Benefit:Risk Evaluation

Scott Evans

Harvard University

Midwest Biopharmaceutical Statistics Workshop

May 18, 2009

Special Thanks

- LJ Wei
- Rich Gelber
- Bob O'Neill
- Christy Chuang-Stein

Outline

- Design and Study Conduct
- Challenges and Issues
- Analyses
  - Methods
    - Some old
    - Some new (sneak previews prior to publication)
  - Concepts to Consider
    - Within-Patient Analyses
    - Personalize Medicine and Subgroup Analyses
- Reporting
- Final Thoughts

Benefit:Risk Evaluation

- Fundamental goal in clinical trials
  - Applies to drugs, biologics, and devices
- Assessed by
  - Patients
  - Clinicians
  - Regulators
  - Sponsors
  - IRBs
  - DSMBs or DMCs
  - FDA Advisory Committees

Need for Systematic Assessment

- Need for a more structured and transparent assessment
  - Limited published guidance
    - EMEA CHMP Working Group
    - Benefit:Risk Action Team (BRAT) of PhRMA
- 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

April 2009, *NEJM*

- EMEA leaders called for regulator refinement of methods to assess benefit:risk
  - Implicit → Explicit descriptions of decision criteria and data interpretation with valuations (e.g., weighting factors for treatment outcomes)
  - Qualitative → Quantitative description of net health benefit
  - Development of consensus definition of "tolerable risk"
    - Zero risk is not possible
    - Risk regulation exists for carcinogenic residue in food or of nuclear power
  - Incorporation of patient values/preferences
    - Patients willing to accept some risk for a given benefit
      - E.g., Survey result suggests that MS patients would definitely or probably use a therapy that was more effective than currently available drugs even with a 1 in 1000 chance of a fatal side effect
- Transition
  - Ensuring drug safety → ensuring a positive benefit:risk profile
  - Communication of risk → communication of benefit:risk

Questions

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

Design and Conduct Issues

- Temporal difference between observable benefits and risks (and the duration of the studies to detect them)
  - Risks often present later (e.g., cumulative exposure may be important)
    - Perhaps even after study completion or treatment discontinuation
      - Thus absence of observed safety event does not imply absence of risk
- Censoring and competing risks
  - Safety data can be censored by the end of the study or by early study discontinuation
  - Efficacy data can be censored by SAEs
- Safety data often gathered post-approval (e.g., case reports)
  - Difficultly estimating incidence
    - What is the denominator?
  - Difficulty with interpretation
    - What is the background (reference) rate?
  - Challenge interpreting benefit:risk
    - Data on efficacy?

Design and Conduct Issues

- Asymmetry of attention to benefits vs. risks
  - Phase III designed for efficacy endpoints
    - Particularly problematic when events are rare but serious
  - Most statistical attention is toward efficacy analyses
  - Passive collection of some safety data
    - Report only if abnormal
    - Cannot distinguish between normal (indicating no safety issue) vs. missing data
  - Efficacy outcomes are often blinded and adjudicated, but safety endpoints are often:
    - Investigator reported
      - False positive/negative rates?
        - » Miss events that an investigator did identify
    - Not adjudicated
    - Limited to specific events
    - Causality/attribution can be unclear

CONSORT: Better Reporting of Harms (*AIM*, 2004)

- Common poor reporting practices (from 11)
  - Reporting AEs with only a certain frequency (e.g., > 3%)
  - Reporting only AEs with a significant p-value for a between-arm comparison
  - Not providing data for all randomized participants
- Recommendations (from 10)
  - Describe the statistical plan for analyzing/presenting harms data
  - Present denominators for analyzing harms
  - Provide balanced discussion of benefits and harms with emphasis on limitations and generalizability

Noninferiority Trials to Rule Out Harms

- Event driven
  - Based on relative risk
- Can we do better?
  - Also need to evaluate absolute risk difference
    - Take advantage of denominator
    - Interpretation of relative risk depends on absolute risk

Group	#Event/#Subj.	Prop.
A	1/500	0.2%
B	3/500	0.6%

Measure	Point est & 95%CI
Risk Diff.(B-A)	0.4% [-0.4%, 1.2%]
Relative Risk (B/A)	3.0 [0.3, 158]

Group	#Event/#Subj.	Prop.
A	1/5,000	0.02%
B	3/5,000	0.06%

Measure	Point est & 95%CI
Risk Diff.(B-A)	0.04% [-0.04%, 0.12%]
Relative Risk (B/A)	3.0 [0.3, 158]

