BIOSTATISTICS METHODS

FOR TRANSLATIONAL & CLINICAL RESEARCH

EARLY PHASE CLINICAL TRIALS

Part B: Selected Statistical Issues

Standard Design Characteristics

STANDARD DESIGN

- Start at lowest dose (of 5-7 doses), enroll groups of 3 pts

(i) Move up if none of first 3 have toxicity;

(ii) If two or three patients have dose-limiting toxicity (DLT), stop.

(iii) If one of 3 has toxicity, enroll 3 more at same dose and move up if none of the second cohort have toxicity. Otherwise, stop

- Moving up, repeat the process at new dose

If r is the toxicity rate of the current dose, the probability of escalating after only 3 patients is (none with tox):

$$UP_3 = (1\text{-}r)^3$$

If r is the toxicity rate of the current dose, the probability of stopping after only 3 patients is (2 or 3 with tox):

$$STOP_3 = 3r^2(1-r) + r^3$$

And probability to stop (3 or 6 patients):

$$STOP_{3/6} = [3r^2(1-r) + r^3] + [3r(1-r)^2][1-(1-r)^3]$$

The probability that the second cohort of 3 patients enrolled/needed is

$$NEED_6 = 1 - (UP_3 + STOP_3)$$

Rate, r	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$STOP_{3/6}$	0.03	0.09	0.29	0.51	0.69	0.83	0.92	0.97
UP_3	0.86	0.73	0.51	0.34	0.22	0.13	0.06	0.03
$STOP_3$	0.01	0.03	0.10	0.22	0.35	0.50	0.65	0.78
$NEDD_6$	0.13	0.24	0.39	0.44	0.43	0.37	0.29	0.19

(1) The cohort is limited to 3 pts each with ≥56% probability: It requires few patients

(2) For more extreme DLT probabilities (5% & 70%), the cohort is only expanded to 6 patients with probability of less than 20%

(3) When the rate is 30% or higher, the probability to stop is ≥51%: It's rather safe

Rate, r	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$STOP_{3/6}$	0.03	0.09	0.29	0.51	0.69	0.83	0.92	0.97
UP_3	0.86	0.73	0.51	0.34	0.22	0.13	0.06	0.03
$STOP_3$	0.01	0.03	0.10	0.22	0.35	0.50	0.65	0.78
$NEDD_6$	0.13	0.24	0.39	0.44	0.43	0.37	0.29	0.19

(1) If 2 or 3 pts show DLT ($STOP_3$), we are 90% sure that $r \geq 20\%$

(2) If no patients show DLT (UP_3), we are 90% sure that $r \leq 55\%$

(3) If r=10%, 91% chance to escalate; if r=60%, 92% chance to stop

Fast-Track Design

- This design was created to move through low doses using fewer patients. The design uses cohorts of one or three patients, escalates through the sequence of doses using a one-patient cohort until the first DLT is observed. After that, only three-patient cohorts are used.
- When a DLT is observed in a one-patient evaluation of a dose, the same dose is evaluated a second time with a cohort of three new patients, if no patient in this cohort experiences a DLT, the design moves to the next higher dose with a new cohort of three patients. From this point, the design progresses as a standard design.

If one-patient cohort is used at each dose level throughout, then six (6) patients are always tested at the very last dose.

ABOUT FAST-TRACK DESIGN

The use of a fast-track design seems attractive because most clinicians want to proceed, as fast as they can. The fast-track design quickly escalates through early doses, thereby reducing the number of patients. On the other hand, the fast-track design might allow a higher percentage of patients to be treated at very high toxic doses; and it uses a single-patient cohort until the first DLT is observed seems too risky for some investigators.

MORE RESEARCH QUESTIONS

Here are 3 basic more three research questions:

(1) What is the actual expected toxicity rate of the MTD selected by the standard design,

(2) Is the expected toxicity rate of the selected MTD by the standard design robust, and

(3) Can improvement be made to the standard design so that we can get to the next phase quicker but still safe enough? What are alternatives to the Standard Design which could help to accelerate the process?

OTHER CRITERIA

Different designs maybe judged on:

(1) Toxicity rate of the selected MTD
(2) Expected trial size, and
(3) Expected sub-optimal trial sizes.

