CER VISIONS AND REALITIES:
A METHODOLOGIC PERSPECTIVE

Constantine Gatsonis
Department of Biostatistics and
Center for Statistical Sciences
Brown University

Topics

1. Definitions and scope of CER
   - From CER to PCOR
2. Methodologies and standards for CER
3. Diagnosis vs therapy
4. Conclusions
   - Goals and available methods are still quite apart
   - We are only at the very beginnning

IOM report on national priorities

Developed list of 100 top national priorities for CER

http://www.iom.edu/en/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx

What is CER?

- Several definitions of CER have been proposed.
- According to the 2009 IOM Committee , ***Comparative effectiveness research (CER) is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care.***
- The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care ***at both the individual and population levels.***

Effectiveness vs Efficacy

- Efficacy trials are conducted to assess interventions under "ideal" circumstances.
- Effectiveness trials are conducted to assess interventions under "real world" clinical settings
- Actual studies exist in the continuum between efficacy and effectiveness.
- However, effectiveness trials are expected to formulate their aims and design based on the realities of routine clinical practice and to assess outcomes that are directly relevant to clinical decisions.

First Quartile Recommendations (sample)

Compare the effectiveness of upper endoscopy utilization and frequency for patients with gastroesophageal reflux disease on morbidity, quality of life, and diagnosis of esophageal adenocarcinoma.

Compare the effectiveness of genetic and biomarker testing and usual care in preventing and treating breast, colorectal, prostate, lung, and ovarian cancer, and possibly other clinical conditions for which promising biomarkers exist.

Compare the effectiveness of management strategies for *localized prostate cancer* (e.g., active surveillance, radical prostatectomy [conventional, robotic, and laparoscopic], and radiotherapy [conformal, brachytherapy, proton-beam, and intensity-modulated radiotherapy]) on survival, recurrence, side effects, quality of life, and costs.

The promise of CER

- CER *generates and synthesizes evidence that compares benefits and harms of alternative methods to prevent, diagnose, treat, and monitor clinical conditions, or to improve the delivery of care*
- CER evidence is intended to support

  clinical and policy

  decision making

  at the individual and the population level

Implications

- ***clinical and policy***

  Covers almost all biomedical research

- ***decision making***

  Transparent decision making = Quantitative decision making.

- ***at individual and population level***

  CER is expected to generate definitive and granular information

Tenets of CER

- Focus on **effectiveness** (in contrast to **efficacy**)
- Study populations representative of clinical practice
- Focus on the **individual** rather than the average patient
- Study **two or more alternative** interventions in direct comparison
  - Multiple arm studies
  - Placebo arms rarely useful

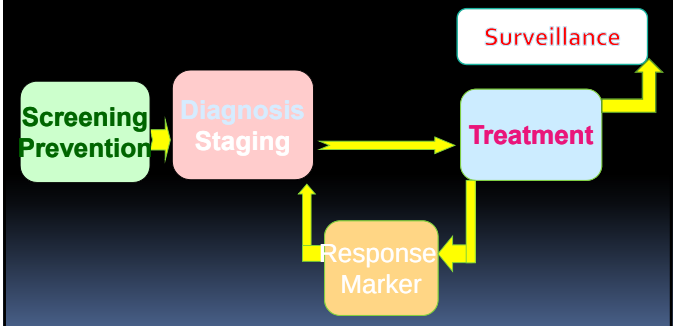
Six characteristics of CER (IOM report)

1. CER has the objective of directly informing a specific clinical decision from the patient perspective or a health policy decision from the population perspective.
2. CER compares at least two alternative interventions, each with the potential to be "best practice."
3. CER describes results at the population and subgroup levels.
4. CER measures outcomes—both benefits and harms—that are important to patients.
   - Patient Centered Outcomes Research(PCOR)

Six characteristics of CER, cont.

5. CER employs methods and data sources appropriate for the decision of interest.
6. CER is conducted in settings that are similar to those in which the intervention will be used in practice

Example: CER in cancer care



Areas for CER studies

- Prevention
- Early detection
- Diagnosis and staging
- Biomarkers for choosing and guiding therapy
- Symptom control
- Outcomes and costs of therapy
- Surveillance post therapy
- Underserved populations

Recent perspectives on CER for cancer care

CER in cancer should

- "... engage patients and the general public.."
- ".. link data from public and private entities.."
- " .. (promote) development of new research methods …"
- " (ensure that) results are translated into clinical practice .." so that ".. (it) better informs decisions made among patients, their health care providers, and payers .."
- "…support the development of personalized or stratified medicine .."
- ".. (give central role to) genomics and personalized medicine .."

