Examining Benefit and Risk Simultaneously at the Individual Patient Level

Christy Chuang-Stein

Statistical Research and Consulting Center

Nov 8 2013

ASA NJ Chapter Workshop

Outline

- Classify Benefit and Risk Jointly at the Patient Level
- Discount Benefit by Risk at the Patient Level
- Clinical Utility Index
- QSPI Benefit-Risk Working Group
- SCT/FDA Benefit-Risk Workshop
- Summary

Major References

- Chuang-Stein C, Mohberg NR, Sinkula M. (1991) Three measures for simultaneously evaluating benefits and risks using a categorical data from clinical trials. *Stat in Med*, 10:1349-1359.

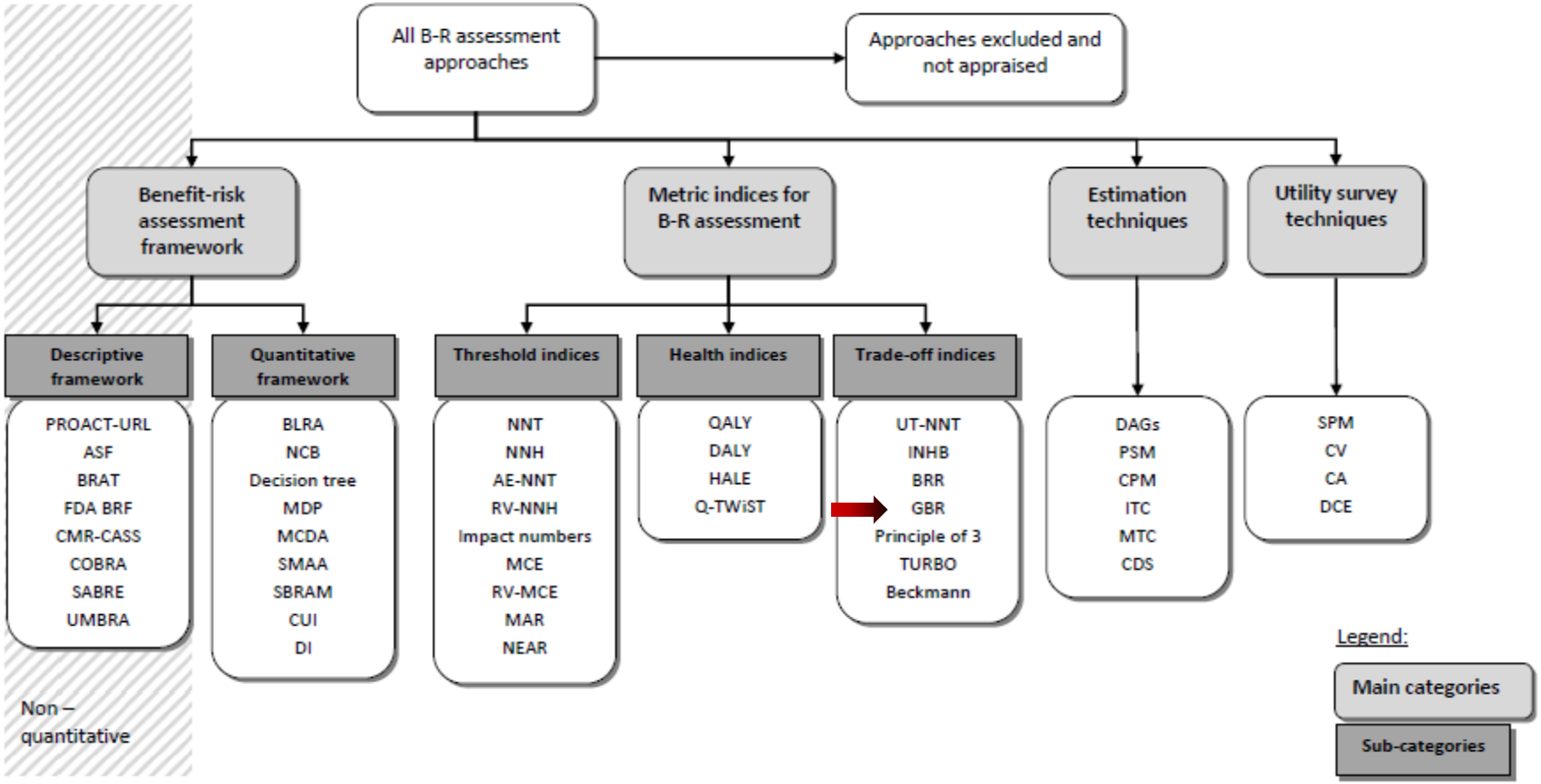
- Chuang-Stein C. (1994) A new proposal for benefit-less-risk analysis in clinical trials. *Control Clin Trials*, 15:30-43.