If r_i is the toxicity rate of the dose i, the probability of stopping at dose i, after 3 or 6 patients is :

$$STOP_{3/6} = [3r^2(1-r) + r^3] + [3r(1-r)^2][1-(1-r)^3]$$

$$= p_{[i]}$$

If r_i is the toxicity rate of the dose i and $p_{[i]}$ the probability of stopping at dose i, the probability of escalating to dose i,

$$p_{(1)} = 1$$

$$p_{(2)} = 1 - p_{[1]}$$

$$p_{(i)} = p_{(i-1)}[1 - p_{[i-1]}] \text{ for } i = 2,3,\ldots$$

Then we have,

$$\text{Expected Toxicity Rate of MTD} = \sum_i r_i p_{[i+1]}$$

(Recall: $p_{[i+1]}$, the probability of stopping at dose (i+1), is also the probability that dose i is selected as MTD)

and we have,

$$\text{Expected Trial Size} = \sum_i p_{(i)}[(3)(UP_3 + STOP_3) + (6)(NEED_6)]$$

AN ILLUSTRATION

For simplicity, let consider seven scenarios in which each scenario is assumed to be composed of a sequence of seven doses with toxic rates increasing by 5%, the performances of the designs with each are evaluated through exhaustive enumeration. The scenarios differ by the toxic rate of the initial dose.

AN ILLUSTRATION

Consider sequences of seven doses with toxic rates increasing by 5%. For example,

(1) Toxic rates for the first scenario is 40%, 45%, 50%, 55%, 60%, 65%, and 70%;

(2) Toxic rates for the first scenario is 35%, 40%, 45%, 50%, 55%, and 65%; and …

(7) Toxic rates for the last scenario is 10%, 15%, 20%, 25%, 30%, 35%, and 40%.

Standard Design: TOXICITY

Scenario	Toxicity Rates	Expected Rate
1	40%-70%	42%
2	35%-65%	38%
3	30%-60%	34%
4	25%-55%	30%
5	20%-50%	27%
6	15%-45%	25%
7	10%-40%	23%

- **Expected values of the toxic rate range from 23%-42%; most are at/below the lower end of the 30%-40% range.**
- **The standard design is not robust; expected rate of the selected MTD is strongly influenced by the doses used; these doses are selected arbitrarily and received very little attention from reviewers or DSMB.**

Standard Design: SIZES

Scenario	Toxicity Rates	Trial Size	Over- Size	Under- Size
1	40%-70%	6.0	1.7	0.0
2	35%-65%	6.7	0.7	0.0
3	30%-60%	7.6	0.3	0.0
4	25%-55%	8.8	0.2	4.3
5	20%-50%	10.4	0.1	7.2
6	15%-45%	12.4	0.1	9.8
7	10%-40%	14.8	0.0	12.6

- **Expected trial size has a range of 6.0-14.8. Expected numbers of patients being over-treated are low; majority of patients are safe. Expected under-treated trial sizes are a large; an unsatisfactory use of patient resources.**
- **Given these results, the standard design can be regarded as a conservative design; maybe it's "too conservative"**

Fast-track Design: TOXICITY

Scenario	Toxicity Rates	Expected Rate
1	40%-70%	47%
2	35%-65%	43%
3	30%-60%	40%
4	25%-55%	37%
5	20%-50%	34%
6	15%-45%	32%
7	10%-40%	30%

- **Expected values for the toxic rate range from 30% to 47%; the range is narrower than that of the standard design.**
- **Expected rate of the selected MTD is still not quite robust;**
- **In addition, this design has higher expected toxic rates in two high dose level scenarios (i.e., 43% and 47%).**

Fast-track Design: SIZES

Scenario	Toxicity Rates	Trial Size	Over- Size	Under- Size
1	40%-70%	6.1	3.9	0.0
2	35%-65%	6.6	2.7	0.0
3	30%-60%	7.2	2	0.0
4	25%-55%	8.0	1.5	1.8
5	20%-50%	8.9	1.2	3.4
6	15%-45%	9.9	1	5
7	10%-40%	10.9	0.0	6.5

- **Expected trial size ranged from 6.1-10.9, and more stable. This design decreases the patients needed, especially in the low-dose-level scenarios.**
- **Expected under-treated trial size are smaller than those of the standard design; an improved use of patient resources.**
- **Expected over-treated trial size are larger than those seen in the standard design; safety could be a problem with inexperienced investigators.**

Design For Selection

The Setting:

(1) There are no Standard/Placebo; both treatments (or modes of treatment) A & B are experimental.

