How to get the best out of large databases

Murray Stewart GSK

Large Databases – How to get the best from them

- Why do we need them
- What databases are available
- What is the best use of databases within the cardiovascular area
- What are the advantages and disadvantages

Why we need to explore data from large observational databases – RCT's can't provide all the answers

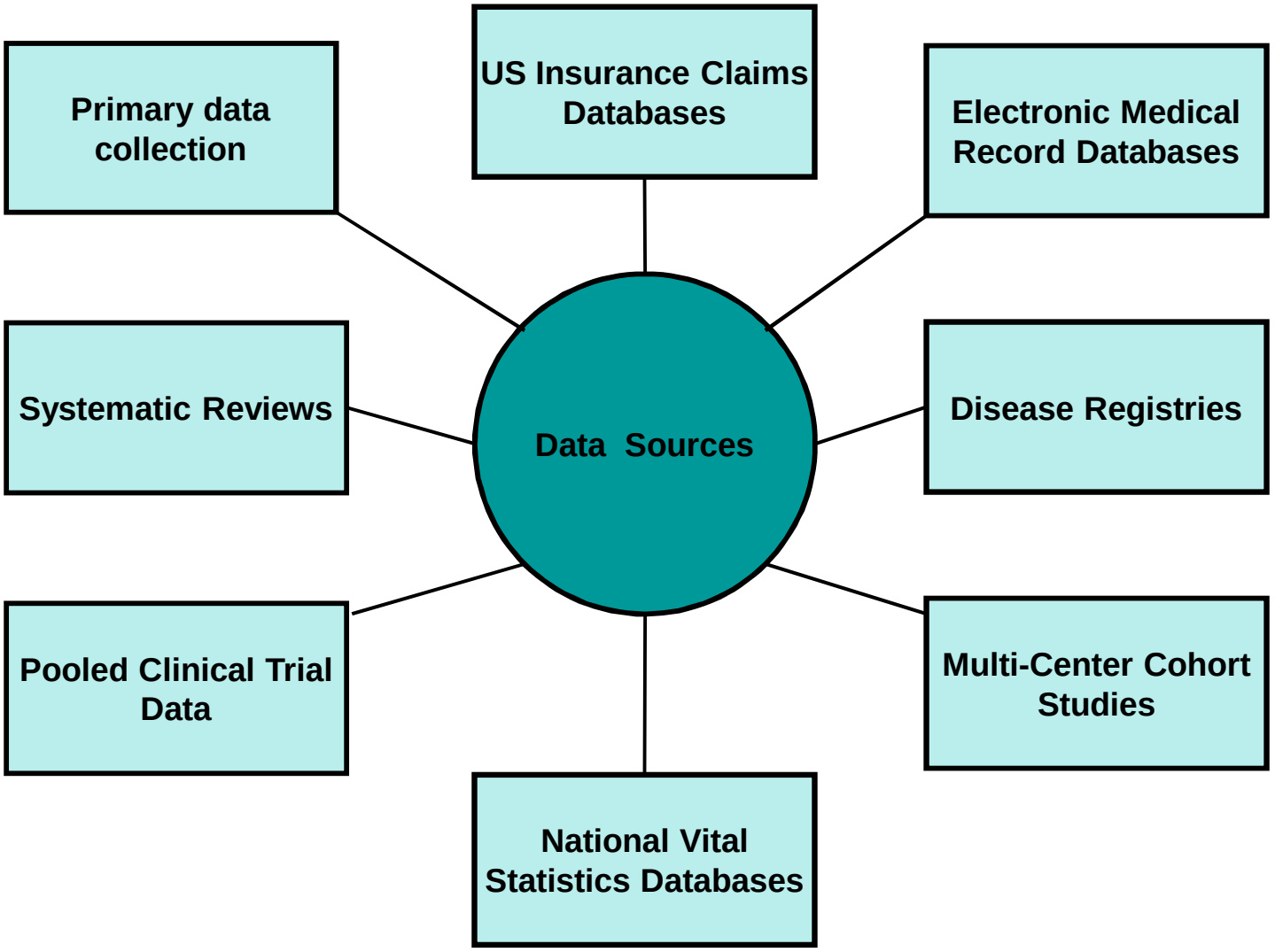
Randomized Clinical Trials are (the 5 Ss):

- Possibly too **SMALL** to detect rare outcomes
- Possibly too **SELECTED** to be generalizable to all users, all indications
- Possibly too **SHORT** to detect long-term effects
- Too **SIMPLE** to detect interactions
- Too **SPECIFIC** to assess all relevant outcomes

- Active post market risk identification and analysis requires studying large populations of medication users in "real world" clinical practice.

Source: Til Stürmer, MD, MPH Head, Pharmacoepidemiology program, Department of Epidemiology, UNC Gillings School of Global Public Health

Sources of data for observational studies



Uses of observational databases within the cardiovascular area studies

- Descriptive epidemiology studies
  - Estimate incidence or prevalence of a disease/drug/co-morbidity/outcome in a patient population – understand CV event rates in real world setting
- Association studies (e.g., drug safety studies)
  - Characterize the association between a particular exposure (medication) and outcome (adverse event) – QTc changes
    - Case control studies, Cohort studies
    - Protocol driven/reportable (POL408, TSS)
- Molecular epidemiology studies
  - Characterize the association between the distribution of molecular markers and disease etiology, prognosis – troponin levels
- Comparative effectiveness research
  - Compare the effectives of 2 drugs (from same class or different classes) in real world setting.

