Why A Bayesian Be?

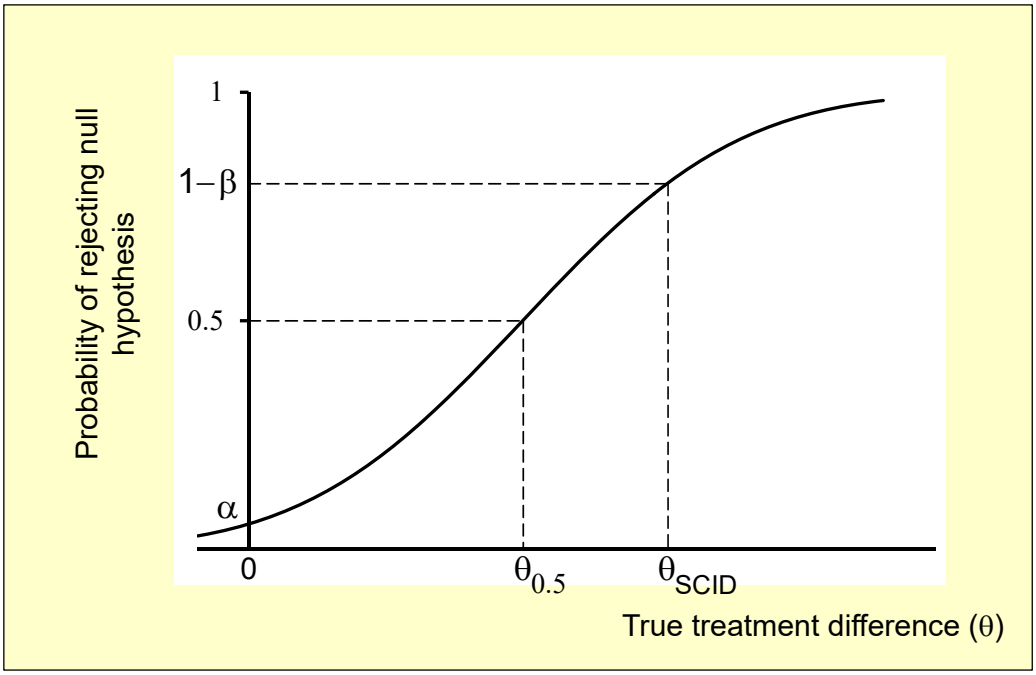
Andrew R. Willan, PhD

Senior Scientist Emeritus
SickKids Research Institute, Toronto

Professor of Biostatistics
Dalla Lana School of Public Health
University of Toronto

“Ask your doctor if taking a pill to solve all your problems is right for you.”

Sample Size and Power Curves



θ_{SCID} = smallest clinical important difference

Sample Size and Power Curves

Power curves will pass through $(0,\alpha)$ because we're testing the hypothesis

H: $\theta = 0$ vs. A: $\theta > 0$ at the level α

(*i.e.* probability of reject H when it's true is α)

Steepness of the power curve will increase with sample size

Problems with Standard Approach

α is almost always set to 0.05 (occasionally 0.01)

Which means that "all" trials have the same probability of falsely rejecting the null hypothesis and adopting the new intervention when it is no better than the standard intervention.

Example: Argon green laser (*Standard*) vs. krypton red laser (*Treatment*) for age-related macular degeneration. Number of letters read.

No difference to the patient, except the colour of the light. No difference in cost.

Could argue that if there's no difference in patient outcomes (*i.e.* null hypothesis is true), there is no real issue with adopting the new intervention

Perhaps α should be set to 0.5

Problems with Standard Approach

α is almost always set to 0.05 (occasionally 0.01)

Example: Trial of planned labour (*Standard*) vs. cesarean section (*Treatment*) for pregnant women in the breech position. Neonatal death or disability.

Huge difference to the patient. Potential difference in cost.

If there is no benefit in patient outcomes (*i.e.* null hypothesis is true), would really like to avoid adopting cesarean section

Perhaps α should be set to 0.001 or lower

Problems with Standard Approach

$1 - \beta$ is almost always set to 0.8 (occasionally 0.9)

Therefore, there is a 20% probability failing to reject the null hypothesis even though the true treatment difference is clinically important.

But also, the probability of rejecting the null hypothesis is greater than 50% for a range of values for the true treatment difference that is less than the smallest clinical important, namely $\delta_{0.5}$ to δ_{SCID}

In practice the stated δ_{SCID} used is often back solved after determining the sample size based feasibility and budget constraints, and is almost always larger than the true δ_{SCID}.

Problems with Standard Approach

Basic problem is the failure to quantify and assign utilities/disutilites to patient outcomes and interventions

Interventions: green vs red light
major surgery vs planned vaginal birth

Outcomes: number of letters read from eye chart
neonatal death and disability

Three Main Reasons to Prefer Bayesian Approach

Permits simple, intuitive and relevant statements of statistical inference regarding the parameters of interest directly
(Bayes Lite)

Provides a transparent framework for combining new information with current knowledge
(Bayes)

Facilitates decision theory (value of information methods) for optimal decision-making and research design
(Full-on Bayes)

Simple, Intuitive, Relevant Statements of Statistical Inference

Frequentist definition of probability of an event: the limiting relative frequency of its occurrence in a series of repeated observations of a chance outcome in which it could occur.

For the Bayesian probability is the (subjective) expression of the uncertainty or “degree of belief” regarding the unknown.

Frequentist Statistical Inference

Frequentist definition of probability leads to the use of test of hypothesis, with associated *p*-values and confidence intervals, to characterize uncertainty regarding model parameters.

Working hypothesis

Null hypothesis (*i.e.* working hypothesis is not true)

If observations refute null hypothesis, the working hypothesis is "proven"

In empirical research, "refute" means observations are "unlikely" if null hypothesis is true. "Unlikely" usually means a probability less than 5%

Bayesian Statistical Inference

Bayesian definition of probability leads to the use of probability statements regarding model parameters to characterize the uncertainty.

Bayesian inference provides probability statements about the truth, given the data. Frequentist inference provides probability statements about the data, given the truth.

Consider a clinical trial comparing *T* and *S* with respect to the relative risk for a bad outcome, where the frequentist's *p*-value is 0.035 and where a one-sided test of hypothesis is applied at the 5% level

Simple, Intuitive, Relevant Statements of Statistical Inference

The frequentist statement of inference is:

"We can reject the null hypothesis that the relative risk is equal to or greater than one (*i.e. T* is equivalent or inferior to *S*) in favour of the alternative (working) hypothesis that the relative risk is less than one (*i.e. T* is superior to *S*) with a probability of being wrong is less than 5%.

This means that if the null hypothesis is true (*i.e. T* is equivalent or inferior to *S*) and the trial was repeated many, many times, the proportion of times that the results of these replications will be at least as inconsistent with the null hypothesis as the data from the trial under consideration is less than 5%."

Simple, Intuitive, Relevant Statements of Statistical Inference

This is not a statement about falsely rejecting the null hypothesis for this particular trial, but rather a statement about the proportion of many, many null hypothesis that would be falsely rejecting using the same criterion.

The Bayesian statement of inference is:

"The probability that the relative risk is less than one (*i.e. T* is superior to *S*) is 96.5%."

Simple, Intuitive, Relevant Statements of Statistical Inference

Frequentist 95% confidence interval:

"The 95% confidence interval for the relative risk is (0.493, 0.917), meaning that if the trial was conducted many, many times, then (in the limit) the proportion of the confidence intervals from these replications that include the true relative risk is 95%." The inference does not say anything about the specific confidence interval based on the data from this trial, and whether or not it includes the true value of relative risk.

What you can say, using the data from this trial, is that the hypothesis H: RR = *x* cannot be rejected at the two-sided, 5% level for any value of *x* in the interval (0.493, 0.917).

Simple, Intuitive, Relevant Statements of Statistical Inference

On the other hand, inference based on a Bayesian credible interval with the same limits is stated as, "there is a 95% probability that the relative risk lies in the interval (0.493, 0.917)."