(2) Decision (i.e. selection) has to be made; does not fit framework of "statistical test of significance" where "not statistically significance" is a possibility. We cannot afford it!

CRITERION

- If the "observed outcome" (e.g. response rate, but could some sample mean) of one arm is greater than "d units" than the other, the arm with "better observed outcome" (larger proportion or larger sample mean) will be selected for use in the next phase III trial.
- If the difference is smaller than d units ("d" may or may not be 0), selection may be based on other factors
- For example:

$$\text{If } "p_A - p_B > d\%"; \text{ treatment A is selected}$$

Or

$$\text{If } "\bar{x}_A - \bar{x}_B > d"; \text{ treatment A is selected}$$

CORRECT OUTCOME

- **Suppose that the outcome variable is response rate and Treatment A is assumed to be better:**

$$\pi_A - \pi_B = \delta$$

- **The "probability of correct outcome" is:**

$$P_{corr} = \Pr[p_A - p_B > d \mid \pi_A, \pi_B]$$

CORRECT SELECTION

- If the "observed outcome is ambiguous", i.e. difference is less than "d", treatment A could still be chosen (by factors other than efficacy), with – say - probability ρ;
- The probability of correct selection is:

$$P_{corr} = \Pr[\, p_A - p_B > d \mid \pi_A, \pi_B \,]$$

$$P_{Amb} = \Pr[\, p_B - d \le p_A \le p_B + d \,]$$

$$\lambda = P_{CorrSel} = P_{Corr} + \rho P_{Amb}$$

WHAT DO STATISTICIANS DO?

- The size of "d" is a clinical decision; at the end of the trial, compute ($p_A - p_B$) and compare to d.
- Statistician is responsible for "the design", to find sample size n (per arm) to ensure that "the probability of correct selection" exceeding certain threshold; say $\lambda \geq .90$ (similar to power).
- Population parameters (such as π_A and π_B, or π_A and δ) are in "Alternative Hypothesis"; ideas from separate phase II trials.
- We cover the case of response rate but method is applicable to continuous outcome variables.

The difference $(p_A - p_B)$ is distributed as Normal with Mean μ and Variance $= \sigma$,

$$\mu = (\pi_A - \pi_B) = \delta$$

$$\sigma^2 = \frac{1}{n}[\pi_A(1-\pi_A) + \pi_B(1-\pi_B)]$$

$$P_{corr} = \Pr[p_A - p_B > d \mid \pi_A, \pi_B]$$

$$= 1 - \Phi[\frac{d - \delta}{\sigma}]$$

$$P_{Amb} = \Pr[-d \le p_A - p_B \le d]$$

$$= \Phi[\frac{d - \delta}{\sigma}] - \Phi[\frac{-d - \delta}{\sigma}]$$

$$\lambda = P_{CorrSel} = P_{Corr} + \rho P_{Amb}$$

Example: Let take $\pi_A = .35$ and $\pi_B = .25$ (or $\delta = .10$) and d = .05, n = 50 ($\sigma = .09$)

$$P_{corr} = 1 - \Phi[\frac{d-\delta}{\sigma}]$$

$$= 1 - \Phi(-.55) = .71$$

$$P_{Amb} = \Phi[\frac{d-\delta}{\sigma}] - \Phi[\frac{-d-\delta}{\sigma}]$$

$$= \Phi[-.55] - \Phi[-1.65] = .24$$

$$\lambda = P_{CorrSel} = P_{Corr} + \rho P_{Amb}$$

$$= \begin{cases} .71 \; for \; \rho = 0 \\ .83 \; for \; \rho = .5 \end{cases}$$

