NIH Collaboratory

Health Care Systems Research Collaboratory

Data and Safety Monitoring in Pragmatic Clinical Trials

Susan S. Ellenberg, PhD
Greg Simon, MD, MPH
Jeremy Sugarman, MD, MPH, MA

Overview

- The need for DSMBs (Jeremy Sugarman)
- Special considerations for DSMBs in PCTs (Susan Ellenberg)
- Challenges for investigators (Greg Simon)
- Discussion

Jeremy Sugarman, MD, MPH, MA

Harvey M. Meyerhoff Professor of Bioethics and Medicine
Johns Hopkins Berman Institute of Bioethics

Background

- Blinding and randomization are powerful research techniques used to minimize bias
- However, emerging experiences and data can pose ethical quandaries for investigators in meeting their obligations to minimize risk to participants

Why have an external approach to monitoring?

- To ensure that participants are not exposed to undue risk
- To ensure that trial will yield usable results
- To balance the interests of patients within the trial with those outside the trial
- To guard trial integrity

What is a Data and Safety Monitoring Board?

- An independent group charged with reviewing the progress, conduct and outcomes of an ongoing clinical trial

Who is on a DSMB?

- Not set in stone
- Clinical experts
- Biostatistician/trialist
- Ethicist? Patient advocate? Investigators? Representative of sponsor?

Susan S. Ellenberg, PhD

Professor of Biostatistics
Professor of Medical Ethics and Health Policy
Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine,
University of Pennsylvania

DSMBs for pragmatic trials

- What are the special issues for DSMBs for pragmatic clinical trials?
- (ARE there any special issues for DSMBs for pragmatic trials?)

Issues for discussion

- Need for a DSMB
- What a DSMB will monitor
- Participant follow-up
- Data analysis
- DSMB composition

Issue 1: Do PCTs need DSMBs?

- All clinical trials require some monitoring of interim data
- General guidelines for requiring a DSMB apply to pragmatic trials
  - Trials in which participant safety requires regular review of comparative safety and efficacy data
  - Trials intended to have substantial public health impact
- Since pragmatic trials will typically be addressing questions intended to impact health practices, an expert oversight group will be important for most PCTs

Issue 2: what gets monitored?

- Traditional trials: monitor data on safety, efficacy, and quality of study conduct
- These are important in pragmatic trials also
- Possible special issues in pragmatic trials
  - Study outcomes
  - Protocol adherence
  - Eligibility
  - Design factor in cluster randomized trials

What gets monitored: study outcomes

- Study outcomes
  - PCTs may be more likely to include subjective outcomes as primary or key secondary endpoints
  - PCTs may be less likely to incorporate central adjudication of outcomes
  - DSMBs will have to recognize that data may be more variable than in more restrictively designed trials

What gets monitored: adherence

- Protocol adherence
  - A basic tenet of PCTs is to evaluate treatments as they would be given in practice
  - To some, this means no great effort to promote, or even monitor, adherence to protocol
  - DSMBs typically consider monitoring study quality as one of its mandates; may be uncomfortable making recommendations based on observed treatment effects without any sense of how effectively interventions are being administered
  - If adherence is very poor and there is no apparent treatment difference, 2 possibilities
    - Treatments produce similar effects
    - Protocol not followed by investigators and participants
  - If you don't know anything about adherence, may not be able to conclude anything about relative treatment effects

DSMBs and protocol adherence

- Should a DSMB ignore data on protocol adherence in a PCT? Should these data not even be reported?
- Poor adherence could lead to safety issues in some studies
- Important to distinguish between
  - Lack of adherence as reflecting how a treatment would be used in practice
  - Lack of adherence as reflecting insufficient understanding of trial on part of investigators and/or participants
- DSMBs need to pay some attention to this issue
- May be particularly important to review adherence data by site, to assess need for re-training

What gets monitored: cluster-randomized trials

- For cluster-randomized trials, design often used in pragmatic trials, also important to monitor the "design factor"
  - Intra-cluster correlation coefficient (ICC)—the extent to which results within a cluster will be more similar than results across clusters—is a component of sample size calculation
  - Typically, hard to estimate ICC from prior data—estimates used to design trial may be way off
  - Interim estimates of ICC important to see whether study will have expected power

Issue 3: Participant follow-up

- Pragmatic approaches to follow-up may create challenges for DSMBs
- Follow-up information will likely be derived from electronic health records (EHRs) in some trials which may be updated on different schedules if different systems are used
- Follow-up frequency may vary by institution according to local policies
- Interim comparisons will be more difficult without standardized follow-up schedules

Issue 4: Data analysis

- Analytical issues
  - Cluster randomization
  - Decentralized analysis
- Philosophical issues
  - Early termination criteria

Data analysis

- Use of cluster designs
  - Many PCTs currently underway with NIH collaboratory or PCORI funding randomize clusters rather than units
  - Analysis of such trials requires accounting for intra-cluster correlation
  - Differing practices among clusters will have to be accounted for in interim analyses
    - Example: minimally restricting usual practice may mean patients in different clusters are followed on different schedules

Potential analytical issue

- Need for de-centralized analysis
  - Privacy concerns may preclude merging data from multiple EHR systems at a central site
  - In such cases, interim analyses may need to be done separately for each site, with summary data only delivered to central statistical group
  - Such arrangements will raise challenges in terms of timeliness of data, quality control and assurance that all analyses have been conducted in identical manner

Interim monitoring strategy

- Early termination for efficacy
  - Since PCTs will be designed to influence practice, could be argued that early termination criteria should be extremely stringent (or maybe not even considered)
  - Will be important to ensure that DSMB and trial leadership are in agreement on criteria
- Early termination for futility
  - When studies compare two “standard-of-care” regimens, questionable whether early stopping for futility should be considered at all
  - As with efficacy, DSMBs and trial leadership must have common understanding of criteria for early termination
- Early termination for safety

