A Bayesian Up-and-Down Design for Clinical Trials with Multiple Adverse Events in Opposite Directions

Yiyi Chen, Zunqiu Chen, and Matthew F. Halsey

Abstract:

Displaced supracondylar humerus (SCH) fractures are common in children, typically treated by closed reduction percutaneous pinning (CRPP) if displaced. While the timing of pin removal is critical for treatment success, no guideline or recommendation on this is available. This is partially because it is not ethical to randomize patients to removing pins at a time that is different from convention. Three potential "adverse events" are associated with the timing of pin removal for SCH fractures in children: Infection, Displacement and Stiffness. The risk of having infection and/or stiffness increases with longer durations of pin placement while early pin removal may lead to displacement. We propose a two-stage Bayesian group sequential design to walk up-and-down based on the most up-to-date observations from the study until concluding in the optimal time for pin removal. Five candidate time windows (19-21, 22-24, 25-27, 28-30, 31-35 days) are considered, and a weighted loss function is used to determine the next stochastic walk movement: window escalation (E), de-escalation (D) or staying (S) at the current window. The trial is allowed to stop early once "equilibrium" window has been found. Simulation study is conducted to find the expected sample size and the operating characteristics of the design.

Introduction

Displaced supracondylar humerus (SCH) fractures (Gartland types II and III) represent one of the most common injuries in children[1]. As such, the American Academy of Orthopedic Surgeons (AAOS) has developed a Clinical Practice Guideline (CPG) outlining best practices for these injuries based on the available evidence in the medical literature.[2] Timing of pin removal after closed reduction percutaneous pinning was one of the critical aspects of the SCH fracture treatment that was evaluated: too early removal may lead to subsequent fragment movement and loss of reduction; later removal may increase the risk of pin infection and/or elbow stiffness. Several studies have documented timing of pin removal. Gaston et al. reported a mean of 26.5 days for pin removal time[3], while the mean time to removal was 30.3 days with a range of 20-46 days in Brauer et al study[4]. In another prospective study conducted by Lobst et al., a follow up radiograph was obtained 3-4 weeks post op, and if the fracture had healed, the pins were removed[5]. Despite these studies, no optimal time for removal was given, nor were there any criteria for removal of pins. As pointed out by the CPG, there is no conclusive evidence for the optimal timing of pin removal in patients with displaced SCH fractures.

We propose to conduct a clinical trial to find the optimal timing of pin removal. A standard randomized clinical trial (RCT) is neither feasible nor ethical to establish appropriate timing for pin removal. For example, consider a RCT that randomizes subjects to one of the three treatment groups (early removal time, mid removal time, late removal time), it would be difficult to recruit subjects because patients would be reluctant to be randomized to the early or late removal time due to potential excessive risks associated with early or late pin removal. Such a trial will also be hard to accept for clinical investigators from ethical consideration. To solve the problem, we propose a Bayesian outcome-adaptive design that starts from the conventional time points, and then

"move" patients up or down the time scales adaptively based on the most up-to-date observations from the study until concluding in the optimal time points for pin removal. This innovative design should help to establish a reasonable time period for pin removal within the "conventional" timeframe while allowing for movement to a shorter or longer time period as determined by the incoming data points.

It is worth noting that unlike a traditional clinical trial with only single primary endpoint, this study has three primary endpoints (fragment movement, pin infection and elbow stiffness) to be considered simultaneously. In addition, the chance of fragment movement decreases monotonically with increased time, while the probabilities for pin infection and elbow stiffness increase monotonically with time. These constraints should be modeled appropriately in the design. The optimal pin removal time should be one with minimal joint risk of the three events.

Methods

Clinical Trial design

This is a non-randomized Bayesian two-stage outcome-based adaptive clinical trial to find optimal pin removal time for patients with displaced SCH fractures. We hypothesize that the optimal time is in the range of 19-35 days. Due to the scheduling difficulty to assign subjects to a particular date for pin removal after injury, we will assess five pin removal time windows instead: 18-21 days, 22-24 days, 25-27 days, 28-30 days and 31-35 days. Patients will be enrolled sequentially once consent is obtained, and be assigned to a window that is adaptively determined by most up-to-date observations.

There are three primary endpoints for this study, all reflect potential complications associated with timing of pin removal simultaneously.

- Fragment movement and loss of reduction (X)
- Pin infection (Y)
- Elbow stiffness (Z)

We use X, Y and Z to represent the three primary outcome variables. For any subject i, we observe (X_i, Y_i, Z_i, j), where (X_i, Y_i, Z_i) is the outcome for this subject, and j is the time window to which patient i is assigned to. X_i, Y_i and Z_i are all binary variables, which means they can take only one of the two values: 0 (represents NO) or 1 (represents YES). For example, if the ith subject had infection and stiffness, but no displacement, then $X_i = 0$, $Y_i = 1$ and $Z_i = 1$.