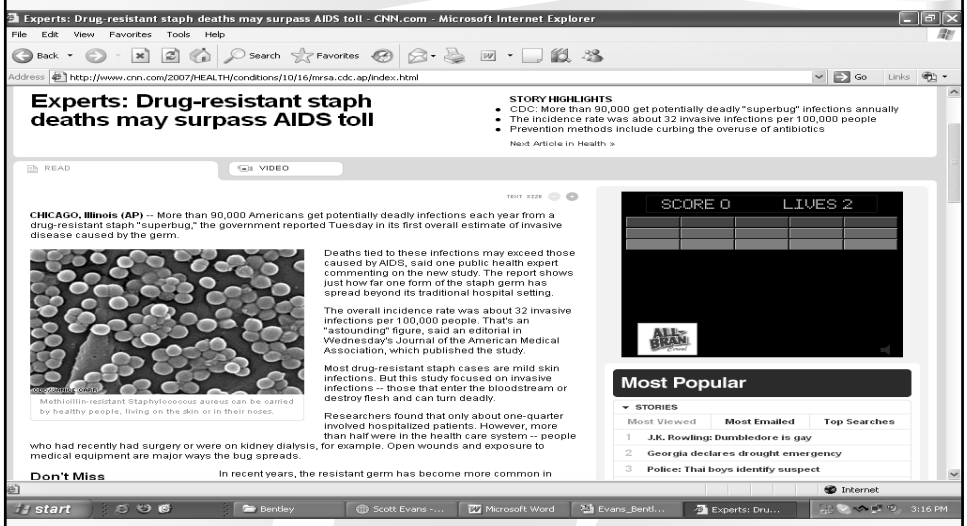
Lessons

- Diligent consideration of the duration of treatment and the follow-up time after treatment discontinuation (Lagakos, *NEJM*)
  - Applies to end of study and premature discontinuation
  - Count events that occur in this window
- ITT principle needs more strict enforcement
  - To observe outcomes after treatment discontinuation
    - Distinction between off-study vs. off-treatment
  - Helps control for informative censoring
- Need more critical thought about the collection and analyses of safety data
- Need for "vision" in during design and conduct
  - What's the question?
  - What does "benefit:risk" mean?
    - Terminology and definitions are too vague
  - How will data be put together to assess benefit:risk?

Challenges

- Definition of "benefit"?
  - Primary endpoint?
  - Composite?
- Definition of "risk" (or harms)?
  - Data reduction
    - Multivariate problem reduced to limited dimensions
    - AEs: Labs (chemistries and hematology's), signs/symptoms, QOL, new diagnoses
- 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
- Dynamic benefit:risk profile
  - E.g., Data Monitoring Committee discussion regarding evaluation of antimicrobials as resistance evolve

MRSA (The “Superbug”)



It’s All Relative

- 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
- Options of alternative therapies
- Benefit:Risk profile varies by product usage
  - Implications for expanding the label

Benefit:Risk Tradeoff: FDA Advisory Committee

- CDRH: NeuroStar™ for the treatment of Major Depressive Disorder
  - Repetitive transcranial magnetic stimulation (rTMS) vs. Electroconvulsive Therapy (ECT)
    - ECT causes a seizure via electronic stimulation
- Substantial Equivalence
  - From the 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?
- Incorporation of non-trial based evidence
  - Testimonial data (FDA Advisory Committee experience)

Benefit:Risk Ratio

- Benefits and risks are each treated as binary
  - Risk: often a significant clinical outcome
- Compute the ratio (incidence of benefit over incidence of risk) and associated precision
  - Population-level measure
    - E.g., for every death, how many lives were saved?
- Then determine a threshold at which benefits and risks are neutralized
- Advantages
  - Easy to understand and communicate
- Limitations
  - 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
  - Benefit:risk is not the same for all patients

Number Needed to Treat (NNT)

- Expected number of subjects that must be treated with a new therapy to experience one additional occurrence of an event versus a control
  - NNTB: NNT for benefit
  - NNTR: NNT for risk (or harm)
- Primarily used for binary measures of benefit and risk
  - Some extensions for other scales
- Advantage: simple interpretation
- Disadvantage: issues with statistical inference