Changing n will change σ and, therefore, the probability of correct selection λ; that's key to sample size determination

If it was a "test of significance', at the conclusion of the trial, one would compute the sample proportions and reject the Null Hypothesis if:

$$\frac{|p_A - p_B|}{\sigma} \geq z_{\alpha/2}$$

The statistical power of this test would be, which is very different from the probability of correct selection:

$$1-\beta = 1-\Phi[z_{\alpha/2} - \frac{\delta}{\sigma}] + \Phi[-z_{\alpha/2} - \frac{\delta}{\sigma}]$$

Example:

$\pi_A = .35$, $\pi_B = .45$ ($\delta = .10$), d = .05 & n = 100 $P_{corr} = .88$; $P_{corr} + (.5)P_{Amb} = .93$

$$\sigma = \sqrt{\frac{.35(1-.35)+.45(1-.45)}{100}}$$

$$= .07$$

$$Power = 1 - \Phi[z_{\alpha/2} - \frac{\delta}{\sigma}] + \Phi[-z_{\alpha/2} - \frac{\delta}{\sigma}]$$

$$= 1 - \Phi(1.96 - \frac{.10}{.07}) + \Phi(-1.96 - \frac{.10}{.07})$$

$$= 1 - \Phi(.53) + \Phi(-3.39)$$

$$= 1 - .7019 + .0003$$

$$= .2984$$

Adverse Effects Monitoring

SEQUENTIAL STOPPING RULE

- In testing for null the hypothesis H_0: $\pi = \pi_0$ against the alternative H_A: $\pi = \pi_A$, the decision is:

(i) to stop the trial and reject H_0 if $LR_n \geq \ln(1-\beta)-\ln\alpha$

(ii) to stop the trial and accept H_0 if $LR_n \leq \ln\beta-\ln(1-\alpha)$

(iii) continue the study otherwise

- In (i) there are too many events and in (ii) there are too few events - enough to make a decision.

SIDE EFFECTS MONITORING

- We do not stop the trial because there are too few events; we only stop the trial early for an excess of side effects, that is when:
  $e(\ln\pi_A - \ln\pi_0)+(n-e)[\ln(1-\pi_A)- \ln(1-\pi_0)]\geq \ln(1-\beta)-\ln\alpha$
- The lower boundary is ignored; trial continues
- Solving equation for “e” yields for upper boundary
- We can also solve the same equation for n .

RESULT

Stop the trial as soon as n, as a function of e, satisfies the following equation:

$$n(e) = \frac{\ln(1-\beta) - \ln\alpha + e[\ln(1-\pi_A) - \ln(1-\pi_0) - \ln\pi_A + \ln\pi_0]}{\ln(1-\pi_A) - \ln(1-\pi_0)}$$

n(e) is the number of evaluable patients for having e of them with adverse effects.

Rule: To stop the trial when we have “e” adverse effects before reaching a total of “n(e)” patients.

EXAMPLE

- **Consider a simple case where we know that the baseline rate is π_0 = .03 or 3% and investigator sets a ceiling rate of π_A = .15 or 15%.**
- **If we pre-set the level of significance at α= .05 and plan to reach of statistical power of 80% (β=.20), the the trial should be stop as soon as: n(1)=-7.8, n(2)=5.4, n(3)=18.6, n(4)=31.8 etc… rounding off to {-, 5,18, 31,…}.**
- **The "-" sign indicates that the first event will not result in stopping; the trial is stopped if "2 of the first 5, 3 of 18, or 4 of 31 patients have side effects"**

WEAKNESSES

The hypothesis testing-based approach has two problems/weaknesses:

(i) At times, the result might appear to be "over aggressive"; the trial is stopped when the "observed rate" of adverse events (i.e. p=e/n) is below the ceiling rate π_A.

(ii) The statistical power falls short of the pre-set level because we apply the rejection rule for a two-sided test to a one-sided alternative.

IS IT REALLY OVER AGGRESSIVE?

