Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 15 (10): 1251-1255, 2015
ISSN 1812-5654

RESEARCH ARTICLE OPEN ACCESS

DOI: 10.3923/jas.2015.1251.1255

Bayesian Monitoring For Experimental Study in Women with Abnormal Uterine Bleeding

[1]Busaba Supawattanabodee, [1]Lily Ingsrisawang and [2]Siriwan Tangitgamol
[1]Department of Statistics, Faculty of Science, Kasetsart University, 10900, Bangkok, Thailand
[2]Department of Obstetrics and Gynecology, Faculty of Medicine Vajira Hospital, Navamindradhiraj University, Bangkok, Thailand

ARTICLE INFO
Article History:
Received: September 13, 2015
Accepted: October 05, 2015

Corresponding Author:
Lily Ingsrisawang
Department of Statistics,
Faculty of Science, Kasetsart University,
10900, Bangkok, Thailand
Tel: 6622525555/3812-3

ABSTRACT
The aim of this study is to evaluate the use of Simple Bayesian Adaptive Randomization Design (SBARD) and predictive probability in clinical trial. The trial was planned to randomize 418 women with abnormal uterine bleeding to have pathologic evaluation by 2 devices. Total of 5,000 simulations were generated to evaluate the performance of SBARD and predictive probability under three different scenarios. The results from the SBARD were compared with the actual results. We found that the results of SBARD under the first scenario required 92 subjects in control arm and 61 subjects in treatment arm with total sample size of 253 (60% of actual total number) were similar to the actual results (135 subjects in control arm and 140 in treatment arm with total sample size of 275) . However, subjects are equally assigned in the real study, while SBARD assign higher subjects in control arm (92 subjects), indicating treatment might inferior than control. Moreover, if the trial is continued until the end of study, we can not conclude that treatment is better than control under all scenarios due to the predictive probabilities and frequentist p-value ($p<0.001$, $p = 0.162$) do not reach pre specific cutoff points too.

Key words: Bayesian adaptive design, interim analysis, monitoring, predictive probability

INTRODUCTION

Among several types of studies, randomized clinical trial is a major point in the top of a pyramid of evidence based medicine (Burns *et al*., 2011). It minimizes the potential for bias by randomization and blinding the subjects into control or study group when different interventions are given. Data are collected and the analysis is performed when the total number of subjects is recruited. The conduct requires budget, time and a certain number of women. These result in a high cost, time consuming and potentially being futile if an intervention is not beneficial or more harmful compared to the control group. Thus, interim analysis while the study is on-going is ethically important to assess toxicity and difference of treatment benefit to make a consideration of an early termination of the study. The Data Monitoring Committee (DMC) or Data Safely Monitoring Board (DSMB) is the group of experts responsible for the review and evaluation of data for safety, progress, efficacy and provide a recommendation to the investigators or sponsor to proceed or stop the trial (Sydes and David, 2010).

Two common approaches are used in data monitoring: group sequential and stochastic curtailment. Both methods can modify the trial plan. Group sequential procedure examines the accumulated data at specific time points and interim analysis is performed by fixing the overall type I error (Jennison and Turnbull, 2000; Emerson *et al*., 2007). In contrast, stochastic curtailment also assesses accumulated data at a certain time point to predict the probability of successful outcomes when there are events which may impact the final outcome. If the probability falls below a pre-specified threshold, the DMC may consider stopping the trial.

The stochastic curtailment methods commonly used in practice are based on conditional power and predictive approaches (Dmitrienko and Wang, 2006; Snapinn *et al*., 2006; Spiegelhalter *et al*., 1986; Lee and Liu, 2008). Conditional power is a frequentist probability of a statistically significant outcome at the end of the trial when data are completely collected. Conditional power is often criticized for relying on the expected probability of success and ignoring new current information identified during the trial. The predictive approach may involve with the power (predictive power) or probability (predictive probability). The predictive power which is widely used is hybrid or mixed Bayesian and frequentist probability. This method allows the investigators to adjust the conditional power function by estimating the outcomes differences during the interim analysis. However, the use of hybrid or mixed Bayesian has been criticized due to various possibilities of interpretation. Furthermore, it is inconsistent with principle of the Bayesian theory that focuses only to the posterior probability (Jennison and Turnbull, 1990; Geisser and Johnson, 1994; Greenhouse and Wasserman, 1995; Bolstad, 2007). The other approach, Predictive Probability (PP), is a pure or fully Bayesian approach which was introduced by Geisser in 1998. The PP is a continuous Bayesian monitoring, obtained by calculating probability of reject null hypothesis should the trial be conducted to the plan maximum sample size, given the current information, the decision making to continue or stop the trial is made based on the PP. Moreover, the PP approach is observed more closely by projecting into the future observed data when comparing with posterior probability (Lee and Liu, 2008). Dmitrienko and Wang (2006) who performed multiple simulation models to evaluate several types of prior probability recommended that predictive approach for futility stopping rules based on 'Weak priors' may be less reliable because the priors were too sensitive leading to negative results. It may be more appropriate in a trial expected that the study arm would yield superior outcome or in a large confirmatory phase III clinical trial. 'Stronger priors' were preferred for futility monitoring in the proof of concept studies. However, the choice of prior distributions should be determined mainly by the objective of the trial (Dmitrienko and Wang, 2006; Snapinn *et al*., 2006; Spiegelhalter *et al*., 1986).

