## 533 HW1

Mo Tang mo.tang@rice.edu S01276578

## PART ONE: Relational Calculus & Relational Algebra

- 1.Because MRN might be different for one patient at different time.
- $2.(a) \{pa.MRN \mid PATIENT(pa) \land pa.AGE = 25\}$
- (b) {v.MRN | VISIT(v) ^ v.DATETIME = 'December,2017'}
- (c)  $\{v.MRN \mid VISIT(v) \land \exists (pr)(PROCEDURE(pr) \land PROCEDURE(pr).VISIT_ID = VISIT(v).ID \land PROCEDURE(pr).NAME = 'flu shot'\}$
- (d)  $\{v.MRN \mid VISIT(v) \land \forall (v2)(VISIT(v2) \land v.VISIT_ID = v2.VISIT_ID =$
- ¬∃(pr)(PROCEDURE(pr) ^ v2.VISIT\_ID = pr.VISIT\_ID ^ pr.NAME = 'flu.shot'}
- (e) {p.FIRSTNAME,p.LASTNAME | PATIENT(p)  $\land \forall (v, pr, c)(VISIT(v) \land PROCEDURE(pr) \land CLINICIAN(c) \land p.MRN = v.MRN \land v.VISIT_ID = pr.VISIT_ID \land pr.CLIN_ID = c.CLIN_ID \land c.CERT = 'MD' <math>\land c.FIRSTNAME = 'Paula' \land c.LASTNAME = 'Jones')$ }
- 3.(a)  $\pi_{MRN}(\sigma_{AGE=25}(PATIENT))$
- (b)  $\pi_{FIRSTNAME, LASTNAME}(\sigma_{CERT=PA'}(CLINICIAN) \rtimes PROCEDURE \rtimes VISIT \rtimes PATIENT)$
- (c)  $\pi_{MRN}(PATIENT \bowtie \sigma_{FIRSTNAME=FIRSTNAME, LASTNAME=LASTNAME, AGE=AGE}(PATIENT))$
- (d)  $\rho(template1, (\sigma_{NAME='flu\ shot'}(PROCEDURE) \bowtie VISIT))$   $\rho(template2, (\sigma_{CERT='PA'}(CLINICIAN) \bowtie PROCEDURE \bowtie VISIT))$  $\pi_{MRN}(template1 \cap template2)$
- (e)  $\rho(template1, (\sigma_{CERT=MD'}(CLINICIAN) \bowtie PROCEDURE \bowtie VISIT))$   $\rho(template2, (\sigma_{CERT=PA'}(CLINICIAN) \bowtie PROCEDURE \bowtie VISIT))$  $\pi_{MRN}(template1 - template2)$

## **PART TWO: Queries**

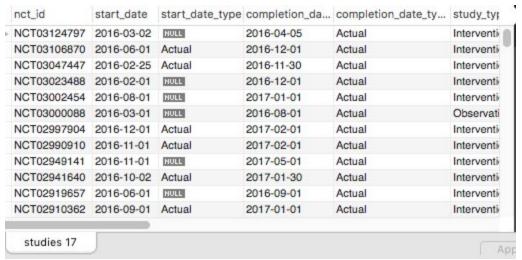
 select nct\_id, study\_type from studies where brief\_title = 'Autologous Cell Therapy After Stroke';

nct_id	study_type
NCT00908856	Interventional
HULL	NULL

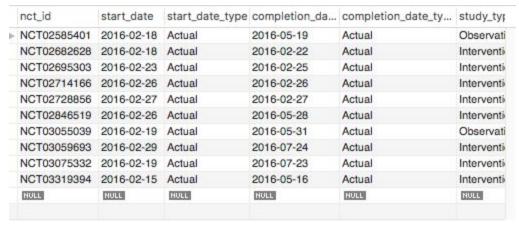
2. select DISTINCT study\_type from studies sort order by study\_type;

study_type	
Expanded Access	Ť
Interventional	
N/A	
Observational	
Observational [Patient Registry]	

 select \* from studies where year(studies.start\_date) = 2016 and studies.nct\_id in (select reported\_events.nct\_id from reported\_events);



4. select \* from studies where studies.start\_date <= '2016-2-30' and studies.start\_date >= '2016-2-15' and datediff(studies.completion\_date, studies.start\_date) < 180;</p>



## PART THREE: Reading / Short answer

- 1.Data efficiently is organized; Redundancy is eliminated; Logical data dependencies are ensured by storing only related data within a given table.
- 2.The DESIGNS table stores the data from Study Design, which was parsed into components. All components of study design and their respective enumerated values are stored by additional data elements (Design Name and Design Value). Values related to masking/blinding (e.g., Single; Double-Blind) were further parsed into their components, along with the list of corresponding masking subjects (Participant, Investigator, Outcome Assessor, and Caregiver). 3.Observational\_model
- 4. The downcase\_name doesn't contain capital letters. The name and downcase name are useful for the unification, and it's convenient for users to search for data. And it's also for data

utilization because we don't need letter-transforming when utilizing data. Space-time trade-offs is involved.

5. For instance, downcase name 'rice' is confusing, because users don't know whether it is 'RICE' as our school name or 'Rice' as food.