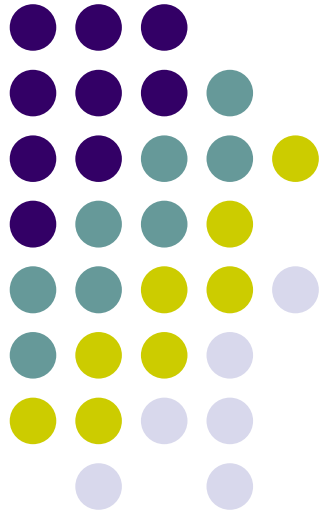
OCRe:
Ontology of Clinical Research



11th International Protégé Conference

June 28, 2009

Samson W. Tu[1], Ida Sim[2], Simona Carini[2],
Steve Harris[3], Peter Maccallum[3],
Alan Rector[4], Igor Toujilov[3]

[1]Stanford, [2]UCSF, [3]UK CancerGrid, [4]U Manchester

Acknowledgement

- The OCRe work has been funded in part by National Library of Medicine grant LM06780, and by MRC-G0100852
- Principal Investigator Ida Sim (who contributed a number of slides)

Outline

- Background
- Characterization of OCRe
- Examples
- Status

“Epidemiology” of Human Studies

- Human study
  - Any study on data collected from or about humans
- Data about human studies
  - Little data available about observational and qualitative studies
  - Interventional studies
    - ~ 20-60,000 new studies worldwide annually
    - US Pharma spent $16b on trials in 2006
      - clinical research organizations spent $14b in 2007
    - Avg pharma trial takes 6-7 years

Difficulties in Human Studies: Design Challenges

- What's been studied already?
  - what interventions, outcomes and timepoints
  - early phase studies especially hard to find
- What hasn't been studied that I should?
  - e.g., data mining results for new hypotheses
- What is most effective/efficient design?
  - sample size calculation: base rates, simulating effect sizes
  - feasibility: % eligible, % recruitable, time to accrue? cost?
- What can I reuse?
  - biologic agents, questionnaires, case report forms

Difficulties in Human Studies: Executing & Using Studies

- Years to run studies, large minority don't finish
- Results publishing is systematically biased
- Difficult to practice evidence-based medicine
  - can't find "best relevant evidence"
  - relevant questions often not studied
  - can't put related evidence together
  - can't apply evidence to patient at time of need

Vision (Ida Sim)

- Interoperable federated database system of study design and results data from all human studies worldwide
  - sufficiently detailed to support care and discovery
  - in which all data elements are standardized to controlled vocabularies and common ontologies to enable cross-study comparison and synthesis
  - integrated with electronic IRB, clinical research management systems, reporting systems, etc.

OCRe: Ontology of Clinical Research

- Description of human studies to enable cross-study comparison and synthesis
- Ontology about
  - Meta-data of a study
  - High level design of a study
  - Analytic results of a study
- Not about
  - Domain ontologies
    - Bindings to domain ontologies & vocabularies
  - Detailed design of study
    - e.g., schedule of activities
  - Schema of individual-level data

