Synthesizing Information from Different Study Designs

Dean Follmann
National Institutes of Health

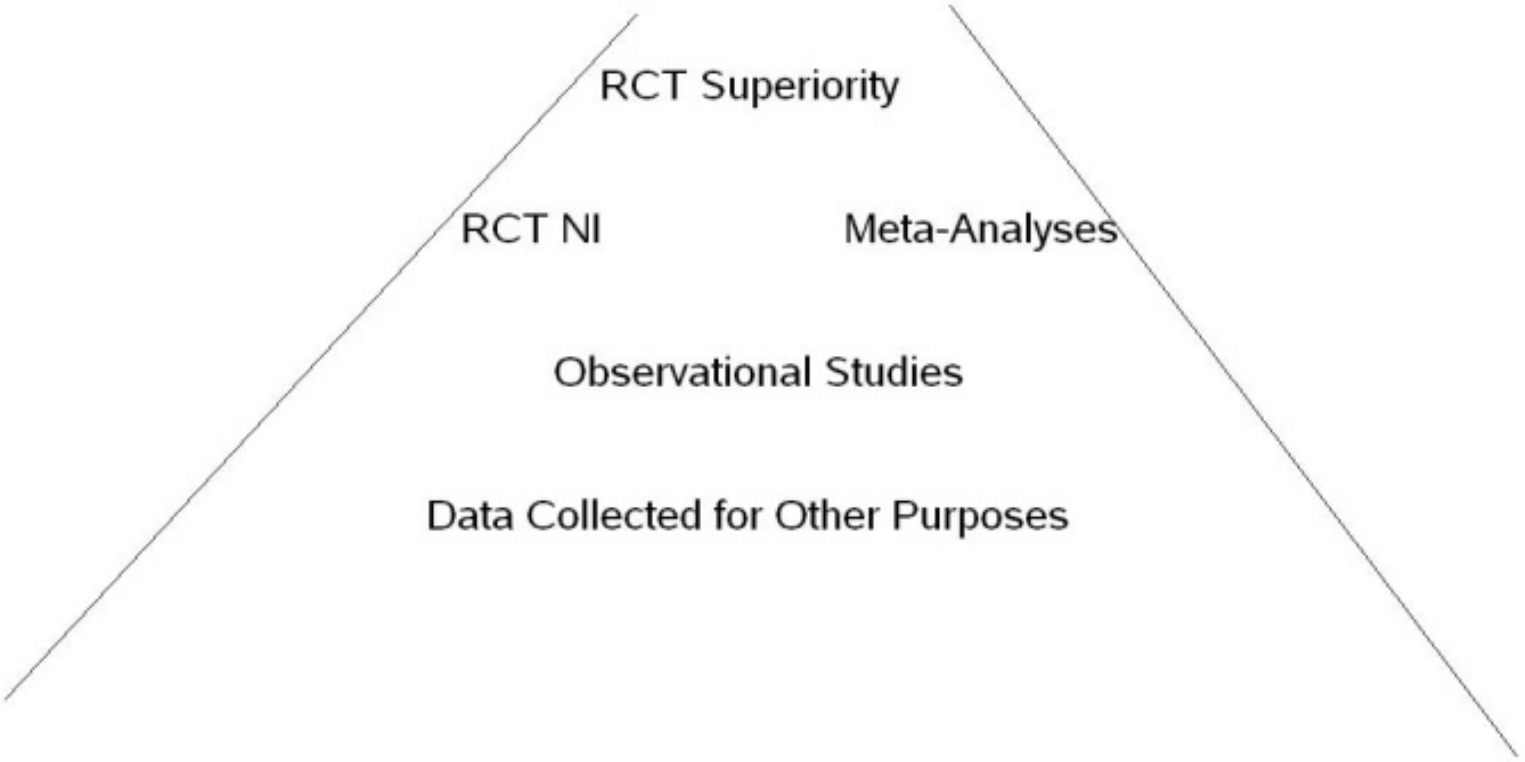
Study Designs

- Randomized double blind *superiority* clinical trial
- Randomized *noninferiority* trial
  - double blind
  - open label
- Meta-analyses of Randomized Trials
- Observational studies
- Data collected for other purposes

Endpoints

- Clinical endpoint: a measure of how a patient feels, functions, or survives, e.g. MACE, microvascular events.
- Surrogate endpoint: an endpoint chosen to stand in place of a clinical endpoint, e.g. HbA1c.
  - Depends on drug given (Muraglitazar)
  - Depends on level: 10 -> 8.5 v 7 -> 5.5 (ACCORD)

Study Hierarchy



Less reliable studies can suggest/support hypotheses

More reliable studies settle hypotheses

RCT double blind superiority trial

- Greatest medical invention ever.
- Randomization ensures similar groups at start
- Double blinding ensures a level playing field
  - Can't favor one arm over another
- Incentives encourage good study conduct
  - Sloppiness makes arms more similar.