- Chuang-Stein C, Entsuah R, Pritchett Y. (2008) Measures for Conducting Comparative Benefit:Risk Assessment. *Drug Information J*, 42:223-23.

- Norton J. (2011). A longitudinal model and graphic for benefit-risk analysis, with case study. *Drug Information Journal*, 45:741-747.

3.3.5 Snapshot of classifications

Figure 2 Classifications of benefit-risk assessment approaches



Source: IMI Protect Consortium Report, 14 Feb 2012

Classify Benefit and Risk Jointly

- Use pre-specified rules, categorize patient overall (benefit, risk) experience into 5 categories

C1: Benefit and no pre-specified risk (π_1)	C2: Benefit and pre-specified risk (π_2)
C3: No benefit and no pre-specified risk (π_3)	C4: No benefit and pre-specified risk (π_4)
C5: Adverse experience leading to serious complications or withdrawal from the study (π_5)	

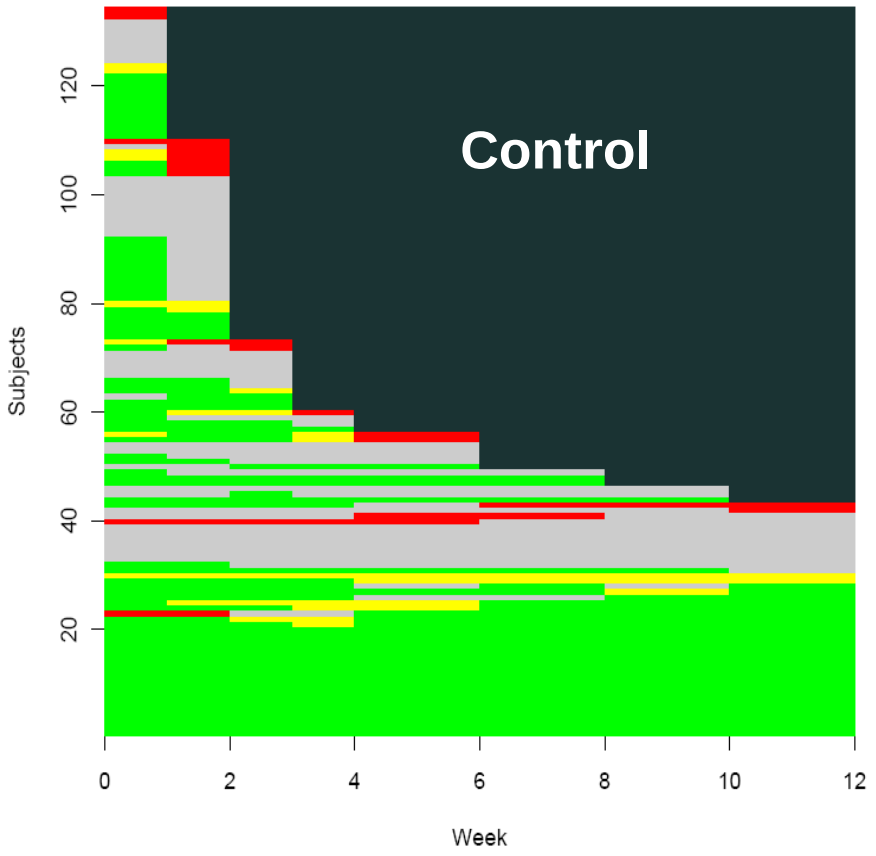
Source: Chuang-Stein, 1991.

