

# Five challenges of using nonrandomized studies in systematic reviews of effectiveness of interventions

**Maria Ospina; BSc, MSc<sup>\*, +</sup>**

**Brian H. Rowe; MD, MSc, CCFP(EM), FCCP<sup>\*, ¥</sup>**

**\*University of Alberta Evidence-Based Practice Center**

**+Alberta Research Centre for Child Health Evidence**

**¥Department of Emergency Medicine, University of Alberta**



# Outline

- Introduction.
  - 20 minutes.
- Group Discussions.
  - 5-10 minutes.
- Presentations.
  - 5 minutes/group.
- Summary/Questions.

# Acknowledgements

- Dr. Terry Klassen; Director, EPC and ARCHE;
- Lisa Hartling; EPC and ARCHE;
- Dr. Finlay McAlister; Division of General Internal Medicine, University of Alberta;
- EPC/ARCHE staff.

## Evidence support:

- CADTH:
  - ED Overcrowding.
- AHRQ Evidence Reports:
  - Meditation;
  - Cardiac resynchronization therapy (CRT);
  - Implantable cardioverter defibrillator (ICD).
- Existing literature.

# Systematic reviews

- Definition.
- Therapy >>> diagnostic.
- Quality of SRs.
- Therapy ~ RCT (Cochrane approach).
- Selected concerns (relative to this workshop):
  - Real world vs. sterile environment;
  - Variable reporting of outcomes;
  - Poor reporting and low incidence of SEs.

# Why to include nonrandomized studies in SRs of effectiveness?

- RCTs are unnecessary, impossible or unethical or infrequently performed (surgery) for many research questions.
- RCTs are not available and decisions still need to be made by clinicians and policy makers using available “evidence”.
- Nonrandomized studies are often used to evaluate diagnostic and prognostic interventions.
- Nonrandomized studies are used to evaluate the harm of interventions (especially rare/unexpected events).
- **The funding agency/boss/supervisor/other parties insist!!!**

# Personal experiences

- **University of Alberta Evidence-Based Practice Center**
  - Interventions to reduce overcrowding in the emergency department;
  - Meditation in health care;
  - Cardiac resynchronization therapy;
  - Implantable cardioverter defibrillator.

**Other???**

**What do you think are the challenges of including nonrandomized studies in SRs of effectiveness?**

# Audience Response

- Quality of the observational studies (OS);
- Heterogeneity in OS populations;
- Definition of outcomes vs. RCT's;
- Defining the intervention in OS vs. RCT's;
- Adjusting for confounders in OSs and eliminating bias;
- Methods for subgroup analyses;
- Minimum criteria for OSs designs;
- Searching for OSs;
- Primary vs. secondary outcomes (efficacy vs. effectiveness);
- Synthesizing the results from OSs;
- Assessing for the selection bias in individual OSs.

# Some of the challenges we have found.....

- **CHALLENGE #1:** What study design terminology to use?
- **CHALLENGE #2:** When to include nonrandomized studies?
- **CHALLENGE #3:** How to search the literature on nonrandomized studies?
- **CHALLENGE #4:** How to assess and report on the quality of nonrandomized studies?
- **CHALLENGE #5:** How to combine data across all study designs? How to deal with heterogeneity and bias?

# CHALLENGE #1: Study design terminology

- The terminology to describe nonrandomized studies in the clinical research literature is inconsistent.
- No single taxonomy of study design.
- Reviewers often do not clearly state the designs to be included:
  - *“we will include longitudinal/prospective studies, cross-sectional studies, case-control studies, and observational studies”* (from a Cochrane protocol).
  - *“prospective investigations, retrospective cross-sectional studies, and before-after studies are of interest. Other single-subject case studies and studies without control groups are not of interest”* (from an AHRQ request for Task Order)
- What does the term “Non-randomized studies” mean?
  - CCTs?;
  - Non-experimental studies?;
  - Observational studies?.

## **CHALLENGE # 2:** When to include nonrandomized studies in a systematic review?

- No established guidelines on when and how to include nonrandomized studies in a SR.
- Sometimes nonrandomized studies are included in the review, but most of the evidence is derived from RCTs.
- Sometimes nonrandomized studies are not considered when RCTs are unavailable. The conclusion is that **THERE IS NO EVIDENCE** on the effectiveness of the intervention.