Over the last 20 years, Bayesian Adaptive Designs (BAD) were proposed to enable investigators to modify trials in midcourse. The Bayesian Adaptively Randomization Design (BARD) can assign the patients to a better therapy, early stopping the trial, adding or dropping treatment arms and extending the accrual beyond from the original number when the results was still unsatisfactorily (Berry, 2006; Berry *et al*., 2010; Chow and Chang, 2008). The BARD can be simple or complex with single or multiple outcomes respectively. Although, predictive probability was widely used in both clinical trials phase II and phase III (Lee and Liu, 2008; Sambucini, 2010), the combined use of predictive probability with Bayesian adaptive randomization has been rarely reported (Yin *et al*., 2012).

This study was conducted by using Bayesian stochastic curtailment by a simple Bayesian Adaptive Randomization Design (SBARD) to obtain the predictive probability of a clinical outcome. The clinical outcome was the efficacy of two devices (conventional device and new device) to obtain endometrial tissue in women with abnormal uterine bleeding. The efficacy was a binary endpoint of tissue adequacy. Bayesian predictive probability was compared to the frequentist conditional power approach.

MATERIALS AND METHODS

Background of experimental study in women with abnormal uterine bleeding: A monitoring design was determined for its feasibility in a genuine randomized controlled study. The study was approved by the institutional ethics committee and was registered to the Thai clinical trials registry (study ID TCTR201401080001). Women aged age over 35 years with Abnormal Uterine Bleeding (AUB) in a single institution between January-September 2014 were randomized to have endometrial pathologic evaluation by 2 devices in an out-women setting. Control arm (arm 1) applied conventional device while study arm (arm 2) used a new device. Name of specific devices were withheld for confidentiality reasons. The primary endpoints were binary data of endometrial tissue adequacy (adequate vs inadequate) and pain (pain vs no pain). The conventional sample size calculation was based on the difference of tissue adequacy between the 2 devices. Criterion for significance (alpha) was set at 0.05. For a statistical power at 80%, 194 women were required. Adding with a 20% drop out rate, total sample size in each arm was 214. All authors declared no conflict of interest.

Simple Bayesian Adaptive Randomization Design (SBARD): In this study, tissue adequacy was selected to test the SBARD under PP method.

Assuming p_i were success rates (percentages of tissue adequacy), x_i was number of success and n_i was the number of subjects in each arm. Define arm_i i = 1 was control arm using conventional device, i = 2 was treatment arm using a new device. Based on standard binomial distribution, we had $X_{i\sim}$ binomial (n_i, p_i) with beta prior distribution for p_i.

To illustrate how Bayesian adaptive randomization may work, we first studied a SBARD to test binary response of tissue adequacy by two devices. Women were assigned to receive either treatment (conventional or new device), using an adaptive procedure that based on assignment probabilities. Three scenarios of simulation was conducted to evaluate the posterior probabilities according to the study of Thall and Wathen (2006) and (Yin *et al*., 2012) as in Eq. 1.

$$\pi = \frac{P(p_1 \rangle p_2 \,|X)^{\lambda}}{P(p_1 \rangle p_2 |X)\lambda + P(1-p \,\rangle p_2 |X)^{\lambda}} \quad (1)$$

where, π was posterior probability, l was tuning parameter

The next women was assigned to arm 2 (new device) with probability of π and to arm 1 (conventional device) with probability of 1-π. When l was ∞, the posterior probability by the "Play the winner rule" would allow the next women whom to be assigned to arm 2 would be alternatively assigned to the 'Winner-treatment arm' based on the available current data without randomization. The larger l was the more imbalance of randomization it would be.

A decision rule can be set to compare tissue adequacy between control and treatment arms with various randomization ratios. At an interim phase, predicting the probability of successful outcome of either arm would be considered. If the $p_1 > p_2$ probability was >0.999, the trial could be stopped early with a conclusion that a conventional device was better or the new device was better when $p_2 > p_1$ probability was >0.999. The detail of predictive probability would be described in the following section.

