HOG Clinical Trials Tracker

Project Step 3 Draft
Group 117
Team: Thomas Gathman, John Wong
URL: https://web.engr.oregonstate.edu/~gathmant/

Feedback by Peer Reviewers (Step 1):

The step 1 draft proposal for the HOG Clinical Trials Tracker received the following feedback.

- **Does the overview describe what problem is to be solved by a website with DB back end?**
    - Yes, the problem is clearly stated.
    - Yes, the team has a great idea with detailed solution that can be applied in the healthcare system; It is a very applicable solution to a big problem in the health care. Their approach can be solved using website with databases.
    - Yes, the Health Oncology Group (HOG) is a great idea and has a definite need in the healthcare space. The problem is laid out very well and described in detail.
    - Yes, it was made very clear that the Health Oncology Group is in need of a website with a DB back end after expanding their network and converting to an electronic system.

- **Does the overview list specific facts?**
    - Yes, there are multiple specific numbers used.
    - Yes, the tables have some facts; for example, clinical_trail entity has clinical_trail_id and cancer_type.
    - Yes, there is a lot of specific information in their overview.
    - Yes, the overview lists specific facts and statistics on the number of trials, participating hospitals and patients which allows us to infer the scale of the project.

- **Are at least four entities described and does each one represent a single idea to be stored as a list?**
    - Yes, there are five listed.
    - Yes, there are 5 entities.
    - Yes, five of them.
    - Yes, the entities are Clinical Trials, Patients, Employees, Patients and Hospitals and each one of them represents a single idea to be stored as a list.

- **Does the outline of entity details describe the purpose of each, list attribute datatypes and constraints and describe relationships between entities? Does the outline clearly indicate which entities (tables) will be implemented and which team member is primarily assigned to the associated page(s)?**
    - The entities have attributes and relationships specified and the purpose of each entity is clear. As for team members being assigned to pages, I'm not sure what 'pages' is referring to, but this part doesn't seem to be in the rubric so I don't think not including this should be a problem.
    - The purpose of each entity was described clearly with a separation section for each one; and each has its own attributes. The relations were well explained between entities.
    - Everything was described well, relationships included. I can clearly understand the need for each.
    - Yes, the entity outlines are very detailed with attribute names, datatypes and constraints. Additionally, all the relationships are clearly laid out. The outline does not indicate which tables will be implemented by which team member but I did not think this was a requirement of the assignment...

- **Are 1:M relationships correctly formulated? Is there at least one M:M relationship? Does the ERD present a logical view of the database?**

- Yes, the relationships are correctly formulated and there are 2 M:M relationships.
- The team has multiple 1:M relationship; such as the relationship clinical_trails and patients. There are 2 M:M relationship.
- Yes everything seems to be correct and there are two M:M relationships.
- Yes, the relationships are correctly formulated as far as I can tell and there are multiple M:M relationships. The ERD is easy to follow and aligns well with the entity outline section.

- **Is there consistency in a) naming between overview and entity/attributes b) entities plural, attributes singular c) use of capitalization for naming?**
    - Yes, naming is very consistent.
    - I am amazed by how the team efforts to keep everything consistent and clearly explained. All the names of entities and attributes are written in same format.
    - Yes. Everything is named very well. This team went above and beyond. I think the idea is awesome and named very well. Great job!!!
    - Yes, there is consistency in the naming of entities and attributes, and the format is snake case.

Actions Based on Feedback (Step 1):

Because we did not include the team member primarily associated with each page, we added that to the proposal.

Feedback by Peer Reviewers (Step 2):

The step 2 draft proposal for the HOG Clinical Trials Tracker received the following feedback.

- **Does the schema present a physical model that follows the database outline and the ER logical diagram exactly?**
    - Yes, the schema exactly followed the database ERD.
    - Yes, but...and I could be totally wrong here, but I believe the Schema should only include PK and FK that illustrate the relationships between the entity tables.
    - To this day I don't fully understand the difference between an ER diagram and a schema, but I definitely don't see any discrepancies between the schema here and the ER diagram/rest of the document.
    - Yes, the schema follows the outline and the ER exactly.