Framework for Combining New Information (Data) with Current Knowledge

Bayes Theorem: $p(\theta \mid y) = \frac{p(y \mid \theta)}{p(y)} p(\theta)$

θ are the population parameters

$p(\theta)$ expresses the current knowledge as a probability distribution function (*pdf*) for the model parameters, referred to as the prior distribution

y is the new data

$p(y \mid \theta)$ is the likelihood for the data, *i.e.* the *pdf* for the data, given θ

$p(\theta \mid y)$ is the posterior *pdf* for θ given the data

$p(y)$ is a normalizing factor to ensure that $p(\theta \mid y)$ integrates to 1

Framework for Combining New Information (Data) with Current Knowledge

Bayes Theorem: $p(\theta \mid y) = \frac{p(y \mid \theta)}{p(y)} p(\theta)$

The inference is made with $p(\theta \mid y)$ is conditional on of y (the data) and $p(\theta)$ (prior pdf for θ).

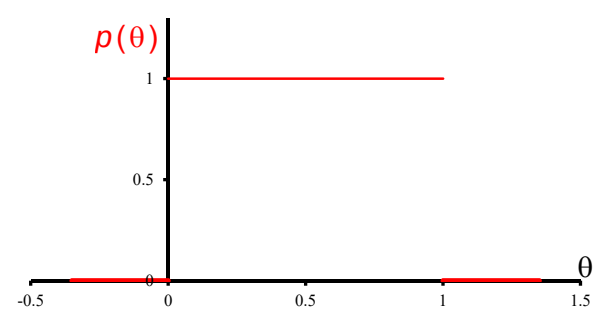
Therefore, the inference is conditional on the data and opinions that were used to form $p(\theta)$.

If an uninformative prior is used for $p(\theta)$ then it is implicit that the inference is conditional only on y.

Bayes Theorem for Binomial Probability

Let θ be the probability that a patient experiences failure

If we have no knowledge of θ, then $p(\theta) = 1$ for $0 \le \theta \le 1$ and 0, othewise



That is, $\theta \sim Beta(1,1)$ Uninformative Prior

Bayes Theorem for Binomial Probability

If $X \sim Beta(a,b)$ then $E(X) = \frac{a}{a+b}$ and $V(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Therefore $E(\theta) = \frac{1}{2}$ and $V(\theta) = \frac{1}{12}$ prior, *i.e.* before any data/opinions

In a sample of *n* patients, suppose *r* failures are observed

Then the posterior distribution for θ is $Beta(1+r, 1+n-r)$

Therefore $E(\theta) = \frac{r+1}{n+2}$ and $V(\theta) = \frac{(r+1)(n-r+1)}{(n+2)^2(n+3)}$ posterior

Early (*T*) vs. Late (*S*) External Cephalic Version for Breech

Frequentist Approach

Pregnant women in the breech position randomized between early (33-36 weeks) vs. late (37+ weeks) external cephalic version to manipulate the fetus into the vertex position and avoid a caesarean delivery,
Hutton *et al. BJOG* 2011; **118**(5):564-577.

Late (*S*): *n* = 767; non-caesarean delivery = 337 (43.9%)
Early (*T*): *n* = 765; non-caesarean delivery = 367 (48.0%)
Two-sided *p*-value = 0.12

Conclusion: Early external cephalic version does not increase the probability of a non-caesarean delivery

Early (*T*) vs. Late (*S*) External Cephalic Version for Breech

Bayesian Approach

Pilot trial
Hutton *et al. AJOG* 2003; **189**(1):245-254.

Late (*S*): *n* = 116; non-caesarean delivery = 33 (28.4%)
Early (*T*): *n* = 116; non-caesarean delivery = 41 (35.3%)

	Prior for Pilot	Posterior to Pilot Prior for Trial	Posterior to Trial
Prob. of non-CD, Late	*Beta*(1,1)	*Beta*(1+33,1+[116-33]) = *Beta*(34,84)	*Beta*(34+337,84+[767-337]) = *Beta*(371,514)
Prob. of non-CD, Early	*Beta*(1,1)	*Beta*(1+41,1+[116-41]) = *Beta*(42,76)	*Beta*(42+367,76+[765-367]) = *Beta*(409,474)

Early (*T*) vs. Late (*S*) External Cephalic Version for Breech

Bayesian Approach

Prob. non-CD	Distribution	Mean	Variance
Late: θ_L	*Beta*(371,514)	0.4192	0.0002748
Early: θ_E	*Beta*(409,474)	0.4622	0.0002813
$\theta_E - \theta_L$	Approx. Normal	0.04399	0.0005561

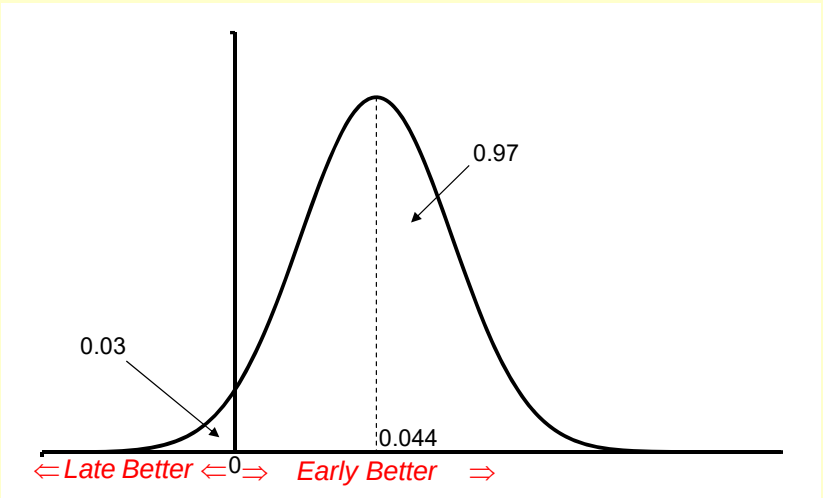
prob. difference **difference** **sum**

$$X \sim Beta(a,b) \qquad E(X) = \frac{a}{a+b} \qquad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Early (*T*) vs. Late (*S*) External Cephalic Version for Breech

Bayesian Approach

Probability distribution function for $\theta_E - \theta_L$ (prob. difference)

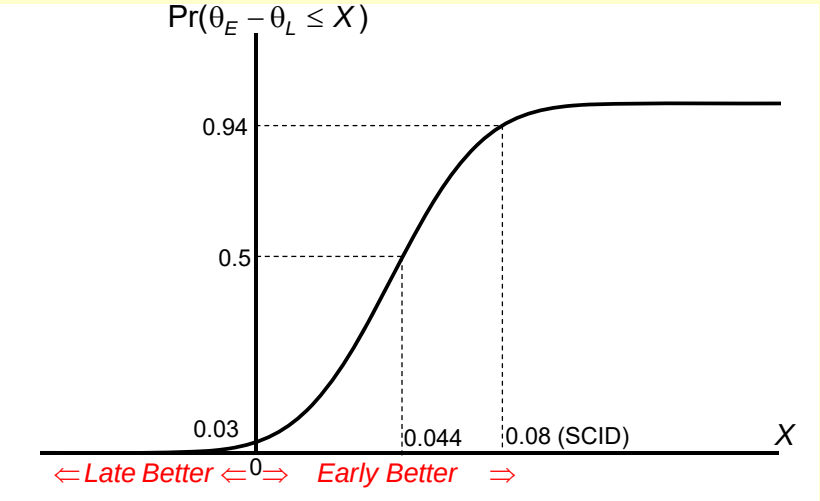


The probability that Early ECV increases the chance of a non-CD is 97%

Early (*T*) vs. Late (*S*) External Cephalic Version for Breech

Bayesian Approach

Cumulative distribution function for $\theta_E - \theta_L$ (prob. difference)



Topical Paromomycin and Gentamicin for Cutaneous Leishmaniasis

Frequentist Approach

Of the 125 patient randomized to active, 101 achieved the threshold for cure (80.8%), compared to 73 of the 125 patients randomized to control (58.4%); 2p=0.0001.

Based on these results the authors rejected the null hypothesis using a two-sided test at the 0.05 level.

Ben Salah *et al. NEJM* 2013; **368**(6):524-532

Conclusion: "This trial provides evidence of the efficacy of paromomycin–gentamicin *(with a 5% probability of error)*."