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

$$NNTB = 1/(\pi_B - \pi_A)$$

  - E.g., if $\pi_B - \pi_A = 0.5$ then 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 decrease as the difference in incidence rates increase
- Similarly calculate benefit:risk ratio = NNTB/NNTR
- Issue: when the probability of benefit and risk in the control is 0.5 and 0.5 respectively, then benefit:risk ratio is the same when for the new therapy, the:
  - Probability of benefit is 0.8 and the probability of risk is 0.7, and
  - Probability of benefit is 0.2 and the probability of risk is 0.3

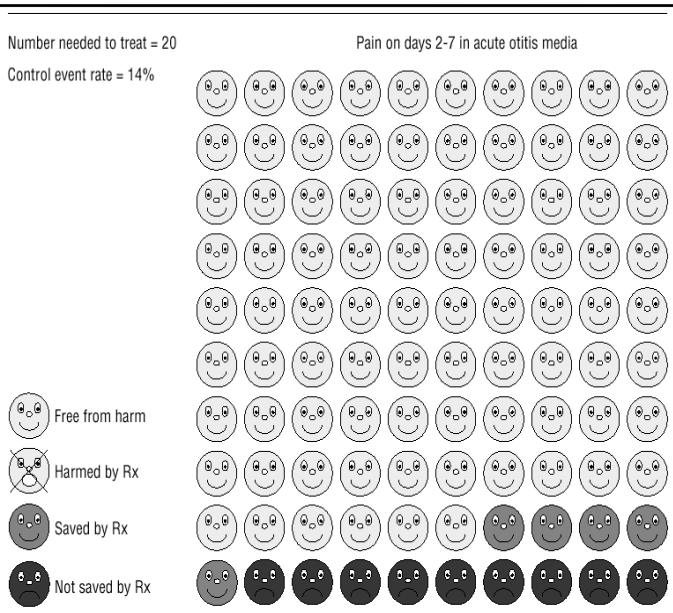


Fig 2 Portrayal of risks and benefits of treatment with antibiotics for otitis media designed with Visual Rx, a program that calculates numbers needed to treat from the pooled results of a meta-analysis and produce a graphical display of the result

Benefit:Risk Index

- Women's Health Initiative (WHI)
  - Randomized trial designed to evaluate effects of hormone replacement therapy (HRT) on heart disease, osteoporosis, breast cancer, endometrial cancer, and mortality (due to other reasons)
- HRT could have preventative effects but possibly increase risk for other events
  - 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

- Weighted index
  - $\mathbf{W = w_1 d_1 + w_2 d_2 + \ldots + w_5 d_5}$
- 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

Benefit:Risk Interaction

- Benefit:Risk is not the same for every person
  - Some patients benefit
  - Some patients experience toxicity
- Consider 2 scenarios:

1. Benefit is occurring in patients with toxicity
   - Need to think about the weighting of benefits and risks and possibly subgroup identification
2. Benefit is occurring in patients without toxicity
   - Need to think about identifying subgroups (predictive markers)

Within-Patient Analyses

- Current Practice
  - Separate and marginal analyses of efficacy and safety
    - Then aggregation of the two marginal analyses are conducted
    - Drug-development programs
      - E.g., Integrated Summary of Efficacy and Safety (ISES)
    - How does the current practice distinguish between the two benefit:risk interaction scenarios?
- Order of operations are important?
  - Clinicians treat individual 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 analyses?

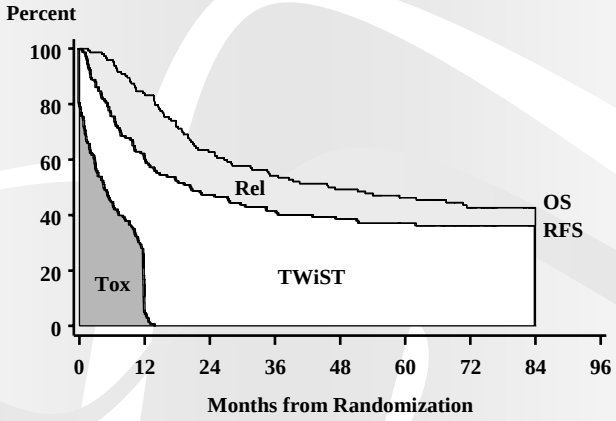
Personalized Medicine and Subgroup Analyses