Advantages and Disadvantages of exploring large observational databases

Pros

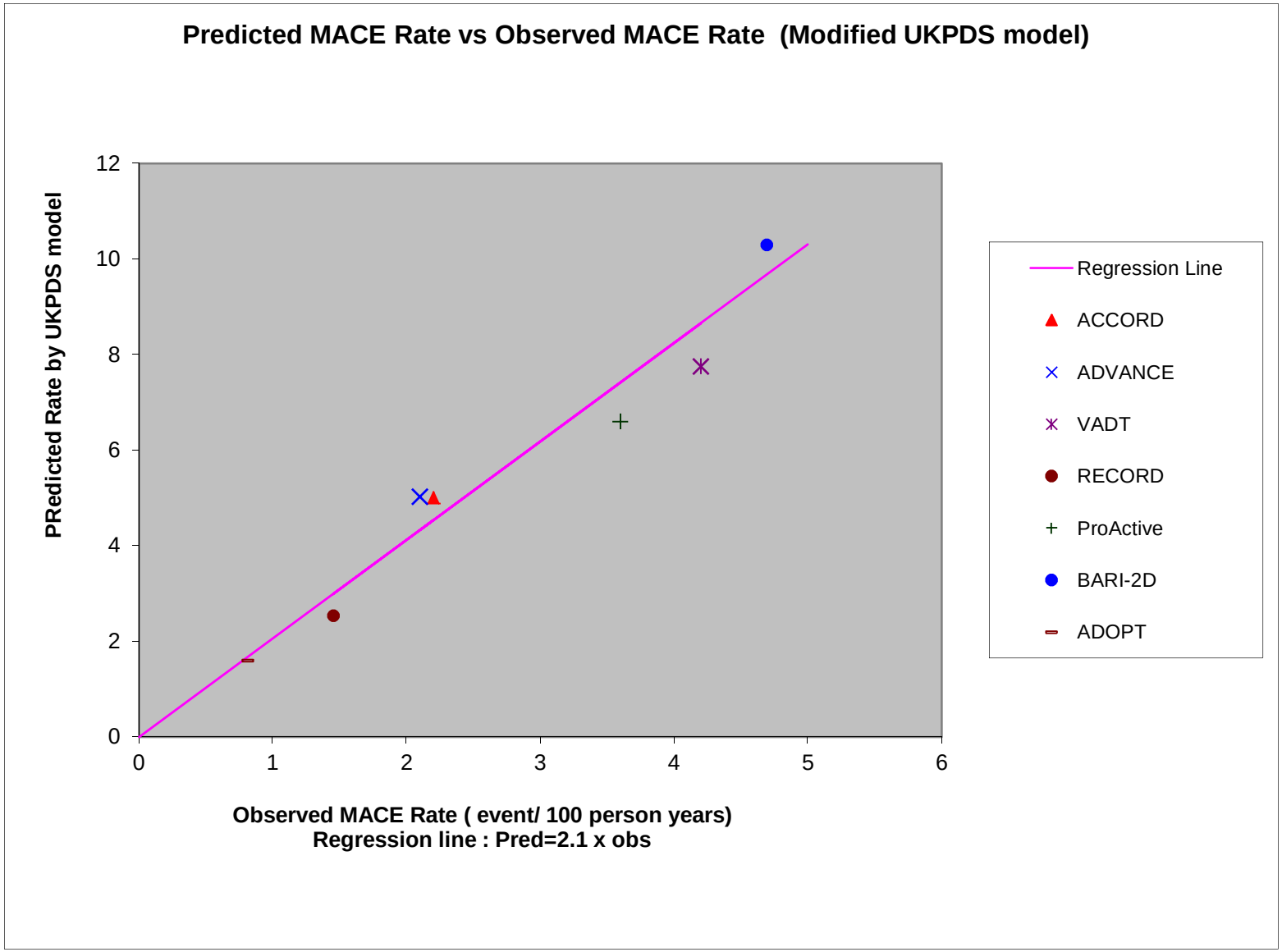
- Hypothesis generation at lower cost prior to embarking on expensive and time consuming RCT
- Can be conducted quickly (especially retrospective studies)--faster time to publication
- Real world setting
- Large sample size; increased ability to do sub group or country specific analyses

Cons

- Lack of randomization leads to a potential for confounding
  - Key variables which affect both the probability of exposure and the outcome may be imbalanced
- Data on key exposures, outcomes or covariates may be missing/incomplete in medical records or databases
- Medical databases not readily available for some emerging markets
- Outcome of interest may not be well captured and ability to validate the outcome is not always possible

Potential collaboration of databases across industry

- Pool data across all clinical trials in diabetes with CV adjudicated-endpoint
- Understand the epidemiology of CHD in diabetes
- Create up to date risk equation – UKPDS
- Understand relationship of glucose and other metabolic parameters to CV outcome


Predicted MACE Rate vs Observed MACE Rate (Modified UKPDS model)
PRedicted Rate by UKPDS model
0
2
4
6
8
10
12
0
1
2
3
4
5
6
Observed MACE Rate (event/ 100 person years)
Regression line : Pred=2.1 x obs
Regression Line
ACCORD
ADVANCE
VADT
RECORD
ProActive
BARI-2D
ADOPT