DATA AND SAFETY MONITORING

Pamela Shaw, Ph.D.

Assistant Professor
Division of Biostatistics and Epidemiology
University of Pennsylvania

IPPCR Jan 5, 2016

Acknowledgements

- Laura Lee Johnson (NCCAM,NIH)
- Dennis Dixon (formerly NIAID, NIH)
- Michael Proschan (NIAID, NIH)
- Susan Ellenberg (UPenn)

Presentation Outline

- Introduction
- Examples – monitoring by independent Data and Safety Monitoring Board (DSMB)
- Data and safety monitoring basics
- When to use a DSMB
- Monitoring without a formal DSMB
- Sample items for monitoring
- Summary

Fundamental Concept: Caution

- In a randomized clinical trial studying the safety/efficacy of a new drug, you don't want to wait until the end of the trial before you consider whether
  - Something fundamentally went wrong with the conduct of the trial
  - The new drug has unexpected harms
  - There was convincing evidence of efficacy much earlier than expected

Fundamental Concept: Integrity

- Looking at data multiple times, i.e. doing a formal statistical test to compare treatment arms, increases your chance of having a false positive result
- Data monitoring plans are a fundamental part of the trial design/protocol
  - Special methods are used in the statistical analysis to maintain the desired false positive (type I error) rate
- DSMBs will require scientific expertise to know what questions to ask and how to interpret the study data with right amount of caution

Fundamental Concept: Objectivity

- One way to help ensure the group charged with monitoring the clinical trial can do their job of evaluating safety and integrity of the trial they should have
  - No vested interest in the trial outcome
  - No relationship with study investigators
  - Sufficient clinical expertise and scientific expertise, including understanding of clinical trial design, conduct and data-analytic issues

A Definition

A **data and safety monitoring board (DSMB)** is a group of independent experts that reviews the ongoing conduct and evolving data of a clinical trial to ensure continuing patient safety as well as the validity and scientific merit of the trial.

Note on terminology

- FDA and European boards called DSMC
- NIH generally uses DSMB

Example 1. ACTG 076: Preventing Mother-Infant HIV Transmission

- Zidovudine (AZT) able to slow progression of HIV in adults with advanced disease
- AIDS Clinical Trials Group Protocol 076 designed to assess both safety and efficacy of Zidovudine in preventing transmission of HIV from infected (not advanced) women to their babies
  - NEJM 1994; **331**, 1173-1180

ACTG 076: Ethical Considerations

- If partway thru trial determine AZT prevents vertical transmission, must stop and give to all mothers in trial (individual ethics)
- Don't want to stop before convincing medical community that AZT prevents vertical transmission

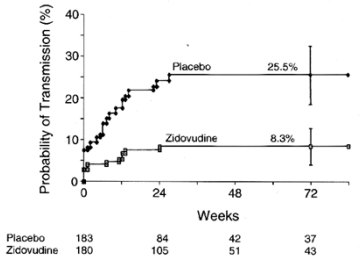
ACTG 076: Study Design

- Powered (80%) to detect a 33% reduction of transmission rate (through 78 weeks of baby's life) relative to projected rate of 30%
- Target sample size (N) was 748
- Accrual began April 1991
- Projected accrual to take at least 5 years, planned for 15% dropouts

ACTG 076: Monitoring

- DSMB met twice a year to monitor safety
- Efficacy reviews planned after each 1/3 of projected infant infections
- 1st efficacy review took place in February 1994, based on mothers enrolled up to December 1993 and their babies

At First Interim Analysis



P = 0.00006

Interim Analysis Considerations

– At time of DSMB meeting 46 babies w/unknown HIV status
  - Concern: might stop because met stopping criteria, but when 46 infants pending are evaluated, evidence could weaken
  - Conditional power can be used for this: Given results in 363 babies, and making different assumptions about the 46 pending, examine probability that test statistic still meets stopping criteria

Outcomes After 1st Interim Review

- DSMB recommended stopping (after careful review of data quality and completeness, toxicity, transmission rates)
- Trial leaders and NIH agreed
- Zidovudine provided to those in control group
- United States Public Health Service (USPHS) national guidelines modified

Example 2: CAST

- Cardiac Arrhythmia Suppression Trial (CAST) was designed to evaluate hypothesis of whether suppressing ventricular arrhythmias in patients with recent MI reduced sudden death and mortality
  - Compared 3 active drugs known to suppress arrhythmias with placebo: encainide, flecainide, and moricizine
  - Designed to randomize 4000 patients with 90% power to detect a reduction in sudden death with a 0.05 one-tailed significance level