- Benefits and risks vary person to person
  - Baseline covariates (e.g., demographics, baseline status)
    - GWAS in Clinical Trials
      - To identify gene*treatment interactions
      - Does treatment effect depend on genomic information?
- Subgroup Analyses
  - Subpopulation Treatment Effect Pattern Plot (STEPP)
    - Examines the pattern of treatment differences across a sequence of overlapping patient subpopulations defined by a covariate
  - Multivariable risk prediction
    - Kent & Hayward (*JAMA*, 2007) cite limitations of applying clinical trial results to individual patients and of traditional subgroup analyses
    - Risk-stratified analyses where several risk factors are combined into a single dimension-of-risk score (developed/validated using prior data) along which treatment effects can vary
    - Reduced multiplicity concern
    - Greater power

Q-TWIST

- Endpoint: survival time
  - E.g., evaluation of cancer chemotherapy
- Step 1: Partition overall survival into 3 clinical disease states:
  1. Time with toxicity due to therapy
  2. Time without toxicity or progression
  3. Time after progression

Q-TWiST: Partition Overall Survival



Q-TWiST

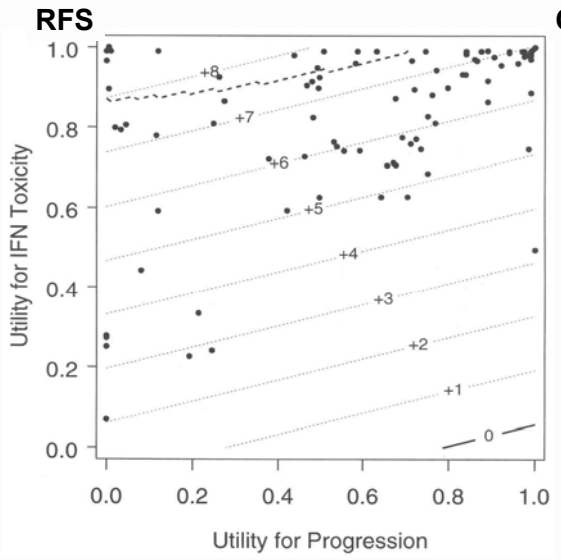
- Step 2
  - Compare treatments with respect to time-to-death appropriately weighting the clinical disease states
    - Time without toxicity or progression gets full weight (full credit)
    - Time after progression gets partial credit (weight<1)
    - Time with toxicity due to treatment gets partial credit (weight<1)

***Q-TWiST = u_{Tox} x Tox + TWiST + u_{Rel} x Rel**

Sensitivity Analyses

- Personalized medicine
  - Weights can vary depending on the patient (e.g., old vs. young)
- Sensitivity analyses to varying weights are necessary
- Threshold Utility Plot
  - Plots contours of similar treatment effects as a function of the weights assigned to the clinical disease states
    - Vertical axis
      - Weight for time with toxicity
    - Horizontal axis
      - Weight for time after relapse

Threshold Utility Plot



Diagonal lines = months gained of quality-adjusted survival

Points = individual patient weight values

Estimating Patient-Specific Benefit:Risk Profiles

Work in progress (Yi Li, Lu Tian, LJ Wei)

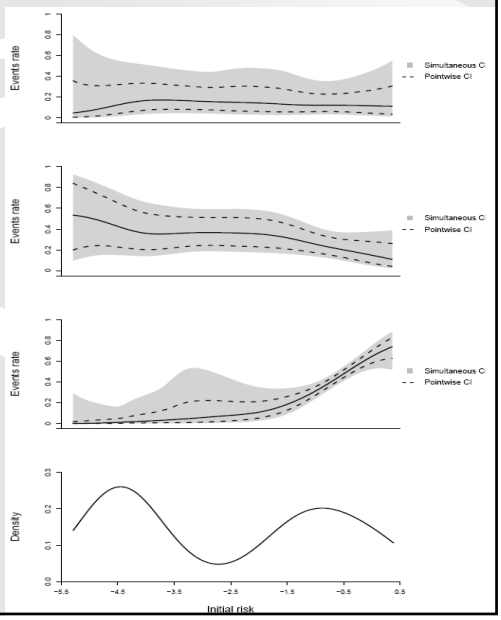
- Step 1
  - Construct a univariate risk index scoring system
    - Parametric model based on baseline covariates
    - Usually based on primary endpoint
- Step 2
  - Estimate consistently average probability of several outcomes ("benefits" and "risks" including competing risks) as a function of index risk score to make joint inferences
    - Nonparametric function estimation procedure
- Plot probability estimates of several outcomes simultaneously as a function of baseline risk

