An Evidential Paradigm for Experiments[1]

Qing Liu, Ph.D.
`Liu.Qing@QRMedSci.Net`

Quantitative & Regulatory Medical Science, LLC

Amicus Therapeutics, Inc.

BASS XXIV

October 24, 2017, Savannah, Georgia

Outline

Introduction

Scope

- Designed experiments or observational studies
- Statistical inference for assessing rival hypotheses
- Paradigm that assimilates frequentist and Bayesian theories and methods

Traditional Setting

- Control **false positive rates**
- Address controversies and substantial difficulties of the Neyman-Pearson-Fisherian paradigm

Big Data Setting

- Control **false discovery rates**
- New procedures for existing and new areas of applications

The Evidential Paradigm

Table: Anatomy of the evidential paradigm

	Question	Criterion	Freq./Bayes
I.	What is the evidence?	Replication	**fff/b**
II.	What should one believe?	Credibility	**ff/bb**
III.	What should one do?	Optimal Decision	**f/bbb**

Foundation - Five Basic Evidential Principles

(a) Measures of strength of statistical evidence for **rival hypotheses**
(b) Thresholds for assessing the strength of statistical evidence
(c) Consistent and objective interpretation of statistical evidence
(d) Indifferent to small effects of no practical interest
(e) Credible belief in alternative hypothesis against sources of bias

Canonical Likelihood Ratio Test

Canonical Z Statistic

- Normal distribution with $E(Z) = I^{1/2}\Delta$ and $Var(Z) = 1$ for **information size** I and effect size Δ
- H_0: $\Delta = 0$ versus H_δ: $\Delta = \delta$ for $\delta > 0$

Construction of Z

- Existing asymptotic likelihood theory following the conditionality and sufficiency principle
- Conditional, profile, partial likelihoods, etc.
- Wald's maximum likelihood statistic
- Rao's efficient score statistic
- Directed Neyman-Pearson likelihood ratio statistic

Other forms of Z

- Nonparametric statistics
- Permutation or randomization test statistics

Canonical Likelihood Ratio Test

Canonical Likelihood Ratio (CLR)

- Measure of the strength of statistical evidence of H_δ vs H_0

$$LR(\delta; Z, \mathcal{I}) = \exp\{Z(\mathcal{I}^{1/2}\delta) - (\mathcal{I}^{1/2}\delta)^2/2\}$$

- $\mathcal{I}$ is the **information index**, adjustable to ensure consistent and objective probabilistic interpretation
- $LR(\delta; Z, \mathcal{I})$ is the **likelihood ratio** when $\mathcal{I} = I$
- $\mu = \mathcal{I}^{1/2}\delta$ is the **experimental precept**

Evidential Threshold

- Type 1 and **maximum** type 2 error rates α and β_*, both $< .5$
- The **minimum experimental precept** given by $\mu_* = z_\alpha + z_{\beta_*}$
- **Evidential Threshold** given by

$$K_\alpha(\mu_*) = \exp(z_\alpha \mu_* - \mu_*^2/2)$$

Canonical Likelihood Ratio Test

Canonical Likelihood Ratio Test or CLR Test

- Assessment of **relative** strength by

$$LR(\delta; Z, \mathcal{I}) \geq K_\alpha(\mu_*)$$

- Maintain sufficient Z and identical $K_\alpha(\mu_*)$ in all applications
- Adjust the information index $\mathcal{I}$ for consistency

Determination of $K_\alpha(\mu_*)$ as a function of α

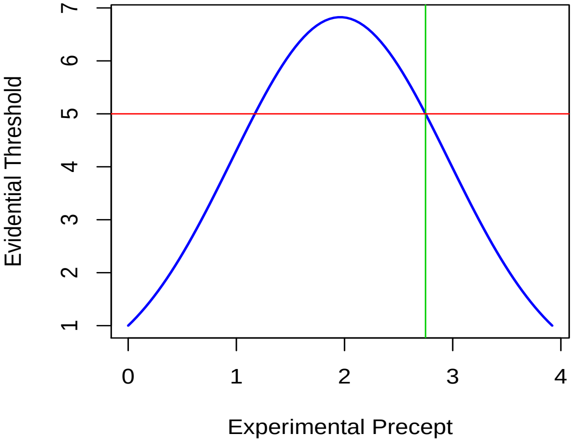
- Naive experiment
- Proportional function of $1/\alpha$
- Natural cubic spline

