

of database research without transmitting patient-level  
databases



2

- ❖ PI, Harvard-Brigham & Women's Hospital Drug Safety Research Center (FDA)
- ❖ Chair, Methods Core of the FDA Sentinel System
- ❖ Member, national PCORI Methods Committee
- ❖ Consulting in past year:
  - WHISCON LLC, [Aetion Inc. \(incl. equity\)](#)
- ❖ Investigator-initiated research grants to the Brigham & Women's Hospital from Novartis, Genentech, Boehringer Ingelheim
- ❖ Grants/contracts from NIH, AHRQ, PCORI, FDA, IMI

## 3

## 4

## 6

## 1

# Reliable Causal Analyses

PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2008; 17: 260-268  
Published online 17 September 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pds.1471

SPE COMMENTARY

Guidelines for good pharmacoepidemiology practices (GPP)<sup>1</sup>

EUROPEAN MEDICINES AGENCY  
SCIENCE · MEDICINES · HEALTH

18 June 2011  
EMA/100000/2011 Rev.2  
European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP)

GRACE Principles: Recognizing High-Quality Observational Studies of Comparative Effectiveness

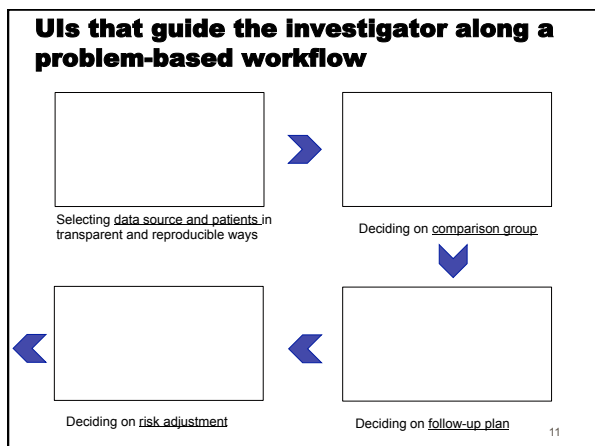
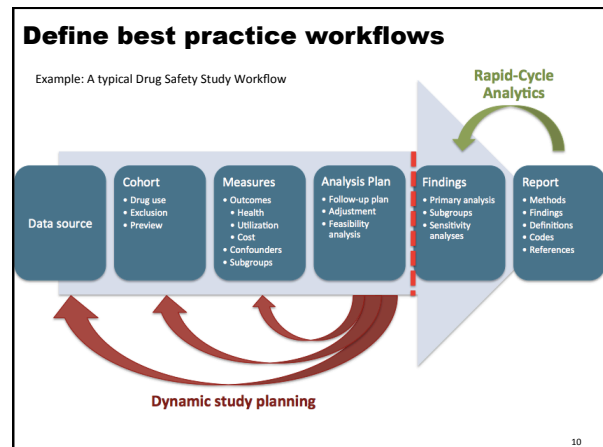
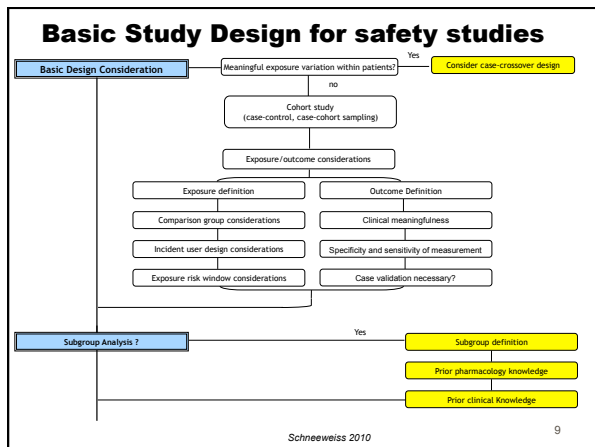
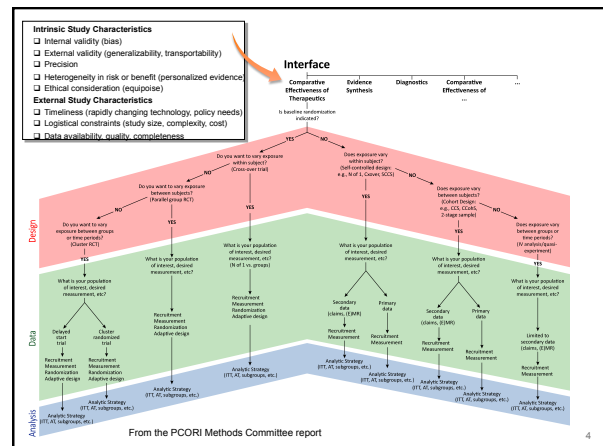
The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP)

Guide on Methodological Standards in Pharmacoepidemiology (Revision 2)

Guidance for Industry and FDA Staff  
Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data

Nancy A. Cross, PhD; Sebastian Schneeweiss, MD; Barbara J. Milrod, MD; Marc L. Berger, MD; Alan M. Walker, MD; Daniel A. Olfendick, MPH; and Richard E. Glatch, MD, for the GRACE Initiative

7



Define the analytic approach (ITT vs. AT), covariate identification period, follow-up time period, censoring etc....