In this model study, 100 subjects were Equally Randomized (ER) into each arm. The next cohort was Adaptively Randomized (AR) by the SBARD (Fig. 1).

Predictive Probability (PP) approach: Assuming p_i was success rates (percentages of tissue adequacy), x_i was number of success and n_i was the total number of subjects in each arm. Defining treatment arm (arm_i, i of 1 was control arm or conventional device and i of 2 was treatment arm or new device. Based on a standard binomial distribution, we had X_i ~ binomial (n_i,p_i) with beta prior distribution for p_i. Posterior distribution of p_i was: $p_i \mid X_i$ ~beta ($\alpha+x_i$,$\beta+n_i- x_i$), If the maximum sample size in arm i was N_i, number of tissue adequacy in the future would be $N_i- n_i$, Y_i~ beta-binomial ($N_i- n_i$, $\alpha+x_i$,$\beta+ n_i- x_i$).

The posterior distribution of tissue adequacy given the current data and future data was in Eq. 2:

$$p_i \,|X_i, Y_i \sim \text{beta}\,(\alpha+x_i+y_i, \beta+ N_i - x_i - y_i\,), \text{ when } Y_i = y_i \quad (2)$$

The decision rules based on the PP were set according to the report of Lee and Liu (2008):

- If $PP<P_L$, the trial could be stopped with the conclusion that the new device was not better than conventional device
- If $PP>P_U$, the trial could be stopped with the conclusion that the new device was superior than the conventional device

A simulation study: The total sample size was set as N= 428 (214 in each arm). The first 50 women were equally randomized to each arm (total 100 subjects). Subsequent women (164 in each arm) were adaptively randomized based on posterior probabilities of success rates which were currently observed (total of 328 subjects). Total of 5,000 simulations were generated to evaluate the performance of SBARD under 3 scenarios. Probability of lower limit (P_L) and upper limit (P_U) were set: $P_L = 0.025$, $P_U = 0.975$, so that the trial would not be terminated early.

From the pilot study among 69 women, the success rates of conventional device and the new device were 78.4% ($p_1 = 0.784$) and 88.9% ($p_2 = 0.889$), respectively. This

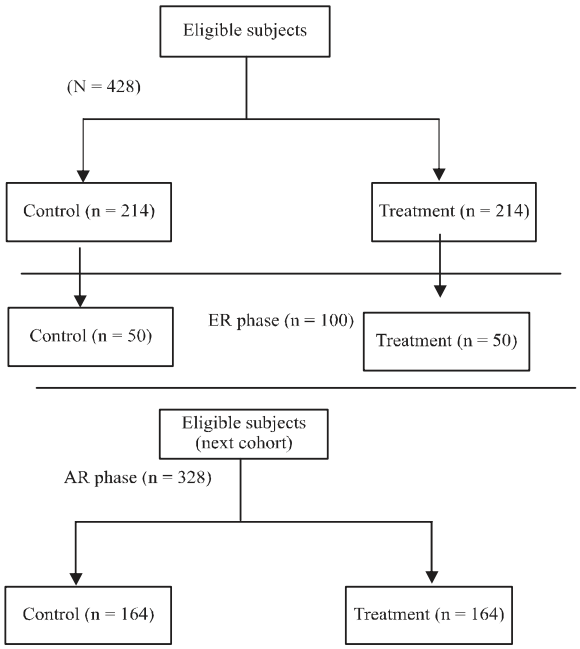

Fig. 1: A schematic of SBARD procedure

information was included into the model as prior data. 5,000 simulations are done to evaluate the performance of SBARD and PP under 3 different probability of response rates (p_i). We set $p_1 = 0.7$ and $p_2 = 0.85$ under scenario 1, $p_1 = 0.8$ and $p_2 = 0.85$ under scenario 2 and $p_1 = 0.8$ and $p_2 = 0.9$ under scenario 3. The performances of the SBARD were evaluated by comparing the average number of subjects used in each arm with their confidence intervals in the SBARD and actual total sample size which including ER phase. In addition, the probability of treatment selection and predictive probabilities were also assessed.