- Take the example where we know that the baseline rate is π_0 = 3% and investigator sets a ceiling rate of π_A = 15%; the stopping rule is: {-, 5,18, 31,…}.
- But, at the 4th event, the observed rate is 4/31 or 12.9%, still below the ceiling set at 15%.
- In the context of the statistical test, at that point, even though the observed rate is only 12.9% but <u>enough</u> to reject H_0 (3%) and "accept" H_A(15%), a rate at which the trial should be stopped.

Still kind of unsettling to a clinician to stop trial when the observed rate is still not yet considered unsafe (to him/her). Actually, the rule {-, 5,18, 31,…} is not very aggressive. In addition, the problem only appears so when the clinician is "too aggressive" to "go on" by setting the ceiling rate ways over the baseline rate (15% versus 3%). It would not appear as a problem when the "gap" is set smaller; for example, if know that the baseline rate is π_0 = 3% and investigator sets a ceiling rate of π_A = 10%; the stopping rule would be: {-,-,10,23}. Here, we did not stop before the ceiling rate.

ABOUT STATISTICAL POWER

- The problem with statistical power, that it falls short of the pre-set level because we apply the rejection rule for a two-sided test to a one-sided alternative, is real!
- We can compute the actual/achieved power and compare to the pre-set power.
- For example, we decide to enroll a total of N patients and came with the rule LR_N; the true power is 1-Pr(N; π_A ,LR_N) where Pr(N; π_A ,LR_N) is the probability of reach N patients without having stopped the trial.

EXAMPLE

Suppose the rule is LR_N = {-, n(2), n(3), N} and let u, v, and w be the numbers of adverse events that occur in each of the three segments of the trial [0,n(2)], [n(2),n(3)], and [n(3),N]. The probabilities for the three segments are b[u;n(2), π_A], b[v;n(3)-n(2), π_A], and b[w;N-n(3), π_A] where b[i;n, π_A] is the binomial probability to have exactly "i" events in n trials when the true rate is π_A. Reaching N patients without stopping the trial means that u<2, v<3-u, and w<4-(u+v). The true power is:

$$1 - \Pr(N; \pi_A, LR_N) = 1 - \sum_{u=0}^{1} \sum_{v=0}^{2-u} \sum_{w=0}^{3-u-v} b[u; n(2), \pi_A] b[v; n(3) - n(2), \pi_A] b[w; N - n(3), \pi_A]$$

By a similar calculation, but replacing π_A by π_0, we can calculate and check for the "size" of the test (type I error rate). For example:

$$1 - \Pr(\mathrm{N}; \pi_0, LR_N) = 1 - \sum_{u=0}^{1} \sum_{v=0}^{2-u} \sum_{w=0}^{3-u-v} \mathrm{b}[\mathrm{u}; \mathrm{n}(2), \pi_0] \mathrm{b}[\mathrm{v}; \mathrm{n}(3) - \mathrm{n}(2), \pi_0] \mathrm{b}[\mathrm{w}; \mathrm{N} - \mathrm{n}(3), \pi_0]$$

SOLUTION?

- The problem of being under-powered is correctable; since the power falls short, the boundary needs to be pulled downward to retain the pre-set level.
- For example, with π_0 = 3% and π_A = 15%; the stopping rule found for 80% power was: {-7, 5,18, 31,…}; the true power is only 74%; we need to stop - say - for the 4th event <u>before</u> n(4) = 31.
- But when? Or How?

A SOLUTION

- **Goldman (1987) described an algorithm for computing exact power (and type I error rate).**
- **Goldman and Hannan (2001) proposed to repeatedly use that algorithm to "search" for a stopping rule which almost achieve the pre-set levels of type I error rate and statistical power; they also provided a FORTRAN program allowing users to set their own size and power (and design parameters); called G&H algorithm.**

ABOUT G&H ALGORITHM

- Goldman and Hannan's algorithm works but choosing one between many rules found sometimes is not an easy job; several found could be "odd"!
- The gain may be small; it is true that the power falls short without a correction, but it's only a few percentage points.
- It does not solve the perceived problem that the observed rate may be below the pre-set ceiling rate.
- May be it would be more simple just to set the power higher, say 85% when we want 80%.