Topical Paromomycin and Gentamicin for Cutaneous Leishmaniasis

Bayesian Approach

Pilot trial

Ben Salah *et al. PLoS Negl. Trop. Dis.* 2009; **3**(5), e432.

Placebo (P): n = 42; cure = 30 (71.4%)

Active (A): n = 50; cure = 47 (94.0%)

	Prior for Pilot	Posterior to Pilot Prior for Trial	Posterior to Trial
Prob. of cure, Placebo	*Beta*(1,1)	*Beta*(1 + 30, 1 + [42 - 30]) = *Beta*(31, 13)	*Beta*(31 + 73, 13 + [125 - 73]) = *Beta*(104,65)
Prob. of cure, Active	*Beta*(1,1)	*Beta*(1 + 47, 1 + [50 - 47]) = *Beta*(48, 4)	*Beta*(48 + 101, 4 + [125 - 101]) = *Beta*(149, 28)

Topical Paromomycin and Gentamicin for Cutaneous Leishmaniasis

Bayesian Approach

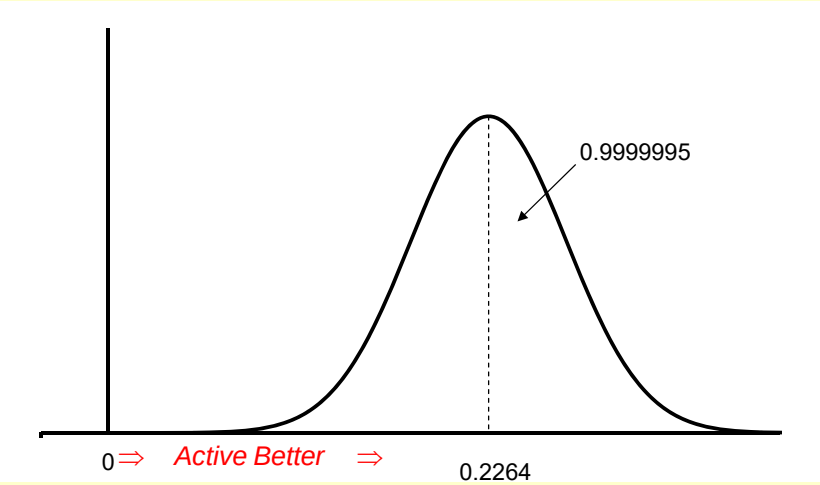
	Distribution	Mean	Variance
Cure, Placebo: θ_P	*Beta*(104,65)	0.6154	0.001393
Cure, Active: θ_A	*Beta*(149,28)	0.8418	0.0007481
$\theta_A - \theta_P$	Approx. Normal	0.2264	0.002140

cure difference **difference** **sum**

$$X \sim Beta(a,b) \qquad E(X) = \frac{a}{a+b} \qquad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Topical Paromomycin and Gentamicin for Cutaneous Leishmaniasis

Probability distribution function for $\theta_A - \theta_P$ (cure difference)

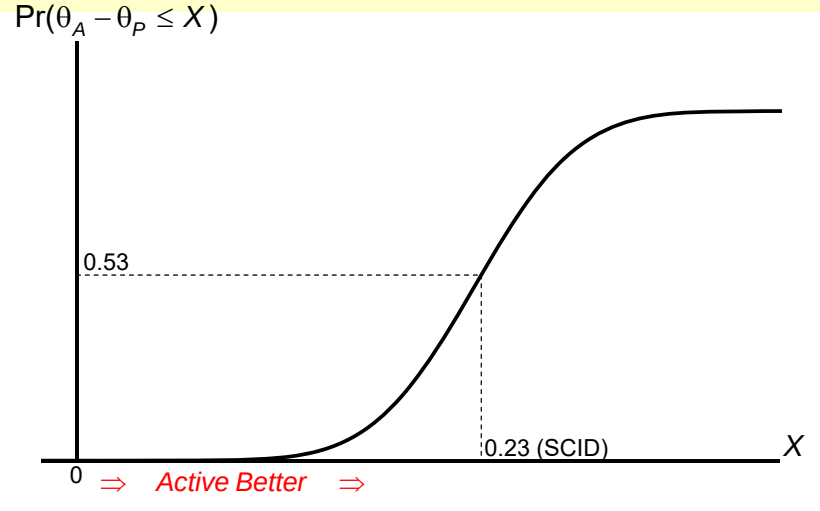


The probability that antibiotics increase the chance of cure is 99.99995%

Topical Paromomycin and Gentamicin for Cutaneous Leishmaniasis

Bayesian Approach

Cumulative distribution function for $\theta_A - \theta_P$ (cure difference)



Electrolyte Maintenance Solution versus Usual Fluids in Children with Acute Gastroenteritis

Frequentist Approach

Non-inferiority trial of EMS (*S*) vs UF (*T*). Binary outcome was a composite measure of treatment failure (prob(failure) = θ).

Null hypothesis: $\theta_{EMS} - \theta_{UF} \geq -0.075$; tested at 2.5% level, one-sided

Observed $\theta_{EMS} - \theta_{UF} = 0.083$; lower 95% confidence interval = 0.02

Conclusion: "usual liquids is at least as good as EMS at preventing treatment failures *(with a 2.5% probability of error)*."

Electrolyte Maintenance Solution versus Usual Fluids in Children with Acute Gastroenteritis

Bayesian Approach

EMS: *n* = 324; failure = 81 (25.0%)
UF: *n* = 323; failure = 54 (16.7%)

	Prior for Trial	Posterior to Trial
Prob. of failure, EMS	*Beta*(1,1)	*Beta*(1 + 81, 1 + [324 - 81]) = *Beta*(82, 244)
Prob. of failure, UF	*Beta*(1,1)	*Beta*(1 + 54,1 + [323 - 54]) = *Beta*(55, 270)

Electrolyte Maintenance Solution versus Usual Fluids in Children with Acute Gastroenteritis

Bayesian Approach

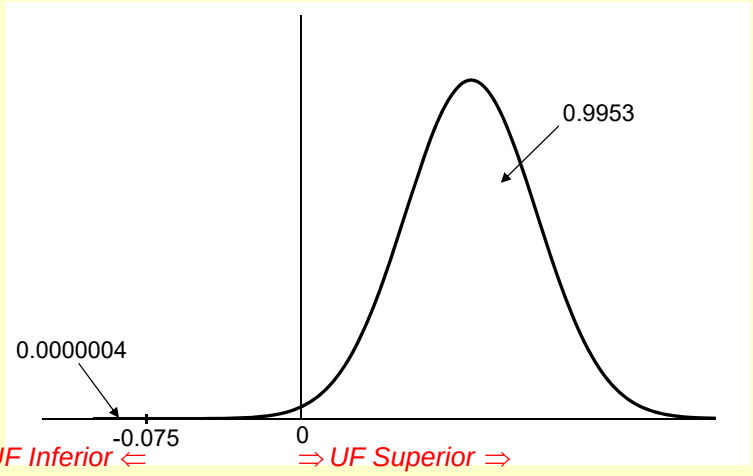
	Distribution	Mean	Variance
Failure, EMS: θ_{EMS}	*Beta*(82,244)	0.2515	0.000576
Failure, UL: θ_{UF}	*Beta*(55,270)	0.1692	0.0000431
$\theta_{EMS} - \theta_{UF}$	Approx. Normal	0.0823	0.001007

failure difference **difference** **sum**

$$X \sim Beta(a,b) \qquad E(X) = \frac{a}{a+b} \qquad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Electrolyte Maintenance Solution versus Usual Fluids in Children with Acute Gastroenteritis

Probability distribution function for $\theta_{EMS} - \theta_{UF}$ (failure difference)



The probability that UF is inferior to EMS is 0.0000004

The probability that UF is superior to EMS is 99.53%

Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis

Frequentist Approach

Of the 487 patient randomized to active, 160 progressed (32.9%), compared to 96 of the 244 patients randomized to placebo (39.3%); p=0.05.

Based on these results the authors rejected the null hypothesis.

Montalban *et al. NEJM* 2017; **376**(3):209-220

Conclusion: "Ocrelizumab was associated with lower rates of clinical and MRI progression than placebo"

Sample size based on a probability of progression on placebo of 0.43 and a smallest clinically important difference (SCID) of 0.13.

Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis

Bayesian Approach

Ocrelizumab: n = 487; progression = 160 (32.9%)
Placebo: n = 244; progression = 96 (39.3%)

	Prior for Trial	Posterior to Trial
Prob. of prog., Ocrelizumab	*Beta*(1,1)	*Beta*(1 + 160, 1 + [487 - 160]) = *Beta*(161, 328)
Prob. of prog., Placebo	*Beta*(1,1)	*Beta*(1 + 96, 1 + [244 - 96]) = *Beta*(97, 149)

Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis

Bayesian Approach

	Distribution	Mean	Variance
Prog., Ocreliz: θ_O	*Beta*(161, 328)	0.3292	0.000451
Prog., Placebo: θ_P	*Beta*(97, 149)	0.3943	0.000967
$\theta_O - \theta_P$	Approx. Normal	-0.0651	0.001418

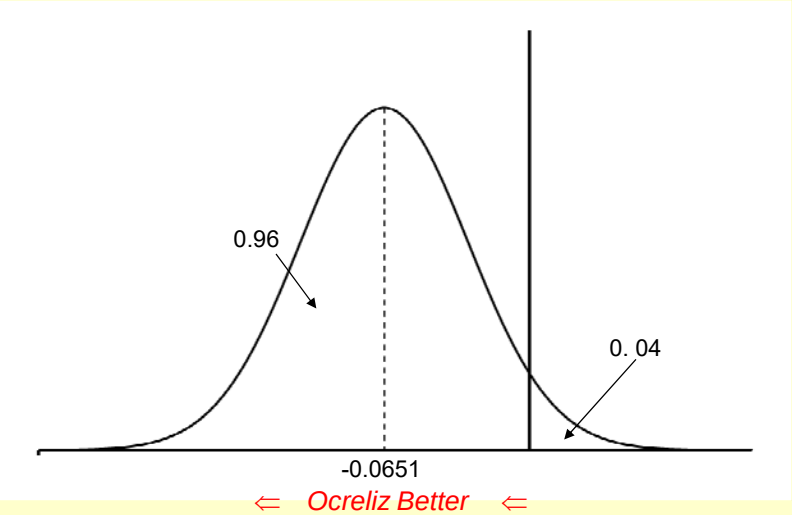
progression difference **difference** **sum**

$$X \sim Beta(a,b) \qquad E(X) = \frac{a}{a+b} \qquad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis

Bayesian Approach

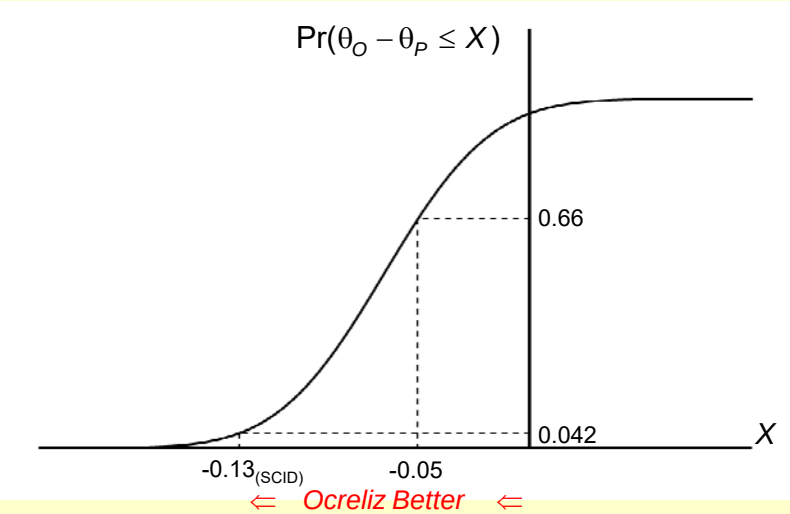
Probability distribution function for $\theta_O - \theta_P$ (progression difference)



Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis

Bayesian Approach

Cumulative distribution function for $\theta_O - \theta_P$ (progression difference)



Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Frequentist Approach

Of the 133 patient randomized to hypothermia, 48 survived (36.1%), compared to 48 of the 124 patients randomized to normothermia (38.7%); p=0.71.

Based on these results the authors stopped trial for futility.

Moler *et al. NEJM* 2017; **376**(4):318-329.

Conclusion: "therapeutic hypothermia, as compared with therapeutic normothermia, did not confer a significant benefit in survival"

Sample size based on a probability of survival on normothermia of 0.45 and a smallest clinically important difference (SCID) of 0.15.

Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Bayesian Approach

Hypothermia: n = 133; survival = 48 (36.1%)

Normothermia: n = 124; survival = 48 (38.7%)

	Prior for Trial	Posterior to Trial
Prob. of survival, Hypo	*Beta*(1,1)	*Beta*(1 + 48, 1 + [133 - 48]) = *Beta*(49, 86)
Prob. of survival, Normo	*Beta*(1,1)	*Beta*(1 + 48, 1 + [124 - 48]) = *Beta*(49, 77)

Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Bayesian Approach

	Distribution	Mean	Variance
Survival Hypo: θ_H	*Beta*(49, 86)	0.3630	0.001700
Survival Normo: θ_N	*Beta*(49, 77)	0.3889	0.001871
$\theta_H - \theta_N$	Approx. Normal	-0.02593	0.059762

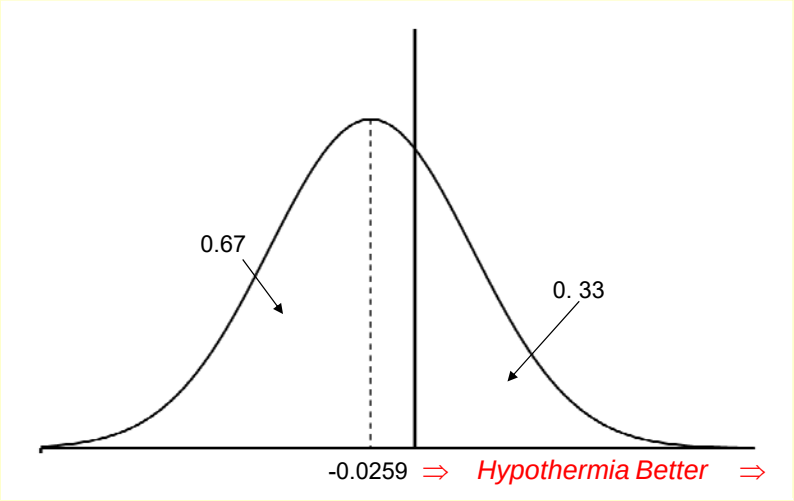
survival difference **difference** **sum**

$$X \sim Beta(a,b) \qquad E(X) = \frac{a}{a+b} \qquad V(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Bayesian Approach

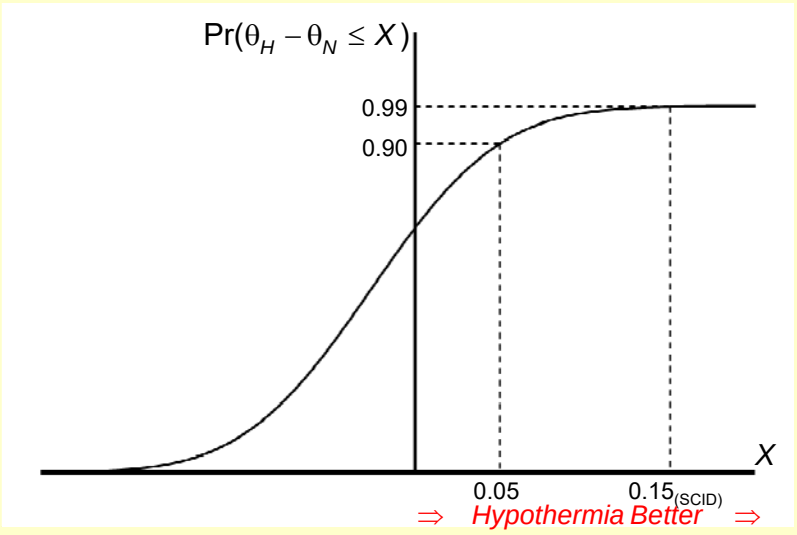
Probability distribution function for $\theta_H - \theta_N$ (survival difference)



Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Bayesian Approach

Cumulative distribution function for $\theta_H - \theta_N$ (survival difference)



Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Adaptive Design Approach

Planned an addition 140 patient per arm when trial stopped.

