

Benefit:Risk

Scott Evans

Harvard University

Graybill Conference

June 11, 2008

Outline

- Challenges and Issues
- Methods
 - Combined Marginal Analyses
 - Benefit-Risk Ratio
 - NNTB and NNTR
 - Within-Patient Analyses
 - Q-TWiST
 - Composite Variables
 - Patient Measures
 - Global Benefit:Risk Score
 - Benefit-Less-Risk-Measure
 - Benefit-Risk Index
 - Nonparametric Approach
 - Bivariate Approach
 - Other Ideas
- Reporting Benefit:Risk

Benefit:Risk

- Fundamental concept in clinical trials
 - Weighed by regulators in product approval decisions
 - Evaluated by sponsors to aid in development decisions
 - Assessed by data monitoring committees during interim analyses to make recommendations

Benefit:Risk

- Need more effective and transparent methods
- Are we collecting the right data in the right way?
- Do we need to revise our traditional approaches to designing, monitoring, analyzing and reporting clinical trials for better assessment?

Benefit:Risk

- September, 2006
 - Institutes of Medicine recommended that FDA develop and continually improve a *systematic* approach to benefit:risk
- December, 2006
 - European Committee for Proprietary Medicinal Products (CPMP) called for improved methodology leading to a more *systematic* approach to benefit:risk analysis

Benefit:Risk

- Limited published guidance
 - EMEA CHMP Working Group
 - Reflection paper on benefit-risk assessment methods in the context of the evaluation of marketing authorization applications of medicinal products for human use
 - Benefit:Risk Action Team (BRAT) of PhRMA

BRAT - PhRMA

- **Goal**
 - Develop a structured, transparent benefit:risk framework and integrate it into the regulatory review process
 - Quantitative and qualitative elements
 - Pilot studies being planned

BRAT - PhRMA

- **Working Groups**
 - Benefit Definition
 - Risk Definition
 - Framework
 - Quantitative Elements
 - Stakeholder Engagement
 - Communications

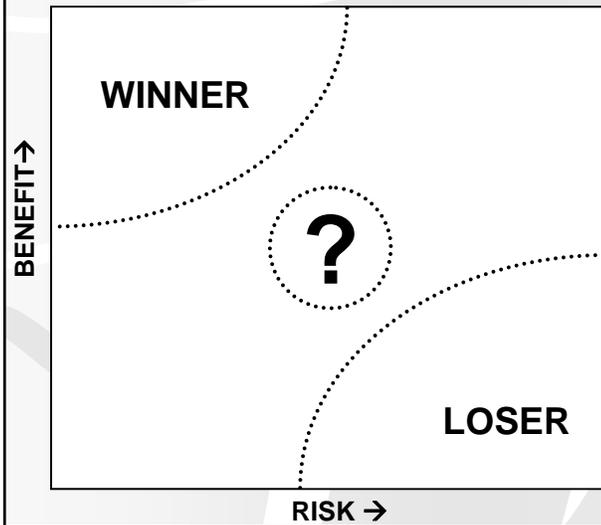
BRAT - PhRMA

- Separate B:R balance by each indication
- Consider all data
- Incorporate uncertainty
- Evaluate temporal effects
- Transparency via systematic, consistent, reproducible model

Challenges

- Definition of “benefit”?
 - Primary endpoint?
 - Composite?
- Definition of “risk” (or harms)?
 - Data reduction
 - Multivariate problem reduced to limited dimensions
 - Labs (chemistries and hematology's), AEs, signs/symptoms, QOL
- Development of a metric to characterize both benefits and risks so that comparisons can be made
- Subjective weighting
 - Of various safety evaluations
 - Of benefit vs. risk

Benefit: Risk



? May be acceptable with a very serious disease with no known cure

? May not be acceptable with a disease that is not life-threatening and other effective and safe treatment options are available

Benefit:Risk Interaction

- Some patients benefit
- Some patients experience toxicity
- If benefit is occurring in patients without toxicity, then we need to think about identifying subgroups (predictive markers)
- If benefit is occurring in patients with toxicity, then we need to think about identifying subgroups and the weighting of benefits and risks

Challenges

- Clarity of objective
 - Benefit:risk of the population vs. the trial cohort vs. an individual patient
- Generalizing benefit:risk assessment in clinical trials to a population
 - Trials may not be representative of real patient usage

Challenges

- Assessment of causality
- Evolving options of alternative therapies
 - Absolute benefit:risk vs. considerations of alternatives
- Incorporation of non-trial based evidence
 - Epidemiological studies
 - Testimonial data (FDA Advisory Committee experience)

Challenges

- Disease-specific context
 - Toleration of toxicities for some indications but not others
- Patient-specific context
 - E.g., benefits and risks are different for old vs. young
- Population-specific context
 - Multinational trials with differing cultures, ethics, and availability of medicines

Challenges

- Need to continually evaluate benefit:risk as the profile changes
 - E.g., Data Monitoring Committee discussion regarding evaluation of antimicrobials as resistance evolves

MRSA (The “Superbug”)

- Methicillin-resistant Staphylococcus aureus
 - Bacterial infection that has become resistant to antibiotics (such as penicillin, amoxicillin, methicillin)
 - Lives on the skin
 - Community acquired
 - Spread through contact

CNN – October 17, 2007

Experts: Drug-resistant staph deaths may surpass AIDS toll - CNN.com - Microsoft Internet Explorer

Address: <http://www.cnn.com/2007/HEALTH/conditions/10/16/mrsa.cdc.ap/index.html>

Experts: Drug-resistant staph deaths may surpass AIDS toll

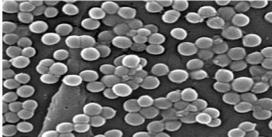
STORY HIGHLIGHTS

- CDC: More than 90,000 get potentially deadly “superbug” infections annually
- The incidence rate was about 32 invasive infections per 100,000 people
- Prevention methods include curbing the overuse of antibiotics

Next Article in Health >

READ VIDEO

CHICAGO, Illinois (AP) — More than 90,000 Americans get potentially deadly infections each year from a drug-resistant staph “superbug,” the government reported Tuesday in its first overall estimate of invasive disease caused by the germ.