- **Is there consistency in a) naming between overview, outline, ER and schema entity/attributes b) entities plural, attributes singular c) use of capitalization for naming?**
    - Yes, it's consistency un naming between overview, outline, ER and schema entities. However, table name is not capitalized.
    - Yes, as Griffin already stated in his review, snake_case were used and attributes are singular while entities are plural.
    - Yes, looks like you used snake_case for everything and were consistent about singular attributes/plural tables
    - There is consistency in naming, pluralization and singularity, and capitalization between overview, outline, ER, and schema.

- **Is the schema easy to read (e.g. diagram is clear and readable with relationship lines not crossed)?**
    - The schema is easy to read, and the diagram is clear with specific primary keys and foreign keys.
    - Schema is very readable and relationship lines are not crossed.
    - I think it's a very well-organized schema, nice job. The entire pdf seems very readable to me.
    - The schema is very easy to follow and well-laid-out with no crossing lines.

- **Are intersection tables properly formed (e.g. two FKs and facilitate a M:N relationship)?**

    - Yes, there are two FKs in patient table. One is from clinical_trials, another one is from hospitials table.
    - Yes, the intersection tables are properly formed and comprised of two FKs which facilitate a M:N relationship.
    - Yes, employees_supporting_clinical_trials and hospitals_supporting_clinical_trials (the only intersection tables I see) include foreign keys and facilitate M:N relationships.
    - The two intersection tables are properly-formed with two FKs each and facilitating a M:N relationship

- **Does the sample data suggest any non-normalized issues, e.g. partial dependencies or transitive dependencies?**

    - In my opinion, the sample data is standard and there is no obvious problem.
    - Looks like sample data does not show any clear non-normalized issues. Everything seems to be in its place.
    - There are no obvious dependencies to me. I'll admit to still being pretty confused about this, particularly transitive dependencies. It doesn't help that most examples online seem too trivial and simplistic to apply to more complex cases.
    - employee_role (in employees_supporting_clinical_trials)/position(in employees) and employer (in employees) are ones to maybe watch out for? I'm not exactly sure if position and employee_role are supposed to be the same thing based on the sample data, but if they are it'd be worth noting that if one of them were to change it'd need to be changed in both locations, and also in both cases that if an entry were deleted it'd be possible to get rid of the last record of a particular position/employee_role. Similarly, an employer could also be erased altogether if the last employee in the database listed were to be deleted.

- **Is the SQL file syntactically correct? This can be easily verified by using PhPMyAdmin and your CS 340 database (do not forget to take backup of your own database before you do this!)**

    - Yes, the SQL file works well.
    - It is, was able to load the provided DDL file into and view it in Designer with no errors.
    - The DDL.sql imports to PhPMyAdmin smoothly, everything looks like it works to me.
    - The SQL file is syntactically correct -- it imported correctly in PhPMyAdmin.

- **In the SQL, are the data types appropriate considering the description of the attribute in the database outline?**

    - I think data types are appropriate. For example, int id ,varchar name. I am not sure it's ok to use char(1) to represent the patient_sex.
    - Yes, all data types are appropriate for their description.
    - All the data types look appropriate to me.
    - The data types appear to be appropriate, based on the outline.

- **In the SQL, are the primary and foreign keys correctly defined when compared to the Schema? Are appropriate CASCADE operations declared?**

    - Yes, the primary and foreign keys are correctly defined. No cascade update or delete operations.
    - No cascade operations are present, but PK and FK are correctly defined according to the Schema.

    - I don't see cascade operations, but the primary and foreign keys are defined consistently in the SQL code and in the schema.
    - Yes, the primary and foreign keys appear correctly defined. Appropriate CASCADE operations are declared.

- **In the SQL, are relationship tables present when compared to the ERD/Schema?**

    - Yes, correct PKs and FKs and correct tables.
    - Yes, the tables are present and viewable in PhPMyAdmin.
    - All the tables are present in the SQL code.
    - Yes, relationship tables are present and appear correctly-formed.