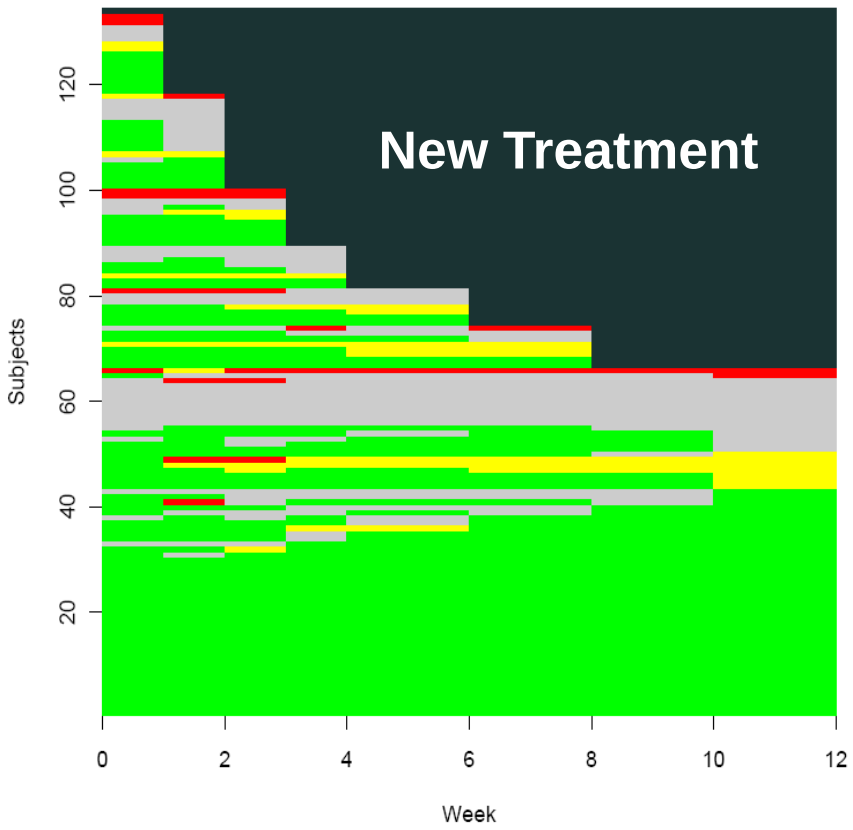
Comparing Two Distributions

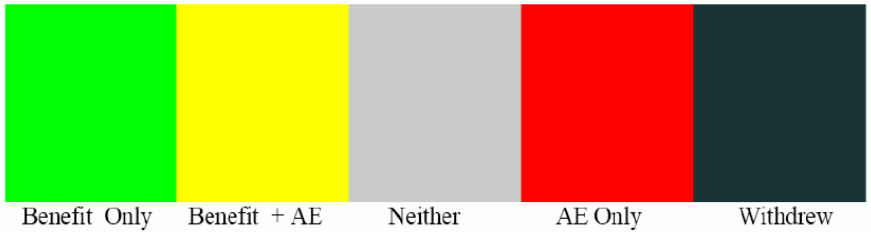
- If we can assume C1 > C2 ≥ C3 > C4 > C5, we can test if the distribution of the response under the new treatment is stochastically smaller than that for the control (i.e. the new treatment has a higher % in the *lower* categories).

- The above can be done using a test based on an association measure, Jonckheere-Terpstra test, order-restricted test treating the response as continuous, fitting models such as a proportional odds model etc.

Display Outcome Longitudinally







Source: Norton (ICSA Symposium, 2010)

Statistics from Longitudinal Display

- We can calculate % of the area occupied by different colors for each treatment group.
- A hypothetical table:

	C1	C2	C3	C4	C5	N
Control	**0.34**	**0.01**	**0.10**	**0.05**	**0.50**	**150**
New Treatment	**0.55**	**0.03**	**0.07**	**0.05**	**0.30**	**150**

Extensions

- If there are 2 mostly independent AEs of equal concerns and that having 2 AEs is substantially worse than having 1, we could construct a net measure of “Benefit Response (0 or 1) - # of AEs”. This measure can take values of 1, 0, -1, -2 and withdrawal.