Example: DES for Stage 3/4 Prostate Cancer

- Participants to high (N=242) and low doses (N=241)
- Primary endpoint
  - Time to death due to prostate cancer
- High doses
  - Appear to reduce prostate cancer death
  - Potential cardiovascular (CV)-related toxicity
  - Deaths (prostate cancer, CV disease, other)
    - High dose group (48, 78, 34)
    - Low dose group (77, 61, 46)
- Challenge
  - Clinical decisions regarding DES use depend on:
    - Evaluation of all important risks (e.g., prostate cancer death, CV death, etc.)
    - Individual baseline risk
- Baseline risk scoring system
  - Constructed using age, weight index, performance rating, history of CV disease, serum hemoglobin, size of primary lesion, and Gleason score

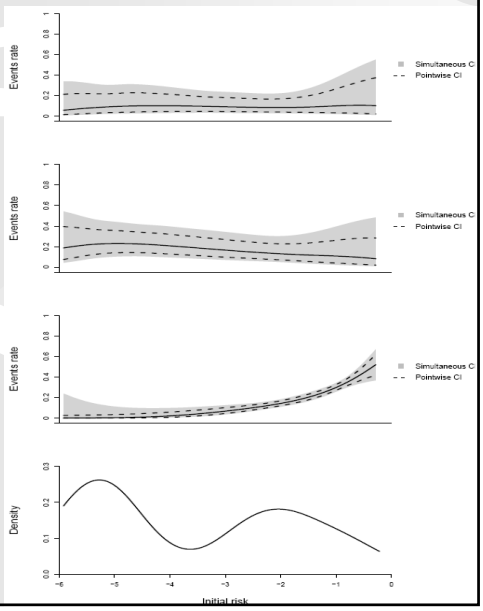
5 Year Event Results

- Death from other causes
  - Little influence of baseline risk
- CV related death
  - High w/ initial risk < -4
    - Monitor these patients
- Prostate cancer death
  - Low w/ initial risk < -2
  - If initial risk > -2, then DES may not be a good option
- Density for risk scoring index



2 Year Event Results

- Death from other causes
  - Little influence of baseline risk
- CV related death
  - Little influence of baseline risk
- Prostate cancer death
  - Low w/ initial risk < -2
- Density for risk scoring index



Composite Endpoints

- Attractive idea for a more complete characterization of treatment effects
- Be wary of differing importance of components
- General wisdom
  - 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

Example: ACTG HIV Trial

- Time to "treatment failure" defined from randomization to any of:
  - Death
  - Disease progression (i.e., AIDS-defining OI or malignancy)
  - Virologic failure (plasma HIV-1 RNA ≥1000 copies/mL)
- Difficult to interpret:
  - Arm A: longer time to event but events are deaths
  - Arm B: shorter time to event but events are virologic failure

Global Benefit:Risk (GBR) Score

- Pre-specify a multinomial outcome based on efficacy and safety data
  - E.g., 5 ordered categories (descending desirability)
    - Efficacy without serious side effects
    - Efficacy with serious side effects
    - No efficacy with no serious side effects
    - No efficacy with serious side effects
    - Side effects leading to drop-out
- Assign weights to reflect the relative importance of each category when evaluating treatment
- Compute a summary measure
- Conduct between-treatment comparisons of the summary measure distributions

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

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

- 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 safety events according to a prespecified algorithm

$$BLR = B_j - f\,R_j$$

- Advantages
  - Intuitive
  - Patient specific interpretation
- Disadvantages
  - Difficult to reducing safety experience into a single score
  - Difficult to choose a conversion factor
- Chuang-Stein, *CCT*, 1994.