The parameters of primary interest are p_j, q_j and r_j , j = 1,…, 5. These correspond to the average probabilities of having X, Y and Z, respectively, if pins are removed at Window j for a typical patient.

The trial is composed of two stages. Stage 1 is a rule-based stage where an initial cohort of 4 patients will start from Window 2(22-24 days). The outcomes for the first cohort will be observed, and the window assignment for the next cohort is determined based on the rules below:

- If $\sum_i w_1 X_i - \sum_i w_2 Y_i + \sum_i w_3 Z_i = 0$ for patients in this cohort, then assign the next cohort of patients to the current window.

- If $\sum_i w_1 X_i - \sum_i w_2 Y_i + \sum_i w_3 Z_i < 0$ for patients in this cohort, then assign the next cohort of patients to Window $W_{\max(1,j-1)}$.
- If $\sum_i w_1 X_i - \sum_i w_2 Y_i + \sum_i w_3 Z_i > 0$ for patients in this cohort, then assign the next cohort of patients to Window $W_{\min(5,j+1)}$.

for all i in the current cohort. The w_1, w_2 and w_3 are the weights assigned to events X, Y and Z to reflect the corresponding relative seriousness of each events. They can be easily elicited from clinical investigators. For the young population in this study, it is expected that the seriousness of infection(X) and displacement (Y) are about the same, while the stiffness(Z) is less serious. Therefore, we set $w_1 : w_2 : w_3 = 1 : 1 : 0.4$ for this study. The assignment of the next cohort is completely determined by the observations in the four subjects of the current cohort.

The goal of Stage 1 is to collect necessary information to reliably estimate posterior means for p_j, q_j, and r_j . It also helps to find a nice starting window for Stage 2. Stage 1 will end after 7 cohorts (e.g., 28 subjects). The trial will then enter Stage 2, the model-based stage.

In Stage 2, the first cohort will be assigned to the window suggested by the last cohort in Stage 1. Afterwards, the posterior probability of having each outcome at different dose levels will be estimated upon the observations for the current cohort become available. All up-to-date outcomes from subjects enrolled in the trial (including both Stages) will be used for estimating the posterior distributions. The window assignment for the subsequent cohort is determined based on the loss function.

For Stage 2, the loss function for time window j is defined as:

$$L_j = w_1 E(p_j) + w_2 E(q_j) + w_3 E(r_j)$$

where $E(p_j), E(q_j)\ and\ E(r_j)$ are the posterior estimates for p_j, q_j, and r_j based on the most up-to-date observations.

For subjects (except for the 8th cohort, the first cohort in Stage 2) in Stage 2, in principle, the cohort of four patients will be assigned to the Window with the minimum expected loss based on most up-to-date observations. However, we impose a non-skipping rule for assignment to avoid aggressive jumps. The rule prevents any jumping across windows. That means if the current cohort of patients was assigned to Window j, then the next potential cohort could only be $W_{j-1}, W_j, or\ W_{j+1}$ for j = 2, 3, 4. For j = 1, the option will be limited to $W_j, or\ W_{j+1}$. Similarly, for j = 5, the option will be limited to W_{j-1} or W_j. Therefore, patients will be assigned to the window with minimum expected loss among potential options.

Stopping rule

This trial will be stopped if one of the following conditions is satisfied:

- Four successive cohorts of patients (i.e., 16 patients) were assigned to the same window in a row for Stage 2 from the 9th cohort. This suggests that the model-based optimal window assignment has reached equilibrium state.

- The maximum sample size of 100 has been reached. This is the maximum resource the study group could afford for this clinical trial.

Statistical model

let $\pi_{abc} = \Pr(X = a, Y = b, Z = c)$, where $a \in \{0,1\}; b \in \{0,1\}; c \in \{0,1\}$, the likelihood function is $l(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{5} \prod_{a=0}^{1} \prod_{b=0}^{1} \prod_{c=0}^{1} (\pi_{abc}^{j})^{I(x_i^j=a, y_i^j=b, z_i^j=c)}$.

As we expect the outcomes across the time windows will be low, we choose not to model the potential association among outcomes to keep the model simple.