Issue 5: DSMB composition

- Clinical and statistical expertise needed
- Will probably be more common to include patient representative
  - PCORI-funded studies require patient partners as members of research teams
  - Studies aimed at questions intended to influence clinical practice may particularly benefit from patient insights
- Expertise in medical informatics may be desirable for some PCTs
  - Use of electronic health data
  - Complex database linkages
  - Natural language processing

The DSMB Charter

- The charter is essentially an agreement between the DSMB, the trial sponsor and the trial investigators about the responsibilities and operation of the DSMB
- The charter will address issues such as
  - Meeting format and frequency
  - Conflicts of interest
  - Statistical approach to monitoring
  - Preparation of meeting minutes
- A Charter is not a formal contract; it guides DSMB actions but the DSMB must be free to exercise its judgment

Greg Simon, MD, MPH

Senior Investigator
Kaiser Permanente Washington Health Research Institute

What's different in pragmatic trials

- Variable fidelity of or adherence to interventions
- Inference regarding adverse events
- Limited access to outcome data
- "Actionability" of interim analyses

Monitoring intervention fidelity or adherence

- The question: Will this trial support valid inference regarding the benefits and risks of the intervention(s) being tested?
- In a traditional clinical trial, gaps in fidelity or adherence are threats to validity
- In a pragmatic trial, gaps in fidelity or adherence are signal rather than noise
- BUT….Is there any limit to that?

SPOT study example

- Outpatients reporting frequent suicidal ideation on routine depression questionnaires randomly assigned to outreach programs or continued usual care (no contact)
- Outreach invitation via online messaging in EHR patient portal
- Up to three cycles of outreach – patients free to decline or ignore invitation
- Analysis by initial treatment assignment, regardless of intervention participation
- Pilot studies found that 40-45% actively “accepted” invitation to program
- BUT invitations themselves have “active ingredients” of proven interventions
- AND we don’t know how participation is related to actual risk of suicide attempt

Is there a lower bound to adherence? It depends…

- Can we define and measure exposure (or non-exposure) to the intervention?
- What proportion of participants would need to be exposed to detect benefit?
- What do we know about beneficial or adverse selection into participation?

Monitoring adverse events

- The traditional question: Do adverse events signal some risk or potential harm of study intervention(s)?
- The traditional process:
  - Review of individual events (especially unexpected events) for “relatedness”
  - Comparison of event rates for serious adverse events (SAEs)
- In a pragmatic trial:
  - Study teams may not have access to clinical data to assess “relatedness”
  - For treatments in widespread use, a “signal” of harm may be only noise
  - SAEs may be indistinguishable from study outcomes

SPOT study example

- Suicide attempts and suicide deaths would usually be considered SAEs requiring immediate review.
- BUT:
  - those are the study outcomes
  - we are expecting to observe about 700 suicide attempts and 70 suicide deaths
  - they may be ascertained by nothing but an ICD10 code
- SO how would we assess whether suicide attempts or deaths are related to study intervention(s) except by finishing the trial?

Should we monitor adverse events? It depends…

- Unlikely that any “unexpected” event would signal a previously unrecognized risk?
- And comparison of event rates often overlaps with interim analyses (more later)
- But there still may be important questions regarding conflict of interest:
  - “Adverse events” may signal a risk that’s important, whether or not it’s “related”
  - Investigators and study staff must place duty to participants over duty to protocol
  - So monitoring may be indicated – but it’s about a different question

Limited access to outcome data

- Traditional trials rely on data collected and recorded by study staff, so:
  - The study protocol can dictate content and process of data collection
  - Data are available (almost) immediately
  - Study staff control the chain of custody
- Pragmatic trials often rely on the “data exhaust” of health care operations, so:
  - Data collection is controlled (or not controlled) by clinical and business needs
  - Access to data may be delayed by weeks or months
  - No single chain of custody is possible

SPOT study example

- Suicide attempts ascertained through EHR and insurance claims data
  - Often delayed up to 3 months
  - Clinical information to validate or adjudicate may be limited
- Suicide deaths ascertained through state mortality data
  - May be delayed by 18 months or more!
  - Clinical information to validate or adjudicate will be absent

Living with limited access to outcome data

- In many cases, prompt reporting of deaths is neither possible nor useful
- May affect timing (or even feasibility) of any interim analyses

When can we act on interim analyses?

- In a traditional clinical trial – once the question is answered, we should:
  - Stop assigning research participants to an inferior treatment
  - Advise clinicians and policy-makers regarding new evidence
- In a pragmatic trial
  - Outcome data may accumulate slowly (at different rates from different sources)
  - The threshold for action may be less clear

SPOT study example

- Delayed (and complex) schedule for outcome data:
  - Outcomes accumulate over 12-18 months
  - Greater delay for suicide deaths than non-fatal suicide attempts
  - Should not over-value early over sustained intervention effects
- Threshold for health system action is not clearly established (and will likely depend on cost as well as benefit)
- Consider practical consequences of stopping recruitment or intervention delivery:
  - For evidence of benefit:
    - Patients in participating health systems would no longer be offered effective programs
    - Timing of widespread implementation uncertain
  - For evidence of harm:
    - Patients in participating health systems would no longer be offered harmful programs
    - Widespread implementation would be avoided

Are interim analyses actionable? It depends…

- Consider timeline of data availability:
  - How soon could you detect meaningful or important difference?
  - What biases could be introduced by using incomplete data?
- Consider effects of early termination on potential study participants and others affected by condition of interest
- Consider different thresholds for detecting benefit and harm

DISCUSSION