Methodologist's overview of CER

- Experimental studies
  - Controlled trials
  - Randomized controlled trials
- Observational studies
  - Prospective observational studies, (e.g. registries, EMR bases studies, post-marketing safety studies)
  - Cohort and case-control studies
  - Cross sectional studies
  - Case series
- Research synthesis
  - Systematic reviews and meta-analyses
  - Modeling studies
  - Technology assessment

From CER to

P(atient) C(entered) O(utcomes) R(esearch)

pcori

Patient-Centered Outcomes Research Institute

Draft Methodology Report:
"Our Questions, Our Decisions: Standards for Patient-centered Outcomes Research"

PCORI Methodology Committee

Patient centered outcomes: PCORI definition

1. Given my personal characteristics, conditions and preferences, what should I expect will happen to me?
2. What are my options and what are the benefits and harms of those options?
3. What can I do to improve the outcomes that are most important to me?
4. How can the health care system improve my chances of achieving the outcomes I prefer?

Comment:
Patient-Centered Outcomes ≠ Patient Reported Outcomes

Standards for Patient Centeredness (PCORI)

1. **Engage Patient Informants,** Persons Representative of the Population of Interest, in All Phases of Patient-Centered Outcomes Research (PCOR)
2. Identify, Select, Recruit, and Retain Study Participants **Representative of the Spectrum of the Population of Interest** Facing the Health Decision of Interest and Ensure that Data Are Collected Thoroughly and Systematically from All Study Participants
3. Use **Patient-Reported Outcomes (PROs)** When Patients or People at Risk of a Condition Are the Best Source of Information
4. Develop and Implement a **Dissemination Assessment** to Achieve Broad Awareness of Study Results

Clinical trials in the CER era

- Continued need for the usual evaluation of therapeutic and diagnostic interventions (e.g. Phase 1,2,3 studies)
- However, CER ushers in emphasis on
  - Patient centered outcomes
  - Efficiency in study design, conduct, analysis (e.g. adaptive designs, Bayesian approaches)
  - Generalizability to "real world" setting
    - "pragmatic" or "practical" trials
    - meaningful involvement of stakeholders
  - Assessment by subgroup

Efficient designs

- No disagreement here!
- Flexible, adaptive designs
- Bayesian methods particularly suitable

Standards for adaptive designs

1. Specify Planned Adaptations and Primary Analysis
2. Evaluate Statistical Properties of Adaptive Design
3. Specify Structure and Analysis Plan for Bayesian Adaptive Randomized Clinical Trial Designs
4. Ensure Clinical Trial Infrastructure Is Adequate to Support Planned Adaptation(s)
5. Use CONSORT Statement, with Modifications, to Report Adaptive Randomized Clinical Trials

Generalizability

- Generalizability to "real world" setting
  - Study interventions in diverse institutions and care settings.
  - Will all CER studies be multi-center?
- "Pragmatic" or "practical" trials
  - How prescriptive do CER study protocols need be?
  - Is focus on studying/comparing effectiveness between *defined* clinical approaches/pathways or on observing what clinical practice is?
  - Rhetoric in the context of scientific realities
- Meaningful involvement of stakeholders

Assessment by subgroup

- Major implication on study size
- What level of statistical precision is needed? Are 70% CI's acceptable?
- Alternative clinical trial designs may accommodate some types of subgroup comparisons

Standards for evaluating heterogeneity (PCORI)

1. State the Goals of HTE Analyses
2. For Confirmatory and Descriptive HTE Analyses, Pre-specify Subgroups and Outcomes; for Confirmatory HTE Analyses, Pre-specify Hypotheses for Each Subgroup Effect
3. For Confirmatory HTE Analyses, Report a priori Statistical Power.
4. For Any HTE Analysis, Perform an Interaction Test and Report Sufficient Information on Tx Effect Estimates
5. For Exploratory HTE Analyses, Discuss Findings in the Context of Study Design and Prior Evidence
6. For Any HTE Analysis, Report All Pre-specified Analyses and, at Minimum, the Number of Post-hoc Analyses, Including Number of Subgroups and Outcomes Analyzed

Trials in community setting

- Especially well suited to evaluate effectiveness of care
  - in diverse settings and care delivery systems
  - in diverse populations and patient subgroups
- Can make effective use of
  - flexible methods for data collection and linking of sources of information
  - group randomized designs
- Ideal for promoting and enhancing meaningful involvement of patients and other stakeholders
- Essential for assessing generalizability to "real world" setting

Principles for engaging patients

1. Engagement process is transparent and includes a conflict-of-interest statement.
2. Process used to invite and select participants is inclusive and balanced in terms of ethnicity, gender, age, disease burden, and socioeconomic status.
3. Roles and relationships for researchers and lay participants are clarified at the beginning of each project.
4. Public is engaged using appropriate, validated, and diverse methods by staff experienced in PCOR or patient-centered care.
5. Process is sustainable and establishes a culture of improvement. There are measures for quality control of patient participation to ensure integrity of process of patient involvement is maintained over time and across different projects.