RESULTS

The operating characteristics of the SBARD, probabilities of treatment selection and predictive probabilities are shown in Table 1. Although the treatment arm had higher chance to be selected under three scenarios (p = 0.55, 0.74, 0.58), all scenarios did not yield statistical power of 90%. Scenario 1 was selected due to similar rate of outcome with the actual study. The SBARD required 92 subjects in control arm and 61 subjects in treatment arm with total sample size of 253 from both ER (100) and AR (153) phase (60% of actual total number (253/428)). Unfortunately, the actual study stopped prior to recruitment completion due higher percentages of pain in the study arm (33 women (27.0%) vs. 17 women (14.4%), p = 0.016) without a benefit over a control arm in terms of tissue adequacy (p = 0.162). The actual number of women enrolled was 275 in total (65% of the intention): 135 in control and 140 in study arms. This figure of 275 women was close to the figure in scenario 1 of the SBARD (253 or 60%). One difference was that the women were approximately assigned into each arm in the actual study while SBARD assigned more number of women to the control arm (92 subjects). This indicated that the study arm was inferior than the control in all scenarios (probabilities to select treatment arm = 0.55, 0.74, 0.58) even when the trial was continued until the end of study due to the predictive probabilities did not reach pre-defined cutoff points (p<0.001).

DISCUSSION

To monitor clinical trials by a conventional monitoring design to evaluate the benefit of treatment (Pocock and O'Brien Flemming), there is an increased need which is concerning on cost and time simultaneously. Many investigators and sponsors in clinical trial are aware of these limitations and frequently attempt to modify their trials while maintaining the quality of conduct . Many authors proposed the advantage of BARD and predictive probability in an effort to reduce these problems and encouraged the use of BARD and predictive probabilities in clinical studies (Dmitrienko and Wang, 2006; Snapinn *et al*., 2006; Spiegelhalter *et al*., 1986; Yin *et al*., 2012). The first BARD trial which illustrated the strength of adaptive randomization to avoid ineffective treatment was a report from the university of Texas M.D. Anderson Cancer Center (Yuan *et al*., 2011). The patients with Acute Myeloid Leukemia (AML) underwent adaptive randomization based on dynamic success probabilities measured as complete response from either one of the three chemotherapy regimens. Although, the BAR was proven to be effective to reduce number of subjects to inferior treatment or early stopping of the unsuccessful trial, this trial provoked a considerable controversy over the decision to drop arm one for failing to achieve success after enrolling five patients. The second report of adaptive randomization was published in 2007 by the researchers from Memorial Sloan-Kettering Cancer Center (Maki *et al*., 2007). The objective of the trial was to determine whether the addition of docetaxel to gemcitabine improved the clinical outcome of patients with metastatic soft tissue of sarcoma. The protocol was specified for a dynamic adaptive randomization that was based on the accrual response rate. The adaptive randomization assigned 73 patients (60%) to gemcitabine plus docetaxel and 49 patients to gemcitabine alone, median Progression Free Survival (PFS) was 6.2 months for gemcitabine plus docetaxel and 3 months for gemcitabine alone, indicating gemcitabine

Table 1: Operating characteristics of response rate , predictive probabilities between control and treatment using Simple Bayesian Adaptive Randomization Design (SBARD)

			SBARD				
Scenarios/arms	Category response rates	Sample size ratio (SBARD)	Mean No. of women (95% CI)	Chance to be selected	Predictive probabilities	Sample size required by SBARD	Actual sample size used in both ER and AR phase (% of actual number of women)
Scenario1							
Control	0.70	1.5:1	92 (0, 287)	0.21	<0.0001	153	253 (60%)
Treatment	0.85		61 (1, 113)	0.55	<0.0001		
Scenario2							
Control	0.80	1:1.6	46 (0, 238)	0	<0.0001	120	220 (51%)
Treatment	0.85		74 (1, 348)	0.74	<0.0001		
Scenario3							
Control	0.80	1:1	82 (1, 296)	0.2	<0.0001	165	265 (62%)
Treatment	0.90		83 (1, 243)	0.58	<0.0001		

ER: Equal randomization, AR: Adaptive randomization

plus docetaxel was superior. The author concluded that adaptive randomization was an effective method to reduce the number of subjects receiving inferior therapy. The proposed design proof that uses of SBARD is not only efficient in a simulation study (Supawattanabodee and Ingsrisawang, 2015; Chen *et al*., 2012) but it is feasible in the real trial too. The result of this study also reveal that SBARD requires smaller number of subjects than equally randomization design and correspondence with the previous studies (Yuan *et al*., 2011; Maki *et al*., 2007). Moreover, the PP is efficient in continuous monitoring the trial outcomes has a higher early stopping probability under null hypothesis but the rejection region has a smoother transition when compare with the posterior probability and in common with the previous studies too (Lee and Liu, 2008; Yin *et al*., 2012).