- **In the SQL, is all example data shown in the PDF INSERTED?**

    - Yes, all data shown in the PDF file and each number of table data is more than 3.
    - Yes, the example data shown is present in their insertions.
    - Yes, it's all present.
    - Yes, all exampled data is INSERTED properly.

Actions Based on Feedback (Step 2):

- Group 117 acknowledges the comment regarding CHAR(1) for patient_sex. Without additional clarification, it is unknown why it may not be okay to use that datatype. In these clinical trials, sex will be declared with either an "M" or an "F" and is strictly related to biological sex.
- To avoid confusion with the naming of *employee_role* in the *employees_supporting_clinical_trials* table, *employee_role* has been renamed to *employee_trial_role* to indicate this references their role in support of the trial, not their position with their employer.
- Group 117 has decided to change the employer attribute in the *employees* entity to a FK*, employer_id*, and to make new entity named *employers*. This allows for better 1NF normalization. This reduces employer name input for each employee and ensures an easy update should an employer's name change.
- Group 117 notes that no CASCADE operations were set (set to NO ACTION) in the DDL.sql file. CASCADE ON DELETE will be added to the *employees_supporting_clinical_trials* so that the employee role associated with the employee id will also delete should an employee be deleted. Additionally, group 117 will seek additional clarification on project requirements and whether other CASCADE operations are required.

Project Outline:

Health Oncology Group (HOG) is a National Cancer Institute (NCI) funded organization that conducts clinical trials of cancer treatments in adults. Increased participation from sites and patients in their clinical trials increased funding from the NCI and allowed HOG to expand their network. During the transition from paper to electronic records, a group of clinical trial recruiters have asked the software engineers to implement a database driven web application that centralizes the records of patients, personnel and sites associated with the clinical trials to keep up with the expansion.

Currently there are 25 clinical trials, with 100 participating hospitals/sites. On average each clinical trial plans to accrue a total of 300 patients over the length of approximately 2 years. Employees of multiple organizations can support the trials, and each clinical trial has at least one employee assigned to the trial; however, an employee may be assigned to work on multiple clinical trial projects. Some hospitals may also have multiple clinical trials occurring at their site; however, a patient at a hospital may only be enrolled in one clinical trial at a time.

HOG hopes to maintain a timeless database system that can track the accrual of patients for clinical trials and that can provide a current employee directory for staff to use over the life of the clinical trials. The electronic database will decrease the amount of paper records required and provide a central hub of information that will be immensely helpful for clinical trial recruiters.

Database Outline:

As a result of the normalization process, a previously included attribute under clinical_trials, *investigational_treatment*, has been removed. Ultimately, the treatment is comprised of multiple drugs and multiple phases, which would have resulted in repeating groups. Instead of breaking this into additional entities, we decided to remove the attribute altogether, as we considered the needs of a recruiter, and ultimately, this would be extraneous information for them.

clinical_trials:

Assigned Team Member: Thomas Gathman

Purpose:

An entity that stores information about the clinical trial including the trial id, type of cancer, investigational drug being used, trial description, and accrual goal (which is the maximum number of patients allowed on a trial).

Attributes:

- clinical_trial_id — Primary Key: VARCHAR — NN
- cancer_type — VARCHAR — NN
- trial_description — VARCHAR — NN
- maximum_patients — INT — NN

Relationship:

- 1:M relationship between clinical_trials and patients with *clinical_trials_clinical_trial_id* as a FK inside of patients; A clinical trial may have 0 or more patients.
- M:M relationship between clinical_trials and hospitals; *clinical_trials_clinical_trial_id* and *hospitals_hospital_id* are FK in hospitals_supporting_clinical_trials, an intersection table. Many clinical trials may be conducted by many hospitals.
- 1:M relationship between clinical_trials and employees_supporting_clinical_trials with *clinical_trials_clinical_trial_id* as a FK inside of *employees_supporting_clinical_trials*; A clinical trial must have at least one employee in some supportive role.
- M:M Relationship between clinical_trials and employees with *clinical_trials_clinical_trial_id* and *employees_employee_id* as FKs in employees_supporting_clinical_trials, an intersection table turned entity due to the addition of an attribute; Many clinical trials must have at least one employee in a supportive role.