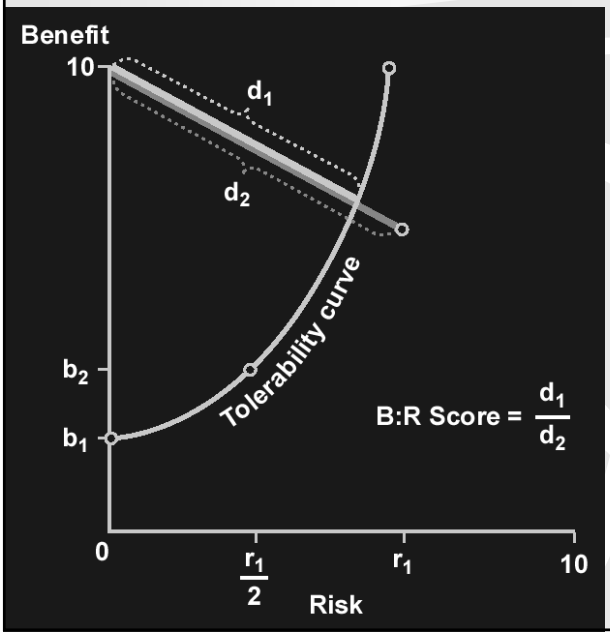
Patient Level Measures

- Patients rate their **overall experience** with respect to ***perceived*** benefit and risk
- Possibly useful for therapies to treat symptoms (e.g., pain) when "risks" are recognizable to the participant
  - E.g., A5265, a randomized study of nystatin vs. GV for treatment of OC
    - Risks are primarily "bad taste" and stigmatization due to staining
- Problematic when symptoms do not equate with risk
  - i.e., silent risk of abnormal labs (e.g., LFTs, bilirubin)

Bivariate Approaches

- Assume benefit and risk each can be measured on a continuous (e.g., 0-10) scale or can be transformed as such
- Create a scatterplot (benefit vs. risk) based on patient results
- Fit a smooth "tolerability boundary" through 2 or more points
  - E.g., based on 3 points: 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)
  - All points on the boundary have equivalent benefit:risk
- Reduce to one-dimensional analyses by defining a distance function
  - Use the tolerability boundary to standardize the distance
  - Closer to (risk =0, benefit=10) → Benefit:risk ↑

Example: Bivariate Approach



A5252: Study of therapies for neuropathic pain

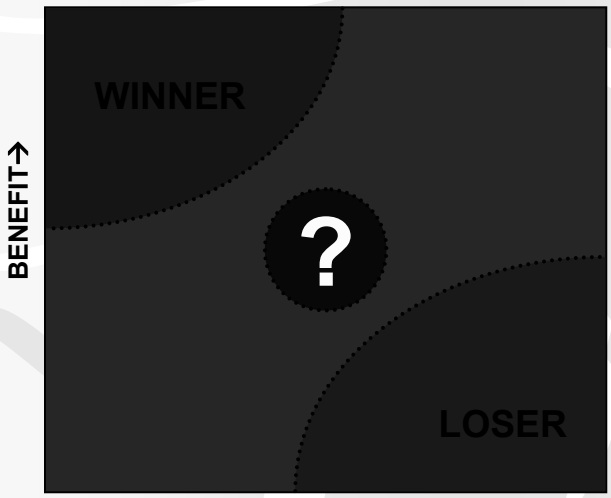
Benefit: Pain (measured 0-10)

Risk: "how bothersome were the side effects?" (measured 0-10)

Bivariate Approach

- Can create "confidence regions" for benefit:risk assuming bivariate normal
  - Produces an ellipse
- Can partition the plane into regions of interest and summarize the proportion of individuals that fall into these regions

Partition plane into regions of interest and summarize the proportion that fall into these regions



? 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

Comparative Bivariate Approach

- Construct confidence region for difference in benefit vs. difference in risk
  - Bivariate normal assumption, or
  - Bootstrap and nonparametric density estimation
    - Draw repeated samples with replacement
    - Obtain benefit difference and risk difference for each sample
    - Obtain 2-dimensional kernal density estimate
    - Identify contour of kernal density estimate that includes (1-α)*100% of the bootstrap estimates
    - Can construct regions of interest and summarize the proportion of bootstrap samples that fall into these regions
- Example: PROPHET
  - Randomized trial of hydrocortisone vs. placebo for very low birth-weight babies to prevent chronic lung disease
    - Benefit: survival without supplemental oxygen at 36 weeks postmenstrual age
    - Risk: GI perforation

