

Introduction to Data Mining : AI2002

Clinical Trials Database Management System



Entity Relationship Diagram



Tables (.schema)

```
CREATE TABLE IF NOT EXISTS Study (
    NCTID TEXT PRIMARY KEY,
    OrgFullName TEXT,
    BriefTitle CHAR(300),
    Title CHAR(600),
    StudyType TEXT CHECK(StudyType IN
    ('Interventional', 'Observational', 'Expanded
    Access', 'Patient Registry')),
    Status TEXT CHECK(Status IN ('Not yet
    Recruiting', 'Recruiting', 'Active, Not
    Recruiting', 'Completed', 'Withdrawn',
    'Terminated', 'Suspended')),
    StartDate DATE,
    PrimaryCompletionDate DATE,
    StudyCompletionDate DATE,
    IsDeleted BOOLEAN DEFAULT 0
);
```

Type Affinities

Study

Prevent Redundancies

```
CREATE TABLE IF NOT EXISTS Investigator (
    InvestigatorName TEXT PRIMARY KEY,
    Affiliation TEXT,
    Role TEXT CHECK(
        Role IN ('STUDY_CHAIR', 'STUDY_DIRECTOR',
        'PRINCIPAL_INVESTIGATOR', 'SUB_INVESTIGATOR',
        'CONTACT')),
    IsDeleted BOOLEAN DEFAULT 0
    -- Enabling Soft Deletions );
```



Eligibility

Tables (.schema)

```
CREATE TABLE IF NOT EXISTS Investigator (
    InvestigatorName TEXT PRIMARY KEY,
    Affiliation TEXT,
    Role TEXT CHECK(Role IN ('STUDY_CHAIR',
    'STUDY_DIRECTOR', 'PRINCIPAL_INVESTIGATOR',
    'SUB_INVESTIGATOR', 'CONTACT')),
    IsDeleted BOOLEAN DEFAULT 0 );
```

→ Enabled Soft
Deletions

→ Key Constraints

```
CREATE TABLE IF NOT EXISTS StudyCondition (
    StudyNCTID TEXT,
    ConditionName TEXT,
    IsDeleted BOOLEAN DEFAULT 0,
    PRIMARY KEY (StudyNCTID, ConditionName),
    FOREIGN KEY (StudyNCTID) REFERENCES Study(NCTID),
    FOREIGN KEY (ConditionName) REFERENCES
    Condition(Name) );
```

Investigator

Condition

Tables (.schema)

```
CREATE TABLE IF NOT EXISTS Organisation (
    OrgFullName TEXT PRIMARY KEY,
    OrgClass TEXT CHECK(
        OrgClass IN ('NIH', 'FED',
        'OTHER_GOV', 'INDIV', 'INDUSTRY',
        'NETWORK', 'AMBIG', 'UNKNOWN')),
    IsDeleted BOOLEAN DEFAULT 0 );
```

Organisation

```
CREATE TABLE IF NOT EXISTS Intervention (
    ID INTEGER PRIMARY KEY AUTOINCREMENT,
    StudyNCTID TEXT,
    Type TEXT CHECK(Type IN ('Drug', 'Procedure',
    'Diagnostic Test', 'Other', 'Biological',
    'Device', 'Behavioral', 'Radiation', 'Genetic',
    'Combination Product')),
    Description TEXT,
    IsDeleted BOOLEAN DEFAULT 0,
    FOREIGN KEY (StudyNCTID) REFERENCES Study(NCTID)
);
```

Intervention

Check In

Querying the Trials Database

List of all the studies that are interventional and their status along with the organisation class

```
SELECT s.NCTID, s.OrgFullName, s.Status, o.OrgClass  
FROM Study s  
JOIN StudyOrganisation so ON s.NCTID = so.StudyNCTID  
JOIN Organisation o ON so.OrgFullName =  
o.OrgFullName  
WHERE s.StudyType = "Interventional";
```