employees_supporting_clinical_trials

Assigned Team Member: Thomas Gathman

Purpose:

An entity that stores information about the clinical trial and serves as the intersection table between clinical_trials and employees due to the M:M relationship. This entity will have a composite attribute, *employee_trial_role*, that will allow an employee to have a specific role in one trial, while allowing for another role in another trial. This allows for the role to not be permanently tied to the employee entity.

Attributes:

- employees_employee_id — Foreign Key: INT — NN
- clinical_trials_clincal_trial_id — Foreign Key: VARCHAR — NN
- employee_trial_role — VARCHAR

Relationship:

M:1 Relationship between employees_supporting_clincal_trials and clinical_trials; many employees will have at most one role in any given clinical trial.

M:1 Relationship between employees_supporting_clinical_trials and employees; many employee roles may be assigned to any given employee.

employees:

Assigned Team Member: Thomas Gathman

Purpose:

An entity that stores information about the employee including the employee id, employee first and last name, employee's position, employee's contact information, and employee's employer organization.

Attributes:

- employee_id — Primary Key: INT — NN
- employee_first_name — VARCHAR — NN
- employee_last_name — VARCHAR — NN
- position — VARCHAR — NN
- email — VARCHAR
- desk_phone — VARCHAR
- employers_employer_id — Foreign Key: INT — NN

Relationship:

- 1:M relationship between employees and employees_supporting_clinical_trials with *employees_employee_id* as a FK in employees_suporting_clinical_trials; An employee may have many roles, but only one role in a given clinical trial.
- M:M relationship between employees and clinical_ trials with *clinical_trials_clinical_trial_id* and *employees_employee_id* as FKs in employees_supporting_clinical_trials, an intersection table turned entity due to the addition of an attribute; Many employees may be working on 0 or more clinical trials.
- M:1 relationship between employees and employers; Many employees must work for exactly one employer.

employers:

Assigned Team Member: John Wong

Purpose:

An entity that stores information about the employer including the employer id and employer name.

Attributes:

- employer_id — Primary Key: INT — NN
- employer_name — VARCHAR — NN

Relationship:

- 1:M relationship between employers and employees with *employers_employee_id* as a FK in employees; An employer will have at least one employee.

patients:

Assigned Team Member: John Wong

Purpose:

An entity that stores information about the patient including their id, first and last name, address, and date of birth.

Attributes:

- patient_id — Primary Key: INT — NN
- patient_first_name — VARCHAR — NN
- patient_last_name — VARCHAR — NN
- patient_street — VARCHAR — NN
- patient_city — VARCHAR — NN
- patient_state — VARCHAR — NN
- patient_zip — VARCHAR — NN
- patient_sex — CHAR — NN
- dob — DATE — NN
- hospitals_hospital_id — Foreign Key: INT — NN

- clinical_trials_clinical_trial_ID VARCHAR NN

Relationship:

- M:1 relationship between patients and hospitals; Many patients must each go to exactly one hospital.
- M:1 relationship between patients and clinial_trials; Many patients must each be on exactly one clinical trial.

hospitals:

Assigned Team Member: John Wong

Purpose:

An entity that stores information about the hospital or site including the hospital id, name, and address.

