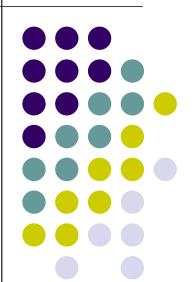
# OCRe: Ontology of Clinical Research

11<sup>th</sup> International Protégé Conference June 28, 2009

Samson W. Tu<sup>1</sup>, Ida Sim<sup>2</sup>, Simona Carini<sup>2</sup>, Steve Harris<sup>3</sup>, Peter Maccallum<sup>3</sup>, Alan Rector<sup>4</sup>, Igor Toujilov<sup>3</sup>



<sup>1</sup>Stanford, <sup>2</sup>UCSF, <sup>3</sup>UK CancerGrid, <sup>4</sup>U Manchester





- 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**

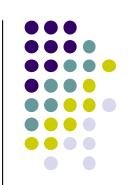
- Background
- Characterization of OCRe
- Examples
- Status

## "Epidemiology" of Human Studies



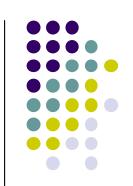
- Human study
  - Any study on data collected from or about humans
- Data about human studies
  - Little data available about observational and qualitative studies
  - Interventional studies
    - ~ 20-60,000 new studies worldwide annually
    - US Pharma spent \$16b on trials in 2006
      - clinical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

# Difficulties in Human Studies: Design Challenges



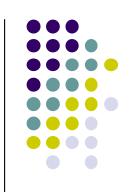
- What's been studied already?
  - what interventions, outcomes and timepoints
  - early phase studies especially hard to find
- What hasn't been studied that I should?
  - e.g., data mining results for new hypotheses
- What is most effective/efficient design?
  - sample size calculation: base rates, 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



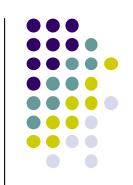
- Years to run studies, large minority don't finish
- Results publishing is systematically biased
- Difficult to practice evidence-based medicine
  - can't find "best relevant evidence"
  - relevant questions often not studied
  - can't put related evidence together
  - can't apply evidence to patient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

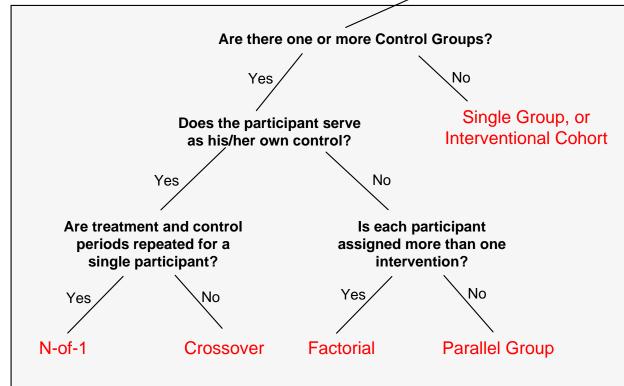
Yes

**Quantitative Human Studies** 

Does investigator assign one or more interventions?







#### Additional descriptors

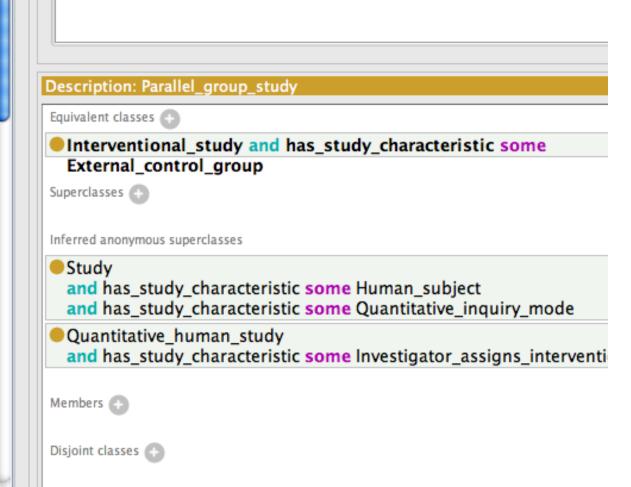
- Statistical intent (superiority, non-inferiority, equivalence)
- Sequence generation (random, blocking, stratification, etc see spreadsheet)
- Allocation method (central, sequentially numbered and identical, etc. see spreadsheet)
- Unit of randomization (single human = not clustered, anything else = 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)

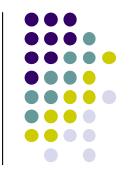
#### Interventional Studies

#### **Study Design Ontology**



- Study
  - Non-individual\_human\_study
  - Qualitative\_human\_study
  - ▼ Quantitative\_human\_study
    - ▼ □ Interventional\_study
      - Crossover\_study
      - N-of-1\_study
      - @Non-controlled\_interventional\_study
        - Parallel\_group\_study
        - Randomized\_clinical\_trial
    - Observational\_study
- Study\_characteristic
  - Recruitment\_status
  - Study\_design\_characteristic
    - Allocation\_concealment\_type
    - Allocation\_type
    - Analysis\_of\_cost\_feature
    - Assignment\_characteristics
    - Blinding\_type
    - Control\_group\_characteristics
      - Control\_group\_locality
        - External\_control\_group
        - Within\_subject\_control\_locality
      - Control\_group\_type
        - No\_control\_group





#### **Statistics Ontology**

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

- Statistical\_concept
  - UnivariateOrMultiVariate
  - Distribution
  - Distribution metric
  - Statistical\_analysis\_type
  - Statistical method
    - Actuarial\_model
    - Categorical\_method
    - General\_linear\_model
    - Non\_parametric\_method
  - ▼ Variable\_type
    - Categorical\_variable\_type
      - ▼ Nominal
        - Dichotomous
        - Polychotomous
      - Ordinal
    - Composite\_variable\_type
      - Difference
      - Proportion
      - Rate
      - Ratio
      - Sum
      - Time\_to\_event
    - Quantitative\_variable\_type
      - Continuous
      - Integer

#### **Statistical Analysis Types**



- Statistical\_analysis\_type
- DepVar\_dichotomous
  - DepVar\_dichotomous\_indVar\_dichotomou
  - DepVar dichotomous indVar nominal car
  - DepVar\_dichotomous\_indVar\_ordinal\_card
  - DepVar\_dichotomous\_indVar\_quantitative
  - DepVar nominal cardinalityGT2
  - DepVar\_ordinal\_cardGT2
  - DepVar\_quantitative
  - DepVar\_survival\_data
- Statistical method
- Actuarial\_model
- Categorical method
- General\_linear\_model
- Non\_parametric\_method
- Variable\_type
- Categorical\_variable\_type
- Composite\_variable\_type
- 🔻 🛑 Quantitative\_variable\_type

Description: DepVar\_dichotomous\_indVar\_dichotomous

Equivalent classes +

DepVar\_dichotomous and has\_independent\_variable\_type some Dichotomous

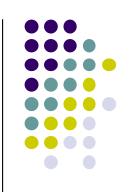
Superclasses (

- has\_possible\_method some Chi-Squared
- has\_possible\_method some Z-test

Inferred anonymous superclasses

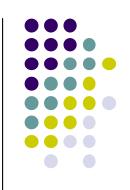
Statistical\_analysis\_type and has\_dependent\_variable\_type some Dichotomous

### 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

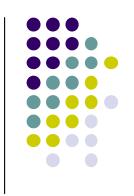
## Annotate free-text criteria with concept expressions

Inclusion: acute myocardial infarction

Exclusion: renal failure

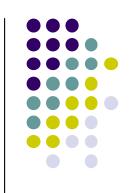
- clarithromycin or azithromycin or clofazimine or ethambutol





- 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)





- 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