Beyond clinical trials

- Observational studies
  - Registry studies
  - Electronic Medical Records
  - Claims datasets
  - Cohort studies
- Research synthesis
  - Systematic reviews and meta-analysis
  - Modeling

Build new, extensive data collection Link data sources into Data Networks

- Major component of CER enterprise
- Establish and utilize Electronic Medical Record (EMR)
- Link data from public and private sources
- Important topic in this Workshop

Comparative studies of drug-eluting and bare-metal stents for acute myocardial infarction

Mauri L,…, and Normand S-L:
NEJM 2008
Circulation 2008

Prospective data collection for state-mandated database of cardiac care in all non-US Government hospitals in Massachusetts

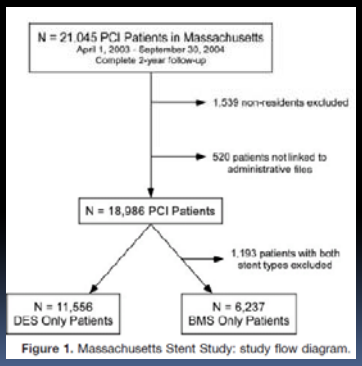


Figure 1. Massachusetts Stent Study: study flow diagram.

Analysis

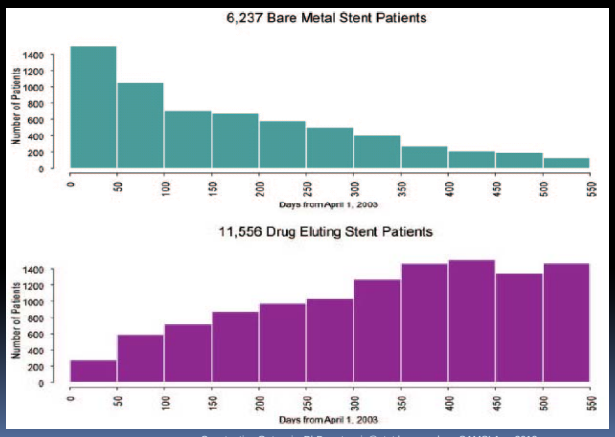
Endpoints
- Mortality
- MI
- Revascularization at 2 years

Statistical approach
- Propensity score with 63 patient & procedure characteristics
- 1-1 matched analysis
- Sensitivity analysis

Table 5. Propensity Score Risks and Matched Risk Differences at 2 Years

Outcome	DES Only (n=5549)	BMS Only (n=5549)	Percentage Difference (DES–BMS) (95% CI)	*P*
Death, % (n/N)				
30 d	1.9 (107/5549)	2.9 (160/5549)	−1.0 (−1.5 to −0.4)	0.0008
1 y	6.0 (334/5549)	8.0 (446/5549)	−2.0 (−3.0 to −1.1)	<0.0001
2 y	9.8 (546/5549)	12.0 (665/5549)	−2.1 (−3.3 to −1.0)	0.0002
MI, % (n/N)				
30 d	2.2 (121/5549)	3.2 (176/5549)	−1.0 (−1.6 to −0.4)	0.0013
1 y	5.5 (306/5549)	8.0 (442/5549)	−2.5 (−3.4 to −1.5)	<0.0001
2 y	8.3 (463/5549)	10.3 (570/5549)	−1.9 (−3.0 to −0.8)	0.0005
Target-vessel revascularization, % (n/N)				
30 d	2.2 (123/5549)	3.1 (172/5549)	−0.9 (−1.5 to −0.3)	0.0039
1 y	7.5 (414/5549)	13.7 (759/5549)	−6.2 (−7.4 to −5.1)	<0.0001
2 y	11.0 (609/5549)	16.8 (930/5549)	−5.8 (−7.1 to −4.5)	<0.0001

Market availability



Study conclusions

- Drug-eluting stenting was associated with lower rates of mortality, MI, and revascularization when compared to bare-metal stenting.
- Sensitivity analyses produced qualitatively similar results.

A cautionary note

Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods.

Stukel,T. et al: *JAMA. 2007;297:278-285*

Objective To compare 4 analytic methods for removing the effects of selection bias in observational studies:

1. multivariable model risk adjustment,
2. propensity score risk adjustment,
3. propensity-based matching,
4. instrumental variable analysis.