RCT NI trial

- Extremely different paradigm.
- Goal is to conclude RSG is not ``unacceptably worse" than comparator.
- Requires a margin
- HR CI = (.90, 1.25) < 1.3
  - HR = Event rate on RSG/Event rate on comparator
  - Formally conclude any potential increase in RSG risk is less than 30%

Concerns with NI Trials

- Hard to think about a margin 20% 30% ?
- Incentive is to show two arms are similar.
  - Designing a trial involves a thousand decisions
- Rigor in conduct paramount
- These concerns are increased in an open label trial.

Open Label NI Trial

- Investigators may subconsciously favor one arm
- RECORD trial
  - Of 549 CRFs reviewed:
    - 10% had ``problems'' favoring RSG
    - 2.5% had ``problems'' favoring control
  - Statin initiation before trigger reached
    - 31% of RSG group start statin with LDL <130
    - 25% of control start statin with LDL <130
- Are these consequential?
- Anything else going on?

Meta-Analyses

- A quantitative synthesis of RCTs. But quality of evidence is a bit less.
  - Negative studies may not be published.
  - Studies can be heterogeneous: population, endpoint, comparator, other therapy
  - Analysis decisions made with some knowledge of results.
    - What to include, how to analyze, which to lump, endpoint definition

Rosiglitazone Meta-Analysis

- Large number of studies, but weren't designed for CV outcomes. Events are rare.
- Heterogeneity is an issue
  - Placebo control, Baseline nitrate use
  - Hard to know if heterogeneity is real
  - Like to test if OR's the same
- Nissan & FDA & GSK, conclusions differ. Why?

Why

GROUP	MI RESULT 95% CI	#	Events	Outcome
FDA	OR: 1.80 (1.03, 3.25)	52	65	MI
Nissen	OR: 1.28 (1.02, 1.63)	56	83+212	MI
GSK	HR: 1.098 (0.89, 1.35)	52	367	MI Serious + nonserious AE

GROUP	MACE RESULT 95% CI	#	events	Outcome
FDA	OR: 1.44 (.95, 2.20)	52	109	MACE
GSK	HR: 1.12 (.79, 1.59)	52	131	MACE + MACE components

GSK: WIDER endpoint FDA, Nissen: narrower endpoint

Nissen: Had Dream, Adopt, RECORD, #100684 large studies less signal on MI

Analysis: GSK one big trial Nissen FDA fixed effects Meta-Analysis

FDA Meta-Analysis

- Pio MACE OR: .85 Rosi MACE OR: 1.44

Pioglitazone	Comparator	Rosiglitazone
?	Placebo	X
?	Placebo	X
X	Active	?

- Rosiglitazone studies different
  - More placebo controlled, fewer therapy naïve, shorter followup
  - Difficult to compare if ORs depend on such factors
  - Can statistically try to level the playing field.