OCRe Ontology: Conceptualization

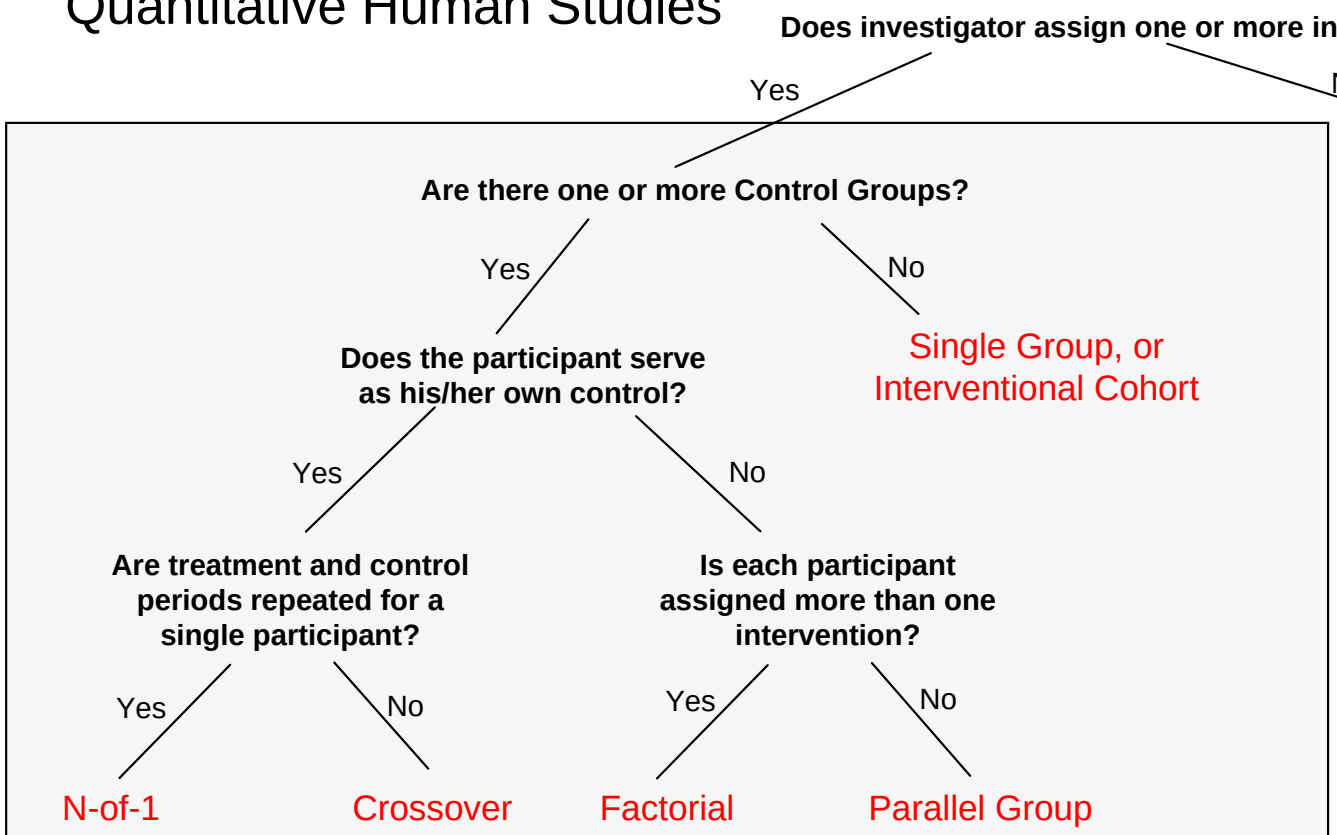
- A study is a real-world entity that includes
  - an *informational* study protocol that defines the design and planned events of the study
  - participating individuals and institutions playing specific roles
  - events carried out during the life cycle of a study
  - data collected and analyzed as part of the study
  - publications resulting from the study

Example Use Cases

- State of investigation
  - e.g., What are the methodological strengths of all current and completed human studies on "percutaneous coronary intervention?"
    - *has_experimental_intervention* some *percutaneious coronary intervention (SNOMEDCT 415070008)*
    - Methodological strength: Study design & statistical methods
- Gaps in literature
  - e.g., Are cardiology trials systematically excluding patients with renal failure?