Number of survivors in 140 patient follows a beta-binomial distribution

Hypothermia arm: BetaBin(140, 49, 86)

Normothermia arm: BetaBin(140, 49, 77)

Simulate completed trial results and determine what proportion would have resulted in "significant" difference in favour of Hypothermia ("*Probability of Success*")

Therapeutic Hypothermia after In-Hospital Primary Cardiac Arrest in Children

Adaptive Design Approach

```
pH ~ dbeta(49, 86)
rH ~ dbin(pH, nH)

pN ~ dbeta(49, 77)
rN ~ dbin(pN, nN)

thetaH <- (48+rH)/(133+nH)
thetaN <- (48+rN)/(124+nN)

thetaDif <- thetaH - thetaN

thetaDifSE <- sqrt( (thetaH*(1-thetaH)/(nH+133)) + (thetaN*(1-thetaN)/(nN+124)) )

z <- thetaDif / thetaDifSE
prob <- step(z-1.65)
```

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
prob	0.01463	0.1201	3.636E-4	0.0	0.0	0.0	10001	90000
thetaDif	-0.02629	0.04338	1.545E-4	-0.1106	-0.02689	0.05848	10001	90000
z	-0.6311	1.042	0.003714	-2.655	-0.6501	1.409	10001	90000

Bayesian Approach for Binary Outcome

Summary

Beta prior → Binomial sampling → Beta posterior

Beta(*a*,*b*) → *r* outcomes of *n* patients → *Beta*(*a* + *r*, *b* + [*n* – *r*])

$\theta \sim Beta(a,b)$ then $E(\theta) = \frac{a}{a+b}$ and $V(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$

For *a* + *b* sufficiently large (> 30) $\theta \sim Normal\left(\frac{a}{a+b}, \frac{ab}{(a+b)^2(a+b+1)}\right)$

Uninformative prior: $Beta(1,1)$

Bayesian Approach for Continuous Outcome

Introduction

Normal prior → Normal sampling → Normal posterior

$\theta \sim Normal(\mu_0, v_0)$ $\quad \hat{\theta}, V(\hat{\theta}) \quad$ $\theta \sim Normal(\mu_1, v_1)$

where $\mu_1 = v_1\left(\frac{\mu_0}{v_0} + \frac{\hat{\theta}}{V(\hat{\theta})}\right)$ $\quad v_1 = \left(\frac{1}{v_0} + \frac{1}{V(\hat{\theta})}\right)^{-1}$

Uninformative prior: $\theta \sim Normal(0, v_0)$ where $\frac{1}{v_0} = 0$

Planned Caesarean vs Vaginal Delivery for Twins

Frequentist Approach

1398 women (2795 fetuses) randomized to planned caesarean delivery (*T*)

1406 women (2812 fetuses) randomized to planned vaginal delivery (*S*)

Primary outcome: death or serious neonatal morbidity

T: 2.2% versus *S*: 1.9%, two-sided p-value=0.49

Odds Ratio (~relative risk): 1.16, 95% confidence limits: 0.77, 1.74

Barrett *et al. NEJM* 2013; **369**(14):1295-1305.

Conclusion: planned caesarean delivery did not significantly decrease or increase the risk of death or serious neonatal morbidity

Planned Caesarean vs Vaginal Delivery for Twins

Frequentist Approach

Sample size was determined to have an 80% probability of rejecting the null hypothesis (power) if planned caesarean section (*T*) halved the risk from 4% to 2% (*i.e.* relative risk = 0.5), with a two-sided Type I error of 0.05.

Since a 2% risk was observed in the planned vaginal group, the question was raised as to whether there was sufficient power.
(*i.e.* Did we miss a difference?)

That is, if the risk in the planned vaginal group was only 2% to start with, was there sufficient sample size to have an 80% probability of rejecting the null hypothesis if planned caesarean section (*T*) halved the risk from 2% to 1% (*i.e.* relative risk = 0.5)?

Feeble answer: 95% confidence interval for odds ratio: 0.77-1.74

Planned Caesarean vs Vaginal Delivery for Twins

Consider Bayesian approach on the log(odds ratio) (= θ) which is assumed to have a normal distribution based on the central limit theorem

Uninformative prior: $\theta \sim Normal(0, v_0)$ where $\frac{1}{v_0} = 0$

From trial, estimate of log(odds ratio): $\hat{\theta} = 0.1456; \quad V(\hat{\theta}) = 0.04381$

Posterior: $\theta \sim Normal(\mu_1, v_1)$ where

$$v_1 = \left(\frac{1}{v_0} + \frac{1}{V(\hat{\theta})}\right)^{-1} = \left(0 + \frac{1}{V(\hat{\theta})}\right)^{-1} = V(\hat{\theta})$$

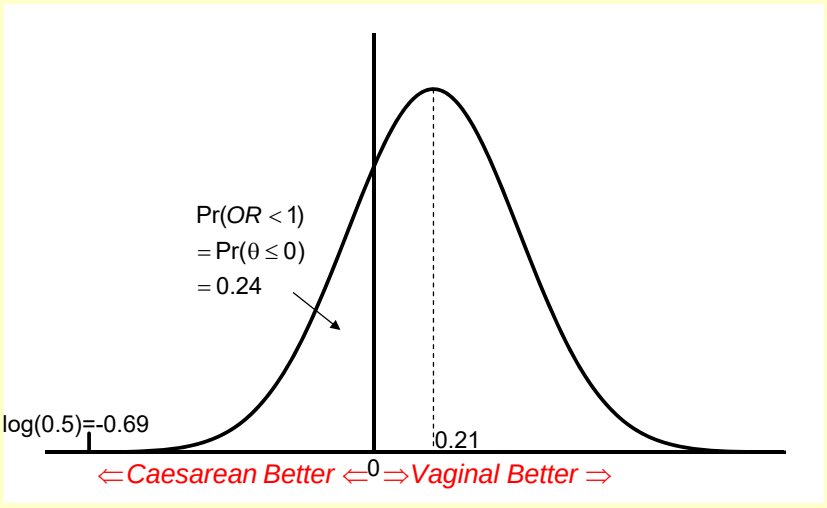
$$\mu_1 = v_1\left(\frac{0}{v_0} + \frac{\hat{\theta}}{V(\hat{\theta})}\right) = V(\hat{\theta})\left(\frac{0}{v_0} + \frac{\hat{\theta}}{V(\hat{\theta})}\right) = V(\hat{\theta})\frac{\hat{\theta}}{V(\hat{\theta})} = \hat{\theta}$$

Posterior distribution for the log(odds ratio): $\theta \sim Normal(0.1456, 0.04381)$

Planned Caesarean vs Vaginal Delivery for Twins

Bayesian Approach

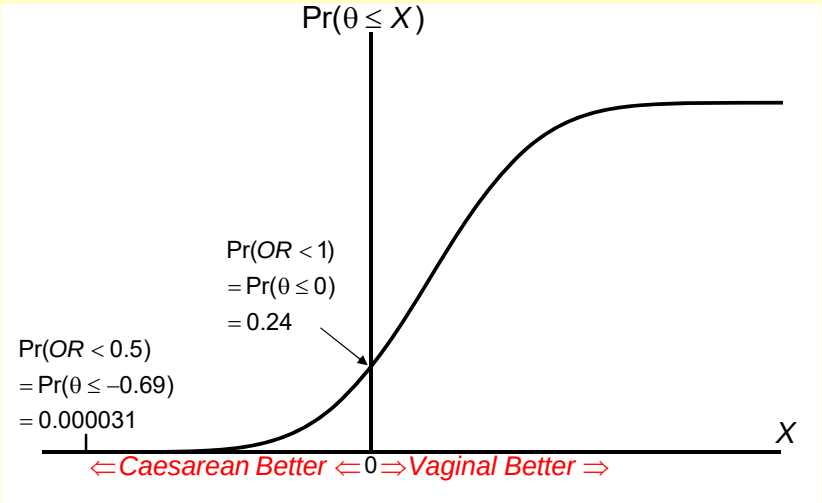
Probability distribution function for θ (logOR)



Planned Caesarean vs Vaginal Delivery for Twins

Bayesian Approach

Cumulative probability distribution function for θ (log*OR*)



The probability that planned C-section halves the risk of a bad outcome is 0.0031%.