Methicillin-resistant Staphylococcus aureus can be carried by healthy people, living on the skin or in their noses.

Deaths tied to these infections may exceed those caused by AIDS, said one public health expert commenting on the new study. The report shows just how far one form of the staph germ has spread beyond its traditional hospital setting.

The overall incidence rate was about 32 invasive infections per 100,000 people. That’s an “astounding” figure, said an editorial in Wednesday’s Journal of the American Medical Association, which published the study.

Most drug-resistant staph cases are mild skin infections. But this study focused on invasive infections — those that enter the bloodstream or destroy flesh and can turn deadly.

Researchers found that only about one-quarter involved hospitalized patients. However, more than half were in the health care system — people who had recently had surgery or were on kidney dialysis, for example. Open wounds and exposure to medical equipment are major ways the bug spreads.

Don't Miss In recent years, the resistant germ has become more common in

SCORE 0 LIVES 2

ALL ABOUT

Most Popular

STORIES

Most Viewed	Most Emailed	Top Searches
1 J.K. Rowling: Dumbdore is gay		
2 Georgia declares drought emergency		
3 Police: Thai boys identify suspect		

start Bentley Scott Evans... Microsoft Word Evers_Bentl... Experts: Dru... 3:16 PM

Benefits and Risks Tradeoff

- FDA Advisory Committee (CDRH)
 - “Substantial Equivalence”
 - Distinct from traditional concept of noninferiority
 - Simultaneous evaluation of benefit and risk
 - NeuroStar™ for the treatment of Major Depressive Disorder
 - Repetitive transcranial magnetic stimulation (rTMS)
 - Vs. Electroconvulsive Therapy (ECT)
 - Causes a seizure via electronic stimulation

Substantial Equivalence

- A 510(k) requires demonstration of substantial equivalence to another legally U.S. marketed device. Substantial equivalence means that the new device is at least as safe and effective as the predicate.
- A device is substantially equivalent if, in comparison to a predicate it:
 - has the same intended use as the predicate; **and**
 - has the same technological characteristics as the predicate;
 - or**
 - has the same intended use as the predicate; **and**
 - has different technological characteristics and the information submitted to FDA;
 - does not raise new questions of safety and effectiveness; **and**
 - demonstrates that the device is at least as safe and effective as the legally marketed device.
- A claim of substantial equivalence does not mean the new and predicate devices must be identical. Substantial equivalence is established with respect to intended use, design, energy used or delivered, materials, chemical composition, manufacturing process, performance, safety, effectiveness, labeling, biocompatibility, standards, and other characteristics, as applicable.

Substantial Equivalence

- From my packet sent by the FDA:
 - “the experimental device does not need to be as effective as the predicate device, if the clinical data demonstrated that any reduction in effectiveness was off-set by an improvement in patient safety/risk”
- What metric/weights?

Temporal Nature of Benefits and Risks

- Temporal difference between the time to development of benefits vs. risks (and the relationship to the duration of the studies conducted to detect them)
 - Risks often present later than benefits
 - E.g., Cox-2 inhibitors
 - Long-term health implications are often unknown
 - Duration of exposure may be important
 - Absence of observed risk during observational period does not necessarily imply the absence of risk
 - Need appropriate follow-up time

Censoring

- Important safety data is often censored by the end of the study or by treatment discontinuation
- Important efficacy data is often censored by SAEs
- Safety data often gathered after approval, but not always the case for efficacy data

Asymmetry of Attention to Benefits and Risks

- Imbalance of attention paid to benefits and risks
 - Later phase studies are designed/powered for efficacy outcomes but not to identify safety effects (particularly when incidence is low)
 - Problematic when events are serious
 - Statisticians spend a lot of time developing methods for efficacy analyses, and less time on safety evaluation

Asymmetry of Attention to Benefits and Risks

- Passive collection of some safety data
 - Report only if abnormal
 - Cannot distinguish between normal (indicating no safety issue) vs. missing data

Adjudication

- Safety endpoints often not adjudicated
 - Causality and attribution is often unknown
 - False positive/negative rates unknown
- When adjudication does occur, then only specific events are examined
 - May miss events that an investigator did not judge to be relevant

ITT Principle

- Important with respect to safety data
- Need to follow patients that withdraw from treatment
 - Otherwise adverse outcomes may not be obtained as they can occur after withdrawal from exposure
 - Potential for informative censoring if subjects that withdraw are dissimilar to subjects that complete the study

Importance of Follow-up

- Clinical trials have a scheduled treatment period T_S and a limited window of time afterward T_W
- Important choices
 - Length of T_W
 - Duration of follow-up for patients that discontinue treatment

Importance of Follow-up

- Carefully consider the length of T_W in design to evaluate safety
- Count events in interval of $T_S + T_W$ to avoid informative censoring and avoid bias
 - More diligent follow-up needed on patients that discontinue the treatment early
 - Else can bias Kaplan-Meier estimates of cumulative incidence
- Lagakos, *NEJM*, 2006.