Example 1

- $K_\alpha(\mu_*) = 5$ and $\mu_* = 2.7490$ for $\alpha = .025$ and $\beta_* = .215$
- Credibility threshold $c_* = (1 - \beta_*)/\alpha = 31.4$

Evidential Thresholds

Figure: Evidential Threshold $K_\alpha(\mu)$ for $\alpha = .025$
(Green line for $\mu_* = 2.749$ with $\beta_* = .215$)



Evidential Thresholds

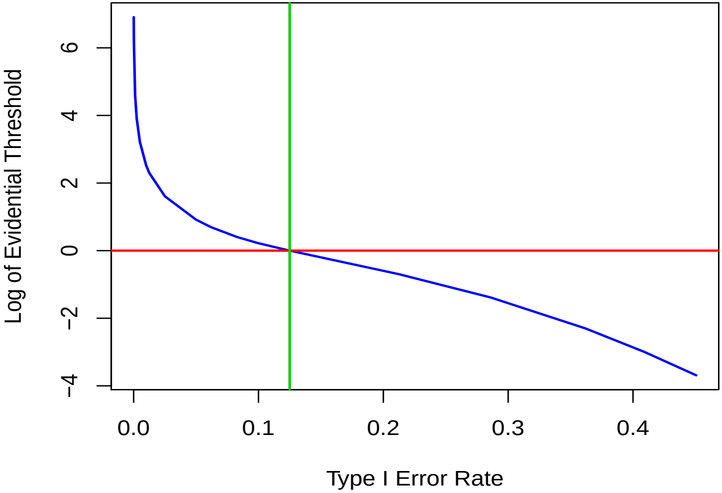
Table: Naive and Proportional Evidential Thresholds

α	$K_\alpha(\mu_*)$	μ_*	β_*	$1-\beta_*$	$c_*^\dagger$
		Naive Experiment			
.005	19.3615	3.4175	.2	.8	160
.010	10.5043	3.1680	.2	.8	80
.025	4.7902	2.8016	.2	.8	32
.05	2.7145	2.4865	.2	.8	16
.1	1.5952	2.1232	.2	.8	8
		Proportional Function			
.005	25.00	3.0198	.3285	.6715	134.30
.010	12.50	2.9267	.2741	.7259	72.59
.025	5.00	2.7490	.2150	.7850	31.40
.05	2.50	2.5792	.1751	.8249	16.50
.1	1.25	2.3752	.1371	.8629	8.63

† $c_* = (1-\beta_*)/\alpha$ is the credibility threshold with α and β_*

Evidential Thresholds

Figure: Full Range Threshold Function of α



Equivalence to the Significance Test

For I_* and δ_* such that $\mu_* = I_*^{1/2}\delta_*$

$$LR(\delta_*; Z, I_*) \geq K_\alpha(\mu_*) \text{ if and only if } Z \geq z_\alpha$$

Lemma

Let $\Psi(\Delta; \delta, \mathcal{I}) = P_{H_\Delta}\{LR(\delta; Z, \mathcal{I}) \geq K_\alpha(\mu_*)\}$ be the power of the CLR test and $\psi(\Delta; I) = P_{H_\Delta}\{Z \geq z_\alpha\}$ be the power of $Z \geq z_\alpha$

(i) $\Psi(0; \delta, \mathcal{I}) \leq \alpha$ for any $\mathcal{I}$ and δ such that $\mu = \mathcal{I}^{1/2}\delta \geq \mu_*$.

(ii) $\Psi(\Delta; \delta, I)$ converges to 0, 1/2, and 1 for $\Delta < \delta/2$, $\Delta = \delta/2$, and $\Delta > \delta/2$, respectively, as $I \to \infty$.