THE BAYESIAN APPROACH

Consider a Binomial distribution B(n,π), if we assume that the probability π has a "prior" distribution say - Beta(α,β); after "e" adverse events having observed, the "posterior" distribution of π becomes Beta (α+e,β+n-e). From this:

$$P(\pi_*) = \Pr(\pi > \pi_*)$$

$$= 1 - \int_0^{\pi_*} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+e)\Gamma(\beta+n-e)} y^{e+\alpha-1}(1-y)^{n-e+\beta-1}\,dy$$

MEHTA AND CAIN'S RULE

- By assuming an "uniform prior" (where $\alpha=\beta=1$), Mehta and Cain (1984) provided a simple formula:

$$P(\pi_*) = Pr(\pi > \pi_*)$$

$$= 1 - \int_0^{\pi_*} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+e)\Gamma)\Gamma+n-e)} y^{e+\alpha-1}(1-y)^{n-e+\beta-1} dy$$

$$= \sum_{i=0}^{e} b[i; n+1, \pi_*]$$

- and proposed a rule for which the trial is stop when $P(\pi_0)$ is large, say exceeding 97%, where π_0 is the baseline side-effect's rate.

EXAMPLE

$$.97 = \binom{n(1)+1}{0}\pi_0^0(1-\pi_0)^{n(1)+1} + \binom{n(1)+1}{1}\pi_0^1(1-\pi_0)^{n(1)}$$

$$.97 = (1-\pi_0)^{n(1)+1} + \{n(1)+1\}\pi_0(1-\pi_0)^{n(1)}$$

$$n(1) \cong 8 \text{ when } \pi_0 = .03$$

EXAMPLE

$$.97 = \binom{n(2)+1}{0}\pi_0^0(1-\pi_0)^{n(2)+1} + \binom{n(2)+1}{1}\pi_0^1(1-\pi_0)^{n(2)} + \binom{n(2)}{2}\pi_0^2(1-\pi_0)^{n(2)-2}$$

$$.97 = (1-\pi_0)^{n(2)+1} + \{n(2)+1\}\pi_0(1-\pi_0)^{n(2)} + \frac{[n(2)+1]n(2)}{2}\pi_0^2(1-\pi_0)^{n(2)-1}$$

$n(2) \cong 21$ when $\pi_0 = .03$

By applying the Mehta and Cain's Bayesian rule, we come up with pairs of numbers [e,n(e)]; it works just as the stopping rule obtained from the hypothesis testing-based approach. The major difference is that this Bayesian rule does not require the setting of a 'ceiling rate". At first it appears reasonable: if the usual normal rate is π_0 then the trial should be stopped when this rate is exceeded because the rate is no longer "normal"

EXAMPLE

With π_0 = .03 or 3%, the Mehta and Cain's rule yields the stopping rule {8, 21, 38,…}; that is to stop at 1 event out of 8 patients, 2 out of 21, 3 out of 38, and so on. As a comparison, with π_0 = 3% and π_A = 15%; the test-based stopping rule found for 80% power was: {-, 5,18, 31,…} - to stop at 2 events out of 5 patients, 3 out of 18, 4 out of 38, and so on.

Goldman (1987), after consulting her collaborators/clinicians, concluded that even though the Mehta and Cain's Bayesian boundaries are philosophically very attractive but rather **liberal**, especially that it allows for the stopping of a trial after a single event. In fact, it seems too aggressive to trial simply because $\pi > \pi_0$; say when $\pi_0 = 3\%$ and $\pi = 3.5\%$ because patients benefit from the treatment as well.

MODIFICATIONS?

To overcome having an over-aggressive Bayesian rule, Goldman (1987) considered to raised the cutpoint ".97" for the posterior probability or formulating rule using $P(\pi_A)$ - instead of $P(\pi_0)$ - where π_A is the ceiling or maximum tolerated rate. For example, "the trial is stop when $P(\pi_A)$ is large, say exceeding 95% or 97%". However, she concluded that "various adjustments did not seem to remedy the problem".

It is true that setting a stopping rule based on large values of $P(\pi_0)$, say when $\pi_0 = 3\%$ and $\pi = 3.5\%$, may be too aggressive; the increase in the rate may not be large enough to be clinically significant (or to outweigh the benefits of the treatment).