# How often are nonrandomized studies included in EPC and HTA reports?

## AHRQ Evidence-based reports (N = 107)

- Report includes  $\geq 1$  clinical effectiveness question (N = 78)
- Included RCTs only (N = 27)
- Included nonrandomized studies (N = 49)
- No studies found (N = 2)

## CADTH Technology reports (N = 107)

- Report includes  $\geq 1$  clinical effectiveness question (N = 87)
- Included RCTs only (N = 37)
- Included nonrandomized studies (N = 29)
- No info on studies/pure economic evaluations/SRs (N = 21).

# CHALLENGE # 3: How to search the literature on nonrandomized studies?

- Searching for nonrandomized studies is more problematic and there has been little methods research published research to date.
- Inefficient use of valuable resources in terms of time involved in screening the titles and abstracts of a large number of irrelevant records.
- Indexing terms are less well established than for RCTs (Haynes 1994). and, when they do exist, they are used inconsistently (Pilkington 2004).
- The uncertainty in identifying appropriate search terms for nonrandomized evidence has meant that a methodology component is often excluded from search strategies.

Haynes et al. J Am Med Inform Assoc. 1994, 1(6):447-58  
Pilkington K et al. J Alt Compl Med 2004, 10:587-90

# CRT example - Search results

	2004	2006 (with non-RCTs)
Number of records identified	3405	7110
Potentially relevant articles	178	530
Included	17	128
RCTs	8	14
Non-RCTs	9 cohorts	114 all designs

## **CHALLENGE # 3:** How to search the literature on nonrandomized studies?

- Fraser et al (2006): Assessment of NICE-commissioned HTA and SR by October 2005 (77 HTA from NICE, seven SRs).
- 36% of the HTA and 100% of the SRs included nonrandomised evidence.
- 89% of these reports used no methodology filter in their search strategies.

Fraser et al. BMC methodology. 2006-Dec;1(6):447-58

## **CHALLENGE # 4:** How to assess and report on the quality of nonrandomized studies?

- Is the quality assessment of nonrandomized studies more difficult than the quality assessment of RCTs?
- Quality assessment methods for observational studies are in early development compared to RCTs.
- Although several assessment scales and checklists exist, none of them have been fully validated or shown to include criteria that are associated with the effect size (outcome) in empiric studies.

# Quality assessment of nonrandomized studies

- **Newcastle-Ottawa Scale, unpublished (8 items)** (*Cochrane Non-Randomized Studies Methods Working Group*);
- **Downs and Black, 1998 (27 items)**;
- **Cowley, 1995 (13 items)**;
- **Reisch, 1989 (57 items)** (*Cochrane Inflammatory Bowel Disease Group*);
- **Thomas, unpublished (21 items)**;
- **Zaza, 2000 (22 items)**; (*Guide to Community Preventive Services: Systematic Reviews and Evidence-based Methods*)
- **TREND statement, 2004** (Transparent Reporting of Evaluations with Nonrandomized Designs) (22 items);
- **Scottish Intercollegiate Guidelines Network, unpublished (26 items-cohort studies, 23 items-case control studies)**;
- **Tooth, 2005 (33 items)**.

## **CHALLENGE # 5:** How to combine data across all study designs? How to deal with heterogeneity and bias?

- Confounding and bias are major concerns with nonrandomized studies and one of the main reasons why they are considered “weaker designs” and excluded from reviews.
- How to address potential confounding from the primary studies in the SR?
- Heterogeneity due to differences in study designs.
- Lack of consistency in the results of experimental and observational studies?