NCTID	OrgFullName	Status	OrgClass
NCT00381225	Johns Hopkins University	Withdrawn	OTHER_GOV
NCT01007279	University of Roma La Sapienza	Completed	INDUSTRY
NCT01110954	Photonamic GmbH & Co. KG	Terminated	INDUSTRY
NCT01522300	Centre Jean Perrin	Terminated	OTHER_GOV
NCT02376244	Liverpool Heart and Chest Hospital	Completed	NIH
NCT02376244	Liverpool Heart and Chest Hospital	Completed	OTHER_GOV
NCT03777852	University of Sao Paulo	Terminated	OTHER_GOV

Retrieve the top 5 organisations with the most number of studies

```
SELECT OrgFullName, COUNT(*) AS TotalStudies  
FROM StudyOrganisation  
GROUP BY OrgFullName  
ORDER BY TotalStudies DESC  
LIMIT 5;
```

OrgFullName	TotalStudies
Royal Marsden NHS Foundation Trust.	2
Centre Jean Perrin	1
EZUS-LYON 1	1
Edge Hill University	1
Johns Hopkins University	1

Querying the Trials Database

Study IDs with their respective locations in the United Kingdom

```
SELECT s.NCTID, l.LocationName  
FROM Study s  
JOIN Location l ON s.NCTID = l.StudyNCTID  
WHERE l.Country = 'United Kingdom';
```

NCTID	LocationName
NCT02069964	London, United Kingdom, SW3 6JJ Royal Marsden Hospital
NCT02376244	Liverpool, United Kingdom, L14 3PE, Liverpool Heart Hospital

Retrieve the study IDs and study type of the condition with the most number of studies

```
SELECT s.NCTID AS StudyID, c.Name AS ConditionName,  
s.StudyType, s.Status  
FROM Condition c  
JOIN StudyCondition sc ON c.Name = sc.ConditionName  
JOIN Study s ON sc.StudyNCTID = s.NCTID  
WHERE c.Name = ( SELECT c.Name FROM Condition c  
JOIN StudyCondition sc ON c.Name =  
sc.ConditionName  
GROUP BY c.Name ORDER BY COUNT(sc.StudyNCTID)  
DESC  
LIMIT 1 );
```

StudyID	ConditionName	StudyType	Status
NCT02060643	Head Tumour	Observational	Completed
NCT02069964	Head Tumour	Observational	Completed

Optimisation : Explain Query Plan and Index

```
CREATE INDEX IF NOT EXISTS idx_study_type ON Study(StudyType);
CREATE INDEX IF NOT EXISTS idx_study_status ON Study(Status);
CREATE INDEX IF NOT EXISTS idx_study_org ON Study(OrgFullName);
CREATE INDEX IF NOT EXISTS idx_eligibility_study_id ON Eligibility(StudyNCTID);
CREATE INDEX IF NOT EXISTS idx_study_condition_study_id ON StudyCondition(StudyNCTID);
CREATE INDEX IF NOT EXISTS idx_study_condition_condition_name ON StudyCondition(ConditionName);
CREATE INDEX IF NOT EXISTS idx_outcome_study_id ON Outcome(StudyNCTID);
CREATE INDEX IF NOT EXISTS idx_study_investigator_study_id ON StudyInvestigator(StudyNCTID);
CREATE INDEX IF NOT EXISTS idx_location_country ON Location(Country);
```

QUERY PLAN

```
|--SEARCH s USING INDEX idx_study_type (StudyType=?)
|--SEARCH so USING COVERING INDEX sqlite_autoindex_StudyOrganisation_1 (StudyNCTID=?)
`--SEARCH o USING INDEX sqlite_autoindex_Organisation_1 (OrgFullName=?)
```

QUERY PLAN

```
|--SCAN sc USING COVERING INDEX sqlite_autoindex_StudyCondition_1
|--SEARCH s USING COVERING INDEX sqlite_autoindex_Study_1 (NCTID=?)
|--SEARCH e USING INDEX sqlite_autoindex_Eligibility_1 (StudyNCTID=?)
`--SEARCH c USING COVERING INDEX sqlite_autoindex_Condition_1 (Name=?)
```

