEXT

1. Systematic review: Futility designs are commonly used in early phase neurological disorder trials to screen out new treatments that are highly unlikely to produce successful results. Adaptive designs are increasingly used in drug development to improve the flexibility and efficiency of trials, having the potential to reduce the cost and save sample sizes.

2. Interpretation: A futility design should not be confused with a design that allows a trial to be stopped due to futility. The number of neurological disorder trials designed by adaptive approaches is not as large as expected.

3. Future directions: Adaptive futility designs should be developed for use in trials, and the associated statistical methods for newly developed designs should be proposed to provide proper statistical inference.

References

[1] Cummings J, Gould H, Zhong K. Advances in designs for Alzheimer's disease clinical trials. Am J Neurodegener Dis 2012;1:205–16.

[2] Cummings J, Lee G, Mortsdorf T, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2017. Alzheimers Dement (N Y) 2017;3:367–84.

[3] Shan G, Ma C, Hutson AD, Wilding GE. Randomized Two-Stage Phase II Clinical Trial Designs Based on Barnard's Exact Test. J Biopharm Stat 2013;23:1081–90.

[4] Wilding GE, Shan G, Hutson AD. Exact two-stage designs for phase II activity trials with rank-based endpoints. Contemp Clin Trials 2012; 33:332–41.

[5] Cummings JL, Raman R, Ernstrom K, Salmon D, Ferris SH, Alzheimer's Disease Cooperative Study Group. ADCS prevention instrument project: Behavioral measures in primary prevention trials. Alzheimer Dis Assoc Disord 2006;20.

[6] Bernick C, Cummings J, Raman R, Sun X, Aisen P. Age and rate of cognitive decline in Alzheimer disease: Implications for clinical trials. Arch Neurol 2012;69:901–5.

[7] Banks SJ, Obuchowski N, Shin W, Lowe M, Phillips M, Modic M, et al. The protective effect of education on cognition in professional fighters. Arch Clin Neuropsychol 2014;29:54–9.

[8] Bernick C, Banks S, Phillips M, Lowe M, Shin W, Obuchowski N, et al. Professional fighters brain health study: Rationale and methods. Am J Epidemiol 2013;178:280–6.

[9] Shan G, Wilding GE, Hutson AD, Gerstenberger S. Optimal adaptive two-stage designs for early phase II clinical trials. Stat Med 2016; 35:1257–66.

[10] Shan G, Wang W. Exact one-sided confidence limits for Cohen's kappa as a measurement of agreement. Stat Methods Med Res 2017; 26:615–32.

[11] Cummings J, Lai TJJ, Hemrungrojn S, Mohandas E, Yun Kim S, Nair G, et al. Role of Donepezil in the management of neuropsychiatric symptoms in Alzheimer's disease and dementia with lewy bodies. CNS Neurosci Ther 2016;22:159–66.

[12] Cummings J, Scheltens P, McKeith I, Blesa R, Harrison JE, Bertolucci PH, et al. Effect size analyses of Souvenaid in patients with Alzheimer's disease. J Alzheimers Dis 2017;55:1131–9.

[13] Bernick C, Banks S. What boxing tells us about repetitive head trauma and the brain. Alzheimers Res Ther 2013;5.

[14] Banks SJJ, Weintraub S. Generalized and symptom-specific insight in behavioral variant frontotemporal dementia and primary progressive aphasia. J Neuropsychiatry Clin Neurosci 2009;21:299–306.

[15] Shan G. Comments on "Two-sample binary phase 2 trials with low type I error and low sample size". Stat Med 2017;36:3437–8.

[16] Cummings J. Disease modification and Neuroprotection in neurodegenerative disorders. Transl Neurodegen 2017;6.

[17] Cummings J. Lessons learned from Alzheimer disease: Clinical trials with negative outcomes. Clin Transl Sci 2017;11:147–52.

[18] Banks SJ, Miller JB, Rissman RA, Bernick CB. Lack of influence of apolipoprotein E status on cognition or brain structure in professional fighters. J Neurotrauma 2017;34:380–4.

[19] Chow SCC. Bioavailability and bioequivalence in drug development. Wiley interdisciplinary reviews. Comput Stat 2014;6:304–12.

[20] Chow SCC, Song F, Bai H. Analytical similarity assessment in biosimilar studies. AAPS J 2016;18:670–7.

[21] Chow SCC, Endrenyi L, Lachenbruch PA, Mentŕe F. Scientific factors and current issues in biosimilar studies. J Biopharm Stat 2014;24:1138–53.

[22] Chow SC. Adaptive Clinical Trial Design. Annu Rev Med 2014; 65:405–15.

[23] Shan G. Exact Statistical Inference for Categorical Data. 1 ed. San Diego, CA: Academic Press; 2015.

[24] Bhatt DL, Mehta C. Adaptive Designs for Clinical Trials. N Engl J Med 2016;375:65–74.

[25] Berry DA. Adaptive clinical trials: the promise and the caution. J Clin Oncol 2011;29:606–9.

[26] Chow SC, Chang M. In: Adaptive Design Methods in Clinical Trials (Chapman & Hall/CRC Biostatistics Series). 2 ed. Boca Raton, FL: Chapman and Hall/CRC; 2011.

[27] Chow SCC, Chang M. Adaptive design methods in clinical trials -a review. Orphanet J Rare Dis 2008;3:11.

[28] Wilding GE, Consiglio JD, Shan G. Exact approaches for testing hypotheses based on the intra-class kappa coefficient. Stat Med 2014; 33:2998–3012.

[29] Food and Drug Administration (FDA). Guidance for Industry Adaptive Design Clinical Trials for Drugs and Biologics. Rockville, MD: Center for Biologics Evaluation and Research; 2010. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf.