CAST: First DSMB Meeting

- DSMB met for first time in March 1987 to review the protocol
- Made recommendation that the one-tailed 0.025 alpha level be used to test for treatment benefit, reducing power to 85%.
  - DSMB argument: one or two-tailed test should not influence the strength of evidence (ie. 1.96 critical value)
  - Very rare to see one-sided 0.05 level tests unless early phase and doing 90% (not 95%) confidence intervals
- **Lesson Learned:** important to have first meeting before study starts to make sure DSMB and investigators agree with scientific design and what will be monitored
  - Two groups need to get along, trust each other, and work together through length of trial

CAST: DSMB 1st data look

- In Cardiac Arrhythmia Suppression Trial (CAST) (NEJM 1989; **321**, 406-412), DSMB chose to be blinded
  - At first look, had 22 of 425 expected events by trial's end; (22/425=.052 = 5% of expected information)
  - Stopping Boundary at first look: 3.22, logrank Z=3.43
  - But boundary not in place at first look
  - Good because actual score at 1st look after pending events adjudicated was 22-10
  - DSMB remained blinded

CAST: DSMB 2nd look

- At second look, 48 events
- Now expected only 300 events by end of study, so 48 represented a higher percent of total events
- Revised information fraction was t=48/300=.16
- Boundary ±2.98
- Logrank Z=3.22
- DSMB unblinded and found drugs harmful

Further investigation

- Verified treatment labels correct and chemically analyzed random sample of medications
- Looked for baseline imbalances of key variables & did analyses adjusting for baseline imbalances
- Looked at results by subgroup to see if harm confined to certain patients
- Looked at secondary endpoints (like total mortality) to see if consistent
- Checked if harm confined to specific drugs

Trial Outcomes

- Harm seemed to be confined to encainide and flecainide
- Overall score: 35-13
- Encainide or flecainide versus placebo: 33-9
- "CAST I" stopped in April, 1989
- "CAST II" stopped in July 1991, also for lack of benefit and possible harm

CAST: Lessons learned

- DSMBs and trial data management system must be in place early on.
  - Negative trend apparent after only 5% of data in, convincing at 15%.
- DSMB and investigators must be able to act/respond quickly
- DSMB/investigators must be ready to make contingency plans

One other lesson demonstrated (if not learned)

- The ability of the DSMB to do their job can be compromised if they are masked
- Human nature: tend to believe an early trend is in favor of new treatment
  - Require less evidence to prove harm vs benefit
- Masked analyses can be inefficient and error prone.

Fundamental concept: competency

Ref: Meinert, Curtis L. "Masked Monitoring in Clinical Trials—Blind Stupidity?." *New England Journal of Medicine* 338.19 (1998): 1381-1382.

Data and Safety Monitoring: *Why?*

- To identify any safety problem rapidly
- To identify logistical problems
- To evaluate continued feasibility of trial
- To determine if trial objectives have been met and trial may be terminated early

Data and Safety Monitoring: *What?*

- Logistics
  - Enrollment
  - Baseline Data, Comparability
  - Protocol Compliance
  - Specimen Collection
  - Data Quality

  Develop specific benchmarks

Data and Safety Monitoring: *What?*

- Outcomes
  - Adverse Events
  - Interim Variables
  - Response Variables (Endpoints)

Data and Safety Monitoring: *Who?*

- Ethics Committee(s)
- Sponsor
- Regulatory Agencies
- Data and Safety Monitoring Board

  (DSMB, DSMC, DMC, External DMB, etc)
- Investigator(s)
- Safety Monitor

Data and Safety Monitoring Board's Purpose Drives Structure

Recall DSMB's purpose is

- To ensure regular and systematic interim monitoring
- To provide an objective assessment of the interim data
- To protect confidentiality of interim treatment comparisons

Generally Accepted Principles

- Certain types of trials should have formal DSMBs
- DSMBs should be multidisciplinary
- A charter should describe the operations and procedures of a committee
- DSMB members should be free of conflicts of interest
- Interim data should be considered highly confidential

Confidentiality of Interim Results

- Interim comparative data generally considered highly confidential, because
- Knowledge of interim data could affect
  - Patient entry
  - Patient care
  - Patient assessment
  - Sponsor action
- When knowledge of interim data potentially could influence trial conduct, interpretation of results could be muddied

An *Independent* DSMB Is One in Which No Member Has

- Any basis for preferring the outcome to be in one or the other direction
- Any ability to influence the trial conduct in a role other than that of DSMB member

Establishing a Committee

- Generally appointed by study sponsor
- Made up of
  - Clinicians (appropriate specialty)
  - Statisticians
  - Others as needed (e.g., bioethicist, subject advocate)
  - Executive Secretary
- Membership should be acceptable to trial leadership: DSMB as any version given major responsibility
- Independent of protocol team
  - May or may not be independent of study sponsor

Structure of Meetings

- Open Session
  - Process data
  - Attended by investigator(s), sponsor representative, data site representatives
- Closed Session
  - Interim and outcome data, adverse events by group
  - Attended by data presenter; any others?
- Executive Session
  - "Private" DSMB member discussion
  - Any other attendees?