Approach

- National cohort of 122 124 patients
- elderly (65-84 years),
- receiving Medicare,
- hospitalized with AMI in 1994-1995,
- eligible for cardiac catheterization.
- Baseline chart reviews linked to Medicare data
- Patients followed for 7 years

Goal

Assess association between long-term survival and cardiac catheterization within 30 days of hospital admission

Study conclusions

- Estimates of the observational association of cardiac catheterization with long-term AMI mortality are highly sensitive to analytic method.
- All standard risk adjustment methods have the same limitations regarding removal of unmeasured treatment selection biases.
- Compared with standard modeling, instrumental variable analysis may produce less biased estimates of treatment effects, but is more suited to answering policy questions than specific clinical questions.

Using EMR data to construct a group randomized design

Electronic Medical Record-Assisted Design of a Cluster-Randomized Trial to Improve Diabetes Care and Outcomes

Love T. et al

J Gen Intern Med 23(4):383–91

Interventions to be compared in two health systems

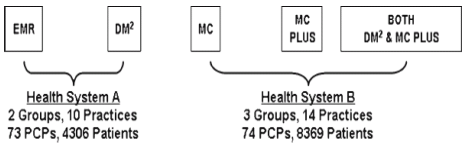
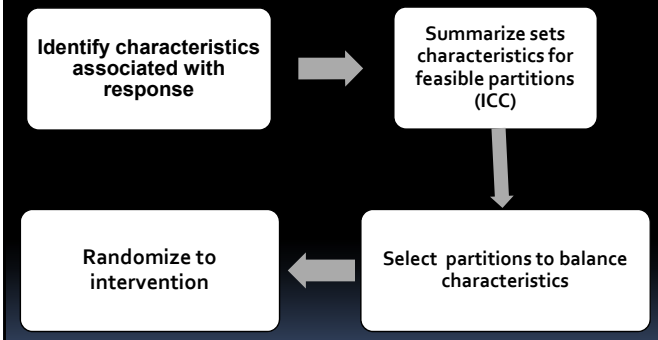


Figure 1. Schematic of DIG-IT design showing the 5 arms, within 2 health systems. *EMR* Usual clinical care using the electronic medical record with standard features, *DM²* EMR amplified by EMR-centered disease management for diabetes mellitus, *MC* EMR amplified patient opportunity to access records through MyChart® web portal, *MC PLUS* MC amplified by patient opportunity to use added diabetes-specific functionalities to aid self-management, *BOTH* DM² and MC PLUS, in combination.

Approach



Balance achieved using EMR information

Variable	System A Study Arms		ICC*	SzDiff (%)†
	DM²	EMR		
Number of practices	5	5	–	–
Number of diabetic patients	2281	2025	–	–
Female (%)	67.1	62.7	0.004	9
African American (%)	48.7	49.1	< 0.001	−1
Hispanic ethnicity (%)	9.7	10.5	< 0.001	−3
Current smoking (%)‡	25.2	22.6	0.001	6
Most recent A1c ≤7.0 (%)‡	49.0	54.6	0.006	−11
Most recent A1c ≥9.0 (%)	18.7	16.9	0.001	5
Prescription for insulin (%)	18.5	19.6	< 0.001	−3

Similarly for System B

Systematic Reviews

Network Meta-analysis or Mixed Treatment Comparisons

- Multiple interventions (therapeutic or diagnostic) are typically available for a particular clinical setting/condition.
- Primary studies may compare two or more but rarely all interventions of interest.
- Network meta-analysis (or MTC) attempts to synthesize evidence from both *direct* and *indirect* comparisons.

Lumley, Stat Med 2002

Graphical representation of three networks of randomized trials

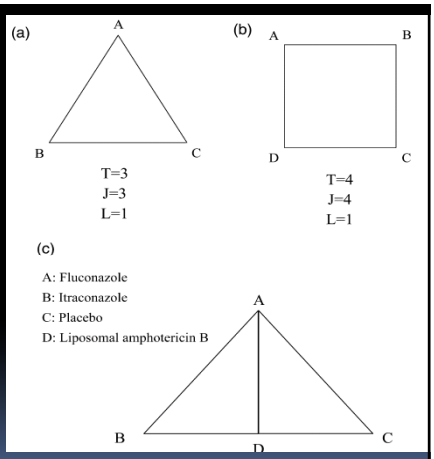
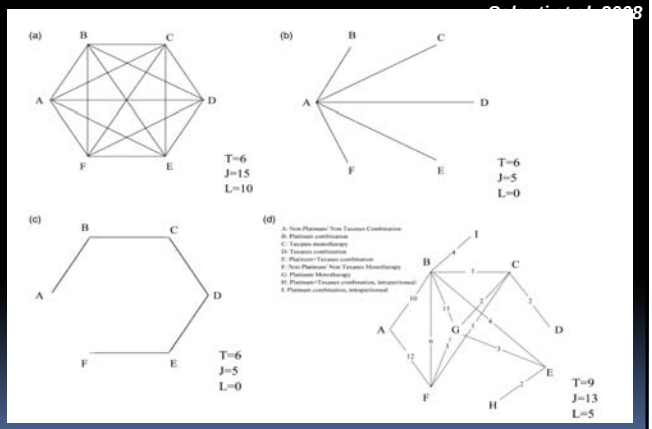