(iii) For any I and δ, let $\psi(\delta, I) \geq 1 - \beta_*$. Then,

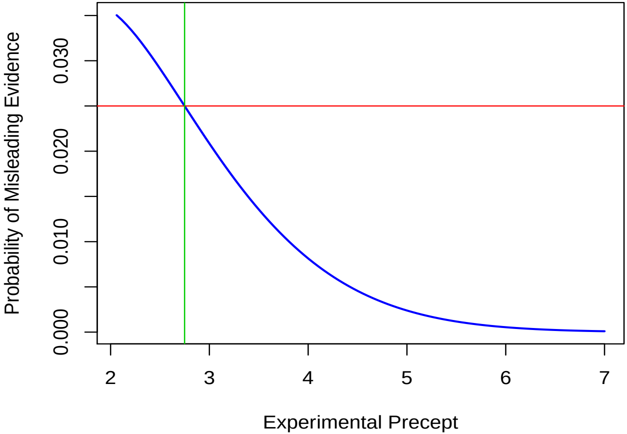
$$\mathcal{C}(\Delta; \delta, I) = \Psi(\Delta; \delta, I)/\Psi(0; \delta, I) \geq \psi(\Delta, I)/\alpha$$

for $\Delta \geq 0$, provided that $K_\alpha(\mu_*) \leq (1 - \beta_*)/\alpha$.

Consistency Principle

Example 2

- $K_\alpha(\mu_*) = 5$, $\mu_* = 2.749$ for $\alpha = .025$ and $\beta_* = .215$
- Let $\delta = .15$, $\beta = .2$ and $I_* = (z_\alpha + z_\beta)^2/\delta^2 = 348.84$, then $\mu = I_*^{1/2}\delta = 2.802 \geq \mu_*$ for $\delta > \delta_* = \mu_*/I_*^{1/2} = .147$.



Consistency Principle - Theorem

Characteristics of Practical Applications

- Specific design features
- Model uncertainty

Consistency Theorem

Let $P_{[d,H_0]}$ be the probability model under H_0 with a certain design **deviation** d and z_{α_d} be the standard normal critical value at α_d such that

$$P_{[d,H_0]}\{Z \geq z_{\alpha_d}\} \leq \alpha.$$

The adjusted experimental precept is

$$\mu_d = z_{\alpha_d} + [z_{\alpha_d}^2 - 2\log\{K_\alpha(\mu_*)\}]^{1/2}.$$

Then,

$$P_{[d,H_0]}\{LR(\delta; Z, \mathcal{I}) \geq K_\alpha(\mu_*)\} \leq \alpha$$

for any $\mathcal{I}$ and δ such that $\mu = \mathcal{I}^{1/2}\delta \geq \mu_d$

Consistency Principle - Adaptive Design

Adaptive Designs

- Probability space $(\Omega, \mathcal{F}, P_\Delta)$, a $\mathcal{F}$-measurable **adaptation rule** g with countable range M
- Collection of test statistics $\{Z_m \colon m \in M\}$ with the global null hypothesis $H_0 = \bigcap_{m \in M} H_m$
- Adaptive statistic $Z_g = \sum_{m \in M} \phi_m Z_m$ where $\phi_m = 1_{\{g=m\}}$

Procedure

- Determine z_{α_g} such that $P_{[g,H_0]}\{Z \geq z_{\alpha_g}\} \leq \alpha$
- Calculate $\mu_g = z_{\alpha_g} + [z^2_{\alpha_g} - 2\log\{K_\alpha(\mu_*)\}]^{1/2}$
- Choose $\mathcal{I}$ such that $\mu = \mathcal{I}^{1/2}\delta \geq \mu_g$

Applications Multiplicity, (group) sequential designs, adaptive designs, treatment selection, etc.