Optimisation : Explain Query Plan and Index

QUERY PLAN

```
|--SCAN si USING COVERING INDEX sqlite_autoindex_StudyInvestigator_1
|--SEARCH s USING COVERING INDEX sqlite_autoindex_Study_1 (NCTID=?)
|--SEARCH o USING INDEX sqlite_autoindex_Outcome_1 (StudyNCTID=?)
`--SEARCH i USING COVERING INDEX sqlite_autoindex_Investigator_1 (InvestigatorName=?)
```

QUERY PLAN

```
|--SEARCH l USING INDEX idx_location_country (Country=?)
`--SEARCH s USING COVERING INDEX sqlite_autoindex_Study_1 (NCTID=?)
```

QUERY PLAN

```
|--SEARCH c USING COVERING INDEX sqlite_autoindex_Condition_1 (Name=?)
|--SCALAR SUBQUERY 1
|   |--SCAN c USING COVERING INDEX sqlite_autoindex_Condition_1
|   |--SEARCH sc USING AUTOMATIC COVERING INDEX (ConditionName=?)
|   `--USE TEMP B-TREE FOR ORDER BY
|--SEARCH sc USING INDEX idx_study_condition_condition_name (ConditionName=?)
|--REUSE SUBQUERY 1
`--SEARCH s USING INDEX sqlite_autoindex_Study_1 (NCTID=?)
```

Optimisation : View and Queries

Active Studies and their details with Organizations

CREATE VIEW IF NOT EXISTS

```
ActiveStudiesWithOrganizations AS
SELECT s.NCTID, s.OrgFullName, s.BriefTitle, s.Title,
s.StudyType, s.Status, s.StartDate,
s.PrimaryCompletionDate, s.StudyCompletionDate,
o.OrgClass
FROM Study s
JOIN StudyOrganisation so ON s.NCTID = so.StudyNCTID
JOIN Organisation o ON so.OrgFullName = o.OrgFullName
WHERE s.Status = 'Active, Not Recruiting';
```

Interventions with Corresponding Study Titles

CREATE VIEW IF NOT EXISTS

```
InterventionsWithStudyTitles AS
SELECT i.*, s.Title
FROM Intervention i
JOIN Study s ON i.StudyNCTID = s.NCTID;
```

ID	StudyNCTID	Type	Description	Title
1	NCT01110954	Drug	PD L 506	Evaluation of the Suitability of PD L 506 for Intraoperative Visualisation of Palpable and Nonpalpable Breast Cancer Tissue
2	NCT01522300	Other	Evaluation of the Tomosynthesis	Etude Tomos- Apport de la Tomosynthèse Dans le Bilan dExtension locorégional préthérapeutique dUne Tumeur du Sein : Recherche de multicentricité
3	NCT01522300	Other	Evaluation of a medical device: the tomosynthesis	Etude Tomos- Apport de la Tomosynthèse Dans le Bilan dExtension locorégional préthérapeutique dUne Tumeur du Sein : Recherche de multicentricité
4	NCT03777852	Diagnostic Test	First Questionnaire application	Emotional Evaluation and Reconstructed Breast Satisfaction After Breast Cancer Treatment: Breast Q Questionary
5	NCT03777852	Diagnostic Test	Second Questionnaire application	Emotional Evaluation and Reconstructed Breast Satisfaction After Breast Cancer Treatment: Breast Q Questionary
6	NCT00381225	Procedure	Ultra-sound guided radio-frequency ablation	Feasibility of Ultra-sound Guided Radiofrequency Ablation in the Management of Head and Neck Neoplasia.
7	NCT02967965	Other	Imagery	CARIM Cohort (CARDioprotection in Myocardial Infarction)
8	NCT01007279	Drug	ROUVASTATIN	Rosuvastatin Pretreatment in Patients Undergoing Elective PCI to Reduce the Incidence of Myocardial Periprocedural Necrosis
9	NCT02376244	Other	High intensity interval training (HIIT)	A Randomised Controlled Trial on the Efficacy of High Intensity Interval Training Compared With Moderate Intensity Interval Training for Patients Undergoing Cardiac Rehabilitation
10	NCT02376244	Other	Standard care	A Randomised Controlled Trial on the Efficacy of High Intensity Interval Training Compared With Moderate Intensity Interval Training for Patients Undergoing Cardiac Rehabilitation