Classification of Study Designs

Quantitative Human Studies

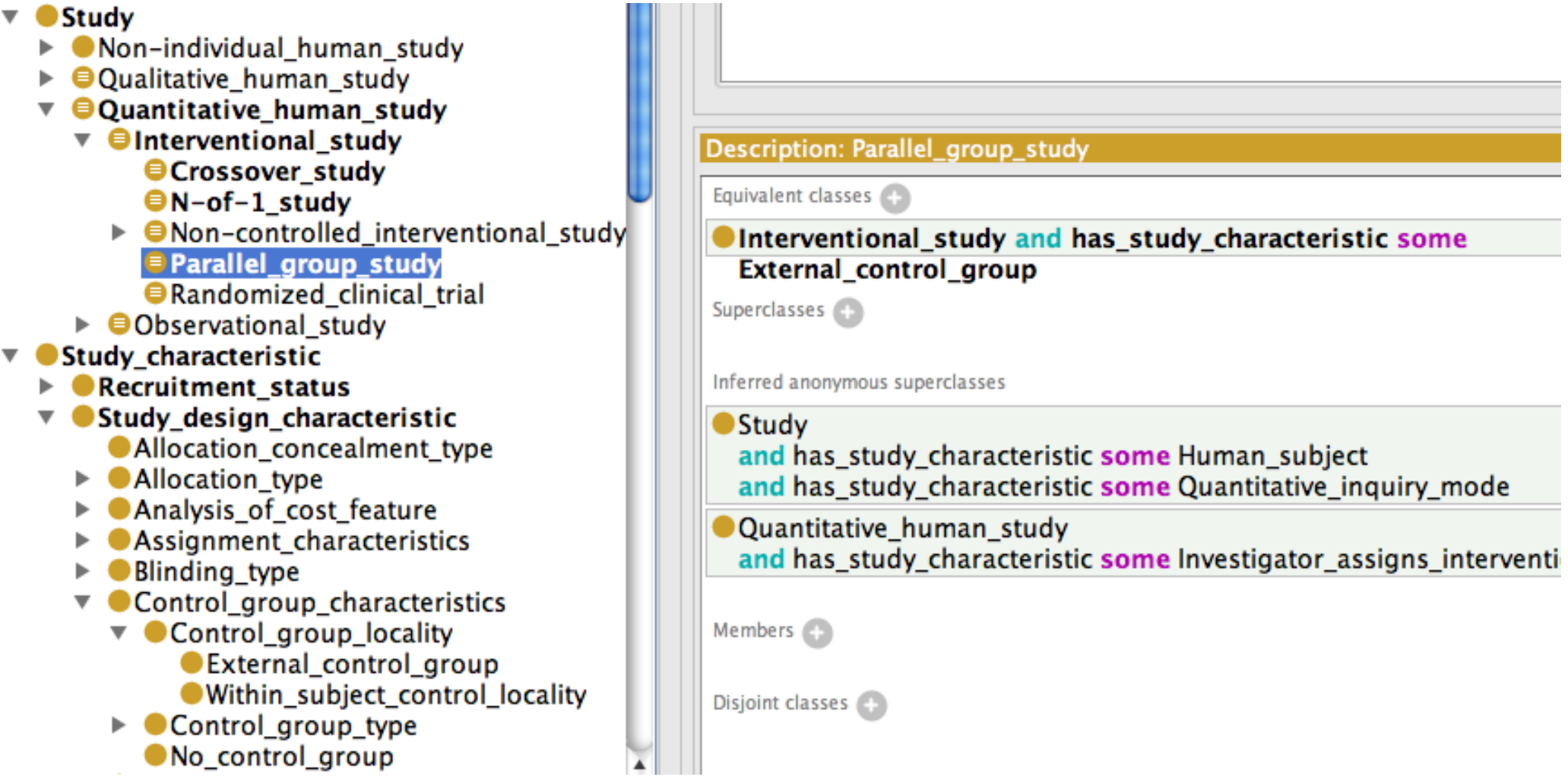


Additional descriptors

- Statistical intent (superiority, non-inferiority, equivalence)
- Sequence generation (random, blocking, stratification, etc see spreadsheet)
- Allocation method (central, sequentially numbered and identical, etc. see spreadsheet)
- Unit of randomization (single human = not clustered, anything else = clustered)
- Blinding/Masking (of participant, investigator, outcomes assessor, statistician)
- Control group type (active, placebo, sham, usual care, dose comparison, historical?)
- Control group timing (historical, contemporaneous) [is this needed?]
- Study phase (Phase 0, I, II, III, IV)