Letting $logit(p_j) = \eta_{p_j}$, $logit(q_j) = \eta_{q_j}$ and $logit(r_j) = \eta_{r_j}$, we assume η_{p_j}, η_{q_j}, and η_{r_j} each follows normal distribution $N(\mu_p, \sigma^2), N(\mu_q, \sigma^2), N(\mu_r, \sigma^2)$, respectively. We further assume that μ_p, μ_q and μ_r follow $N(\mu, \sigma_\mu^2)$, and $\sigma \sim Unif(0,100)$. The hyperparameters μ, σ_μ^2 were set to be $\mu = -2$ and $\sigma_\mu^2 = 10$. Order constraints were placed onto the posterior estimates of the probability parameters to ensure $p_j \leq p_{j+1}$, $q_j \geq q_{j+1}$, and $r_j \leq r_{j+1}$ [6].

Results

Sample size consideration and Design assessment

Due to the adaptive nature of the design, we estimate the expected required sample size and assess the performance of the study design through a simulation study. Five Scenarios are considered (shown in Table 1), based on each we simulated 300 trials to assess the average performance of the design in various situations. The maximum allowed sample size is 100 for all simulated trials.

The primary interest for the design performance is the probability of concluding in optimal and sub-optimal windows by the end of the trial. We are also interested in looking at the expected sample size (defined as the 3rd quartile of the total sample size among all simulated trials for this study). Given acceptable performance, we would like to see more patients be assigned to the optimal or sub-optimal windows within the trial. These performances are shown in Table 2.

The first thing to notice is that the two-stage design reaches its stabilized position fairly quickly, and consistently for all five scenarios: the expected total sample size is 48 subjects, with 28 subjects in Stage 1 and 20 subjects in Stage 2. In all Scenarios, the trial stopped early, with the maximum simulated sample size = 80.

As for the probability of concluding in optimal window, Scenarios 2 and 4 both had greater than 80% chances of concluding in the optimal window. The chances of correctly concluding in the optimal windows for Scenarios 1, 5 and 3 are 72%, 61% and 48%, respectively. If we take into consideration the chance of concluding in the sub-optimal (i.e., second best) window, then Scenarios 1, 2 and 4 had 100%, 99%, and 97% chances of concluding in the best two windows. The chances are 87% and 84% for Scenarios 3, and 5.

The difference in performances for Scenarios results primarily from the distance in loss between the optimal and suboptimal windows. For Scenarios 2 and 4, the distances are

0.16, and 0.14, respectively. The distances are 0.12, 0.09, and 0.05 for Scenarios 1, 5 and 3, respectively.

When looking at the proportion of subjects been assigned to the optimal or suboptimal windows, the majority of subjects are assigned to the two best windows for all five scenarios. In particular, 88%, 94% and 87% of subjects in the trial will be assigned to the best two windows for Scenarios 1, 2, and 4, and 55% and 60% of those will be assigned to Scenarios 3 and 5.

We also compared the performance of the proposed design with two other designs: the up-and-down rule-based design (i.e., the design with Stage 1 component only), and the up-and-down Bayesian model-based design (i.e., the design with Stage 2 component only). The performances of these two designs are shown in Tables 3 and 4.

While imposing the same stopping rules to both designs, the up-and-down rule-based design requires a much bigger sample size when compared with either the one-stage Bayesian design or our proposed two-stage design. In addition, its overall performance is no better (if not worse) than the other two designs. Therefore, it is not our choice. The performance of one-Stage up-and-down Bayesian model-based design is closer to the two-stage design, but with some issues that make us think it is not as favorable as the two-stage design. First, the chance of concluding in the optimal window is much lower for the one-stage design except for Scenario 1. Second, the expected sample size is larger for the one-stage designs. Third, the proportion of patients been assigned to the best two windows is lower for the one-stage design.

Discussion

We have proposed a two-stage "Up-and-Down" Bayesian adaptive design to find an optimal pin removal time for SCH fractures in children. The adaptive design for the study can be easily extended to other pediatric orthopedic treatments such as timing of bracing in adolescent scoliosis or cat removal in pediatric femur fracture management. Outside of pediatric orthopedics, the method could be used to provide evidence for other treatments not amenable to RCT such as timing of physical therapy and surgical interventions where there is some amount of clinical equipoise and lack of comparative evidence. It is conservable that this method could be applied to questions in cardiology, oncology and other surgical subspecialties such as trauma and gynecology.

There are multiple options that could be considered in designing this study. For example, we could model the functional forms for p_j , q_j and r_j differently, and could allow correlation among them. We could also consider different size for Stage 1. An exploration on the effect for Stage 1 size to the performance of the design and to the expected sample size of the design can be interesting. In addition, smaller sized window may be considered, and more scenarios could be assessed.