Optimisation : View and Queries

Studies with Eligibility Criteria and Conditions

```
CREATE VIEW IF NOT EXISTS
StudiesWithEligibilityAndConditions AS
SELECT s.* , e.InclusionCriteria, e.ExclusionCriteria,
c.Name AS Condition
FROM Study s
JOIN Eligibility e ON s.NCTID = e.StudyNCTID
JOIN StudyCondition sc ON s.NCTID = sc.StudyNCTID
JOIN Condition c ON sc.ConditionName = c.Name;
```

Studies with Outcomes and Primary Investigators

```
CREATE VIEW IF NOT EXISTS
StudiesWithOutcomesAndInvestigators AS
SELECT s.* , o.PRIMARYOutcomeMeasure,
o.SecondaryOutcomeMeasure, i.InvestigatorName
FROM Study s
JOIN Outcome o ON s.NCTID = o.StudyNCTID
JOIN StudyInvestigator si ON s.NCTID = si.StudyNCTID
JOIN Investigator i ON si.InvestigatorName =
i.InvestigatorName;
```

Common Table Expression

```
WITH MyCTE AS (
    SELECT * FROM StudiesWithEligibilityAndConditions
    WHERE InclusionCriteria LIKE '%diabetes%'
    ORDER BY StartDate DESC
)
SELECT * FROM MyCTE
LIMIT 10;
```

Optimisation : Soft Deletions and Triggers

```
CREATE TRIGGER Trigger_Study_Soft_Deletion
INSTEAD OF DELETE ON "ActiveStudiesWithOrganizations"
FOR EACH ROW
BEGIN
    UPDATE Study
    SET IsDeleted = 1
    WHERE NCTID = OLD.NCTID;
END;

DELETE FROM "ActiveStudiesWithOrganizations"
WHERE NCTID = 'NCT02967965';

SELECT NCTID, IsDeleted FROM Study;
```

	NCTID	IsDeleted
1	NCT01110954	0
2	NCT01522300	0
3	NCT03777852	0
4	NCT02060643	0
5	NCT02069964	0
6	NCT00381225	0
7	NCT02967965	0
8	NCT01007279	0
9	NCT02376244	0

	NCTID	IsDeleted
1	NCT01110954	0
2	NCT01522300	0
3	NCT03777852	0
4	NCT02060643	0
5	NCT02069964	0
6	NCT00381225	0
7	NCT02967965	1
8	NCT01007279	0
9	NCT02376244	0

Optimisation : Transactions (Commit)

```
BEGIN TRANSACTION;  
INSERT INTO Study (NCTID, OrgFullName, BriefTitle, StudyType, Status, StartDate, PrimaryCompletionDate,  
StudyCompletionDate)  
VALUES ('NCT000001', 'National Cancer Institute (NCI)', 'Study of Tamoxifen and Raloxifene (STAR)', 'Study of  
Tamoxifen and Raloxifene (STAR)', 'Interventional', 'Active, Not Recruiting', '1999-06-01', '2006-06-01',  
'2012-06-01');  
COMMIT;
```