On the other hand, setting a stopping rule based on large values of $P(\pi_A)$ alone seems "unsettling" because it ignores the baseline rate and never reveals the impact of the treatment on having side effects. It is true that setting a ceiling rate is always "subjective"; but by seeing both - π_0 and π_A- one would know how reasonable the parameters are.

To have a fair comparison with the corresponding hypothesis-based stopping rule, may be we should stop the trial based on large values of $P(\pi_A)$, say "the trial is stop when $P(\pi_A)$ is large, say exceeding 80% or 90%" - whatever the number usually used as the pre-set value for statistical power- not 97%. But this would make the resulting Bayesian rule even more aggressive!

The problem was the choice of the 'prior". With the "uniform prior" (where $\alpha=\beta=1$), the mean is .5; we really need some prior distribution with an expected value more in line with the concept of "rare" side effects.

Usually, in Bayesian analysis, the choice of the prior carries only moderate weight - sometimes not that important, a non-informative prior does the job. But here we conduct most very small trials and using sequential rule, it carries very heavy weight. For example, if we observe 3 events from 7 patients then (i) the posterior mean is still .5 (4/8) (leaning to stopping) with choice $\alpha=\beta=1$, but (ii) the posterior mean is .1 (4/40) (leaning to non-stopping) with choice $\alpha=1$ and $\beta=32$

OPTIONS

- There are no perfect choice for a prior
- Uniform prior may be popular but it is biased "toward stopping" (its mean is .5), resulting rule may be too aggressive.
- We should choose so that $(\alpha+\beta)$ is small, eg. take $\alpha=1$, but not easy to set the mean $\alpha/(\alpha+\beta)$
- (i) setting $\alpha/(\alpha+\beta) = \pi_A$ may also be somewhat biased toward stopping- unless want more cautious,
- (ii) setting $\alpha/(\alpha+\beta) = \pi_0$ may be biased toward non-stopping; may be this is the choice when investigator believe that the treatment is safe.

REVISED BAYESIAN RULE

Suppose we choose, as prior, α=1 and β=m (eg. m=32 so that the prior mean is $\alpha/(\alpha+\beta) = \pi_0 = .03$); the revised rule is:

$$P(\pi_*) = Pr(\pi > \pi_*)$$

$$= 1 - \int_0^{\pi_*} \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+e)\Gamma)\Gamma+n-e)} y^{e+\alpha-1}(1-y)^{n-e+\beta-1}\,dy$$

$$= \sum_{i=0}^{e} b[i; n+m, \pi_*]$$

The trial is stop when $P(\pi_A)$ is large, say exceeding 80% or 90%; π_A being the ceiling side-effect's rate.

EXAMPLE

If we choose m=32 so that the prior mean is $\alpha/(\alpha+\beta) = \pi_0 = .03$), then:

$$.80 = \binom{n(1)+32}{0} \pi_A^0 (1-\pi_A)^{n(1)+32-0} + \binom{n(1)+32}{1} \pi_A^1 (1-\pi_A)^{n(1)+32-1}$$

$$.80 = (1-\pi_A)^{n(1)+32} + [n(1)+32]\pi_A (1-\pi_A)^{n(1)+31}$$

n(1) is negative, no stopping - just as in the test based rule

This choice would result in a rule which is even more conservative the test-based one.

EXAMPLE

If we choose m=6 so that the prior mean is $\alpha/(\alpha+\beta) = \pi_A = .15$), then:

$$.80 = \binom{n(1)+6}{0} \pi_A^0 (1-\pi_A)^{n(1)+6-0} + \binom{n(1)+6}{1} \pi_A^1 (1-\pi_A)^{n(1)+6-1}$$

$$.80 = (1-\pi_A)^{n(1)+6} + [n(1)+6]\pi_A (1-\pi_A)^{n(1)+5}$$

n(1) is still negative, no stopping

I believe that this choice would result in a rule which is closer to the hypothesis test-based one.

After a rule is formed, including the Bayesian rule, we can always calculate its type I error rate and check to see if it is over aggressive.

$$1 - \Pr(\mathrm{N}; \pi_0, Rule)$$