Sequencing in cohort studies & large sample collections:

Key lessons and reactions

Rory Collins

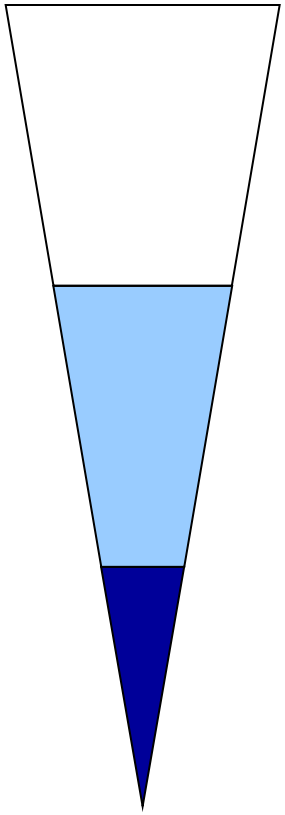
UK Biobank
Principal Investigator,
Oxford University,
United Kingdom

Future strategies: Reflections on previous experience with GWAS in complex diseases

- Many smaller versus few larger experiments
  (e.g. need for meta-analyses of many small studies)
- "Deep" versus "shallow" participant phenotyping
  (e.g. meta-analyses of studies with different definitions)
- Studying few versus many different diseases
  (e.g. prospective cohorts with extensive record linkage)
- "Deep" versus "shallow" disease phenotyping
  (e.g. value of ability to characterise disease subtypes)
- National versus international strategic initiatives

What to do in next few years versus what would we want to have done in 10-15 years?

Approach to disease adjudication: different types of data required at different stages



Approach	Characteristics	Possible examples
Ascertainment of suspected cases		
Confirmation of “caseness”		
Classification of confirmed cases		

UK Biobank: an international resource

- PROSPECTIVE: It can assess the full effects of a particular exposure (such as smoking) on all types of health outcome (such as cancer, vascular disease, lung disease, dementia)
- DETAILED: The wide range of questions, measures and samples at baseline allows good assessment of exposures, and outcome adjudication allows good disease classification
- LARGE: Inclusion of large number of people allows reliable assessment of the causes of a wide range of diseases, and of the combined impact of many different exposures
- ACCESS: General consent for follow-up through all health records for all types of health research, and for re-contact, by academic and commercial researchers worldwide

UK Biobank: Expected numbers of participants developing diseases during long-term follow-up

Condition	2012	2017	2022
Diabetes	10,000	25,000	40,000
MI/CHD death	7,000	17,000	28,000
Stroke	2,000	5,000	9,000
COPD	3,000	8,000	14,000
Breast cancer	2,500	6,000	10,000
Colorectal cancer	1,500	3,500	7,000
Prostate cancer	1,500	3,500	7,000
Lung cancer	800	2,000	4,000
Hip fracture	800	2,500	6,000
Rheum. arthritis	800	2,000	3,000
Alzheimer's	800	3,000	9,000

Value of detailed phenotyping not only of participants at baseline but also of disease cases during follow-up

- Enhancement of power to detect associations between risk factors and disease outcomes
  (false positive diagnoses have main adverse impact)
- Increased specificity of disease classification allows the detection of specific associations
  (e.g. risk factor only linked to disease sub-type)
- “Future-proofing” of the outcome data so that more detailed phenotyping is possible in future
  (e.g. retain data/samples to allow refined sub-typing)

Advantages of PROSPECTIVE cohorts for studying the causes of different diseases

- Risk factors can be measured before disease develops (helping to avoid “reverse causality”)
- Associations can be assessed with a range of diseases (provided sufficient numbers occur)
- Appropriate controls can be selected from within the same population as the disease cases
- Confounding by other factors is typically less extreme and can be allowed for more fully

But prospective cohorts need to be LARGE

Prospective studies need to be LARGE: CHD versus SBP for 5K *vs* 50K *vs* 500K people in the Prospective Studies Collaboration