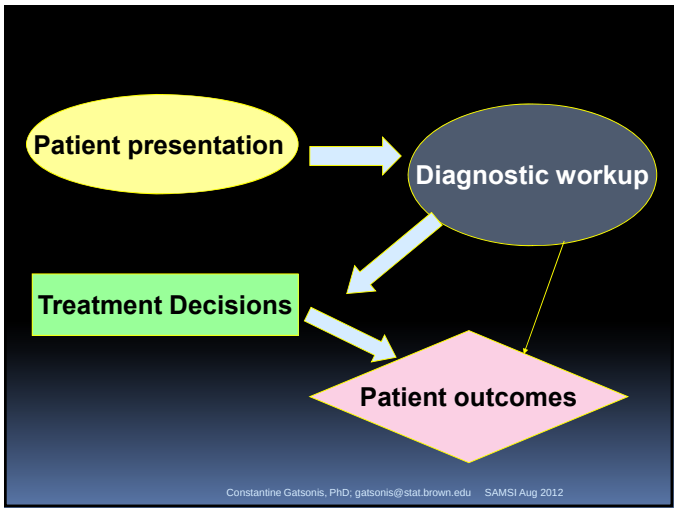
Figure from: Salanti et al, SMMR 2008; 279-301

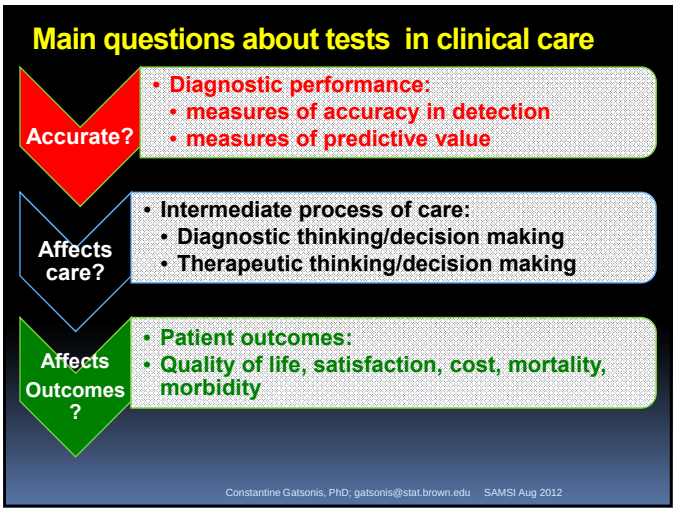
Different geometries for a network

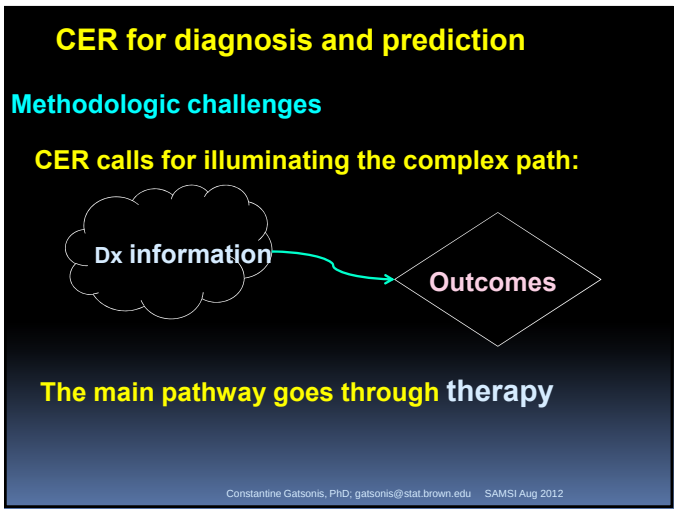


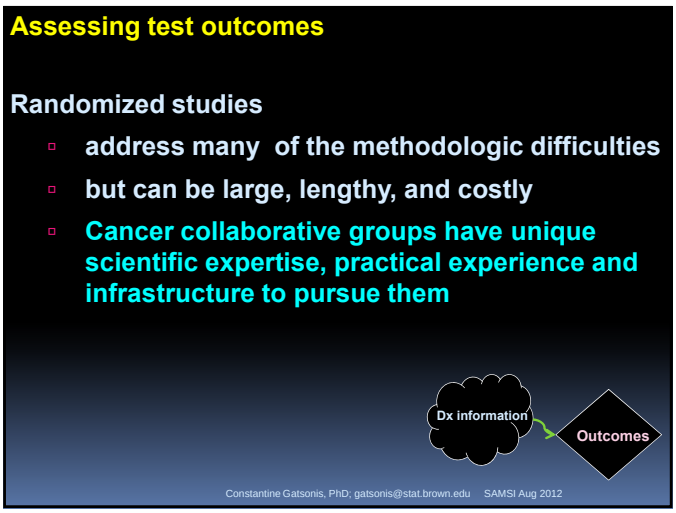
Tests and therapeutic interventions

- Fundamentally tests provide ***information*** for use in selecting course of care.
- Both long- and short-term effects of tests materialize in context of available health care options, including therapeutic interventions.
- Not possible to define and measure test effects outside the particular health care context in which the test will be used.
- However, oftentimes diagnosis may be ***ahead*** of therapy: DCIS is a good example


Patient presentation
Diagnostic workup
Treatment Decisions
Patient outcomes
Constantine Gatsonis, PhD; gatsonis@stat.brown.edu SAMSI Aug 2012


Main questions about tests in clinical care
Accurate?
• Diagnostic performance:
• measures of accuracy in detection
• measures of predictive value
Affects care?
• Intermediate process of care:
• Diagnostic thinking/decision making
• Therapeutic thinking/decision making
Affects Outcomes ?
• Patient outcomes:
• Quality of life, satisfaction, cost, mortality, morbidity
Constantine Gatsonis, PhD; gatsonis@stat.brown.edu SAMSI Aug 2012


CER for diagnosis and prediction
Methodologic challenges