Consistency Principle - Robust Inference

Model Uncertainty

- Failure of asymptotic theory with small size
- Violation of model assumption (e.g., over-dispersion, non-normality)
- Unknown sampling scheme (e.g., Royal vs. Cox tea time stopping rule)
- Unknown differential missing data mechanism

Procedure

- Formulate a collection of probability models $\{P_{[m,H_0]} : m \in M\}$
- Determine z_{α_m} such that $P_{[m,H_0]}\{Z \geq z_{\alpha_m}\} \leq \alpha$
- Set $z_{\alpha^*} = \max_{m \in M} z_{\alpha_m}$
- Calculate $\mu^* = z_{\alpha^*} + [z_{\alpha^*}^2 - 2\log\{K_\alpha(\mu_*)\}]^{1/2}$
- Choose $\mathcal{I}$ such that $\mu = \mathcal{I}^{1/2}\delta \geq \mu^*$

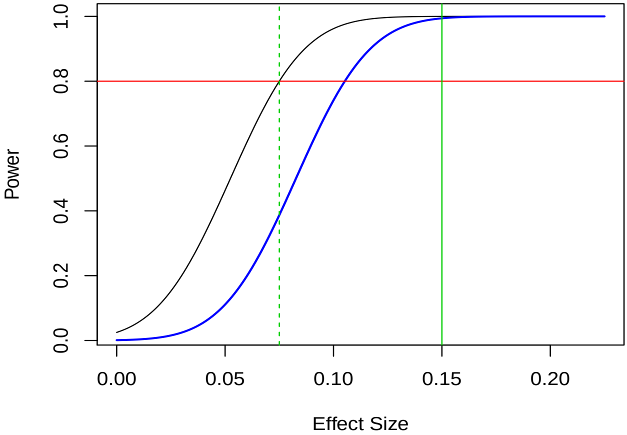
Model Selection Criterion

Consistent with the observed data pattern

Indifference Principle

Example 3

- $K_\alpha(\mu_*) = 5$, $\mu_* = 2.749$ for $\alpha = .025$ and $\beta_* = .215$
- Let $\delta = .075$, $\beta = .2$, then $I_* = 4 \times 348.84$.



Credibility Principle - Epistemic Justification

Credibility of the CLR Test

- The credibility of the CLR test is measured by the ratio

$$\mathcal{C}(\delta, \mathcal{I}) = \Psi(\delta; \delta, \mathcal{I}) / \Psi(0; \delta, \mathcal{I})$$

- $\mathcal{C}(\delta; \mathcal{I})$ is deflated by various sources of bias, misconduct and fraud before, during and after the experiment
- Strong statistical evidence is not sufficient for a favorable conclusion towards the stated alternative hypothesis

Sources of Bias, Misconduct or Fraud

- Imbalance in confounding factors in randomization or sampling
- Missing data pattern in favor of the alternative hypothesis
- Post-hoc analysis, choice of (enriched) subpopulation, etc.
- Selective publication and regulatory submission
- Falsification of data and results; selective reporting; misuse of statistical methods and regulatory guidance

Credibility Principle - Epistemic Justification

Bayesian Belief

- For given priors ξ_0 and ξ_δ for H_0 and H_δ, the posterior probability ratio
$$(\xi_\delta/\xi_0)\mathcal{C}(\delta, \mathcal{I})$$
measures the **belief** in the alternative hypothesis H_δ
- Belief is also altered by bias, misconduct or fraud through increased ξ_δ/ξ_0
- Select adjusted priors π_0 and π_δ towards H_0 away from H_δ such that
$$\pi_\delta/\pi_0 < \xi_\delta/\xi_0$$
- For a **belief threshold** γ_B, require that
$$(\pi_\delta/\pi_0)\mathcal{C}(\delta, \mathcal{I}) \geq \gamma_B \text{ or } \mathcal{C}(\delta, \mathcal{I}) \geq (\pi_0/\pi_\delta)\gamma_B$$
- Misconduct or fraud not **admissible** for a conclusion, irrespective of the strength of statistical evidence