Scope of DSMB Responsibilities

- Evaluating accumulating data with regard to toxicity, and potentially efficacy and/or futility
- Recommending termination or continuation of study
- Recommending other study modifications
- Reviewing study protocol
- Assessing study conduct
- Recommending additional analyses

Statistical Considerations

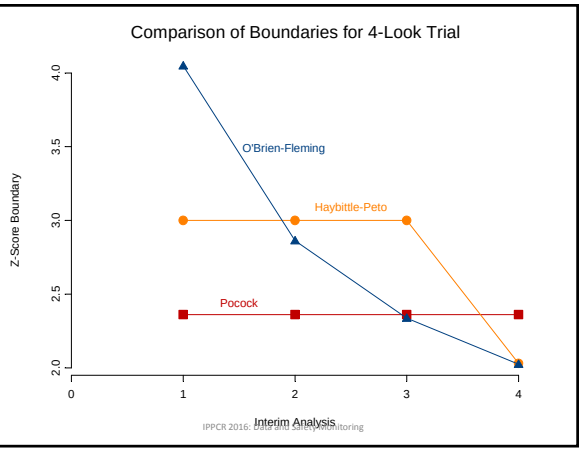
- Repeated testing for efficacy over time inflates Type I (false positive) error rate if no adjustment made
- In "early days" of clinical trials, not uncommon to stop study as soon as p-value reached magic level of 0.05
  - Better statistical methods now exist to ensure more valid monitoring structure
- Currently, many methods available to permit monitoring and potential to stop early for efficacy without increasing error rate
- Monitoring plan should be laid out in protocol or a Statistical Analysis Plan and reviewed by DSMB before study begins
  - Situations can arise that require unplanned monitoring

Actual False Positive Rate Using ±1.96:

# Looks	Equally-Spaced	Worst-Case Spacing
1	.05	.05
2	.08	.10
3	.11	.15
4	.13	.19
5	.14	.24
6	.15	.28
7	.17	.32
8	.18	.37
9	.19	.41
10	.19	.45

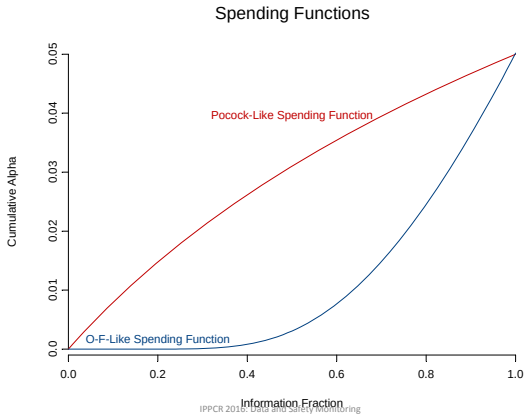
Efficacy and Boundaries

- Several examples of early results reversing by end of trial
  - In the Coronary Drug Project (CDP), the z-score reached -1.96 three times and ended up about 0
  - In the Multicenter Unsustained Tachycardia Trial (MUSTT), treatment of arrhythmia guided by electrophysiological testing appeared to be harmful and then proved beneficial at end



Common Efficacy Monitoring Approach

- Lan-DeMets (Biometrika 1983; **70**, 659-63) spending function approach:
  - Don't need equal amounts of information between successive looks
  - Don't even need to pre-specify number or timing of looks
- Instead you specify a spending function $\alpha^*(t)$ telling how much alpha you will spend by information fraction t.
  - Example information fraction: current # events/N



Futility Monitoring

- Studies are sometimes stopped for futility because of:
  1) Poor recruitment
  2) Lower than expected event rate and trial was powered for given percentage reduction
  3) Higher than expected dropout or crossover rates
  4) No effect of treatment on intermediate outcome through which treatment is assumed to work (e.g., HbA1c in diabetes trial)
  5) ***Poorer than expected treatment effect on primary endpoint?***

Futility Monitoring

- Reason 5 is very different from the others
  - If everything else (recruitment, dropout, effect on surrogate, crossover, etc.) is okay, then
    - Power to detect originally hypothesized treatment effect is still high
    - A small treatment effect answers the question
    - If continuation is ethical, may want to continue to prove that treatment doesn't work (especially if treatment is in widespread use and disease not too serious—glucosamine/chondroitin for arthritis)