One- vs. Two-tailed Tests in RCTs

Frequentist Approach

Most good RCTs are performed to determine in the resulting evidence supports adoption of *Treatment* to replace *Standard*. That is, the research is addressing a one-sided question.

In general, RCTs are not done to determine which of two interventions is best. If neither of the interventions is standard of care, it's very hard to know what to do with the results.

When performing the analysis, you need to protect yourself (*i.e.* limit the probability of) falsely rejecting the null hypothesis in favour of the hypothesis that *Treatment* is superior to *Standard*, because you want to limit the probability of adopting *Treatment* when it is no better than *Standard*.

One- vs. Two-tailed Tests in RCTs

Frequentist Approach

You do not need to protect yourself (*i.e.* limit the probability of) falsely rejecting the null hypothesis in favour of the alternative hypothesis that *Treatment* is inferior to *Standard*, because the health care policy implications of

1. *Treatment* equal to *Standard* (null)
2. *Treatment* inferior to *Standard* (alternative)

are the same.

Namely, **do not adopt *Treatment***

Bayesian Statements of Inference are "One-sided"

Bayesian Approach

Bayesian statement of inference are usually "one-sided" because they are simple and intuitive, and address relevant questions, such as:

What is the probability that early external cephalic version increases the probability of a non-Caesarean delivery?

What is the probability that topical Paromomycin and Gentamicin increases the probability of curing Cutaneous Leishmaniasis?

What is the probability that planned Caesarean section halves the risk of death or serious neonatal morbidity?

Table 1 *p*-values

Table 1. Characteristics of Women and Their Pregnancies at Baseline.*

Characteristic	Planned Cesarean Delivery (N=1393)	Planned Vaginal Delivery (N=1393)
Maternal age ≥30 yr — no. (%)	632 (45.4)	632 (45.4)
Parity ≥1 — no. (%)	857 (61.5)	856 (61.5)
Previous cesarean section — no. (%)	100 (7.2)	97 (7.0)
Gestational age at randomization		
Mean — wk	34.9±1.8	34.9±1.8
<32 wk 0 days — no. (%)	0	1 (0.1)
32 wk 0 days to 33 wk 6 days — no. (%)	475 (34.1)	477 (34.2)
34 wk 0 days to 36 wk 6 days — no. (%)	679 (48.7)	665 (47.7)
37 wk 0 days to 38 wk 6 days — no. (%)	239 (17.2)	250 (17.9)
Estimated fetal weight — g†		
First twin	2238±424	2238±419
Second twin	2223±413	2232±422
Chorionicity — no. (%)‡		
Dichorionic and diamnionic	961 (69.0)	970 (69.6)
Monochorionic and diamnionic	334 (24.0)	326 (23.4)
Unknown	98 (7.0)	97 (7.0)
Not in labor at randomization — no./total no. (%)	1190/1392 (85.5)	1159/1393 (83.2)
Membranes ruptured at randomization — no. (%)	83 (6.0)	76 (5.5)
National perinatal mortality in mother's country of residence — no. (%)§		
<15 deaths/1000 births	724 (52.0)	730 (52.4)
15–20 deaths/1000 births	596 (42.8)	591 (42.4)
>20 deaths/1000 births	73 (5.2)	72 (5.2)

* Plus–minus values are means ±SD. There were no significant between-group differences with respect to any of the baseline variables.

Table 1 *p*-values

Table 1 *p*-values are used mistakenly to measure the extent of treatment comparison confounding

Not suited for assessing confounding of treatment comparison

A *p*-value is a function of observed difference and sample size

In a small trial, a big difference that might confound may not be statistically significant because of small sample size

On the other hand, in a large trial, a small difference may be statistically significant, but unlikely to confound

Table 1 *p*-values

p-values can only be used to test hypotheses, *i.e.* to make inference from samples to populations

In this case the null hypothesis is that the difference in population means (or proportions) between two treatment arms of an RCT is zero

If the *p*-value is sufficiently small, the deductive reasoning is that the difference in the population means is not zero

Table 1 *p*-values

But that is preposterous! The two populations are created by randomization. How could the population means of baseline variables possibly differ?

If we reject the null hypothesis in this situation, we reject the validity of randomization and therefore the results of all RCTs

Table 1 *p*-values

The Bayesian prior for the mean difference for a baseline variable:

Prob(difference = 0) = 1

Prob(difference ≠ 0) = 0

Therefore the posterior is

Prob(difference = 0) = 1

Prob(difference ≠ 0) = 0

regardless of the observed data

In any case, using a Bayesian statement of inference such as, the probability that the mean difference is less than zero is 70%, reveals the fallacy immediately, because the population mean difference has to be zero due to randomization

Children with High risk, Stage 3 Neuroblastoma

ABMT - myeloablative chemotherapy, total-body irradiation and transplantation of purged autologous bone marrow

CC - intensive non-myeloablative continuation chemotherapy

5 years of recruitment: 72 eligible patients, 43 consented and enrolled

Park JR et. al. *Pediatr Blood Cancer* 2009; **52**:44–50

1-sided Fisher exact 0.13	Survival				Total
	No		Yes		
	n	row %	n	row %	N
Treatment Arm					
ABMT	7	35.0	13	65.0	20
CC	13	56.5	10	43.5	23

Bayesian Decision Theory

In the face of uncertainty, decision theory permits optimal decision making, answering the following questions:

Should a new intervention be adopted for future patients?

Is more research needed?

If so, how big should the study be?

Bayesian Decision Theory
Guiding Principles

A new intervention should be adopted if *no* more research is needed

More research is needed if the *value of the information* from the research is greater than its *cost*

The size of the study should maximize the difference between the *value* and the *cost*

Incremental Net Benefit (Utility)

Incremental net benefit of a new intervention defined as:

$b(\lambda) \equiv \lambda\Delta_e - \Delta_c$

Δ_e is the increase in mean effectiveness

λ is the threshold value placed on a unit of effectiveness

Δ_c is the increase in mean cost

Children with High risk, Stage 3 Neuroblastoma

Δ_e is the difference in probability of survival

Mean = 0.196; SD = 0.1402

Δ_c is the increase in mean cost

Mean = 50,000; SD = 50,000 (*i.e.* CV = 1)

λ = 500,000

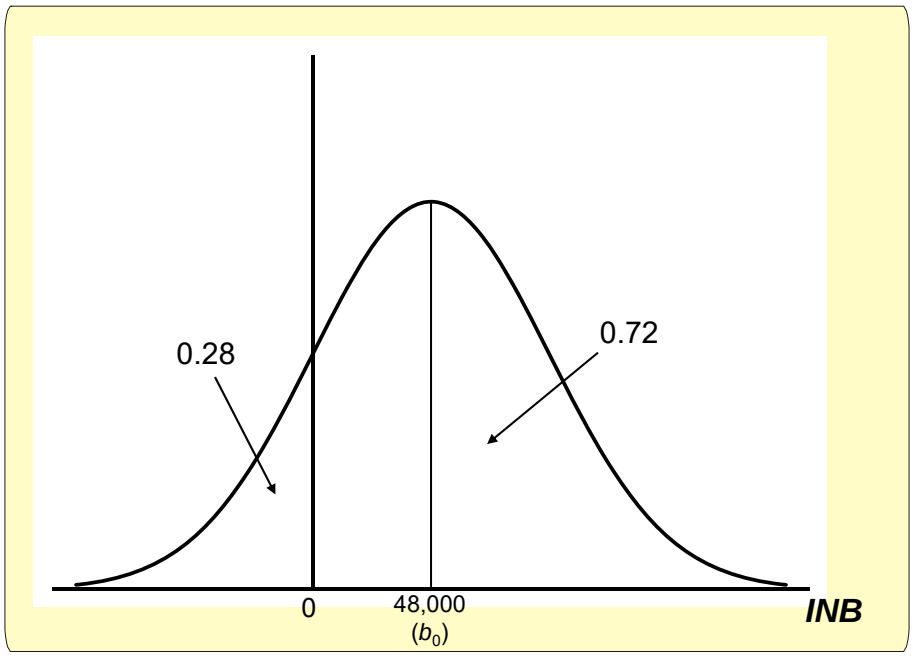
ICER = Δ_c/Δ_e = 50,000/0.196 = 255,102

b(500,000) ~ N(b_0,v_0) = N(48,000, 6,794,410,000)