PROPHET Example: Confidence Region

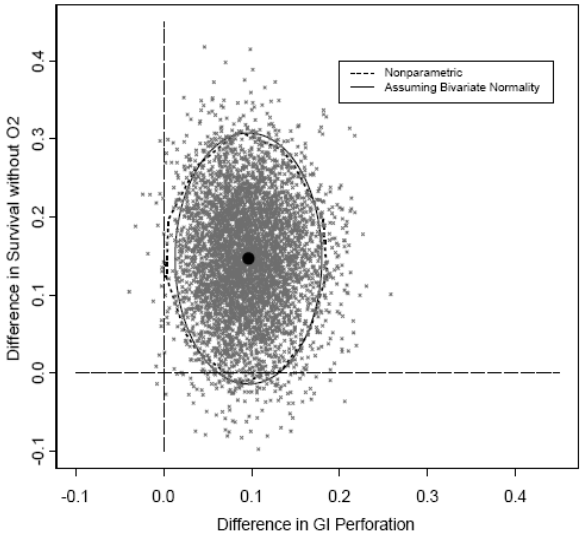


Figure 3
Confidence regions and bootstrap estimates for the PROPHET study. 90% confidence regions and bootstrap estimates for the PROPHET study.

Reporting

- Need for presentations that are interpretable for non-researchers
  - Help patients/clinicians/regulators/industry make better decisions
- Suggestions
  - Visual summaries
  - Include measures of variation
  - Consult CONSORT
  - Presentation by subgroups for which benefit:risk varies
  - Kaplan-Meier curves estimating the cumulative incidence may be appropriate for presenting risk data when risk changes over time

Reporting

- Provide absolute risk information with relative risk summaries
- Provide risk for alternative therapy or risk without therapy for reference (i.e., "baseline risk")
- Provide summaries for each important outcome (benefits AND harms) affected by therapy
- Group information by outcome severity

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

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)

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

For those outcomes included in the WHI "global index", the absolute excess risks per 10,000 women-years in the group treated with PREMPRO were 7 more CHD events, 8 more strokes, 8 more PEs, and 8 more invasive breast cancers, while the absolute risk reductions per 10,000 women-years were 6 fewer colorectal cancers and 5 fewer hip fractures. The absolute excess risk of events included in the "global index" was 19 per 10,000 women-years. There was no difference between the groups in terms of all-cause mortality. (See **BOXED WARNING** , **WARNINGS** , and **PRECAUTIONS** .)

Communicating benefits and risks in the label of a new drug

PDR® entry for

PREMPRO™ (Wyeth)
(conjugated estrogens/medroxyprogesterone acetate tablets)
PREMPHASE®
(conjugated estrogens/medroxyprogesterone acetate tablets)
Rx only

TABLE 8. RELATIVE AND ABSOLUTE RISK SEEN IN THE ESTROGEN PLUS PROGESTIN SUBSTUDY OF WHI [a].

Event [c.]	Relative Risk PREMPRO vs Placebo at 5.2 Years (Nominal 95% CI*)	Placebo n = 8102	PREMPRO n = 8506
Absolute Risk per 10,000 Women-years			
CHD events	1.29 (1.02-1.63)	30	37
Non-fatal MI	*1.32 (1.02-1.72)*	*23*	*30*
CHD death	*1.18 (0.70-1.97)*	*6*	*7*
Invasive breast cancer [b.]	1.26 (1.00-1.59)	30	38
Stroke	1.41 (1.07-1.85)	21	29
Pulmonary embolism	2.13 (1.39-3.25)	8	16
Colorectal cancer	0.63 (0.43-0.92)	16	10
Endometrial cancer	0.83 (0.47-1.47)	6	5
Hip fracture	0.66 (0.45-0.98)	15	10
Death due to causes other than the events above	0.92 (0.74-1.14)	40	37
Global Index [c.]	1.15 (1.03-1.28)	151	170
Deep vein thrombosis [d.]	2.07 (1.49-2.87)	13	26
Vertebral fractures [d.]	0.66 (0.44-0.98)	15	9
Other osteoporotic fractures [d.]	0.77 (0.69-0.86)	170	131
a: adapted from JAMA, 2002; 288:321-333			
b: includes metastatic and non-metastatic breast cancer with the exception of in situ breast cancer			
c: a subset of the events was combined in a "global index," defined as the earliest occurrence of CHD events, invasive breast cancer, stroke, pulmonary embolism, endometrial cancer, colorectal cancer, hip fracture, or death due to other causes			
d: not included in Global Index			