[30] Iasonos A, O'Quigley J. Adaptive dose-finding studies: a review of model-guided phase I clinical trials. J Clin Oncol 2014;32:2505–11.

[31] Shan G, Wilding G. Unconditional tests for association in 2 × 2 contingency tables in the total sum fixed design. Stat Neerl 2015;69:67–83.

[32] Shan G, Zhang H, Jiang T. Minimax and admissible adaptive two-stage designs in phase II clinical trials. BMC Med Res Methodol 2016; 16:90.

[33] Schwid SR, Cutter GR. Futility studies: Spending a little to save a lot. Neurology 2006;66:626–7.

[34] Tilley BC, Palesch YY, Kieburtz K, Ravina B, Huang P, Elm JJ, et al. Optimizing the ongoing search for new treatments for Parkinson disease: Using futility designs. Neurology 2006;66:628–33.

[35] Levy G, Kaufmann P, Buchsbaum R, Montes J, Barsdorf A, Arbing R, et al. A two-stage design for a phase II clinical trial of coenzyme Q10 in ALS. Neurology 2006;66:660–3.

[36] NINDS NET-PD Investigators. A randomized clinical trial of coenzyme Q10 and GPI-1485 in early Parkinson disease. Neurology 2007;68:20–8.

[37] Lesaffre E. Superiority, equivalence, and non-inferiority trials. Bull NYU Hosp Jt Dis 2008;66:150–4.

[38] Constantinescu G, Theodoros D, Russell T, Ward E, Wilson S, Wootton R. Treating disordered speech and voice in Parkinson's disease online: A randomized controlled non-inferiority trial. Int J Lang Commun Disord 2011;46:1–16.

[39] Winblad B, Cummings J, Andreasen N, Grossberg G, Onofrj M, Sadowsky C, et al. A six-month double-blind, randomized, placebo-controlled study of a transdermal patch in Alzheimer's disease–rivastigmine patch versus capsule. Int J Geriatr Psychiatry 2007; 22:456–67.

[40] Mizuno Y, Nomoto M, Hasegawa K, Hattori N, Kondo T, Murata M, et al., Rotigotine Trial Group. Rotigotine vs ropinirole in advanced stage Parkinson's disease: a double-blind study. Parkinsonism Relat Disord 2014;20:1388–93.

[41] Levin B. The futility study–Progress over the last decade. Contemp Clin Trials 2015;45:69–75.

[42] Weintraub D, Mavandadi S, Mamikonyan E, Siderowf AD, Duda JE, Hurtig HI, et al. Atomoxetine for depression and other neuropsychiatric symptoms in Parkinson disease. Neurology 2010;75:448–55.

[43] Shan G, Chen JJ, Ma C. Boundary problem in Simon's two-stage clinical trial designs. J Biopharm Stat 2017;27:25–33.

[44] Shan G, Zhang H, Jiang T, Peterson H, Young D, Ma C. Exact p-values Simon's two-stage designs. Clin Trials 2016;8:351–7.

[45] Shan G, Zhang H. Exact unconditional sample size determination for paired binary data (letter commenting: J Clin Epidemiol. 2015;68:733-739). J Clin Epidemiol 2017;84:188–90.

[46] Barnard GA. Significance tests for 2 × 2 tables. Biometrika 1947; 34:123–38.

[47] Shan G, Ma C, Hutson AD, Wilding GE. An efficient and exact approach for detecting trends with binary endpoints. Stat Med 2012; 31:155–64.

[48] Shan G. More efficient unconditional tests for exchangeable binary data with equal cluster sizes. Stat Probab Lett 2013;83:644–9.

[49] Shan G. Exact confidence intervals for randomized response strategies. J Appl Stat 2016;43:1279–90.

[50] Shan G, Ma C. Unconditional tests for comparing two ordered multinomials. Stat Methods Med Res 2016;25:241–54.

[51] Saville BR, Berry SM. Balanced covariates with response adaptive randomization. Pharm Stat 2017;16:210–7.

[52] Ning J, Huang X. Response-adaptive randomization for clinical trials with adjustment for covariate imbalance. Stat Med 2010; 29:1761–8.

[53] Lenz RA, Pritchett YL, Berry SM, Llano DA, Han S, Berry DA, et al. Adaptive, dose-finding phase 2 trial evaluating the safety and efficacy of ABT-089 in mild to moderate Alzheimer disease. Alzheimer Dis Assoc Disord 2015;29:192–9.

[54] Elkind MS, Sacco RL, Macarthur RB, Peerschke E, Neils G, Andrews H, et al. High-dose lovastatin for acute ischemic stroke: Results of the phase I dose escalation neuroprotection with statin therapy for acute recovery trial (NeuSTART). Cerebrovasc Dis 2009; 28:266–75.

[55] . Ravina B, Cummings J, McDermott M, Poole RM, eds. Clinical Trials in Neurology: Design, Conduct, Analysis. Cambridge University Press; 2012.

[56] Bretz F, Gallo P, Maurer W. Adaptive designs: The Swiss Army knife among clinical trial designs? Clin Trials (Lond) 2017;14:417–24.

[57] Hatfield I, Allison A, Flight L, Julious SA, Dimairo M. Adaptive designs undertaken in clinical research: A review of registered clinical trials. Trials 2016;17.

[58] Freidlin B. Futility Analysis. Wiley StatsRef: Statistics Reference Online; 2014; https://doi.org/10.1002/9781118445112.stat07120.

[59] Shan G, Zhang H, Jiang T. Efficient confidence limits for adaptive one-arm two-stage clinical trials with binary endpoints. BMC Med Res Methodol 2017;17.