Expanding the Label

- Benefit:Risk profile varies by:
 - Product usage
 - Type of exposure
 - Background risk of target population
 - Prevalence of resistance
 - Therapeutic alternatives
- Need more post-marketing evaluation with a labeling of unknowns at the time of approval

Proposed Methods

Within Trial vs. Synthesized Assessment

- Distinction between benefit:risk evaluation within a particular clinical trial vs. meta-analyses and epidemiological methods that may combine results from several studies

Combined Marginal Analyses

- Current Practice
 - Separate and marginal analyses of efficacy and safety
 - Then informal (usually non-quantitative) aggregation of the two marginal analyses are conducted as an assessment of benefit:risk
- Meta-analyses of several trials
 - E.g., Integrated Summary of Efficacy and Safety (ISES)

Benefit:Risk Ratio

- Population-level measure
 - E.g., for every death, how many lives were saved?
- Benefits and risks are each treated as binary
 - For risk, it is usually a significant clinical outcome
- Compute the ratio (incidence of benefit over incidence of risk) and associated precision
- Then determine a threshold at which benefits and risks are neutralized

Benefit:Risk Ratio

- Advantages
 - Easy to understand and communicate
- Disadvantages
 - Does not account for the relative timing of these events
 - Does not account for the censoring by competing events
 - Can be challenging to identify a threshold at which benefits and risks are neutralized

Number Needed to Treat (NNT)

- Expected number of subjects that have to be treated with a new therapy to experience one additional occurrence of an event
 - NNTB: NNT for benefit
 - NNTR: NNT for risk (or harm)
- Primarily used for binary measures of benefit and risk
- Useful for individual clinical trials

NNTB and NNTR

- Let π_A and π_B ($> \pi_A$) be the incidence rates of a beneficial outcome in treatment groups A (control) and B (experimental) respectively, then

$$\text{NNTB} = 1/(\pi_B - \pi_A)$$

- E.g., if $\pi_B - \pi_A = 0.5$ then $\text{NNTB}=2$
 - Implying that on average if 4 patients are treated (2 on each arm), then we expect one more beneficial outcome on treatment B than treatment A

NNTB and NNTR

- NNTB decreases as the difference in incidence rates increases
- Similarly calculate NNTR
 - Note treatment B may have more risk for side effects

Ratio based on NNT

- $R = \text{NNTB}/\text{NNTR}$
 - If the events of benefit and harm are of equal importance, then considering benefits and risks:
 - $R=1$, suggests that A and B are similar
 - $R>1$, suggests that B is better than A since B provides benefits at a faster rate than providing harms (relative to A)
 - $R<1$, suggests that B is worse than A since B provides benefits at a slower rate than providing harms (relative to A)

NNT

- Advantage
 - Practical and simple interpretation
- Disadvantage
 - Results are applicable only in settings similar to the conditions in the clinical trial
 - Issues with statistical inference

NNT and the Ratio based on NNT

- Statistical inference based on NNT or R is challenging
 - Hutton, *JRSS*, 2000.
 - Lesaffre et.al., *CCT*, 1999.
 - Walter, *SIM*, 2001.
 - Walter & Irwig, *SIM*, 2001.
 - Grieve, *PS*, 2003.
 - Thabane, *Biostatistics*, 2003.
 - Duncan & Olkin, *SIM*, 2005.
 - Alemayehu et.al., *JBS*, 2006.

NNT Extensions

- Confidence intervals for NNT
 - Lower bound can be negative when incidents rates are not dissimilar between groups
 - Altman, *BMJ*, 1998.
- Correction for NNT for length of follow-up
 - Observed NNT is multiplied by the ratio of the average duration of FU to the duration of interest
 - Assumes treatment effects and event rates are constant over time
 - Often unrealistic assumption
 - Sackett, et.al., 1991.

NNT Extensions

- NNT for time-to-event data
 - Altman et.al., *BMJ*, 1999.
- “Unqualified success” and “unmitigated failure”
 - Mancini and Schulzer, *Circulation*, 1999.

Within-Patient Analyses

- Less frequently conducted
- Order of operations are important
- Clinicians treat patients based on considerations of combined benefits and risks
- Does within patient analyses (combining benefits and risks) make sense before aggregate analyses over all patients?
 - Perhaps at least as a supporting and informative analyses?

Q-TWiST

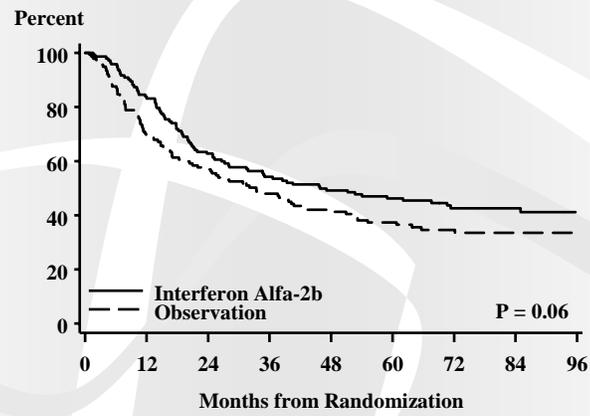
- **Quality-adjusted Time Without Symptoms or Toxicity** (Gelber, et.al., *Biometrics*, 1989)
 - Proposed to evaluate adjuvant therapies for breast cancer
 - Has been used in other disease settings
 - HIV, melanoma, rectal cancer, lymphoma, prostate cancer, childhood AML
 - Evaluates benefit:risk “trade-offs” in clinical trials

Example: ECOG 1684 Interferon for Melanoma

- Population
 - High risk resected melanoma
- Randomized therapy
 - Observation (n=137) vs. high-dose interferon (n=143)
- Median follow-up = 7 years

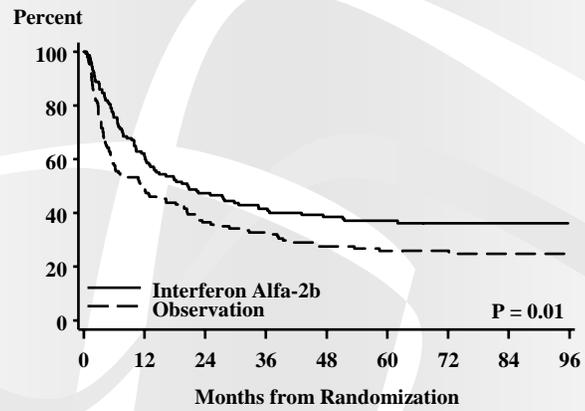
ECOG 1684: Q-TWiST

Overall Survival



ECOG 1684: Q-TWiST

Relapse-Free Survival



ECOG 1684 Q-TWiST

Toxicity summary

- Grade 3 or worse toxicity occurred in 78% of patients
 - Dosing delays or reductions occurred in 62% of patients during the IV phase and for 51% during the SQ phase
 - Included constitutional, myelosuppression, hepatotoxicity and neurological