Improved Assessment of Benefit:Risk

- Thoughtful design
  - Plan appropriate follow-up duration to observe events due to cumulative exposure and after treatment discontinuation
  - Planning for collection and analyses of safety data
  - Better vision and clarity of objectives via clear definitions
    - How will data be reduced and integrated
- Adherence to fundamentals of trial conduct
  - Diligent adherence to ITT and patient follow-up regardless of adherence, etc.
- Creative analyses
  - Awareness to time-varying benefit:risk
  - Within-patient analyses
  - Personalized medicine and subgroup analyses
  - Flexibility and sensitivity analyses to subjective weighting
  - Include measures of variation
  - Validation of risk scoring systems and models
  - Safety analyses in Statistical Analysis Plan
- More transparent and interpretable reporting
  - Absolute risk and relative risk
  - Graphical displays
  - Use of reference groups

Keys to Progress

- Education
  - Thoughtful design and planning
  - Sound trial conduct fundamentals
- Communication between statisticians and clinicians
  - More detailed discussions regarding data reduction, integration, synthesis, and interpretation
  - Develop better vision
    - Develop objectives, associated definitions, and scoring systems to used for assessment
- Thinking creatively
- Continuous re-assessment as information accumulates, therapeutic alternatives change, and diseases and medical advances evolve

CONSORT: Better Reporting of Harms (*AIM*, 2004)

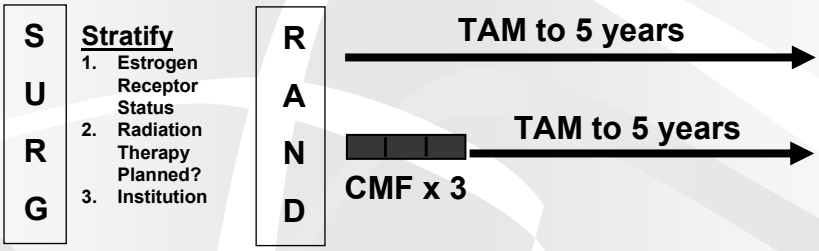
- Common poor reporting practices
  - Generic statements (e.g. generally well-tolerated)
  - Failing to provide data by arm
  - Providing summary numbers for all AEs but not for each type of AE
  - Providing summary numbers for a specific type of AE without severity or seriousness
  - Reporting AEs with only a certain frequency (e.g., > 3%)
  - **Reporting only AEs with a significant p-value for a between-arm comparison**
  - Reporting measures of central tendency without information on extreme values
  - Disregard for timing of the events
  - Not distinguishing between patients with 1 AE vs. several AEs
  - Providing statements of statistical significance without providing event counts
  - **Not providing data for all randomized participants**

CONSORT: Better Reporting of Harms (*AIM*, 2004)

- Recommendations
  - Title/abstract should state harms data were collected
  - Introduction should state harms data were collected
  - List AEs with definitions
  - Clarify how data were collected (mode, timing, attribution, monitoring, stopping rules)
  - Describe statistical plan for analyzing/presenting harms data
  - Describe by arm the withdrawals due to harm and experience with allocated treatment
  - **Present denominators for analyzing harms**
  - Present absolute risk for each AE
  - Describe subgroup analyses conducted
  - **Provide balanced discussion of benefits and harms with emphasis on limitations and generalizability**

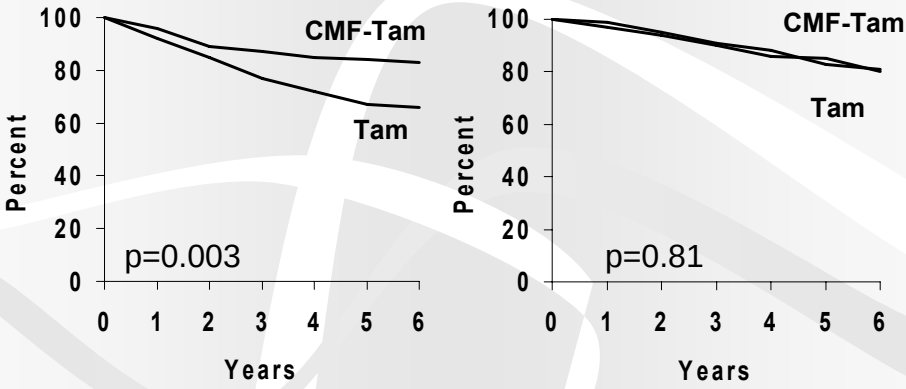
Trial IBCSG IX

Postmenopausal, Node Negative, n=1,715



Disease-Free Survival (Trial IX)

ER- (0-9 fmol/mg) **ER+ (≥10 fmol/mg)**



STEPP: Trial IX by ER