Mendelian Randomization studies and genetic risk scores for causal inference and risk prediction in CAD

Themistocles (Tim) Assimes, MD PhD
Assistant Professor of Medicine
Stanford University School of Medicine
WHI Investigators meeting May 2, 2013
What is a genetic risk score (GRS)?

- A single variable summarizing individual’s exposure to susceptibility alleles
- Constructed on basis of validated GWAS polymorphisms (SNPs)
- Utility
  - Easy to incorporate into established risk prediction (e.g FRS) algorithms once risk per unit measure of GRS is established
  - Powerful instrument for MR studies
Why try to improve on the Framingham Risk Score for CHD?

• Despite high predictive value of traditional risk factors (TRF)
  – CHD → leading cause of death
  – 30-50% of population → 0 or 1 TRF
  – Multiple interventions that
    • Reduce risk regardless of source
    • all subgroups
    • Generally safe
      – Lifestyle including exercise, diet, smoking cessation
      – LDL and BP lowering
      – aspirin
New discrimination tests

• In addition to C-statistic, consider absolute predicted risk of individuals (“reclassification”)
  – Net reclassification index (NRI)
  – Integrated discrimination index (IDI)
  – Others

• US Preventative Services Task Force has endorsed reclassification as important metric for prognostic tests

NRI: some key points often forgotten

• Calculate only for calipers of risk accepted in clinical practice
  – 4 categories always does better than 3 but this is not accepted clinical practice
  – The 3 categories for CAD are 0-10, 11-20, >20

• Only for population with an accepted actionable algorithm
  – Exclude diabetes

• Even small improvement → major public health implication
How can WHI contribute to development and testing of a GRS?

- **GRS are here to stay**
  - Will continue to improve over time, reach threshold of clinical utility, only challenge could be imaging (CAC)

- **For CHD**
  - More populations tested, the better
  - Increase precision of utility of GRS for white women
    - Start with current genetic data, increase to include all cases?
    - Explore subgroup differences in performance of GRS
  - Development of race/ethnic specific scores

- **For other outcomes with potential actionable preventive measures**
  - Primary prevention of breast cancer with tamoxifen
Mendelian Randomization (MR) studies

• The problem with observational epidemiology
  – many high profile failures
  – Cannot randomize some exposures
    • e.g. smoking, alcohol
    • unethical and impractical (long lead time)
    • Safe agents that alter risk factor do not always exist

• Even RCTs may not be generalizable
Causal vs. non causal associations
Why do we care?

• For prognosis
  – We don’t
  – long term survival of marker?

• For therapeutics
  – Big implications
  – Risk of failure of drugs increases dramatically
Why are MR studies more prevalent in last few years?

• Background
  – Very precise and accurate assessment of magnitude of effect between marker and outcome

• Affordable high throughput genotyping
  – Discovery of many polymorphisms influencing biomarkers of interest (locus specific or not)

• Collaboration
  – to overcome issues of low power to detect variants that influence risk factor and/or outcome
Mendelian Randomization Principle

• if genetic variant(s) alters the level of a (modifiable) exposure that itself alters disease risk, then it should be related directly to disease risk to the extent predicted by its influence on the exposure

• Advantages of using genetic marker
  – randomly assigned at conception
  – unlikely to be correlated with wide range of behavioral, social, and physiological factors

MR is an application of general theory of Instrumental Variable (IV) analysis

• What is IV?
  – variable associated with outcome ONLY through robust association of an intermediate phenotype

• Used frequently in econometrics to deal with “endogeneity”

• Endogeneity = confounding/reverse causality/regression dilution bias

Framework of a Mendelian randomization study.

One or multiple
Single nucleotide variants (SNPs)
Genetic Risk Score (GRS)
"Instrument"

Causal relation?
Or due to (residual) confounding and/or reverse causation?


© The Author 2010. Published by Oxford University Press on behalf of ERA-EDTA. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org
Assumptions/needs in Mendelian randomization studies

(a) presence of a robust association between genetic variant(s) and exposure
(b) absence of (direct/indirect) association between genetic variant and confounding factors
(c) absence of other pathways between genetic variants and outcome (pleiotropy).

Adapted from Verduijn M et al. Nephrol. Dial. Transplant. 2010;25:1394-1398
Instrumental Variable (IV) analysis (aka MR study) procedure used to estimate causal effects of exposure

- 2 stage least square (2SLS)
  - Stage 1: perform least squares regression of the exposure on the IV (genetic variant(s), GRS for exposure)
    - F statistics (> 10)
  - Stage 2: perform least-squares regression of the outcome $Y$ on the predicted exposure values from the first regression
    - Compare: 2SLS IV analysis $\beta$ vs. the ordinary least squares (OLS) $\beta$
    - examine overlap of 95% CI.
  - Best to obtain both $\beta$ from same study
    - IV estimate will have smaller variance
    - Can check for violations
    - not always possible

Estimate of the association of genetically raised LDL cholesterol or HDL cholesterol and risk of myocardial infarction using multiplegenetic variants as instruments

<table>
<thead>
<tr>
<th></th>
<th>Odds ratio (95% CI) per SD increase in plasma lipid based on observational epidemiology*</th>
<th>Odds ratio (95% CI) per SD increase in plasma lipid conferred by genetic score†</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL cholesterol</td>
<td>1.54 (1.45-1.63)</td>
<td>2.13 (1.69-2.69), p=2x10⁻¹⁰</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>0.62 (0.58-0.66)</td>
<td>0.93 (0.68-1.26), p=0.63</td>
</tr>
</tbody>
</table>

*Observational epidemiology estimates derived from more than 25 000 individuals from prospective cohort studies as shown in the appendix p 22. †LDL genetic score consisting of 13 single nucleotide polymorphisms (SNPs) as shown in the appendix p 27; HDL genetic score consisting of 14 SNPs as shown in the appendix p 28.

Ordinary Least Squares (OLS)  2 stage least squares (2SLS)

MR and causality of CAD risk factors
Summary of well powered studies to date

• CAUSAL
  – LDL, Lp(a), BMI, HTN

• NON CAUSAL
  – Fibrinogen, CRP, HDL
    – CETP inhibitor trials have failed to date (torcetrapib, dalcetrapib)
How can WHI contribute to MR of CHD

• Replication - Some initial MR studies false positive (e.g. CRP)

• CAUSALITY STILL TO BE DETERMINED FOR CHD risk factors
  – Triglycerides → evidence to date suggest causal
  – IR, Diabetes
  – Variety of markers of inflammation
  – Other markers: lipoprotein-associated phospholipase A2
  – EtoH
  – Homocysteine/folate levels
  – other dietary nutrients

• Large sample: develop instrument & perform MR study

Acknowledgements

• Stanford University
  – Benjamin Goldstein, John Ioannidis, Joshua Knowles, Thomas Quertermous, Marcia Stefanick
• University of Pennsylvania
  – Benjamin Voight
• Broad Institute
  – Sekar Kathiresan
• University of Leicester
  – Nilesh Samani
• University of Luebeck / Technische Universität München
  – Heribert Shunkert
• CARDIoGRAM & CARDIoGRAM+C4D consortium