- If, in addition, there are two equally important and distinct categorical responses to assess benefit, we could construct a measure of “# of Benefit Responses - # of AEs” (2, 1, 0, -1, -2, withdrawal) and display the results graphically.

- Be aware of the assumptions behind these measures.

Classify Benefit and Risk Jointly

C1: Benefit and no pre-specified risk (π_1)	C2: Benefit and pre-specified risk (π_2)
C3: No benefit and no pre-specified risk (π_3)	C4: No benefit and pre-specified risk (π_4)
C5: Adverse experience leading to serious complications or withdrawal from the study (π_5)	

$$GBR\, r = \frac{\left(w_1\pi_1 + w_2\pi_2\right)^e}{w_3\pi_3 + w_4\pi_4 + w_5\pi_5}, \text{ which is a multiple of}$$

$$r^* = \frac{\left(\pi_1 + w_2^*\pi_2\right)^e}{w_3^*\pi_3 + w_4^*\pi_4 + \pi_5}, \text{ if } w_1 w_5 \neq 0$$

Observations

$$r^* = \frac{\left(\pi_1 + w_2^* \pi_2\right)^e}{w_3^* \pi_3 + w_4^* \pi_4 + \pi_5}$$

- Other GBR measures could also be constructed.
- When $\{\pi_i\}$ are replaced by the observed rates $\{p_i\}$, one can use the delta method to derive the asymptotic variance for $\log \hat{r}^*$.
- For given w_i^*, one may compute $\log \hat{r}_T^* - \log \hat{r}_C^*$.

An Example (*Chuang-Stein et al, 1991*)

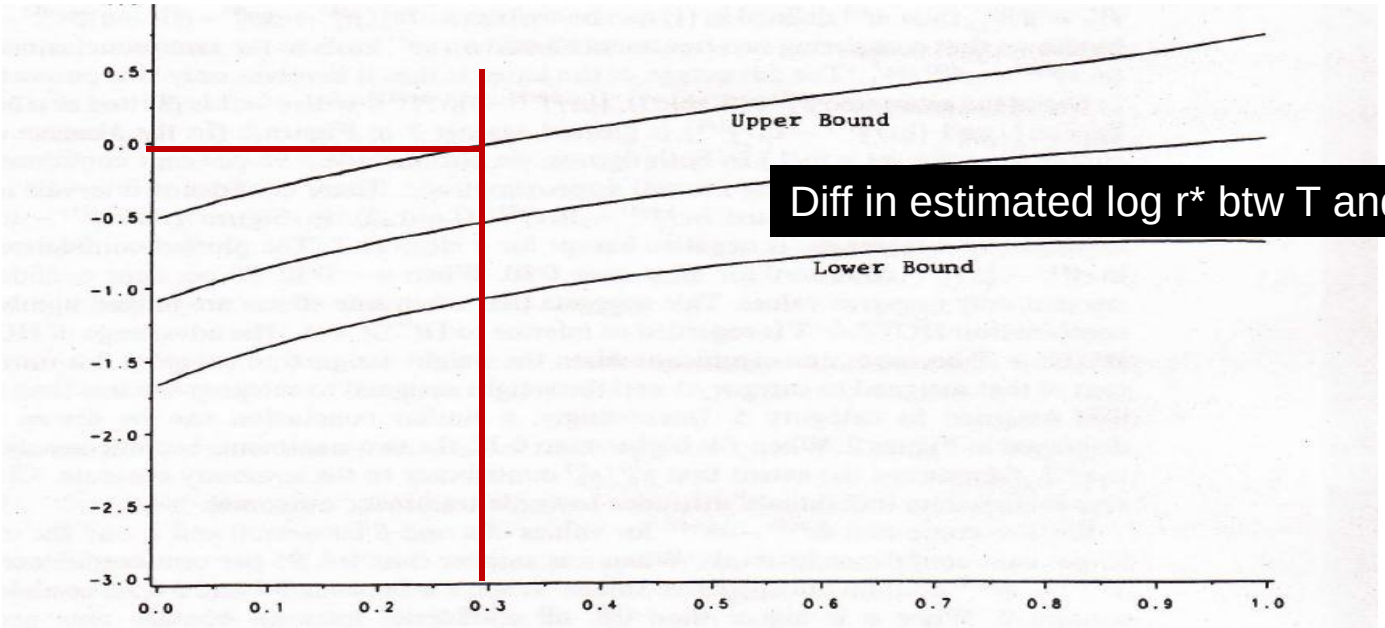
	C1	C2	C3	C4	C5	N
HCTZ + T	50	153	14	24	20	261
HCTZ + S	101	40	82	16	14	253