ECOG 1684: Q-TWiST

Clinical Health States

1. Grade 3 or worse toxicity (Tox)
2. Disease relapse (Rel)
3. Survival (Death)



ECOG 1684: Q-TWiST

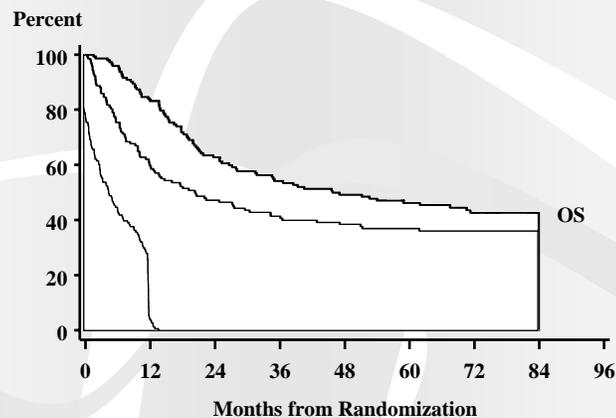
Step 1: Define clinical health states

- Toxicity
 - Time with grade 3 or worse treatment-related toxicity
 - The entire treatment duration was assigned to Tox if any grade 3 or worse toxicity was experienced
- Relapse
 - All time following disease relapse
- TWiST
 - All remaining survival time

ECOG 1684: Q-TWiST

Step 2: Partitioning overall survival

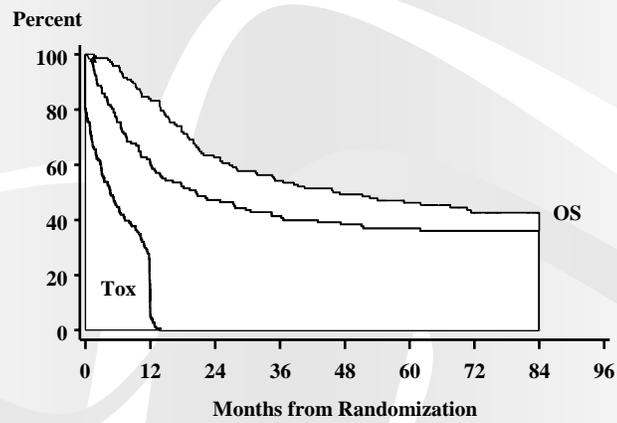
Interferon Alfa-2b



ECOG 1684: Q-TWiST

Step 2: Partitioning overall survival

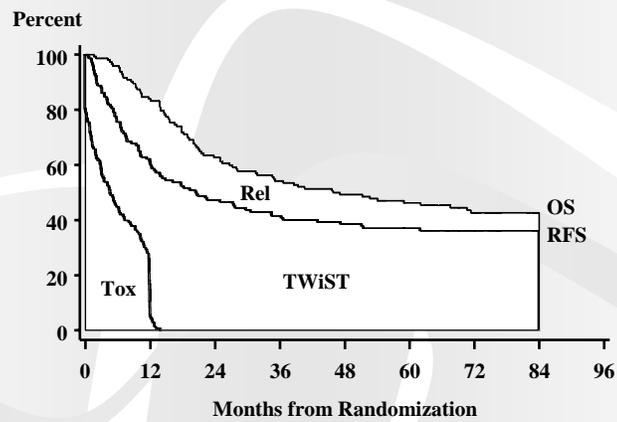
Interferon Alfa-2b



ECOG 1684: Q-TWiST

Step 2: Partitioning overall survival

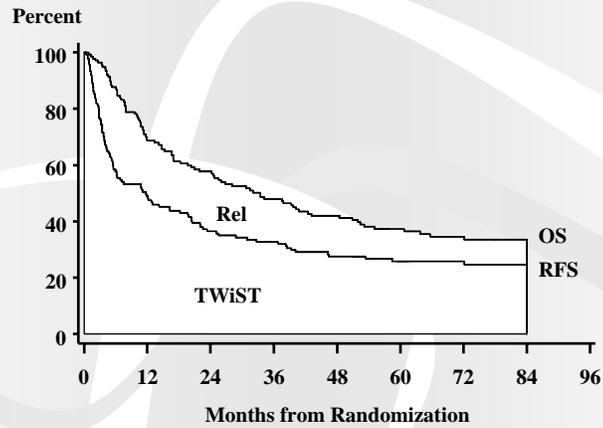
Interferon Alfa-2b



ECOG 1684: Q-TWiST

Step 2: Partitioning overall survival

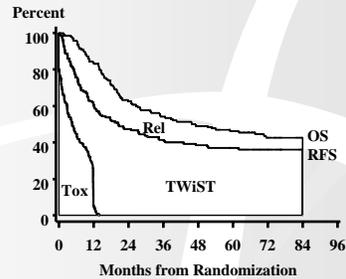
Observation



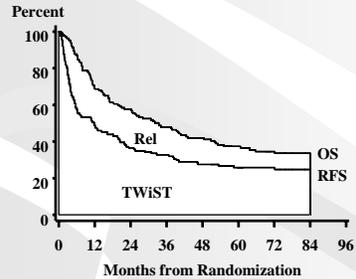
ECOG 1684: Q-TWiST

Step 2: Partitioning overall survival

Interferon Alfa-2b



Observation



ECOG 1684: Q-TWiST

Step 3: Compare treatments using weighted components

Outcome	Treatment Group		Difference	(95% C.I.)
	Obs	IFN		
Tox	0.0	5.8	5.8	(5.0 to 6.7)
TWiST	30.0	33.1	3.1	(-4.8 to 11.0)
Rel	12.4	10.4	-2.0	(-6.2 to 2.3)
Q-TWiST*	34.9	42.5	7.6	(0.0 to 15.1)
$u_{Tox}=.9, u_{Rel}=.4$				
OS	42.4	49.3	7.0	(-0.6 to 14.5)
RFS	30.0	38.9	8.9	(0.8 to 17.0)
Patients	137	143		