NCTID	OrgFullName	BriefTitle	StudyType	Status	StartDate	PrimaryCompletionDate	StudyCompletionDate
NCT03777852	University of Sao Paulo	Emotional Evaluation and Reconstructed Breast Satisfaction	Interventional	Terminated	2019-07-01	2021-02-28	2022-02-28
NCT02967965	EZUS-LYON 1	CARDioprotection in Myocardial Infarction	Observational	Active, Not Recruiting	2013-11-30	2022-12-31	2024-12-31
NCT02376244	Liverpool Heart and Chest Hospital NHS Foundation Trust	The Health Impact of High Intensity Exercise Training With Intervals During Cardiac Rehabilitation	Interventional	Completed	2015-09	2016-03	2016-03
NCT02069964	Royal Marsden NHS Foundation Trust	A Prospective Longitudinal Cohort Study Looking at the Difference in Carotid Intima-medial Thickness	Observational	Completed	2011-10	2017-10	2017-10
NCT02060643	Royal Marsden NHS Foundation Trust	A Cross-sectional Study Looking at the Effect of Radiotherapy on Carotid Intima-medial Thickness in Head and Neck Cancer	Observational	Completed	2011-10	2013-11	2013-11
NCT01522300	Centre Jean Perrin	Etude Tomos- Apport de la Tomosynthèse Dans le Bilan dExtension locorégional préthérapeutique dUne Tumeur du Sein	Interventional	Terminated	2012-01	2013-08	2013-08
NCT01110954	Photonamic GmbH & Co. KG	Fluorescence-guided Resection in Breast Cancer	Interventional	Terminated	2010-05-31	2012-07-19	2012-12-20
NCT01007279	University of Roma La Sapienza	Rosuvastatin in Preventing Myonecrosis in Elective Percutaneous Coronary Interventions (PCIs)	Interventional	Completed	2010-03	2011-09	2011-09
NCT00381225	Johns Hopkins University	Feasibility of Ultra-sound Guided Radiofrequency Ablation in the Management of Head and Neck Neoplasia.	Interventional	Withdrawn	2006-11	2008-03-18	2008-03-18



NCTID	OrgFullName	BriefTitle	StudyType	Status	StartDate	PrimaryCompletionDate	StudyCompletionDate
NCT03777852	University of Sao Paulo	Emotional Evaluation and Reconstructed Breast Satisfaction	Interventional	Terminated	2019-07-01	2021-02-28	2022-02-28
NCT02967965	EZUS-LYON 1	CARDioprotection in Myocardial Infarction	Observational	Active, Not Recruiting	2013-11-30	2022-12-31	2024-12-31
NCT02376244	Liverpool Heart and Chest Hospital NHS Foundation Trust	The Health Impact of High Intensity Exercise Training With Intervals During Cardiac Rehabilitation	Interventional	Completed	2015-09	2016-03	2016-03
NCT02069964	Royal Marsden NHS Foundation Trust	A Prospective Longitudinal Cohort Study Looking at the Difference in Carotid Intima-medial Thickness	Observational	Completed	2011-10	2017-10	2017-10
NCT02060643	Royal Marsden NHS Foundation Trust	A Cross-sectional Study Looking at the Effect of Radiotherapy on Carotid Intima-medial Thickness in Head and Neck Cancer	Observational	Completed	2011-10	2013-11	2013-11
NCT01522300	Centre Jean Perrin	Etude Tomos- Apport de la Tomosynthèse Dans le Bilan dExtension locorégional préthérapeutique dUne Tumeur du Sein	Interventional	Terminated	2012-01	2013-08	2013-08
NCT01110954	Photonamic GmbH & Co. KG	Fluorescence-guided Resection in Breast Cancer	Interventional	Terminated	2010-05-31	2012-07-19	2012-12-20
NCT01007279	University of Roma La Sapienza	Rosuvastatin in Preventing Myonecrosis in Elective Percutaneous Coronary Interventions (PCIs)	Interventional	Completed	2010-03	2011-09	2011-09
NCT00381225	Johns Hopkins University	Feasibility of Ultra-sound Guided Radiofrequency Ablation in the Management of Head and Neck Neoplasia.	Interventional	Withdrawn	2006-11	2008-03-18	2008-03-18
NCT000001	National Cancer Institute (NCI)	Study of Tamoxifen and Raloxifene (STAR)	Interventional	Active, Not Recruiting	1999-06-01	2006-06-01	2012-06-01

Optimisation : Transactions (Rollback)

```
BEGIN TRANSACTION;  
UPDATE Study  
SET Status = 'Erroneous'  
WHERE NCTID = 'NCT000001';  
ROLLBACK;
```

```
sqlite> BEGIN TRANSACTION;  
sqlite> UPDATE Study  
...> SET Status = 'Erroneous'  
...> WHERE NCTID = 'NCT000001';  
Runtime error: CHECK constraint failed: Status IN ('Not yet Recruiting', 'Recruiting', 'Active, Not  
Recruiting', 'Completed', 'Withdrawn', 'Terminated', 'Suspended') (19)  
sqlite> ROLLBACK;
```

Thank You