Women's Health Initiative
Effect of Estrogen on Breast Cancer

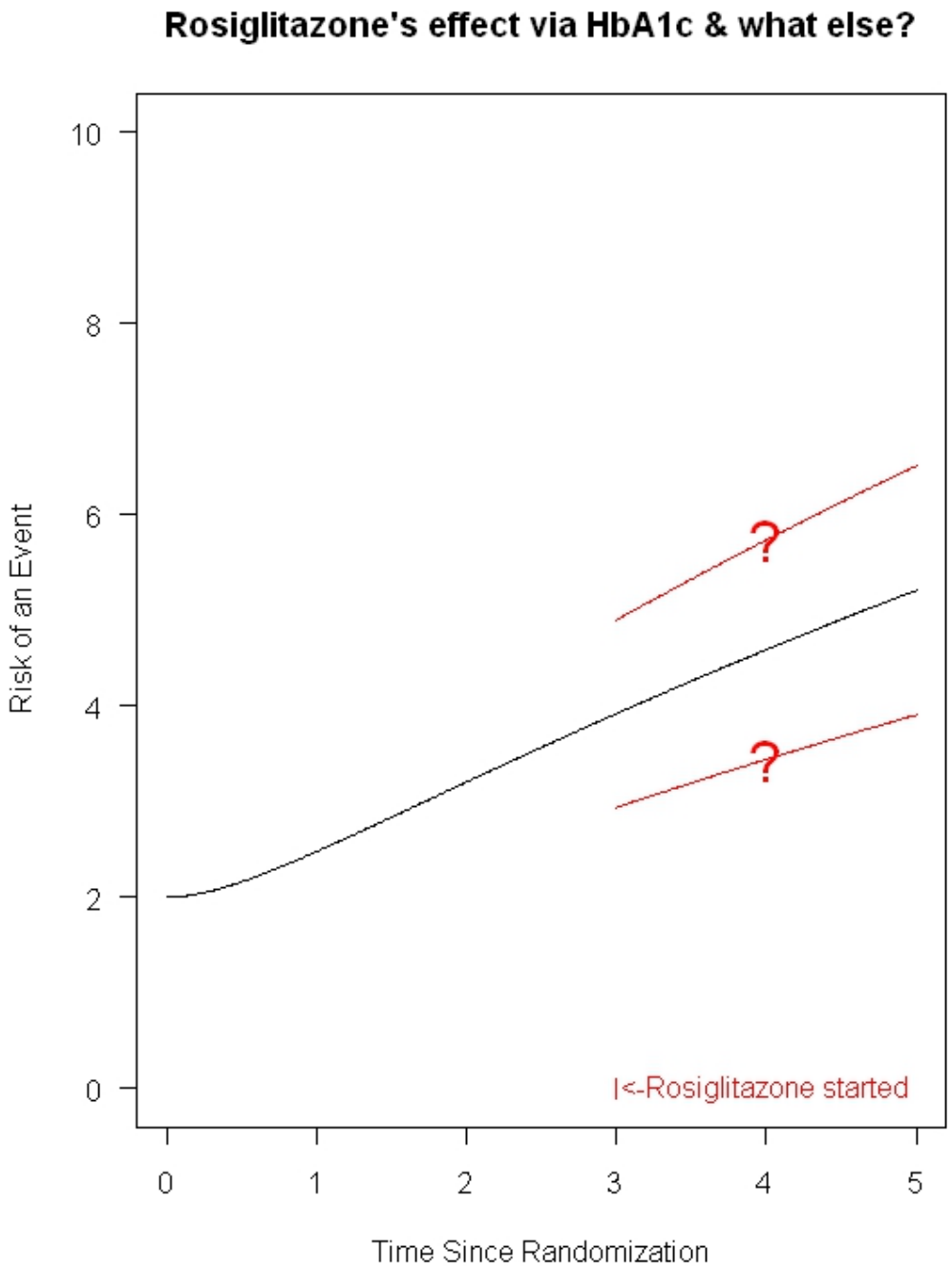
- WHI Observational Study (OS) HR=1.28
- WHI Clinical Trial (CT) HR= .71
  - OS HR 80% larger than CT HR ($p<.01$)
- Adjust for screening, estrogen dose, time from menopause to estrogen.
  - OS HR 7% larger than CT HR ($p=.82$)
- Imbalances can create spurious differences
- Statistical adjustment can make them go away (or give them more credence).

Observational Studies

- Decision to give drug typically based on
  - Patient characteristics, doctor choice, milieu, uncollected/unquantifiable factors.
- Statistical analysis of such studies need to adjust for nonrandomized assignment e.g.
  - Matching of cases & controls
  - Assume within homogeneous strata (e.g. Men with HbA1c 8.0%-8.5%) choice is as if randomized.
- Such adjustment less reliable than a RCT.

Observational Study within a RCT

- RCTs designed to compare two HbA1c targets will have volunteers initiating rosiglitazone post–randomization
- Statistical analyses can estimate the increase/decrease in risk following rosiglitazone initiation.
- Risk change is relative to those who survive up to the time of rosiglitazone initiation.



Curves are illustrative
Not based on data

Interpretation is difficult

- Rosiglitazone initiation is not randomized.
- Rosiglitazone initiators probably have worse HbA1c compared to their fellow survivors
  - Anticipate a worse risk
- Rosiglitazone initiators should get a 1-2% HbA1c benefit compared to their fellow survivors
  - Anticipate a better risk
- Can further try to adjust for current HbA1c, but this is hard.

Observational Studies

- Different outcomes & heterogeneity make quantitative summaries problematic.
- Qualitative summaries e.g. 8 of 8 favor A may overstate the case
  - Drug assignment ``biases" may apply across different studies.

Data Collected for Other Reasons

- Health organization databases have descriptions of what goes on with their patients.
  - pharmacy dispensing data.
  - MI determined by a diagnostic code
- Quality control, standardization, adjudication, unlike RCT.
- Unclear how drug choice is decided.
- Difficult to known if modest effects are due to drug.

TIDE

- TIDE has appeal: N=16,000 to rule out a HR of 1.3. Large double-blind, placebo controlled trial powered for MACE.
- But it is a NI study. Possible conclusion: RSG has at most a ``modest" < 30% increase in MACE compared to placebo.
- Would the decision today be the same no matter what TIDE shows?

Perspective

- Is the body of evidence sufficiently different from 2007 to change that decision?
- Public Health. Should we think of the overall consequences of a market w/ versus w/o rosiglitazone?
  - Unintended consequence---no generic drugs?
- Drug Umpire. For rosiglizazone is the evidence clear-cut to approve or disapprove?

Summary

- RCTs are the gold standard especially when trying to decide on modest effects.
- NI trials need to be conducted with utmost rigor.
- Meta-analysis of good RCTs provide strong evidence but not quite as good as prospective RCT.
- Observational studies are good to raise questions but not to reliably settle ones involving modest effects.