$$*Q-TWiST = u_{Tox} \times Tox + TWiST + u_{Rel} \times Rel$$

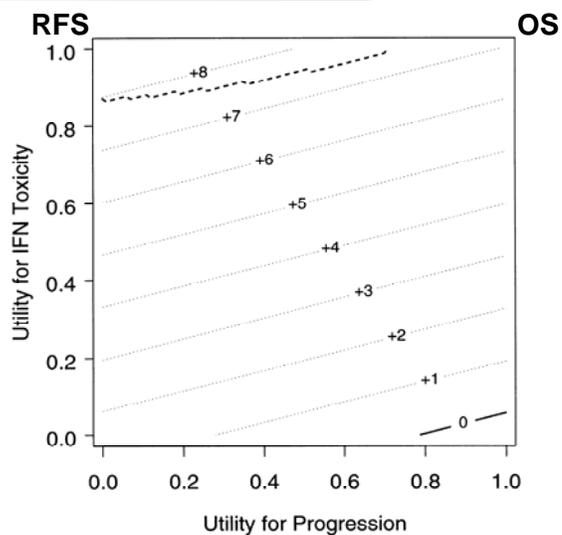
Sensitivity Analyses

- Weights for clinical health states are subjective
 - Perhaps patient specific
 - E.g., different for older vs. younger people
- Sensitivity analyses to varying weights are necessary
 - Allowing for “personalized medicine”

Threshold Utility Plot

- Fundamental outcome of a Q-TWiST analysis
 - Compares treatments across combinations of weights
- Vertical axis
 - Weight for time with toxic effects
- Horizontal axis
 - Weight for time after relapse
- Plots contours of similar treatment effects

Threshold Utility Plot



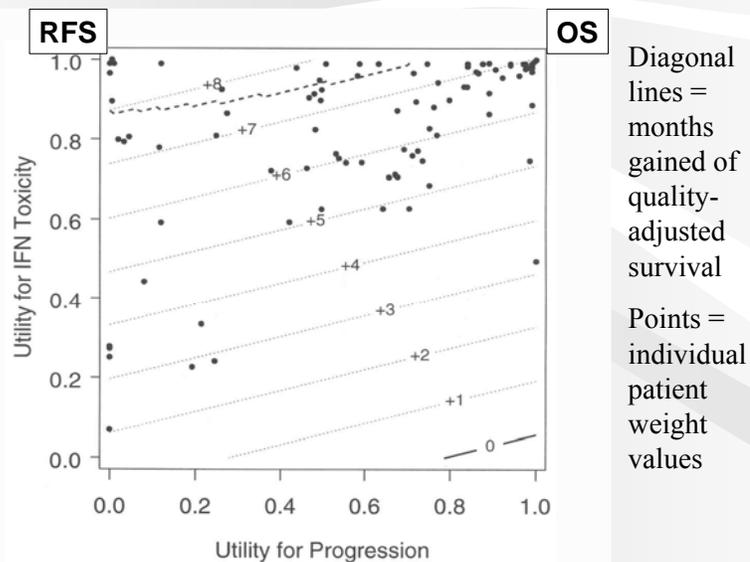
Diagonal lines = months gained of quality-adjusted survival (Q-TWiST) for interferon compared with no interferon

Weight Coefficients

- 95 patients with melanoma provided their patient-specific weights for the clinical health states
- Applied to threshold analysis
- 16% had preferences in the region of significant Q-TWiST gain with interferon
- 84% had preferences in the region of non-significant Q-TWiST gain

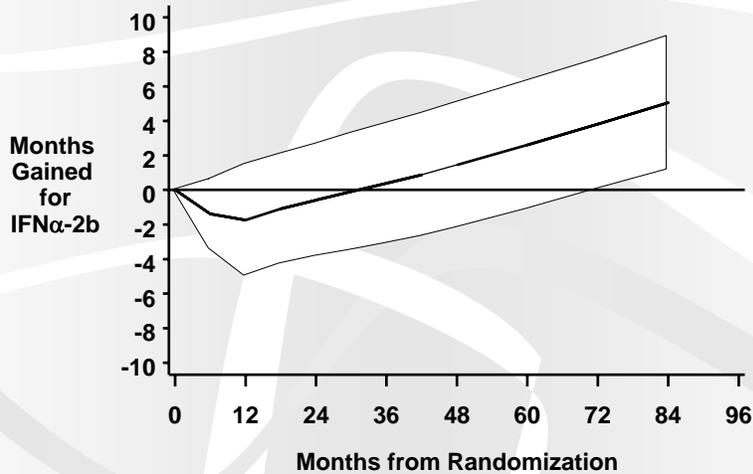
Kilbridge KL, et al. J Clin Oncol 2001;19:812-23.
Kilbridge KL, et al. J Clin Oncol 2002;20:1311-18.