$$r^* = \frac{\left(\pi_1 + w_2^* \pi_2\right)^e}{w_3^* \pi_3 + w_4^* \pi_4 + \pi_5}$$

Set $w_3^* = 0, w_4^* = 1 - w_2^*, e = 1.$

Plot $\log \hat{r}_T^* - \log \hat{r}_S^*$ against w_2^* with asymp 95% CI.

An Example (continuing)



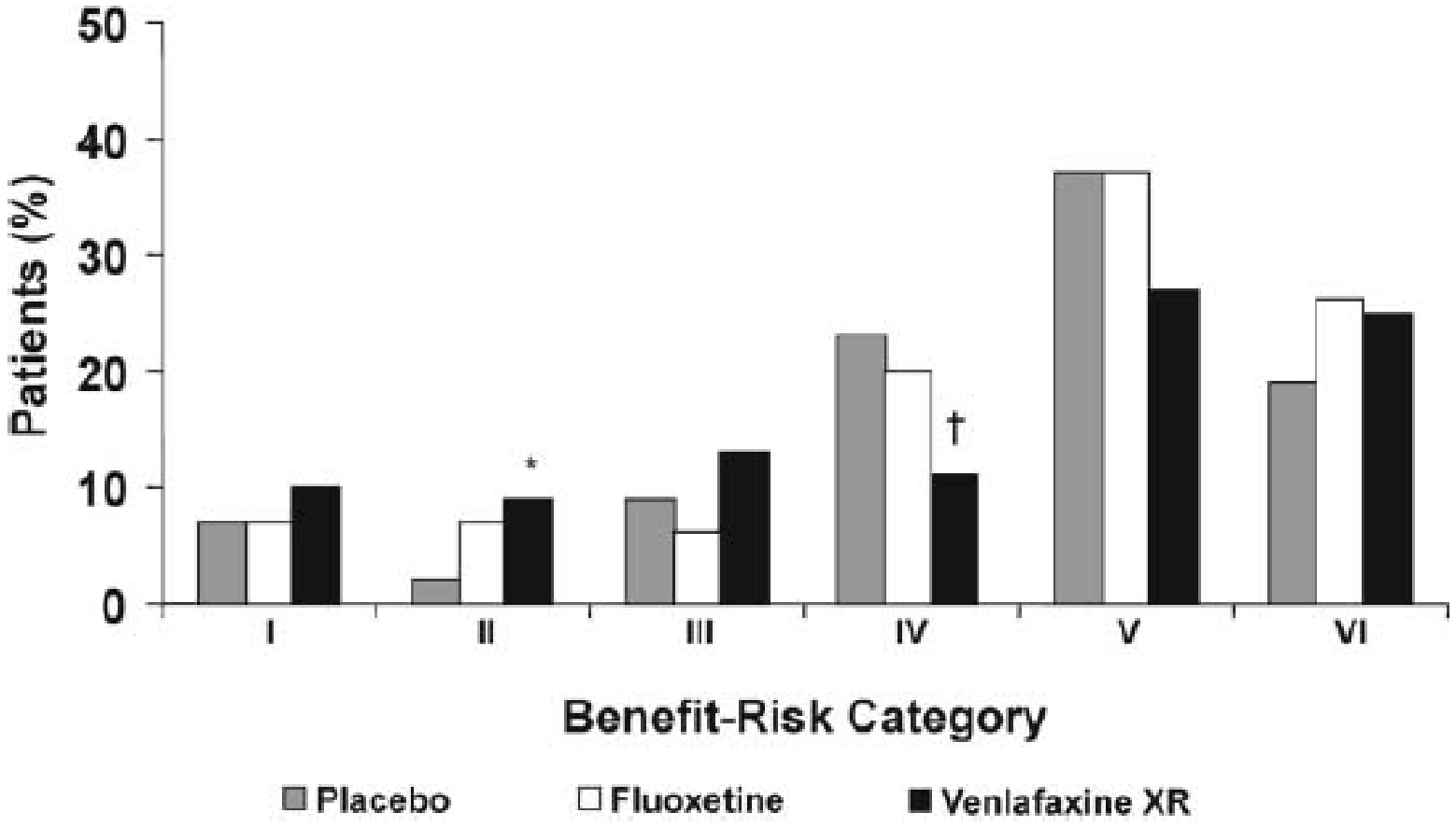
- When $w_2^* < 0.3$, 95% CI contains only negative values.
- The above means that when side effect is highly undesirable, HCTZ+T is worse than HCTZ+S since higher values of the GBR ratio are more desirable.