CER calls for illuminating the complex path:
Dx information
Outcomes
The main pathway goes through therapy
Constantine Gatsonis, PhD; gatsonis@stat.brown.edu SAMSI Aug 2012


Assessing test outcomes
Randomized studies
address many of the methodologic difficulties
but can be large, lengthy, and costly
Cancer collaborative groups have unique scientific expertise, practical experience and infrastructure to pursue them
Dx information
Outcomes
Constantine Gatsonis, PhD; gatsonis@stat.brown.edu SAMSI Aug 2012

CER thinking has long tradition in cancer screening

- Methodologic challenges in extrapolating from accuracy and intermediate outcomes to patient-level outcomes have been studied.
- Fallacies arising from length and lead time bias have been documented and addressed.
- Long debates about benefits and harms of testing.
- Randomized studies of outcomes have become the "gold standard" .

National Lung Screening Trial (LSS/ACRIN 6654)

53,456 participants, at high risk for lung cancer.

Randomized

or

CXR screening

Baseline, plus two annual screens

Helical CT screening

Baseline, plus two annual screens

Participants followed-up in subsequent years

Primary endpoint: lung cancer mortality

Several secondary endpoints

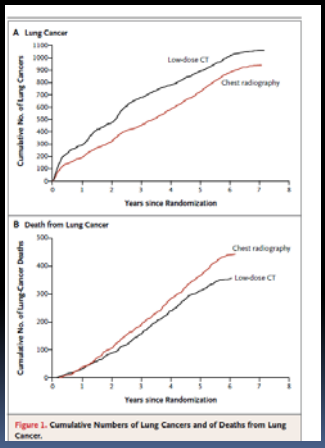


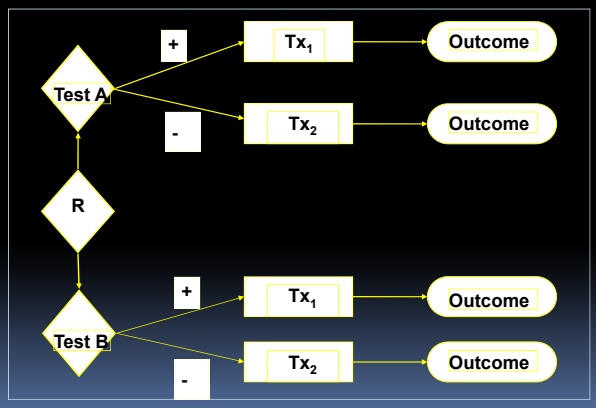
Figure 1. Cumulative Numbers of Lung Cancers and of Deaths from Lung Cancer.