Threshold Utility Plot



ECOG 1684: Q-TWiST

Q-TWiST Gain Analysis for fixed weights



QALY's

- Quality-adjusted life years
- Combine life duration and health status
- Uses weights to reflect importance of health states (death=0, best health=1)

Composite Endpoints

- Desirable Endpoint Characteristics
 - Clinically relevant
 - Addresses the scientific question
 - Easy to interpret
 - Easy to obtain
 - Quantified/qualified in an unbiased manner
 - Sensitive to changes induced by treatment
 - Affordably obtained
 - Results in a reasonable sample size
 - Note: continuous responses usually require smaller sample sizes than binary or time-to-event

Composite Endpoints

- Common
 - CV trials
 - E.g., CV death + MI + revascularization procedure
 - HIV
 - E.g., A5175: Death + VL Failure + AIDS-defining event
 - E.g., virologic failure + treatment discontinuation
 - Studies of co-infections
 - Combining endpoints from more than one diseases

Composite Endpoints: Advantages

- More complete characterization of treatment effect
 - May be interested in a variety of outcomes
 - E.g., Perspective of treating more than a single disease
- Reduces bias due to competing risks and informative censoring
- More events could imply more power
 - If expected effect size is unchanged
- Avoids multiplicity issue and avoids difficult selection between endpoints (ICH E-9)

Composite Endpoints: Disadvantages

- Difficult interpretation if:
 - Relative importance of components differs and the treatment effects on components differ
 - E.g., how do we interpret a study in which one arm has lower event rate, but the events are more serious?
 - Treatment effects for different components go in different directions

Composite Endpoints: Disadvantages

- Competing risk issue
 - If censored patient has a different risk for event than a non-censored patient, then “informative censoring”
 - Occurs when an event (e.g., disease progression) is analyzed while ignoring circumstances (e.g., death) that preclude the occurrence of the event

Composite Endpoints: Disadvantages

- Lose power if there's no effect on some components
 - Dilution of treatment effect
- Combining components measured on separate scales can be challenging

Composite Endpoints: Considerations

- Are the components of similar importance?
- Do the components occur with similar frequency?
- Is the treatment effect similar across components?

Composite Endpoint: Recommendations

- Include more “severe” events (e.g., death) in the composite to avoid the bias induced by competing risks
- Collect/report data on all components
 - Continue to follow patients after experiencing an event
- Consider ordering or weighting the components
- More easy to justify use and interpret results when the components have similar importance

Example: Composite Endpoint in HIV Trial

Time to “treatment failure”

- Time from randomization to:
 1. death due to any cause, or
 2. disease progression - defined as a new or recurrent AIDS-defining opportunistic infection or malignancy occurring after completion of 12 weeks of ARV treatment, or
 3. virologic failure defined as two successive measurements of plasma HIV-1 RNA ≥ 1000 copies/mL

Composite Benefit and Harm Endpoint

- Effectiveness of atypical antipsychotic drugs in patients with Alzheimer’s disease
 - Composite endpoint:
 - min(time to discontinuation of treatment due to lack of efficacy, time to discontinuation of treatment due to toxicity or adverse effects)
 - Schneider et.al., *NEJM*, 2006.

Patient Level Measures

- Patients rate their overall experience with respect to *perceived* benefit and risk
- Possibly useful for therapies to treat symptoms
- Problematic when symptoms do not equate with risk
 - i.e., silent risk of abnormal labs (e.g., LFTs, bilirubin)

Global Benefit:Risk (GBR) Score

- Pre-specify a multinomial outcome based on efficacy and safety data
- E.g., 5 ordered categories (descending desirability)
 - Efficacy with no serious side effects
 - Efficacy with serious side effects
 - Non-efficacy with no serious side effects
 - Non-efficacy and serious side effects
 - Side effects leading to drop-out

Global Benefit:Risk (GBR) Score

- Assign non-negative weights
 - Chosen to reflect the relative importance of each category when evaluating treatment
- Then compute summary measures and compare the distributions between treatment groups

Global Benefit:Risk (GBR) Score

Linear score, $m = w_1\pi_1 + w_2\pi_2 - w_3\pi_3 - w_4\pi_4 - w_5\pi_5$

Ratio score, $r_1 = \frac{(w_1\pi_1 + w_2\pi_2)^e}{w_3\pi_3 + w_4\pi_4 + w_5\pi_5}$, $e \geq 0$

Composed ratio score, $r_2 = \frac{w_1\pi_1}{w_5\pi_5} \left(\frac{w_2\pi_2}{w_3\pi_3 + w_4\pi_4} \right)^f$, $f \geq 0$

- Asymptotic distributions derived
 - Chuang-Stein, *SIM*, 1991.
- Relative measures can be constructed
 - $R_{AB} = r_A/r_B$

Benefit-Less-Risk Measure

- Benefit is discounted for the presence of untoward safety events according to a prespecified algorithm
 - Similar discounting approach to Q-TWiST
- Chuang-Stein, *CCT*, 1994.

Benefit-Less-Risk Measure

- Step 1: consolidate safety data (e.g., Chuang-Stein et.al., *SIM*, 1992)
 - Select a set of classes representing body functions to cover all areas of safety concern
 - Summarize the safety experience for each class with an intensity grade (higher grades imply more adverse experience)
 - Use pre-defined rules
 - Compute a summary score R_j that reflects the overall safety experience for patient j by combining the intensity grades for each class using weights based on relative importance

Benefit-Less-Risk (BLR) Measure

- Step 2: Suppose benefit is measured on a single primary endpoint
 - Denote the benefit for subject j as S_j
 - Discount S_j by a multiple of R_j using a “conversion factor”, f

$$\text{BLR} = S_j - f R_j$$

- Use BLR to compare treatments

Benefit-Less-Risk Measure

- Advantages
 - Intuitive
 - Patient specific interpretation
- Disadvantages
 - Reducing safety experience into a single score requires much deliberation
 - Choice of a conversion factor is challenging

Benefit:Risk Index

- **Women's Health Initiative**
 - 15 year project sponsored by NHLBI
 - >161,000 postmenopausal women
 - 3 components
 - Randomized clinical trials
 - Enrolled > 68,000 women
 - Designed to evaluate effects of hormone replacement therapy (HRT), dietary modification, calcium/vitamin D on clinical endpoints
 - Observational study
 - Community prevention trial

