

## 533 HW1

Mo Tang [mo.tang@rice.edu](mailto: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)  $\{pa.MRN \mid PATIENT(pa) \wedge pa.AGE = 25\}$

(b)  $\{v.MRN \mid VISIT(v) \wedge v.DATETIME = 'December, 2017'\}$

(c)  $\{v.MRN \mid VISIT(v) \wedge \exists (pr)(PROCEDURE(pr) \wedge PROCEDURE(pr).VISIT\_ID = VISIT(v).ID \wedge PROCEDURE(pr).NAME = 'flu shot'\}$

(d)  $\{v.MRN \mid VISIT(v) \wedge \forall (v2)(VISIT(v2) \wedge v.VISIT\_ID = v2.VISIT\_ID \Rightarrow \neg \exists (pr)(PROCEDURE(pr) \wedge v2.VISIT\_ID = pr.VISIT\_ID \wedge pr.NAME = 'flu.shot')\}$

(e)  $\{p.FIRSTNAME, p.LASTNAME \mid PATIENT(p) \wedge \forall (v, pr, c)(VISIT(v) \wedge PROCEDURE(pr) \wedge CLINICIAN(c) \wedge p.MRN = v.MRN \wedge v.VISIT\_ID = pr.VISIT\_ID \wedge pr.CLIN\_ID = c.CLIN\_ID \wedge c.CERT = 'MD' \wedge c.FIRSTNAME = 'Paula' \wedge c.LASTNAME = 'Jones')\}$

3.(a)  $\pi_{MRN}(\sigma_{AGE=25}(PATIENT))$

(b)  $\pi_{FIRSTNAME, LASTNAME}(\sigma_{CERT=PA'}(CLINICIAN) \bowtie PROCEDURE \bowtie VISIT \bowtie PATIENT)$

(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

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

nct_id	study_type
NCT00908856	Interventional
NULL	NULL

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

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

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

nct_id	start_date	start_date_type	completion_da...	completion_date_ty...	study_ty
NCT03124797	2016-03-02	NULL	2016-04-05	Actual	Interventi
NCT03106870	2016-06-01	Actual	2016-12-01	Actual	Interventi
NCT03047447	2016-02-25	Actual	2016-11-30	Actual	Interventi
NCT03023488	2016-02-01	NULL	2016-12-01	Actual	Interventi
NCT03002454	2016-08-01	NULL	2017-01-01	Actual	Interventi
NCT03000088	2016-03-01	NULL	2016-08-01	Actual	Observati
NCT02997904	2016-12-01	Actual	2017-02-01	Actual	Interventi
NCT02990910	2016-11-01	Actual	2017-02-01	Actual	Interventi
NCT02949141	2016-11-01	NULL	2017-05-01	Actual	Interventi
NCT02941640	2016-10-02	Actual	2017-01-30	Actual	Interventi
NCT02919657	2016-06-01	NULL	2016-09-01	Actual	Interventi
NCT02910362	2016-09-01	Actual	2017-01-01	Actual	Interventi

studies 17

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;

nct_id	start_date	start_date_type	completion_da...	completion_date_ty...	study_ty
NCT02585401	2016-02-18	Actual	2016-05-19	Actual	Observati
NCT02682628	2016-02-18	Actual	2016-02-22	Actual	Interventi
NCT02695303	2016-02-23	Actual	2016-02-25	Actual	Interventi
NCT02714166	2016-02-26	Actual	2016-02-26	Actual	Interventi
NCT02728856	2016-02-27	Actual	2016-02-27	Actual	Interventi
NCT02846519	2016-02-26	Actual	2016-05-28	Actual	Interventi
NCT03055039	2016-02-19	Actual	2016-05-31	Actual	Observati
NCT03059693	2016-02-29	Actual	2016-07-24	Actual	Interventi
NCT03075332	2016-02-19	Actual	2016-07-23	Actual	Interventi
NCT03319394	2016-02-15	Actual	2016-05-16	Actual	Interventi
NULL	NULL	NULL	NULL	NULL	NULL

### 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, lowercase name 'rice' is confusing, because users don't know whether it is 'RICE' as our school name or 'Rice' as food.