Stroup, et al. JAMA 2000;283:2008-12

# A classical example.....

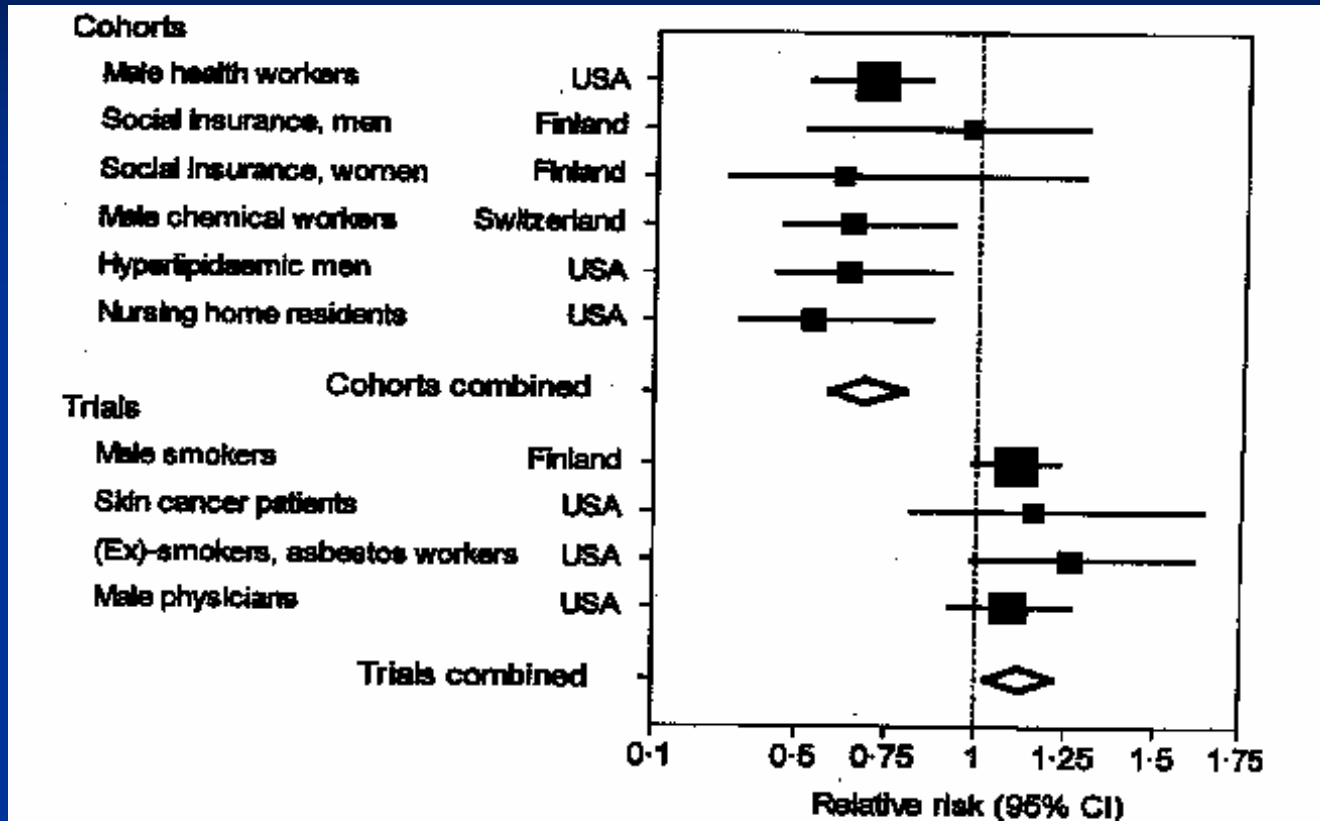
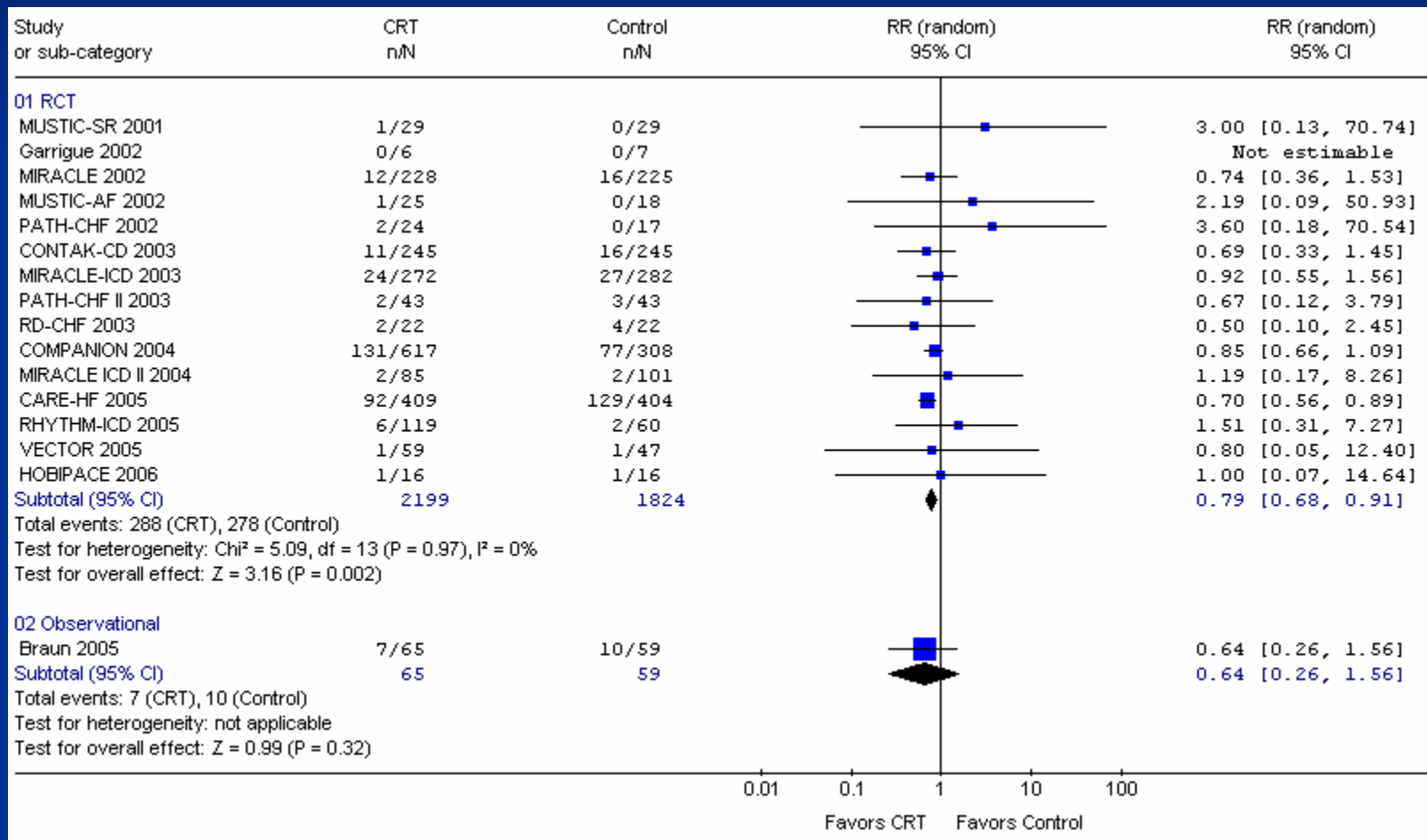