Benefit:Risk Index

- **HRT Trial**
 - Objective
 - To evaluate the effects of HRT on the prevention of heart disease, osteoporosis, breast cancer, endometrial cancer, and mortality (due to other reasons)
 - Primary endpoint
 - Death from a coronary cause or a nonfatal MI

Benefit:Risk Index

- HRT could have preventative effects and possible adverse effects (noted by DMC)
 - Freedman et.al. (*CCT*, 1996) proposed indices to combine the estimated treatment effects on the five endpoints
 - Estimate the effect of HRT on each endpoint
 - Combine the effects to form a composite HRT effect

Benefit:Risk Index

- Unweighted index
 - $U = d_1 + d_2 + \dots + d_5$
- Weighted index
 - $W = w_1d_1 + w_2d_2 + \dots + w_5d_5$
- d_i is the observed difference in proportions for outcome i
- The HRT trial used weights of 0.50, 0.35, 0.15, 0.18, and 1.0 for heart disease, breast cancer, endometrial cancer, hip fracture, and mortality

WHI: PREMPRO

- Drug to help women manage symptoms of menopause
- Product label includes summary of benefits and risks
- Allows individual patients and their physicians to make patient-specific decisions based on benefit:risk
 - Patient-specific weighting and value judgment based on outcomes

Absolute Excess Benefits/Risks per 10,000 Woman-Years

TABLE 8. RELATIVE AND ABSOLUTE RISK SEEN IN THE ESTROGEN-PLUS-PROGESTIN SUBSTUDY OF WHI AT AN AVERAGE OF 5.6 YEARS*

Event	Relative Risk CE/MPA vs. Placebo (95% nCI ^b)	Absolute Risk per 10,000 Women-years	
		CE/MPA n = 8,506	Placebo n = 8,102
CHD events	1.24 (1.00–1.54)	39	33
<i>Non-fatal MI</i>	<i>1.28 (1.00–1.63)</i>	<i>31</i>	<i>25</i>
<i>CHD death</i>	<i>1.10 (0.70–1.75)</i>	<i>8</i>	<i>8</i>
All Strokes	1.31 (1.02–1.68)	31	24
<i>Ischemic stroke</i>	<i>1.44 (1.09–1.90)</i>	<i>26</i>	<i>18</i>
Deep vein thrombosis	1.95 (1.43–2.67)	26	13
Pulmonary embolism	2.13 (1.45–3.11)	18	8
Invasive breast cancer ^c	1.24 (1.01–1.54)	41	33
Invasive colorectal cancer	0.56 (0.38–0.81)	9	16
Endometrial cancer	0.81 (0.48–1.36)	6	7
Cervical cancer	1.44 (0.47–4.42)	2	1
Hip fracture	0.67 (0.47–0.96)	11	16
Vertebral fractures	0.65 (0.46–0.92)	11	17
Lower arm/wrist fractures	0.71 (0.59–0.85)	44	62
Total fractures	0.76 (0.69–0.83)	152	199

Nonparametric Approach

- Useful when patients can be ranked with respect to each of benefit and risk (i.e., continuous or ordinal measures)
- Rank patients with respect to combined benefit and risk based on their benefit-specific and risk-specific ranks
- Compare the combined ranks between groups
- Issues
 - Ranking two patients, each higher on one scale (benefit or risk)
 - Subjective metric or weighting
 - Handling ties
 - Clinical relevance of effects?
- Wittkowski, *CSS*, 2004.

Within Patient (Personalized Medicine)

- Assume benefit and risk can be measured on a 0-10 scale (or can be transformed as such)
- Create a scatterplot (benefit vs. risk) for each patient

Within Patient (Personalized Medicine)

- Define
 - Minimum tolerable benefit when risk = 0 (b_1)
 - Maximum tolerable risk when benefit = 10 (r_1)
 - Minimum tolerable benefit when risk = $r_1/2$ (b_2)
- Fit a smooth “tolerability curve” through these 3 points

Within Patient (Personalized Medicine)

- For each patient i , calculate the B:R Score as the ratio:
 - $D((0,10), (b_{i_t}, r_{i_t}))$ over $D((0,10), (b_i, r_i))$ where:
 - D is the Euclidian distance between two points
 - (b_{i_t}, r_{i_t}) is the point of intersection between the tolerability curve and a straight line from $(0,10)$ and (b_i, r_i)
- Compare B:R Scores between treatments
 - Note that larger B:R Scores are better

Within Patient (Personalized Medicine)

- Analyses combines benefit and risk data within patient
- Allows for flexibility in analyses for varying trade-offs
 - By selecting different points that will define the tolerability curve
 - Assessment of treatment differences can be made on a personalized level

Other Methods

- Decision Analyses
 - MCDA: Mussen et.al., *PEDS*, 2007.
 - Hughes et.al., *Nature*, 2007.
- Principle of Three
 - Based on concepts of seriousness, duration, and incidence as related to disease indication, disease amelioration by the medicine, and adverse effects ascribed to the medicine
 - Parameters are rated as low, medium, or high
- TURBO
 - Transparent Uniform Risk Benefit Overview
 - Quantitative and graphical

Ideas

- PK and biomarkers for benefit:risk evaluation?
 - If safety is dependent upon drug concentration and when predictive biomarkers are available

Benefit:Risk Reporting

- Need for more interpretable presentation
- Suggestions
 - Avoid using probabilities
 - Visual summaries are helpful
 - Brevity (e.g., limit summary to one page)
 - Summarize absolute risk and absolute risk differences in addition to relative risk
 - Avoid use of relative expressions in isolation
 - State benefits in a way that is comparable to risks
 - E.g., potential lives saved by treatment vs. potential lives lost as a result of adverse reactions

Absolute Risk

- More easily interpreted and may be more informative than relative risk
- More effectively summarizes medical impact
- Conveys absolute chance of event
- Describe the expected numbers of events (both beneficial and adverse) in a population of patients of a given size
 - For both the experimental treatment and the control
 - A comparison can be made by defining weights for the events

Example Data Presentation 40 Year-old white women with UTERI. 5-Year risk of invasive breast cancer with and without Tamoxifen.