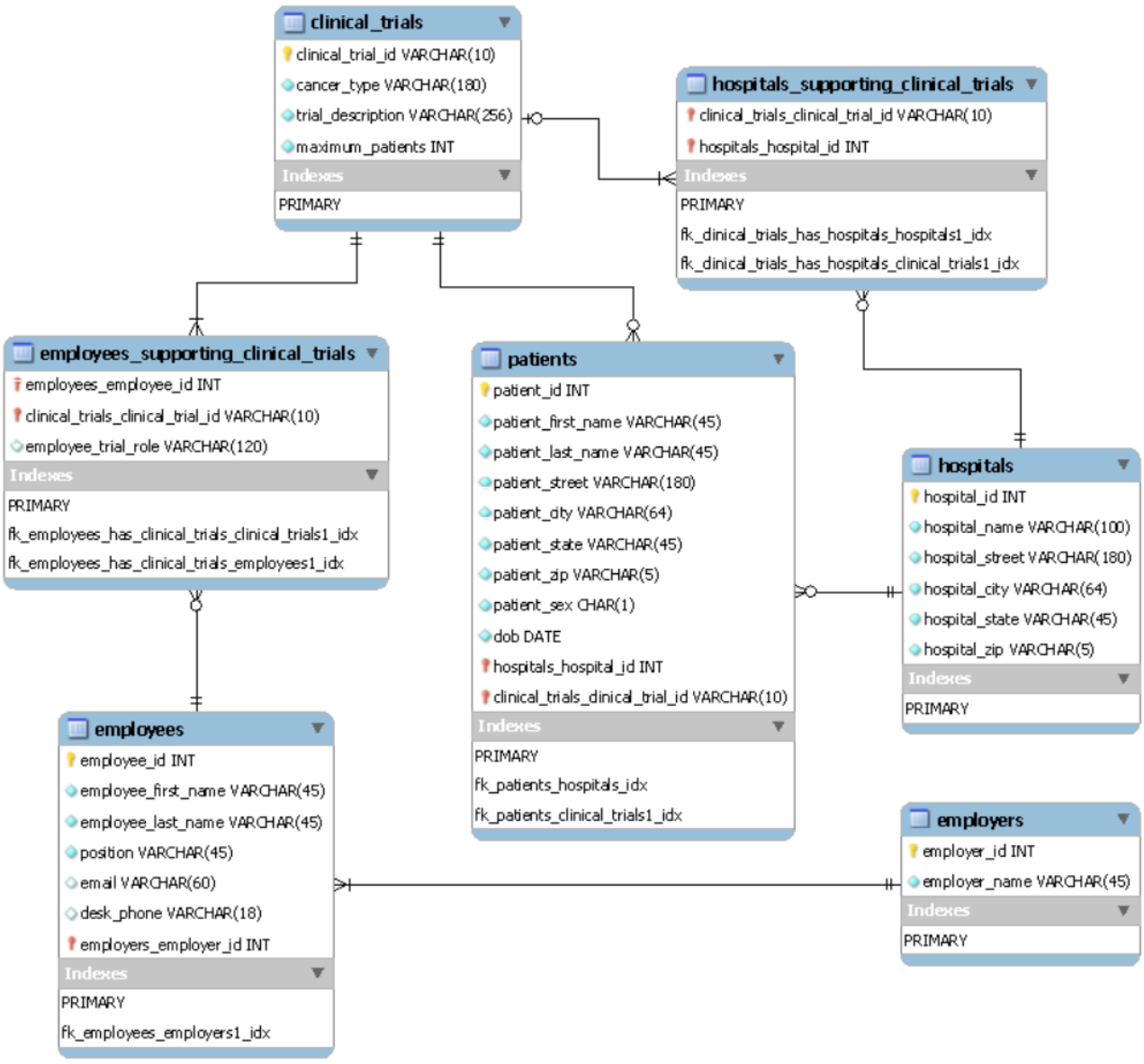
Attributes:

- hospital_id Primary Key: INT NN, AI
- hospital_name VARCHAR NN
- hospital_street VARCHAR NN
- hospital_city VARCHAR NN
- hospital_state VARCHAR NN
- hospital_zip VARCHAR NN