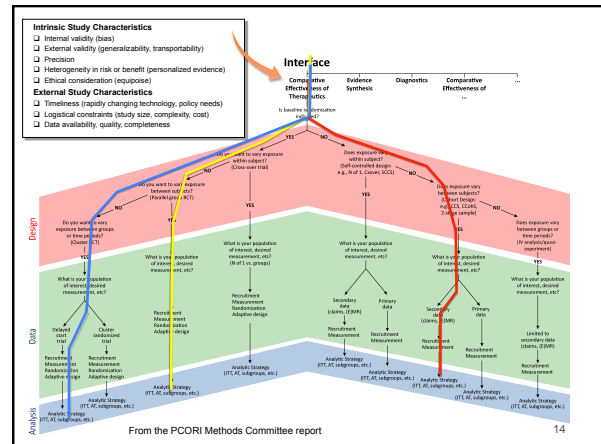
CONFIDENTIAL

12

Specify the outcome model for the primary and secondary outcomes, propensity score matching, trimming, stratifying approaches...

CONFIDENTIAL

13

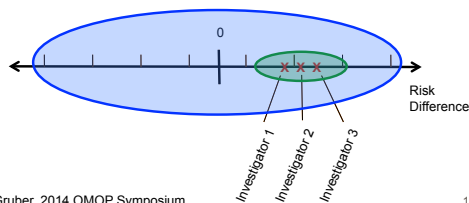


14

## Avoiding obviously wrong choices will reduce heterogeneity of results

Limited heterogeneity from valid design and analysis choices

Extreme and unnecessary heterogeneity from invalid choices (eg some of OMOP's choices)\*



\* Susan Gruber, 2014 OMOP Symposium

15

## 3) Make Studies Reproducible

16

## Reproducibility in Epidemiology

TABLE 2. Results from examining the epidemiologic literature: articles from the *American Journal of Epidemiology* and the *Journal of the American Medical Association* published between January 2005 and May 2005

	No. of papers
Total papers collected	90
Observational studies	69
Cross-sectional	20
Case-control	20
Cohort	29
Source of outcome data	
Original study	31
Ongoing study	29
Government	8
Other	1
Statistical analysis implementation	
Not reported	21
By hand	0
Use of software package	48
Method of processing measured data	
Not reported	43
By hand	1
Use of software package	13
Outcome data reported to be available	0
Exposure data reported to be available	1
Code for statistical analysis available	0
Code for processing measured data available	0

Peng et al. 2006

17

## Reproducibility in Epidemiology (not Replicability)

TABLE 3. Making results from the National Morbidity, Mortality, and Air Pollution Study reproducible\*

Research component	What we have done
Data	The entire NMMAST database is available to the public via the IHAPSS website and the NMMAST data package for R; the data are available under a "full access" class of license.
Methods	A full compendium written in L <sup>A</sup> T <sub>E</sub> X and R is available for download.
Documentation	We have outlined our data-processing pipeline on the IHAPSS website, and papers/technical reports are available for download.
Distribution	We use the World Wide Web to disseminate our data and software.

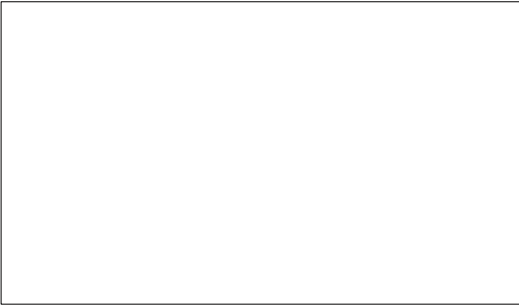
\* Details at <http://www.biostat.jhsph.edu/~rpeng/reproducible/>.

† NMMAST, National Morbidity, Mortality, and Air Pollution Study; IHAPSS, Internet Health and Air Pollution Surveillance System.

Peng et al. 2006

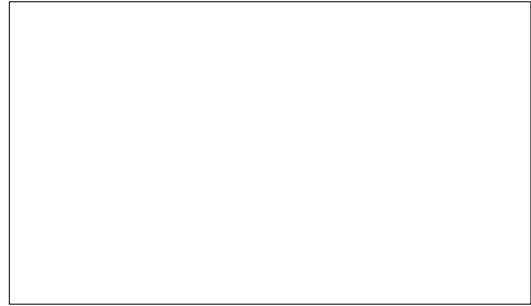
18

### Shared cloud-based analytics



19

### Solution for investigators that cannot store data on their cloud:



20

Every analysis generates a comprehensive and readable report that allows 100% reproduction of the research ...

#### Methods

p3-13

#### Results

p14-30

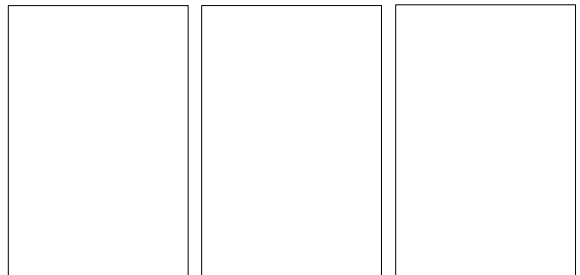
#### Appendix

p31-65



21

... by providing all details regarding coding and methods



22

### Good News

- ❖ Reliability of database research can be improved through structured approaches
- ❖ Reproducibility can be achieved if
  - We completely and precisely record all choices made during design and analysis
  - We share analytic code (R, SAS, etc.)
  - We share data
- ❖ Sharing the analytic environment gets around the inability to freely share most healthcare databases

23