Figure 12.3 Meta-analysis of the association between beta-carotene intake and cardiovascular mortality: Results from observational studies<sup>36</sup> indicate considerable benefit whereas the findings from randomised controlled trials show an increase in the risk of death.<sup>37-40</sup> CI: confidence interval.

Egger et al. Systematic reviews in health care. London: BMJ books, 2001

# Example – CRT (mortality)



# CRT-Safety

	<b>RCTs</b>	<b>Observational</b>
<b>Peri-implant death</b>	0.4% [0.1, 1.0] <sup>4</sup>	0.3% [0.1, 0.6] <sup>24</sup>
<b>Infection</b>	1.3% [0.5, 2.7] <sup>1</sup>	1.9% [1.3, 2.5] <sup>30</sup>
<b>Lead problems</b>	6.4% [5.1, 7.9] <sup>6</sup>	6.6% [5.8, 7.4] <sup>32</sup>

# Problems of combining data across RCTs and Nonrandomized studies

## RCTs

- The exposed and unexposed groups tend to be comparable with respect to confounding variables.
- Combining crude measures of effect (ORs, RR, WMD/SMD) seems reasonable.

## Nonrandomized studies

- Comparability in nonrandomized studies is suspect because of confounding.
- Differences between studies in the adjustment of confounders.
- Selective report of the most significant confounders in the model.
- The adjusted measure of effect may be different from the crude one.
- The crude measure of effect is “naïve” and biased.
- Combining adjusted measures of effect may be difficult.

# RESULTS FROM THE WORKSHOP



**Question #1: In which situation are nonrandomized studies helpful in addressing questions of the effectiveness of interventions?**

## **Audience Response**

- **When there are no RCTs available;**
- **When we have a profound understanding of the underlying mechanisms of the disease;**
- **Ethical considerations sometimes preclude an RCT;**
- **When we are examining long-term effects;**
- **When we are examining rare outcomes;**
- **Underreporting of studies with negative results. Need to look at studies other than RCTs.**

# Q1: When nonrandomized studies are helpful in addressing questions of treatment effectiveness in SRs?

- Areas where it is difficult to conduct RCTs.
- Examining long-term outcomes of treatments assessed in trials.
- Exploring applicability of findings from RCTs.
- Clarifying outcomes for patients.
- Addressing policy issues raised by nonrandomized studies.
- Clarifying research priorities.