Futility Monitoring

- Want to ask two different questions:
  1) Will end result be null?
  2) Will null result still be meaningful?
- Two different tools for the two different questions:
  1) Will end result be null? *Conditional power*
  2) Will null result still be meaningful? *Unconditional power*

Futility Monitoring

- *Unconditional power:* Had we known prior to trial what we know now about dropout, crossover, control event rate, achievable sample size, what would power have been to detect hypothesized treatment effect?
  - Don't need to see the results by arm
- *Conditional Power:* Knowing the above ***and taking into account current results,*** what is probability of getting a significant result at end?
  - Need to take into account current results

Monitoring Recommendations

- Continue protocol unmodified
- Continue and modify protocol
- Study be terminated
- Sponsor makes final decision regarding termination, usually will follow recommendation of DSMB
  - At NIH: Institute director usually makes final decision, but almost always follows recommendation of DSMB

DSMB May Recommend Stopping If

- A safety issue has emerged
- The trial has already demonstrated efficacy
- Interim results preclude a positive finding
- Operational difficulties are insurmountable
- External information undercuts the scientific rationale for the trial

Decision Making Process is Complex

- Internal consistency
- External consistency
- Benefit/risk balance
- Current vs. future patients
- Clinical and public health impact
- Statistical issues

Individual versus Collective Ethics

- Must balance individual & collective ethics
  - *Individual ethics:* primary concern is about welfare of patients in trial
    - Waiting too long to stop jeopardizes their welfare
  - *Collective ethics:* must also be concerned about how trial results affect future patients
    - Stopping too soon may not convince medical community)
- Decisions often difficult!

Downside of Early Stopping for Efficacy

- Early stopping handicaps safety analysis
- Monitoring safety and efficacy are inherently different
- Clinical trials address relative effects much better than absolute ones

Example 3: Male Circumcision to Prevent HIV Acquisition

- Phase III controlled trials began at about the same time in South Africa, Kenya, and Uganda
- Designs similar
- South African trial reported clear evidence of efficacy in July 2005
- NIH DSMB recommended continuing other trials in August 2005, June 2006, stopping in December 2006

ALL TRIALS NEED MONITORING BUT

NOT ALL TRIALS NEED DSMBS

Data and Safety Monitoring Regulations, Policies, Guidelines (1)

- Policies - NIH
  - All trials need a monitoring plan, need to describe in application for funding
  - Independent DSMB required for all multicenter Phase III trials since 1979
  - Notify IRBs of DSMB Recommendations
  - Policies across NIH vary

Data and Safety Monitoring Regulations, Policies, Guidelines (2)

- NIH policy http://grants.nih.gov/grants/guide/notice-files/NOT-OD-00-038.html
  - Type of monitoring commensurate with risks, trial size and complexity
  - Part of Clinical Terms of Award
- FDA Policies – NONE
- FDA Guidance on Data Safety Monitoring Committees (2006)

HHS Office of Inspector General Review

- HHS Office of Inspector Genera conducted a review of NIH DSMBs to identify whether DSMBs of NIH sponsored trials met with NIH guidelines and whether any challenges to DSMB effectiveness existed
- In June 2013 issued report http://oig.hhs.gov/oei/reports/oei-12-11-00070.pdf

HHS Office of Inspector General Review: Results

Main findings:

1) Generally NIH DSMBs met guidance
2) Some issues exist:
   1) IC participation in closed DSMB meetings diminishes the appearance of independence;
   2) Not all Institutes and Centers (IC) policies reference DSMB access to unmasked data;
   3) NIH faces challenges in recruiting and training DSMB members.

HHS Office of Inspector General Review: Recommendations

NIH should

(1) Direct ICs to articulate the circumstances in which IC staff should participate in DSMB meetings.

(2) Direct ICs to explicitly reference DSMB access to unmasked data in their DSMB policies.

(3) Identify ways to recruit and train new DSMB members.

NIH concurred with all three recommendations.

Formal Study Monitoring

- ANY clinical research study involving more than minimal risk to volunteers
  - Does not have to be a "clinical trial" nor a randomized study
  - Data quality and safety monitored

When Are Independent/External DSMBs Needed?

- Large randomized trials with mortality or major morbidity endpoints
  - Trial has implications for clinical practice/ public health
- Trials for which assessment of serious toxicity requires comparison of rates
- Trials of novel, potentially high-risk treatments
- Highly vulnerable patient population

External DSMBs Generally Not Needed for

- Single-arm trials
- Early phase trials
- Short-term trials of treatments to relieve common symptoms
- Any trial for which there is no ethically compelling need to monitor the interim comparisons of safety or efficacy

Are There Disadvantages to Having a DSMB?