Credibility Principle - Theorem

Credibility Theorem

(i) For $\gamma \geq (1 - \beta_*)/\alpha > 1$ and any $\delta \geq \delta_*/2$, there is a unique information size I such that $\mu = I^{1/2}\delta \geq \mu_*$ and

$$\mathcal{C}(\delta, I) = \gamma.$$

(ii) For $\gamma_2 > \gamma_1 \geq (1 - \beta_*)/\alpha > 1$, let $\delta_j \geq \delta_*/2$ and I_j satisfy

$$\mathcal{C}(\delta_j, I_j) = \gamma_j$$

for $j = 1, 2$. Let $\alpha_j = P_{H_0}\{LR(\delta_j; Z, I_j) \geq K_\alpha(\mu_*)\}$, $\beta_j = 1 - P_{H_{\delta_j}}\{LR(\delta_j; Z, I_j) \geq K_\alpha(\mu_*)\}$ and $\mu_j = I_j^{1/2}\delta_j$ for $j = 1, 2$. Then

$$\alpha_2 < \alpha_1 \leq \alpha, \beta_2 < \beta_1 \leq \beta_* \text{ and } \mu_2 > \mu_1 \geq \mu_*,$$

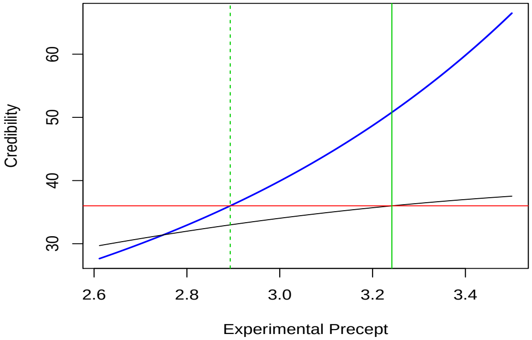
Credibility Criterion

Choose the information size I according to the credibility criterion $\mathcal{C}(\delta, I) = \gamma$ for a desired **credibility threshold**

Credibility Principle - Efficiency of the CLR Test

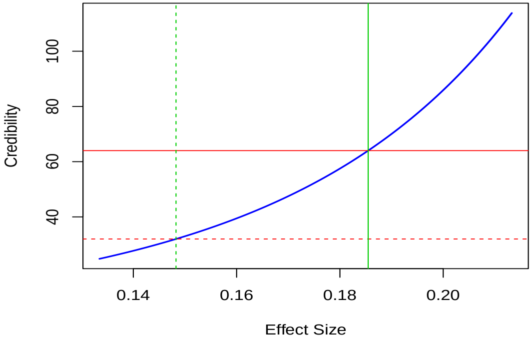
Example 4

- $K_\alpha(\mu_*) = 5$ for $\alpha = .025$
- Let $\delta = .15$, $\beta = .1$. Then for $\gamma = (1-\beta)/\alpha = 36$, $I_* = 467$ and $I = 372$; 20% saving!
- Power and type 1 error rates: .0226 and .813 for the CLR test; .025 and .825 for the NPLR test



Example 5

- $K_\alpha(\mu_*) = 5$ for $\alpha = .025$, $\beta = .2$, and $\gamma = 32$.
- Interim $I_1 = 170.42$ and final $I = 340.84$
- For $\gamma_1 = 64$, $\delta_1 = .1855$, $\alpha_1 = .014$ and $\beta_1 = .1025$



Credibility Principle - Group Sequential Design

D. R. Cox (2004)

> *"Indeed, I believe that many statisticians approaching statistics from a broadly frequentist perspective are uneasy at notions such as 'spending error rates', perhaps because these treat notions of error rates as more than just hypothetical concepts used for calibrating measures of uncertainty against performance in idealized situations. While in some situations there may be compelling quasi-political arguments, as well as cost considerations, pointing against too frequent an analysis, in principle it is hard to see an argument at a completely fundamental level."*

Evidential Group Sequential Design

- More efficient or credible
- Continuous monitoring **without** increasing the maximum information size

Credibility Principle - False Discovery Rate (FDR)

FDR (Benjamini and Hochberg, 1995)

$$FDR = E(V/R \,|\, R > 0)P\{R > 0\} \leq E(V/R \,|\, R > 0)$$

Bayes pFDR (Storey, 2003)

$$E(V/R \,|\, R > 0) = \frac{1}{1 + \{(1-\beta)/\alpha\}\{(1-\pi_0)/\pi_0\}}$$

where α and β are type 1 and 2 error rates of a **single test** and π_0 is the prior probability of the null hypothesis.