Prob. cost effective: Prob[b(500,000) > 0] = 0.72

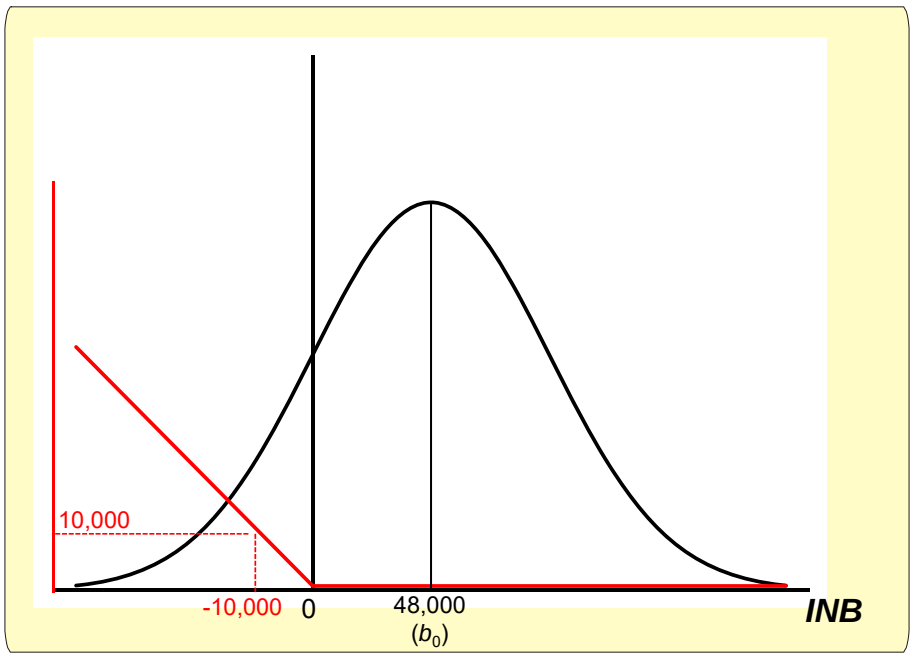
Value of Additional Evidence

Current Distribution of INB



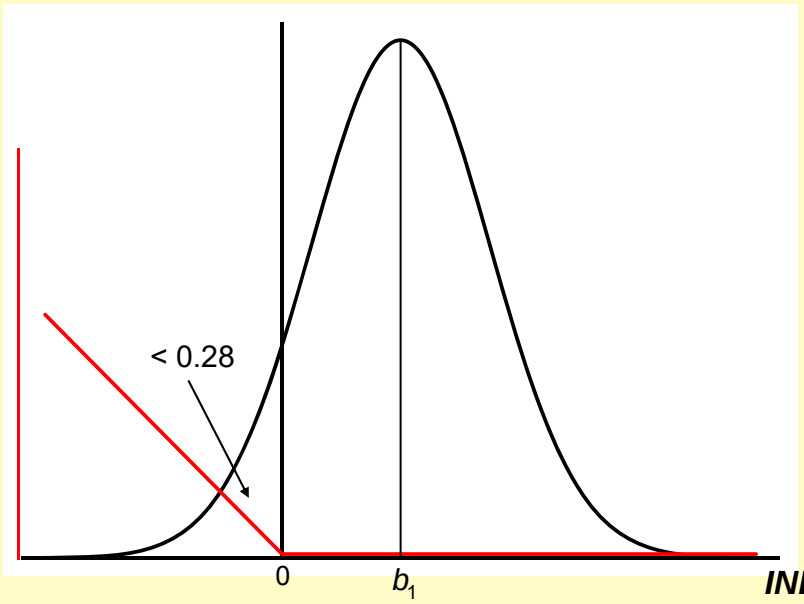
Value of Additional Evidence

Opportunity Loss Function and *Current* Distribution of INB



Value of Additional Evidence

Opportunity Loss and *"Future"* Distribution of INB



Reduction in patient Expected Opportunity Loss (EOL) = $\mathcal{K}(b_0, v_0, n, \sigma_+^2)$

Value of Additional Evidence

Reduction in Population Expected Opportunity Loss

Expected Value of new trial = the reduction in population EOL

$$EV(n) = k(h-t) \times \mathcal{K}(b_0, v_0, n, \sigma_+^2)$$

k = incidence

h = time horizon

t = duration of trial

k ⇑ Value of New Trial ⇑

$$\mathcal{K}(b_0, v_0, n, \sigma_+^2) =$$
$$v_0 \exp\left(-b_0^2(v_0 + \sigma_+^2/n)/(2v_0^2)\right) \Big/ \sqrt{2\pi(v_0 + \sigma_+^2/n)} - b_0 \Phi\left(-b_0\sqrt{(v_0 + \sigma_+^2/n)}/v_0\right)$$

Expected Total Cost

$ETC(n) = C_f + 2nC_v + (kt - n)b_0$

where

C_f = fixed financial cost

C_v = variable financial cost per patient

$(kt - n)$ is the number of patients who are denied intervention (*i.e.* receive standard) because of the trial, each of whom incur an expected opportunity cost of b_0

k ⇑ Cost of New Trial ⇑

Expected Net Gain

$ENG(n) = EV(n) - ETC(n)$

Let n^* maximize $ENG(n)$

If $ENG(n^*) < 0$ then current evidence is sufficient and optimal decision is to adopt the intervention

If $ENG(n^*) > 0$ then current evidence is insufficient and optimal decision is to do a trial with $2n^*$ patients