CONCLUSION

SBARD and predictive probability are more efficient than conventional design in reducing number of women being assigned to inferior treatment. The SBARD mainly focuses on the short term response while predictive probability would predict the future outcome. Future study may apply the SBARD with predictive probability to determine long term outcomes, such as, survival in trials of cancer patients.

ACKNOWLEDGMENT

The authors thank Dr. Arinporn Imruetaicharoenchok for provision of data.

REFERENCES

Berry, D.A., 2006. Bayesian clinical trials. Nat. Rev. Drug Discovery, 5: 27-36.

Berry, S.M., B.P. Carlin, J.J. Lee and P. Muller, 2010. Bayesian Adaptive Methods for Clinical Trials. CRC Press, USA., ISBN: 9781439825518, Pages: 323.

Bolstad, W.M., 2007. Introduction to Bayesian Statistics. 2nd Edn., John Wiley and Sons, New Jersey, pp: 317-332.

Burns, P.B., R.J. Rohrich and K.C. Chung, 2011. The levels of evidence and their role in evidence-based medicine. Plastic Reconstr. Surg., 128: 305-310.

Chen, Z., Y. Zhao, Y. Cui and J. Kowalski, 2012. Methodology and application of adaptive and sequential approaches in contemporary clinical trials. J. Prob. Stat., Vol. 2012. 10.1155/2012/527351

Chow, S.C. and M. Chang, 2008. Adaptive design methods in clinical trials: A review. Orphanet J. Rare Dis., Vol. 3. 10.1186/1750-1172-3-11

Dmitrienko, A. and M.D. Wang, 2006. Bayesian predictive approach to interim monitoring in clinical trials. Stat. Med., 25: 2178-2195.

Emerson, S.S., J.M. Kittelson and D.L. Gillen, 2007. Frequentist evaluation of group sequential clinical trial designs. Stat. Med., 26: 5047-5080.

Geisser, S. and W. Johnson, 1994. Interim analysis for normally distributed observables. Lecture Notes-Monograph Ser., 24: 263-279.

Greenhouse, J.B. and L. Wasserman, 1995. Robust Bayesian methods for monitoring clinical trials. Stat. Med., 14: 1379-1391.

Jennison, C. and B.W. Turnbull, 1990. Statistical approaches to interim monitoring of medical trials: A review and commentary. Stat. Sci., 5: 299-317.

Jennison, C. and B.W. Turnbull, 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman and Hall, New York, USA., ISBN-13: 9780849303166, Pages: 390.

Lee, J.J. and D.D. Liu, 2008. A predictive probability design for phase II cancer clinical trials. Clin. Trials, 5: 93-106.

Maki, R.G., J.K. Wathen, S.R. Patel, D.A. Priebat and S.H. Okuno *et al*., 2007. Randomized phase II study of gemcitabine and docetaxel compared with gemcitabine alone in patients with metastatic soft tissue sarcomas: Results of sarcoma alliance for research through collaboration study 002. J. Clin. Oncol., 25: 2755-2763.

Sambucini, V., 2010. A Bayesian predictive strategy for an adaptive two-stage design in phase II clinical trials. Stat. Med., 29: 1430-1442.

Snapinn, S., M.G. Chen, Q. Jiang and T. Koutsoukos, 2006. Assessment of futility in clinical trials. Pharmaceut. Stat., 5: 273-281.

Spiegelhalter, D.J., L.S. Freedman and P.R. Blackburn, 1986. Monitoring clinical trials: Conditional or predictive power? Controlled Clin. Trials, 7: 8-17.

Supawattanabodee, B. and L. Ingsrisawang, 2015. Bayesian adaptive randomization designs for clinical trial. J. Applied Sci., 15: 374-376.

Sydes, M.R. and D. Neal, 2010. Data monitoring committees in clinical trials: Guidance for research ethics committees. National Patient Safety Agency, May 2010. http://www.hra.nhs.uk/documents/2013/10/data-monitoring-committees-in-clinical-trials.pdf.

Thall, P.F. and J.K. Wathen, 2006. Practical Bayesian adaptive randomization in clinical trials. UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series, Paper 31, University of Texas, MD Anderson Cancer Center, USA. http://biostats.bepress.com/cgi/viewcontent.cgi?article=1030& context= mdandersonbiostat

Yin, G., N. Chen and J.J. Lee, 2012. Phase II trial design with Bayesian adaptive randomization and predictive probability. J. R. Stat. Soc.: Ser. C (Applied Stat.), 61: 219-235.

Yuan, Y., X. Huang and S. Liu, 2011. A Bayesian response-adaptive covariate-balanced randomization design with application to a leukemia clinical trial. Stat. Med., 30: 1218-1229.