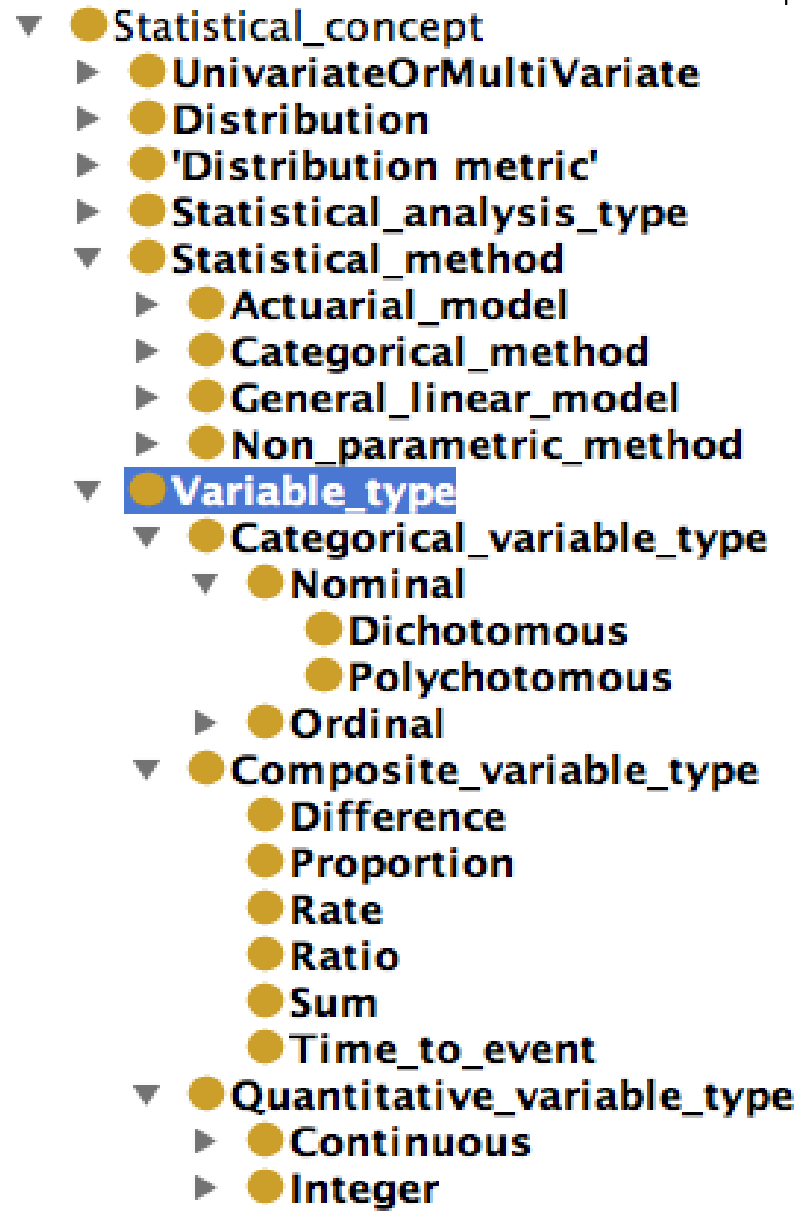
Interventional Studies

Study Design Ontology

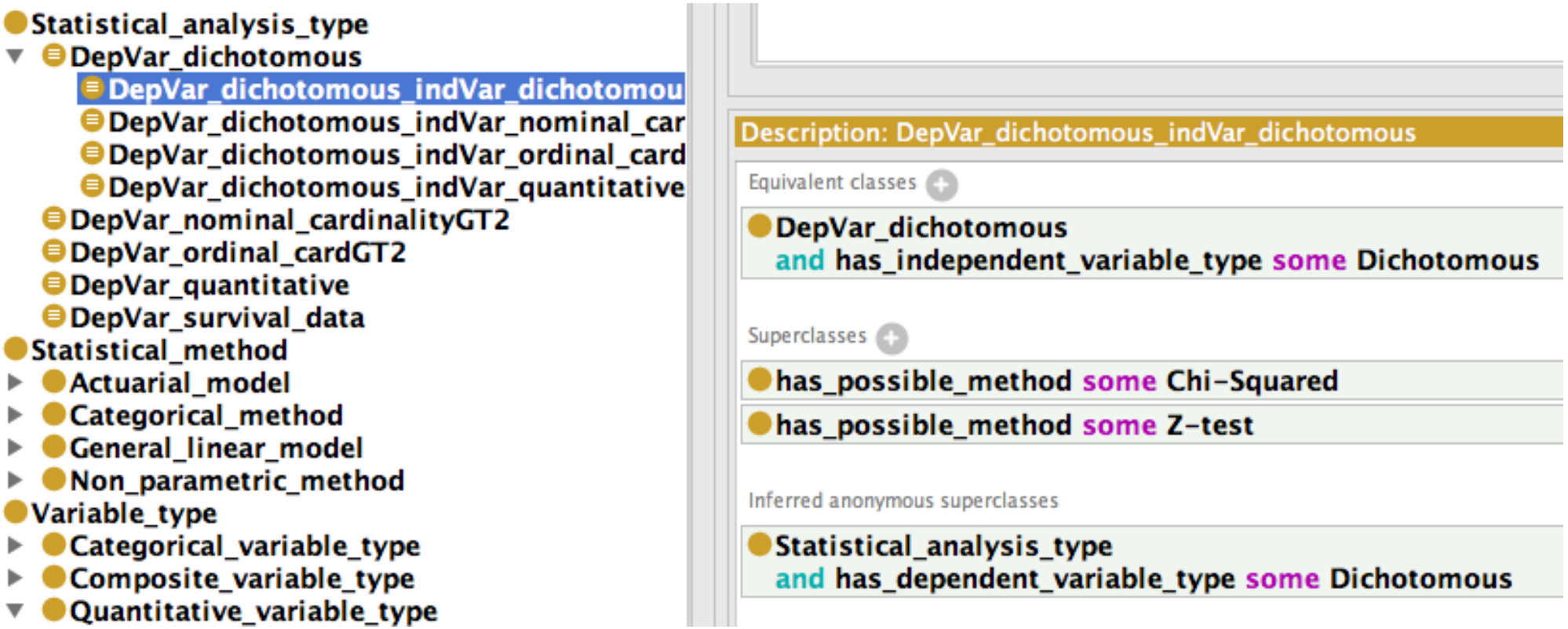


Statistics Ontology

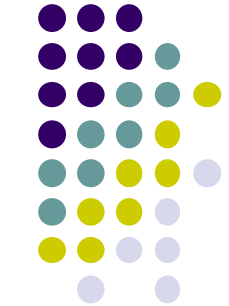
- Small ontology involving
  - Variable types
  - Statistical methods commonly used in human studies
  - Statistical analysis types…



Statistical Analysis Types

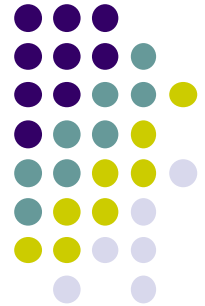


Description of Studies to Answer Questions about Study Methodology



- xyz is a parallel-arm randomized study that
  - *has_study_protocol* that
  - *has_primary_outcome* some *Study_Outcome* that
  - *is_analyzed_by* statistial_analysis0 that
  - *is_instance_of* DepVar_Dichotomous_IndepVar_Dichotomous
  - and *use_statistical_method* some Chi-squared

Are cardiology trials systematically excluding patients with renal failure?



- Inclusion/exclusion criteria

-Inclusion: Presence of myocardial infarction
-Exclusion:
-Acute renal failure syndrome
-No monotherapy with clarithromycin, azithromycin, clofazimine, or ethambutol for more than 1 month prior to enrollment
-Blood pressure higher than normal despite lifestyle changes and treatment with medications

- Annotate free-text criteria with concept expressions

Inclusion: acute myocardial infarction
Exclusion: renal failure
- clarithromycin or azithromycin or clofazimine or ethambutol

Status of OCRe

- Initial core ontologies and import structure
- Preliminary testing using CancerGrid and RCT Bank studies
- Current goals
  - documentation of the ontology
  - alpha release for community use and critique
    - e.g., via NCBO BioPortal
- NCRR project to develop federated databases of human studies (UCSF, Mayo Clinic, and Washington University St. Louis)

Summary

- Need for an ontology for describing human studies
- Description of studies including
  - Study design
  - Statistical methods
  - Characterization of target population
  - Health conditions studied & interventions (when applicable)
  - Use of standard domain terminologies