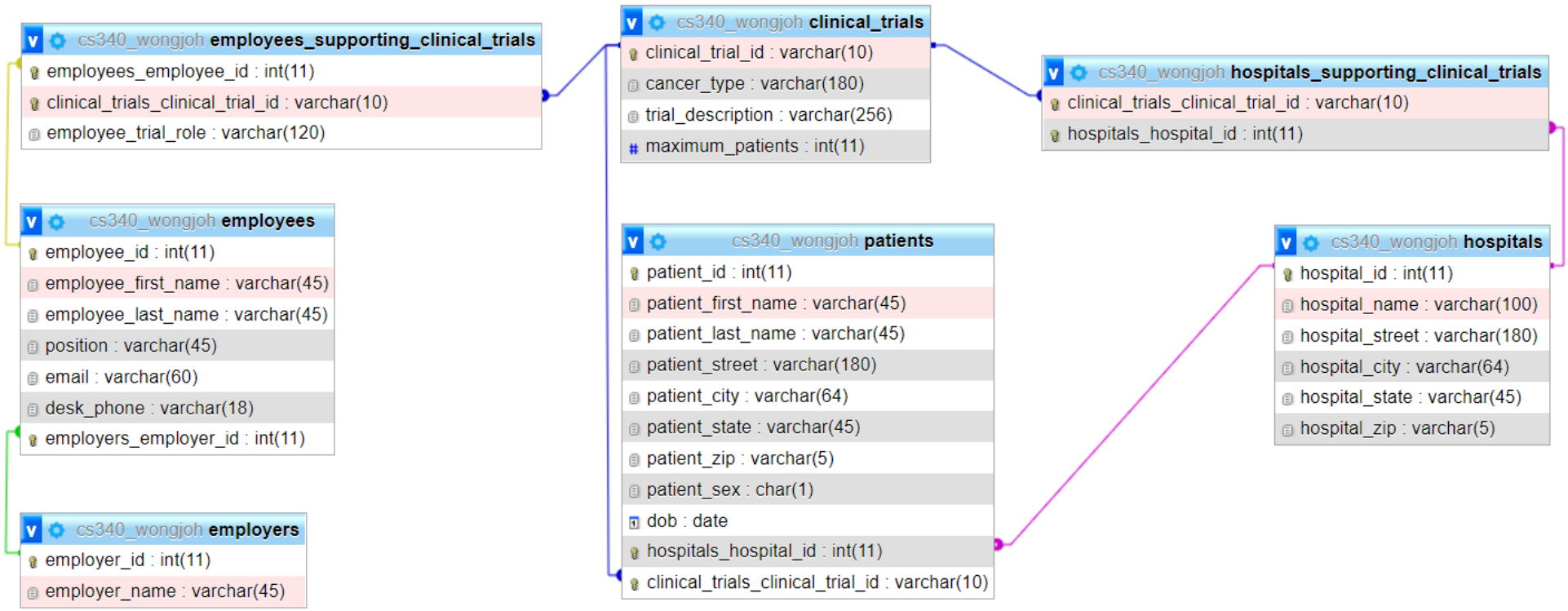
Relationship:

- 1:M relationship between hospitals and patients, with *hospitals_hospital_id* as a FK in patients; A hospital may have many patients.
- M:M relationship between hospitals and clinical trials, with *hospitals_hospital_id* and *clincical_trials_clinical_trial_id* as a FKs in hospitals_supporting_clinical_trials, an intersection table; Many hospitals may have 0 or more clinical trials.

Entity – Relationship Diagram:



Schema:


cs340_wongjoh employees_supporting_clinical_trials
employees_employee_id : int(11)
clinical_trials_clinical_trial_id : varchar(10)
employee_trial_role : varchar(120)
cs340_wongjoh clinical_trials
clinical_trial_id : varchar(10)
cancer_type : varchar(180)
trial_description : varchar(256)
maximum_patients : int(11)
cs340_wongjoh hospitals_supporting_clinical_trials
clinical_trials_clinical_trial_id : varchar(10)
hospitals_hospital_id : int(11)
cs340_wongjoh employees
employee_id : int(11)
employee_first_name : varchar(45)
employee_last_name : varchar(45)
position : varchar(45)
email : varchar(60)
desk_phone : varchar(18)
employers_employer_id : int(11)
cs340_wongjoh patients
patient_id : int(11)
patient_first_name : varchar(45)
patient_last_name : varchar(45)
patient_street : varchar(180)
patient_city : varchar(64)
patient_state : varchar(45)
patient_zip : varchar(5)
patient_sex : char(1)
dob : date
hospitals_hospital_id : int(11)
clinical_trials_clinical_trial_id : varchar(10)
cs340_wongjoh hospitals
hospital_id : int(11)
hospital_name : varchar(100)
hospital_street : varchar(180)
hospital_city : varchar(64)
hospital_state : varchar(45)
hospital_zip : varchar(5)
cs340_wongjoh employers
employer_id : int(11)
employer_name : varchar(45)

Sample Data:

clinical_trials

clinical_trial_id	cancer_type	trial_description	maximum_patients
HG1141	Breast	This randomized phase II trial studies how well abbreviated breast magnetic resonance imaging (MRI) and digital tomosynthesis mammography work in detecting cancer in women with dense breasts.	1500
HG8143	Kidney	This phase III trial compares nephrectomy (surgery to remove a kidney or part of a kidney) with nivolumab to the usual approach of nephrectomy followed by standard post-operative follow-up and monitoring.	750
HGG173	Myeloma	This phase III trial studies how well lenalidomide and dexamethasone works with or without daratumumab in treating patients with high-risk smoldering myeloma.	300

hospitals_supporting_clinical_trials

clinical_trials_clinical_trial_id	hospitals_hospital_id
HG1141	1
HG1141	3
HG8143	2
HGG173	3

hospitals

hospital_id	hospital_name	hospital_street	hospital_city	hospital_state	hospital_zip
1	Baylor University Medical Center	3500 Gaston Ave	Dallas	TX	75246
2	Case Western Reserve University	9501 Euclid Ave	Cleveland	OH	44106
3	Geisinger Medical Center	100 N Academy Ave	Danville	PA	17821

patients

patient_id	patient_first_name	patient_last_name	patient_street	patient_city	patient_state	patient_zip	patient_sex	dob	hospitals_hospital_id	clinical_trials_clinical_trial_id
10001	Gaviin	Jep	9020 Garland Rd	Dallas	TX	75218	M	11/1/2000	1	HG1141
13001	Herbie	Williams	2136 Murray Hill Rd	Cleveland	OH	44106	M	9/21/1995	2	HG8143
38001	Terry	Noel	502 Church St	Danville	PA	17821	F	2/28/2000	3	HGG173
38002	Paden	Heidi	331 W Mahoning St	Danville	PA	17821	F	8/27/1984	3	HG1141

employees_supporting_clinical_trials

employees_employee_id	clinical_trials_clinical_trial_id	employee_trial_role
1	HG1141	Recruiter
2	HG8143	Project Coordinator
3	HGG173	Data Manager
4	HG1141	Study Chair
5	HG8143	Committee Chair
6	HGG173	Study Co-Chair

employees

employee_id	employee_first_name	employee_last_name	position	email	desk_phone	employer
1	Regina	Dutch	Recruiter	dregina@hog.com	402-250-8329	Health Oncology Group
2	Calvin	Jonas	Project Coordinator	cjonas@hog.com	714-338-2340	Health Oncology Group
3	Haylie	Sandra	Data Manager	hsandra@hog.com	832-209-6628	Health Oncology Group
4	Steve	Glenna	M.D.	sglenna@case.edu	775-981-5220	Case Western Reserve University
5	Deitra	Gracie	M.D.	dgracie@baylor.edu	561-552-5564	Baylor University Medical Center
6	Alex	King	M.D.	aking@geisinger.edu	914-774-0513	Geisinger Medical Center

employers

employer_id	employer_name
0	Health Oncology Group
1	Case Western Reserve University
2	Baylor University Medical Center
3	Geisinger Medical Center