Children with High risk, Stage 3 Neuroblastoma

h = 20 years

k = 20 per year

accrual = 0.7k = 14 per year

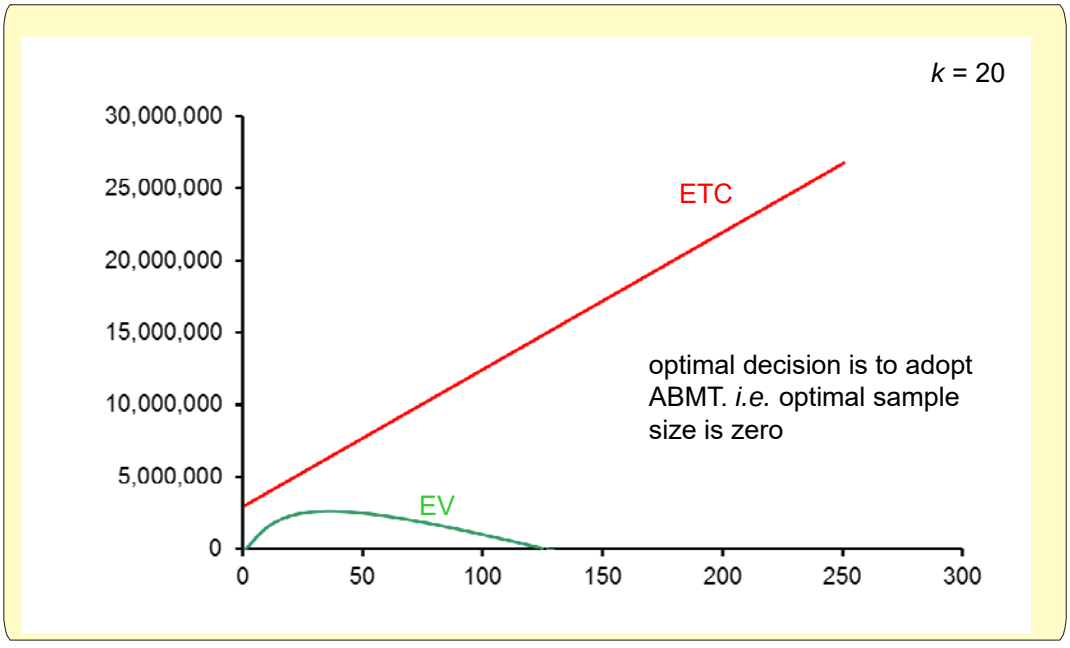
follow-up = 2 years

$t = (2n/14) + 2$

C_f = 1,000,000

C_v = 3000

Children with High risk, Stage 3 Neuroblastoma



Children with High risk, Stage 3 Neuroblastoma

Standard Approach

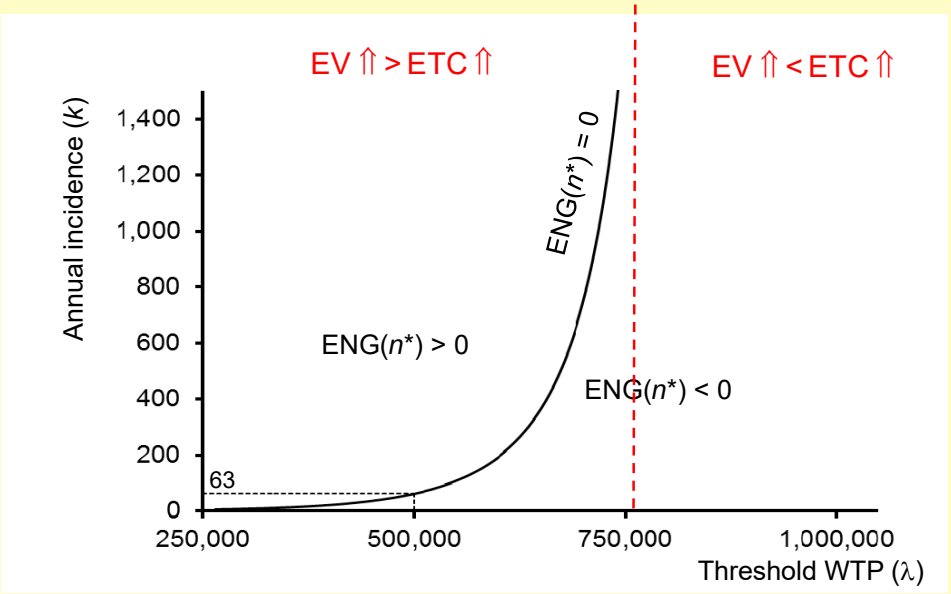
Type I error probability = 0.05, one-sided

Power = 0.8

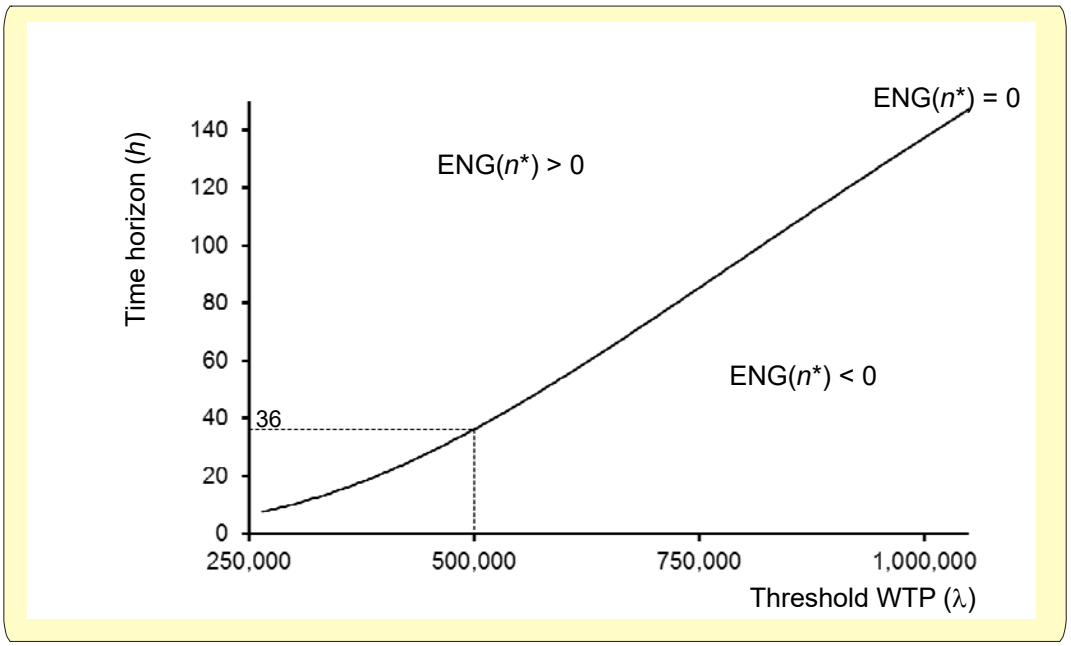
SCID = 0.1, based on an NNT of 10 $(\text{NNT} \times \Delta_c \leq \lambda)$

n/arm = 305

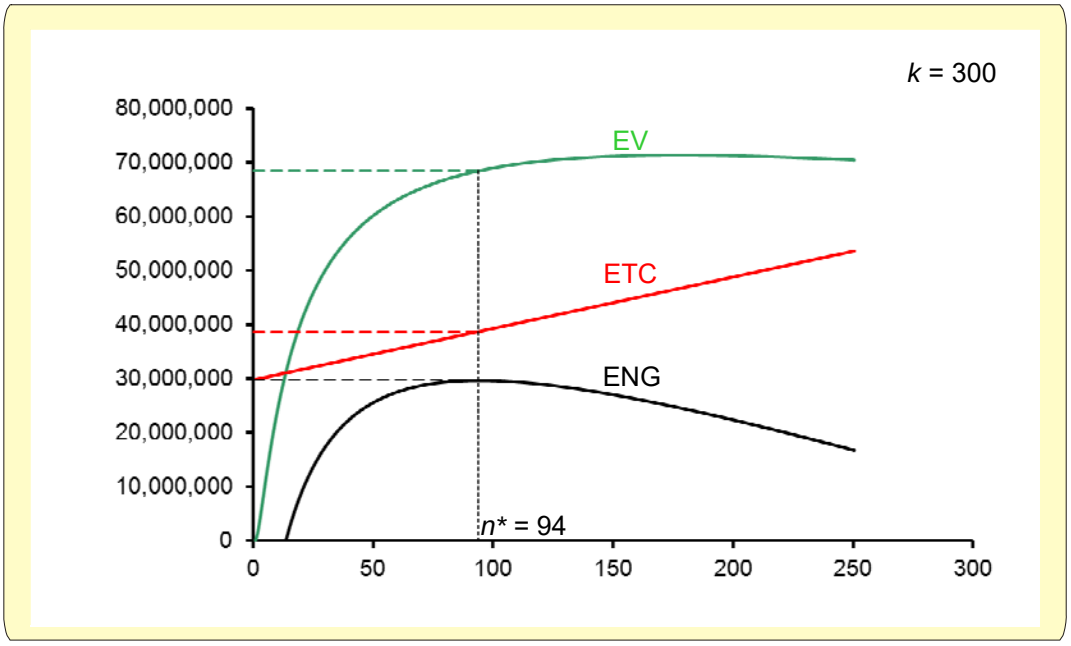
Children with High risk, Stage 3 Neuroblastoma



Children with High risk, Stage 3 Neuroblastoma



Children with High risk, Stage 3 Neuroblastoma



Summary

Bayesian Decision Analysis has advantages in assessing the evidence in support of new health care interventions

Takes into account:

- current evidence
- threshold value for health outcomes
- trial costs (financial and opportunity)
- accrual rate
- duration of follow-up
- time horizon
- incidence (requiring less evidence for rare health conditions)

Allow for comparison of “return for investment” between proposed trials

For rare health conditions, trials are smaller (and cheaper), may lead to less expensive interventions

Summary—Bayesian Advantages

Permits simple, intuitive and relevant statements of statistical inference

Provides a transparent framework for combining new information with current knowledge

Facilitates decision theory for optimal decision-making and research design

References—Bayesian Advantages

Spiegelhalter DJ, Abrams KR, Myles JP. (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester.

Willan AR. (2013) Bayesian methods provide important advantages for the design, analysis and interpretation of clinical studies. In: Berger VW, Zhang X. (Eds.) *Important Considerations for Clinical Trial Methodologies*. Future Medicine, London. (eISBN 978-1-909453-31-9)