The conduct of the adaptive design can also be more demanding compared to RCT as it assigns patients adaptively to a time window based on observations from previous patients. An interactive software is necessary for a timely determination of patient assignment.

References:

1. Hamdi A, Poitras P, Louati H, Dagenais S, Masquijo JJ, Kontio K (2010). Biomechanical Analysis of Lateral Pin Placements for Pediatric Supracondylar Humerus Fractures. *Journal of Pediatric Orthopaedics*, 30(2): 135-139.
2. American Academy of Orthopaeic Surgeons (2011). The Treatment of Pediatric Supracondylar Humerus Fractures. http://www.aaos.org/research/guidelines/SupracondylarFracture/SupConFullGuideline.pdf
3. Gaston RG, Cates TB, Devito D, Schmitz M, Schrader T, Busch M, Fabregas J, Rosenberg R, Blanco J (2010). Medial and Lateral Pin Versus Lateral-Entry Pin Fixation for Type 3 Supracondylar Fractures in Children. Journal of Pediatric Orthopaedics. 30(8): 799-806. doi: 10.1097/BPO.0b013e3181f73d59.
4. Brauer CA, Lee BM, Bae DS, Waters PM, Kocher MS (2007). A Systematic Review of Medial and Lateral Entry Pinning Versus Lateral Entry Pinning for Supracondylar Fractures of the Humerus." Journal of Pediatric Orthopaedics. 27(2):181-186.
5. Lobst CA, Spurdle C, King WF, Lopez M (2007). Percutaneous Pinning of Pediatric Supracondylar Humerus Fractures with the Semisterile Technique. Journal of Pediatric Orthopaedics. 27(1) 17-22.
6. Gunn LH, Dunson DB (2005) A transsformatin approach for incorporating monotone or unimodla constraints. Biostatistics. 6(3): 434-449.

Tables

Table 1: Five Scenarios considered for design assessment						
Scenarios		W1	W2	W3	W4	W5
1	p	0.05	0.08	0.1	0.25	0.3
	q	0.5	0.2	0.05	0.04	0.03
	r	0.05	0.08	0.1	0.2	0.3
	Loss	0.57	0.31	**0.19**	0.37	0.45
2	p	0.05	0.2	0.25	0.27	0.3
	q	0.1	0.09	0.08	0.06	0.05
	r	0.05	0.1	0.15	0.2	0.3
	Loss	**0.17**	0.33	0.39	0.41	0.47
3	p	0.02	0.03	0.04	0.05	0.05
	q	0.35	0.3	0.25	0.15	0.1
	r	0.05	0.1	**0.1**	0.1	0.1
	Loss	0.39	0.37	0.33	0.24	**0.19**
4	p	0.05	0.05	0.15	0.15	0.2
	q	0.3	0.1	0.1	0.08	0.07
	r	0.04	0.05	0.15	0.2	0.3
	Loss	0.37	**0.17**	0.31	0.31	0.39
5	p	0.05	0.06	0.08	0.09	0.14
	q	0.3	0.25	0.11	0.08	0.07
	r	0.05	0.06	0.08	0.1	0.15
	Loss	0.37	0.33	0.31	**0.19**	0.28

Table 2: Performance of the two-stage Design

Scenarios	Assigned to Opt (%)	Assigned to Sub-op (%)	Concluding in Opt (%)	Concluding in Sub-opt (%)	median N (range)	3rd Quartile N
1	47	42	72	28	48(48-80)	48
2	62	32	86	13	48(48-64)	48
3	32	23	48	39	48(48-68)	48
4	63	24	83	14	48(48-76)	48
5	17	43	61	23	48(48-72)	48

Table 3: Performance of the One-Stage Up-and-down Rule-based Design

Scenarios	Assigned to Opt (%)	Assigned to Sub-op (%)	Concluding in Opt (%)	Concluding in Sub-opt (%)	median N (range)	3rd Quartile N
1	37	44	47	39	100(40-100)	100
2	58	36	89	1	56(40-100)	81
3	30	33	72	20	72(40-100)	100
4	51	30	67	20	100(40-100)	100
5	21	38	25	43	100(40-100)	100

Table 4: Performance of the One-Stage Up-and-down Bayesian Model-based Design

Scenarios	Assigned to Opt (%)	Assigned to Sub-op (%)	Concluding in Opt (%)	Concluding in Sub-opt (%)	median N (range)	3rd Quartile N
1	41	31	80	20	44(40-100)	56
2	37	37	79	18	44(40-100)	56
3	23	33	44	51	44(40-100)	60
4	39	44	79	21	48(40-100)	56
5	25	40	38	48	56(40-100)	72