Event Severity	Event Type	Expected cases in 10000 untreated women	Cases prevented (+) or caused (-) in 10000 treated women
Life-threatening			
	Invasive Breast Cancer	200	+97
	Hip Fracture	2	+1
	Endometrial Cancer	10	-16
	Stroke	22	-13
	Pulmonary embolism	7	-15
		NET	+54 (prevented)
Severe			
	In situ breast cancer	106	+53
	Deep Vein Thrombosis	24	-15
		NET	+38 (prevented)

Explaining risks: turning numerical data into meaningful pictures

Adrian Edwards, Glyn Elwyn, Al Mulley

The way in which information is presented affects both how health professionals introduce it and how patients use it

BMJ 2002;324:827-30

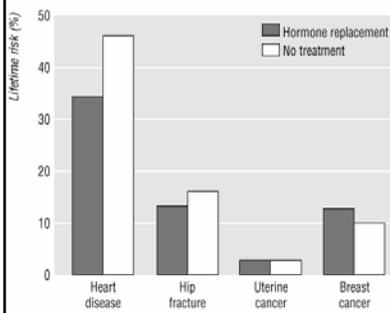


Fig 3 Portrayal of the risks and benefits of hormone replacement taken for five years¹⁵

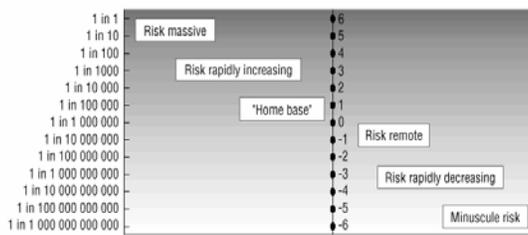


Fig 1 Risk language proposal, derived from Paling¹⁷

Number needed to treat = 20
Control event rate = 14%

Pain on days 2-7 in acute otitis media

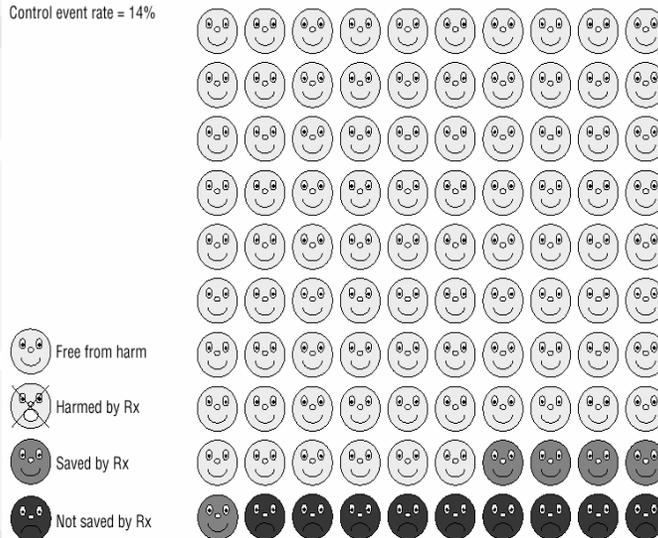


Fig 2 Portrayal of risks and benefits of treatment with antibiotics for otitis media designed with Visual Rx, a program that calculates numbers needed to treat from the pooled results of a meta-analysis and produce a graphical display of the result

Interpretation of Graphic

- 100 patients given treatment for disease
 - 86 patients would have been disease free even if they had not received treatment
 - 9 patients (red faces) are patients that are not cured event with treatment
 - 5 patients (yellow faces) show a benefit
 - They would not have been cured without treatment but are cured when they receive treatment
 - NNTB = 20
 - It is not possible to identify which patients will benefit; thus all 100 patients need to be treated in order for 5 to benefit
 - Can similarly summarize NNTR

Benefit:Risk Reporting

- Provide summaries for each outcome (benefits and risks) affected by therapy
 - Provide detailed information for the most important outcomes
 - Group information by outcome severity
- Provide summaries for the control (or those untreated) for context and comparison
- Consider providing
 - Every day risks (e.g., road crashes) for comparison
 - Life-time risks

Benefit:Risk Reporting

- Remember CONSORT reporting principles
- Describe how benefit:risk varies across subgroups or disease characteristics
 - Often a single aggregate summary description of benefit:risk is not appropriate
 - E.g., by age, race, etc.

Reporting Risks

- Kaplan-Meier curves estimating the cumulative incidence may be appropriate for presenting risk data when risk changes over time

Concluding Remarks

- More formal benefit:risk evaluation is needed to supplement traditional methods of analyses
 - Estimates should include measures of variability
 - Analyses should include sensitivity analyses to subjective weighting and other assumptions
 - Consider analysis by time intervals
- Benefit:Risk is an ongoing process
 - Continuous re-assessment is necessary as information accumulates, therapeutic alternatives changes, medical advances evolve, resistance evolves, etc.

Concluding Remarks

- Analyses of benefit risk requires
 - Detailed discussions between statisticians and clinicians
 - Creative thinking
 - Validation?

Concluding Remarks

- Prospective planning of the collection and analysis of safety data is important
- Enforce the ITT principle
 - Follow all randomized subjects until study completion

Concluding Remarks

- Clear and complete reporting is important
 - Graphical displays
 - Absolute risk
- Flexibility
 - Helpful if the individual patient/clinician can make personalized decisions and apply their own value judgments regarding the balance of benefits and risks
- Within-patient analyses will be informative as it is consistent with the manner in which patients are treated