- YES!
  - Increases complexity of trial management
  - Increases costs

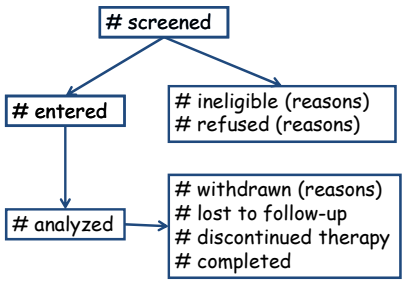
  If the ethical imperatives discussed earlier are not applicable, other (simpler) monitoring approaches are usually acceptable

Independent DSMBs are Not the Only Form of Monitoring

- Independent monitor for minimal risk studies
- Safety Monitoring committees take many forms
  - Protocol Team
  - Independent Safety Monitoring Committee (SMC)
  - IRB
  - DSMB

Sample Items Monitored by the DSMB

CONSORT Diagram



Protocol Adherence

Missed Visits

	Baseline	Visit 1	…	Last Visit
# Expected			…	
# Missed			…	
% Retained			…	

Overdue Forms

	Baseline	Visit 1	…	Last Visit
# Expected			…	
# Received			…	
% Complete			…	

Baseline Characteristics

Gender	% Female
Age	Median, Min, Max
Race	% White, % African American
Stage of Disease	% Early, % Advanced
Prior Therapy	% None, % Surgery, % Chemo

Serious Adverse Events

Line Listing showing

- Event
- Study entry, treatment start dates
- Event start, stop dates, final resolution
- Relationship to research procedures
- Other relevant patient characteristics

Adverse Event Summaries

Lab Abnormalities and Clinical Signs

Tables of frequencies, by adverse event (AE) type and severity Include all those treated

- Sort by body system
- Count 1st occurrence for each volunteer
- Severity of an event summarized by highest grade individual experienced for that event
- Summarize across types

Fundamental Concept: data summaries help identify trends not apparent in line listing

Adverse Event (AE) Summaries

Body System MedDra Term	Mild	Moderate	Severe or Worse	Total
INFECTIONS INFESTATIONS	3(30%)	1(10%)	0(0%)	4(40%)
Gastroenteritis	1(10%)	0(0%)	0(0%)	1(10%)
Sinusitis	1(10%)	0(0%)	0(0%)	1(10%)
Upper Respiratory Infection (URI)	0(0%)	1(10%)	0(0%)	1(10%)
Viral infection	1(10%)	0(0%)	0(0%)	1(10%)

Efficacy Summaries: If Planned

Summarize study endpoints:

- % treatment success
- Average area under the curve (AUC), antibody response, etc.
- Combined or by group?
  - Keep report masked, but not DSMB
- Interim futility monitoring
- Interim efficacy monitoring
  - Important to have independent committee

DSMB Report

Fundamental Concept: Make it as easy as possible for DSMB to do their job

- DSMB report should be provided to the DSMB in enough time for review prior to the meeting
- Report should include a concise (1-2 page) summary of protocol design, including planned N, # of arms, and endpoints
- Should include a clearly written executive summary
- Tables/figures should be concise and easy to understand

Summary

- DSMBs have a difficult job balancing:
  - Individual ethics (most important)
  - Collective ethics
- Interim monitoring of efficacy requires formal monitoring plan to prevent inflation of false-positive rate, need monitoring boundaries
- Sometimes want to stop for futility
- Stopping decisions almost always based on statistical and extra-statistical considerations
- DSMBs make recommendations, final decisions rest with trial and sponsor leadership

Further Reading

- Data Monitoring Committees in Clinical Trials: A Practical Perspective by Susan S. Ellenberg, Thomas R. Fleming and David L. DeMets (Wiley Series in Statistics in Practice, 2002)
- Data Monitoring in Clinical Trials: A Case Studies Approach, Editors: David L. DeMets, Curt D. Furberg, Lawrence M. Friedman (Springer 2005)
- Friedman, Furberg, DeMets. Fundamentals of Clinical Trials, 4th Edition (Springer 2010)

Statistical References

- Jennison C and Turnbull BW. Group Sequential Methods with Applications to Clinical Trials Chapman & Hall/CRC 2000
- Statistical Monitoring of Clinical Trials: A Unified Approach (Statistics for Biology and Health) by Michael A. Proschan, K. K. Gordon Lan and Janet Turk Wittes (Springer, 2010)

Questions?