The traditional randomized studies of how cancer screening affects mortality *are*

blunt instruments

- Typically used to study broad populations
- Typically framed as evaluations of public health interventions, without tight connections to specific diagnostic and intervention algorithms

Simple randomized design, comparing two tests



Studies of test outcomes can be very costly

Difference in success rates between two arms:

$$D=(r_1-r_2)\times p \times (Sens_A - Sens_B)$$

r_1 and r_2 = *success rates* for therapeutic interventions Tx_1 and Tx_2, when performed on cases that have the clinical condition (irrespectively of which test detected them)

p = prevalence of the clinical condition

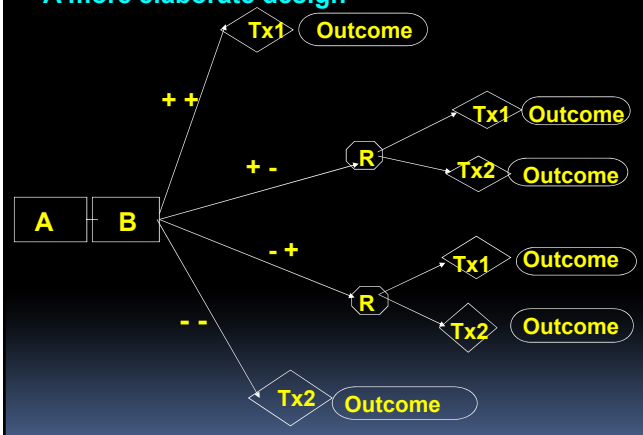
Sens= test sensitivity

Specificities are assumed equal.

Typically, D will be much smaller than r_1-r_2

If r_1-r_2 = 0 or $Sens_A = Sens_B$ then D=0.

A more elaborate design



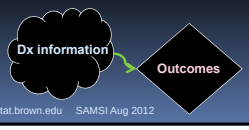
Beyond clinical trials

- Observational studies
  - Registry studies
  - Electronic Medical Records
  - Claims datasets
  - Cohort studies
- Research synthesis
  - Systematic reviews and meta-analysis
  - Modeling

Examples of CER studies in diagnosis

Observational studies

- **Registries (NOPR)**
- **Linkage of clinical trial data and secondary databases (CMS, EMR)**
- Area for current GO grant (Dartmouth, Brown, NOPR, Tufts EPC), built around ACRIN's Outcomes Committee and NOPR



National Oncologic PET Registry: A Nationwide Collaborative Program

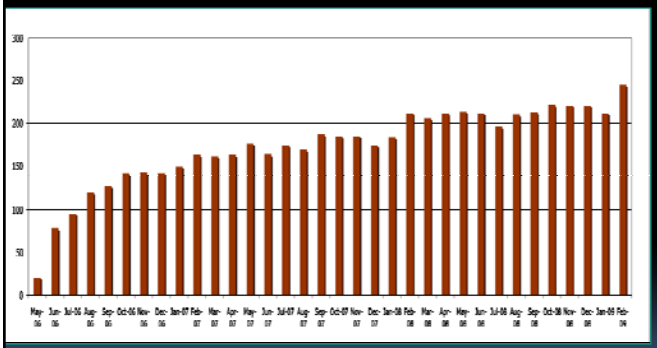
Sponsored by AMI *Advisor* CMS

Managed by ACR ACRIN

Goal:
Assess the effect of PET on referring physicians' plans of intended patient management.

- **across a wide spectrum of cancer indications for PET, not covered currently by the Medicare program,**
- **in relation to cancer-type, indication, performance status, physician's role in management, and scan type**

NOPR Accrual (Cases Completed/Business Day)



More than 200,000 patients registered

PET Changed Intended Management in 36.5% of Cases

		Clinical Indication for PET Study (Percent)				
Pre-Pet Plan	Post-PET Plan	Dx n=5,616	Staging n=6,464	Restaging n=5,607	Recurrence n=5,388	All n=22,975
Treat	Same	16.0	46.5	15.8	20.4	25.5
Non-Treat	Same	52.9	14.0	48.0	40.7	37.9

Non-Treat	Treat	23.2	31.6	28.6	29.2	28.3
Treat	Non-Treat	7.9	7.9	7.5	9.7	8.2
Patients with change post-PET (%)		31.1	39.5	36.1	39.0	36.5

Hillner et al J Clin Onc 2008

NOPR strengths

- Extensive, timely, "real world" data
- Results are consonant with those from more tightly controlled studies.
- Supports the Coverage with Evidence Decisions approach of CMS

Limitations (of NOPR and similar registries)

- Long list !
- Evidence documents change in ***intended*** management, not ***actual*** management
- Evidence not available on whether management changes were appropriate.
- Evidence not available on whether PET improved long-term outcomes
- Registry does not address how PET should be used in the flow of patient care.
- etc etc

Follow up studies to NOPR

Link registry and medical record and medical claims data

- Validate intended vs. actual management.
- Examine long term outcomes (survival, health care utilization)
- Examine regional associations between PET use and intensity of non-PET advanced imaging
- Current research by GO Grant involving Dartmouth, Brown, ACRIN and NOPR.