An Extension

Response and no AE $(C_1; \pi_1)$	**Response and mild AE** $(C_2; \pi_2)$	**Response and moderate/severe AE** $(C_3; \pi_3)$
No response and no AE; or dropout due to LOE or non treatment reason $(C_4; \pi_4)$	**No response and mild AE** $(C_5; \pi_5)$	**No response and moderate/severe AE; or AE-related dropout** $(C_6; \pi_6)$

***Source: Entsuah & Gorman (2002). J Psy Research, 36:111-118*.**

Distribution of the Categories



GBR Measure

$$r = \frac{(w_1\pi_1 + w_2\pi_2 + w_3\pi_3)^e}{w_4\pi_4 + w_5\pi_5 + w_6\pi_6}$$

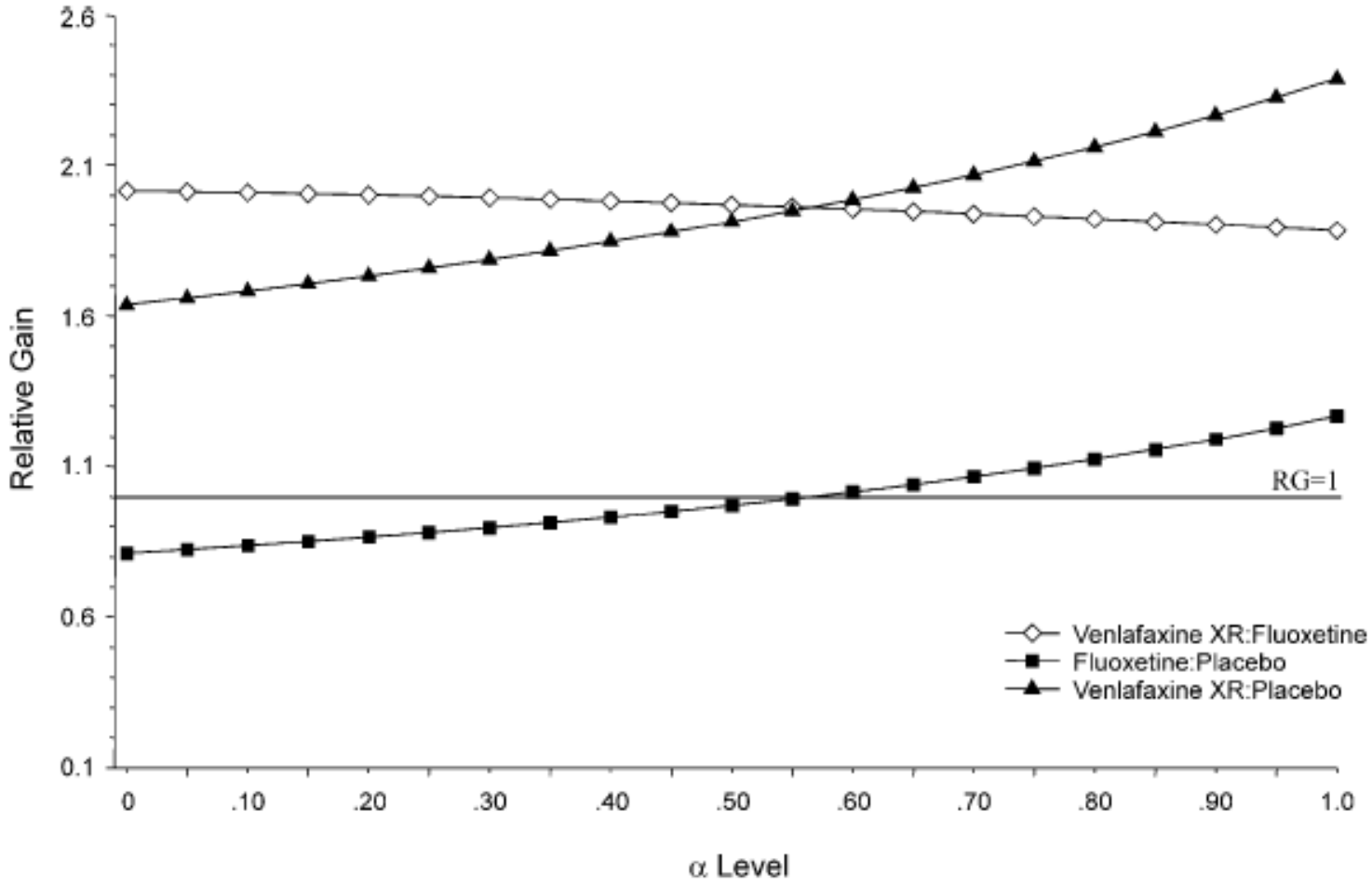
One option:

$$e = 1,\ w_4 = w_3,\ w_5 = w_2,\ w_6 = w_1$$

$$w_2 = \alpha w_1,\ w_3 = (1-\alpha)w_1$$

$$w_1 = 1$$

Three Pairwise Comparisons



Patient Level Measure

- Overarching concept: the original benefit (efficacy) measure is discounted by the presence of untoward safety experience according to pre-specified rules at the individual patient level.
- TWiST: Time without symptoms of disease and toxic effects (Ref: Gelber et al 1989, *Biometrics*)
- Q-TWiST: Quality-adjusted TWiST (Ref: Glasziou et al. 1990, *Stat in Medicine*)
- Quality-adjusted life years (QALY) gained has been used in Health Technology Assessment.

Benefit-Less-Risk Measure

- Suppose e_i is the benefit experienced by subject *i*.
- Suppose RS_i (RS for risk score) is the score summarizing the risk experienced by subject *i*.
- We can discount the benefit by a multiple of the risk score in a linear fashion to obtain an adjusted benefit.

$$e_i^* = e_i - f \times RS_i$$

An Example (*Chuang-Stein, 1994*)

- Situation: Patients with exertional angina were randomized to an investigational medication or a beta blocker for control of their angina pectoris.
- Efficacy: Change in “time to angina during an exercise tolerance test” at the end of 4 weeks of treatment from baseline.
- Safety: Data were collected on clinical signs and symptoms, regular labs, ECG. Within each system organ class (SOC), four grades (0, 1, 2, 3) were used to describe the overall experience of each patient in that SOC (0 = none, 1 = mild, 2 = moderate, 3 = severe).

How to Organize Safety Data

SOC (weight)	Grade 0	Grade 1	Grade 2	Grade 3
CVD (6)	0	6	12	18
Pulmonary (3.5)	0	3.5	7.0	10.5
Neuro/Psy (3)	0	3	6	9
Hematologic (2)	0	2	4	6
Musculoskeletal (2)	…	…	…	…
GI/Hepatic (1.5)	0	1.5	3.0	4.5
GU/Renal (1.5)	…	…	…	…
Metabolic (1.5)	…	…	…	…
Special Senses (1)	0	1	2	3
Dermatologic (1)	…	…	…	…
				Max = 69