Roles of the Evidential Paradigm

- For given $(1-\beta)/\alpha$, choice of α and β
- Information size or effect size determination
- Evidential discoveries as opposed to those of ordered *p*-values

Credibility Principle - False Discovery Rate (FDR)

Evidential Procedure

- For given Bayes pFDR threshold η and (estimated) prior probability π_0, determine the credibility threshold c_* such that

$$\frac{1}{1 + c_*\{(1-\pi_0)/\pi_0\}} = \eta$$

- Choose α and β_* such that $(1-\beta_*)/\alpha = c_*$
- Choose I or δ such that $I^{1/2}\delta = \mu_* = z_\alpha + z_{\beta_*}$
- Perform the CLR test $LR(\delta; Z, \mathcal{I}) \geq K_\alpha(\mu_*)$

Example 6

- For $\eta = .05$ and $\pi_0 = .95$, we have $c_* = 361$, $\alpha = .001644$, $\beta_* = .406476$, $\mu_* = 3.176046$ and $K_\alpha(\mu_*) = 73.121352$
- For $I = 1$, $\delta_* = 3.176046$; for $I = 2$, $\delta_* = 2.245804$.

Credibility Principle - Bayesian Decision Framework

Notations and Assumptions

- Actions: A and R for accepting and rejecting H_δ, respectively
- Loss function: $L(A,0)$, $L(R,0) = 0$, $L(A,\delta)$ and $L(R,\delta)$. It is assumed that $L(A,\delta) < L(R,\delta)$.
- Prior probabilities: π_0 and π_δ for H_0 and H_δ, respectively

Posterior Risk Ratio

$$\frac{PR(A \,|\, LR \geq K)}{PR(R \,|\, LR \geq K)} = \left\{\frac{1}{\mathcal{C}(\delta,\mathcal{I})}\right\}\left\{\frac{\pi_0}{\pi_\delta}\right\}\left\{\frac{L(A,0)}{L(R,\delta)}\right\} + \frac{L(A,\delta)}{L(R,\delta)}$$

where $\mathcal{C}(\delta,\mathcal{I})$ is the credibility of $LR(\delta; Z, \mathcal{I}) \geq K_\alpha(\mu_*)$

Optimal Decision

Accept H_δ given $LR(\delta; Z, \mathcal{I}) \geq K_\alpha(\mu_*)$ if and only if

$$\mathcal{C}(\delta,\mathcal{I}) \geq \left\{\frac{\pi_0}{\pi_\delta}\right\}\left\{\frac{L(A,0)}{L(R,\delta) - L(A,\delta)}\right\}$$

Research History - A Personal Quest

- Accumulated various problems in basic science and clinical research from past 20 years' practice in academic, regulatory and industry settings
- Given up on frequentist approach to adaptive designs in 2006
- Carefully studied works by Edwards (1972) and Royall (1997)
- Interviewed with AnalyticalEdge, Inc. for the CTO position
- Further developed the evidential paradigm, addressed ALL criticisms
- Developing evidential based designs for various clinical trials
- Discovered new applications in big data science
- Compiled a big list of current and future research topics
- Super secrecy, **first time group disclosure in 2016**

About QRMedSci, LLC.

Founded by Qing Liu, Ph.D., ASA Fellow (2014) to serve innovative small and medium sized biopharmaceutical and medical device companies in the following areas:

- Innovative therapies (e.g., immuno-oncology, breakthrough designation, rare disease, individualized medicine, cutting edge pharmaceutical, medical device)
- Strategic clinical development plan and innovative trial designs
- Blinded trial monitoring, efficacy response signature, adaptive statistical analysis planning and trial modifications
- Interim analysis and trial adaptations
- AbacusCloud^{TM} super-computing for clinical trial designs and simulations
- Scientific and regulatory review of clinical plan, trial design, reports and regulatory submission packages
- Competitive basic statistical and programming service through partnership
- Collaborative statistical research