Norris et al, Ann Intern Med 2005; 142:1112-9

**Question #2: If your funding agency request the inclusion of “prospective studies” in your review, what designs would be eligible for inclusion? What approaches would you use to incorporate nonrandomized studies into a SR?**

## **Audience Response**

- **Prospective cohort studies.**
- **Second part of the questions was not answered due to lack of time.**

## Q2. Approaches to incorporate nonrandomized studies into SRs

- “Best evidence” approach: reviewers restrict the review to “controlled studies” when 5 or more such studies addressed a specific question and expand the review to include other designs when too few trials are identified.
- Decision is *a priori*, but reviewers decide to broad or narrow the inclusion criteria over the course of the review.
- To expand inclusion criteria on the basis of a more subjective assessment of the evidence from trials.
- To drop plans of including nonrandomized studies after numerous RCTs are identified.

Norris et al, Ann Intern Med 2005: 142:1112-9

**Question #3: If you are going to include nonrandomized studies into a SR of effectiveness, what would you do to preserve the sensitivity and increase the specificity in the literature search?**

## **Audience Response**

**IT DEPENDS on...**

- **The nature of the SR topic;**
- **The clarity regarding study designs of what might be used;**
- **Close collaboration between content experts and research librarian;**
- **Multi-step process;**
- **Feedback with content experts;**
- **Repeat searches until appears successful;**
- **Broad search to increase sensitivity;**
- **Use filters/strict keywords.**

## Q3. Searching the literature to identify nonrandomized studies

- To establish a reference standard as the set of relevant records against which filters are assessed. (Jenkins 2004):
  - Hand searches of a set of journals;
  - Screening titles and abstracts from a subject-only search;
  - Development and validation of filters.
- Web of science searching.
- Careful literature search planning.
- Reality check!

Jenkins M. Health Info Lib J. 2004; 21:148-63

Question #4: How would you incorporate quality assessment of nonrandomized studies into a SR of effectiveness?

## Audience Response

- Select an appropriate tool (validated?);
- Use an instrument different than tool for QA RCTs;
- Need to decide *a priori* whether to include or exclude OS based on quality;
- Use 3 tools and look for congruence;
- Pooling a sensitivity analysis to quantify the amount of heterogeneity;
- Need to be transparent and describe studies included;
- To ensure internal and external validity.

## Q4. Incorporating quality assessment of nonrandomized studies into a SR

- Avoid reporting of summary quality scores. Unreliable, and hard to interpret (Juni 1999).
- Avoid appraising nonrandomized studies with checklists or scales designed for RCTs (Juni 1999).
- Address methodological issues particular to nonrandomized studies (biases, confounding, and differences in study designs).
- Examine the importance of individual quality components (and rated as 'met, partially met, not met') (Deeks 2003).
- Evaluate the influence of quality on effect estimates (sensitivity analysis stratifying by criteria met or not met).
- Calculate separate summary effect estimates for studies that meet and do not meet the individual quality criterion.
- Meta-regression.

Juni et al. JAMA 1999; 282:1054-60  
Deeks et al. Health Technol Assess 2003; 7:1-173

**Question #5: How would you plan the analysis of a SR of effectiveness that includes both RCTs and nonrandomized studies?**

## **Audience Responses**

- **Mapping heterogeneity/visual display;**
- **Group visualization by population;**
- **Potential subgroup analysis.**

## Q.5. Planning the analysis of a SR that includes both RCTs and nonrandomized studies

- Subgroup analysis by study design: To understand and quantify sources of variability in results across studies (Egger 1998).
- Summary results from both RCTs and nonrandomized studies should be presented separately for each of these two broad categories.
- Stratification should be done further within nonrandomized studies by study design (ie., cohort, case-control, cross-sectional, etc) to explore potential heterogeneity due to study design.
- Discuss the consistency, or lack of consistency, in the results of experimental and observational studies.

Egger et al. *BMJ*. 1998;316:140-144

# Conclusions

- Not all nonrandomized studies are created equally or classified in a consistent fashion.
- Care should be taken when including nonrandomized studies:
  - Nonrandomized studies should be considered for SRs when large numbers or mega-trials are not available.
- Literature searching requires expanded terms (and resources) when nonrandomized studies are to be included (Reality check required).
- Some quality assessment seems appropriate; however, very little evidence exists to guide the researcher.
- Separate pooling of RCT and nonrandomized studies should be considered.

# Thank you!

Questions?