Determine the Discounting Factor

- Each individual has a total score obtained by summing his/her scores from the 10 SOCs.
- Dividing the above total score by the maximum total score (69), we obtain a score between 0 and 1. We define this re-scaled score to be the risk score for an individual.
- Literature suggests that the average placebo effect is about 27 seconds. If it is felt that this benefit is not worthy of grade 2 CVD experience, one can set *e* = 27 and *RiskScore* = 12/69 and solve for the discounting factor *f. f* is about 149.

$$0 = e - f \times RS$$

Comparing between Two Drugs

Discounting Factor *f*	P-value from Comparing between Drug A and Drug B
0	0.045
20	0.047
40	0.049
60	0.051
80	0.054
100	0.056
120	0.059
140	0.063
150	0.065
200	0.075
400	0.156

CUI in Early Development

The Use of a Clinical Utility Index to Compare Insomnia Compounds: A Quantitative Basis for Benefit–Risk Assessment

D Ouellet[1,2], J Werth[1], N Parekh[1], D Feltner[1], B McCarthy[1] and RL Lalonde[1]

The use of a clinical utility index (CUI) was proposed in order to compare two calcium channel α2δ ligands that were in development for the treatment of insomnia. The important attributes included in the CUI were two measures of residual sedation and five measures of efficacy (wake after sleep onset, sleep quality, sleep latency, and sleep stages (stage 1 and stages 3–4)). Dose–response analyses were conducted on each end point, and a sensitivity analysis was conducted to determine a clinically meaningful difference in CUI. Nonparametric bootstrap parameters were used to build confidence intervals (CIs). Peak CUI (80% CI) was 0.345 (0.25–0.43), observed at a dose of ~30 mg with the lead compound and 0.436 (0.35–0.52) observed at >600-mg dose for the backup. Although CUI was slightly greater for the backup, peak CUI values were observed at doses that were not considered viable, and therefore development of the ligand was discontinued. The use of the CUI allowed an efficient, quantitative, and transparent decision.

Source: Ouellet D, et al. Clin Pharmacol Ther 2009;85:277-82.

Normalizing and Weighting

Table 1 List of attributes, weights, and clinical differences used in the calculation of the CUI

	CUI—attribute	Clinical difference	Weight (%)
1	Residual effect (two measures from Leeds questionnaire)	5 points	35
2	Wake after sleep onset (min)	25 min	25
3	Quality of sleep	20 points	17
4	Latency to persistent sleep (min)	15 min	13
5	Sleep architecture (% in stage 1, % in stages 3–4)	5%	10

Source: Based on input from 581 physicians in the insomnia field. Ouellet D, et al. Clin Pharmacol Ther 2009;85:277-82.

QSPI Benefit-Risk Working Group

- Formed in Feb 2012, co-led by Qi Jiang (Amgen) and Weili He (Merck) with members from industry, government and academia.

- Objectives
  - Prepare the WG members for increased use of benefit-risk assessment approaches through training/education.
  - Educate the broader statistical community on this important area via publications, newsletters, conferences, website…, and promote the role of statisticians' leadership through cross-functional collaborations.
  - Develop new methods to address existing B-R challenges .
  - Work with regulators to promote structured B-R assessment.

SCT/FDA Benefit-Risk Workshop

- Theme: Bridging Qualitative and Quantitative Assessments
- When: Dec 5-6 2013
- Where: The Universities at Shady Grove Conference Center, Rockville, MD
- Planned sessions: regulatory background and industry response, benefit-risk evaluation methods/graphics, endpoint selection and weighting, 4 case studies, post-approval benefit-risk evaluation, future direction.
- Registration at http://meeting.sctweb.org/qspi.
- Please join us.

Summary

- Questions are now focusing on “Do the risks outweigh the benefits”.
- The first step in benefit-risk assessment is to agree on the relevant data elements for each specific case.
- It is difficult to settle on a single set of weights/factor. A common approach is to show how the overall conclusion depends on the choice of weights/factor.
- Eventually, a judgment needs to be made at the societal and/or at each individual level.
- We should try out some intuitive quantitative benefit/risk approaches for collective experience.