Standards for registries (PCORI)

1. Describe Data Linkage Plans, if Applicable
2. Plan Follow-up Based on the Registry Objective(s)
3. Describe Data Safety and Security
4. Take Appropriate Steps to Ensure Data Quality
5. Document and Explain Any Modifications to the Protocol
6. Collect Data Consistently
7. Enroll and Follow Patients Systematically
8. Monitor and Take Actions to Keep Loss to Follow-up to an Acceptable Minimum
9. Use Appropriate Statistical Techniques to Address Confounding

The role of modeling

- Modeling and simulation have significant promise as CER methodology for both therapy and diagnosis.
- In the diagnostic context, modeling utilizes and integrates information from *empirical* studies on
  - accuracy
  - course of disease
  - effectiveness of therapeutic interventions
  - patient utilities and costs
- Value of Information Analysis

Examples of modeling for diagnosis and screening studies

- Assess the impact on outcomes using findings from ACRIN trials (DMIST, NLST, NCTCC, 6666)
- Project impact of lung cancer screening (NLST)
- Assess impact of screening in various forms of cancer (CISNET)
- Value of information analysis

PCOR criteria and research questions

Table 1: PCORI Criteria

PCORI Criteria	Statuatory Language	Questions
Impact on Health of Individuals and Populations	*disease incidence, prevalence, and burden in the United States (with emphasis on chronic conditions)*	How many people are impacted by this priority area? How severe are the consequences, in terms of mortality, symptoms, adverse effects of treatment, patient experience and loss of function?
Probability of Improvability via Research	*the potential for new evidence to improve patient health, well-being, and the quality of care*	How likely is additional information in this priority area to make important improvements in patients' health status, the quality of their care, or the public's health?
Inclusiveness of Different Populations	*Research shall be designed, as appropriate, to take into account the potential for differences in the effectiveness of health care treatments, services, and items as used with various subpopulations, such as racial and ethnic minorities, women, age, and groups of individuals with different comorbidities, genetic and molecular sub-types, or quality of life preferences and include members of such subpopulations as subjects in the research as feasible and appropriate.*	Would new information in this priority area be particularly likely to increase understanding of differences in best treatments, prevention strategies, or a personalized assessment of an individual's unique biological characteristics and/or social circumstances?
Current Gaps in Knowledge/Variation in Care	*gaps in evidence in terms of clinical outcomes, practice variations and health disparities in terms of delivery and outcomes of care*	Does medical care in this area currently show wide variations in practice or clinical outcomes, suggesting a lack of clear evidence on effectiveness or a lack of awareness about this evidence?
Impact on Health System Performance	*the effect on national expenditures associated with a health care treatment, strategy, or health conditions*	Will more information in this priority area help [health care systems support] improve health care treatment or get better health outcomes for the money invested?
Potential to Influence Decision-Making	*the relevance to patients and clinicians in making informed health decisions*	Will more information in this priority area be particularly likely to help patients and clinicians address decisions that are currently difficult to make?
Patient-Centeredness	*patient needs, outcomes, and preferences*	Have patients or other key stakeholders explicitly identified a need for more research or is there a lack of resources in this priority area?
Rigorous Research Methods	*The Institute shall make available to the public and disclose ... the process and methods for the conduct of research including ... research protocols, including measures taken, methods of research and analysis, research results and such other information...*	Does proposed research or study in this priority area use or develop optimal methodologic and analytic approaches to addressing patient-centered evidence?
Efficient Use of Research Resources	*taking into consideration the types of research ... and the relative value (determined based on the cost of conducting research compared to the potential usefulness of the information produced by research)*	Will the proposed study use PCORI resources efficiently? Might it create common data or infrastructure that could support future research?

CER Challenges (partial list!)

- Variable availability, standardization, and completeness of data.
- Fragmented regulatory framework.
- Complexity in balancing research needs with effective protection of patient privacy.
- Fragmentary framework for
  - prioritizing research needs and organizing research
  - evaluating the quality and strength of available and necessary evidence

Thank you