# Clinical Trial Database Design

**BMI701 Midterm Project**

**Linzi Yu**

**11/1/2020**

## 1 Clinical Trial Database Overview

Clinical databases systems are evolving at a high speed because of the complexity and rapid expansion of the domain of clinical data[1]. In terms of clinical trial management systems (CTMS), they provide a way to find all the related information about a particular clinical study.

Some websites support clinical trial data exploration. For instance, *ClinicalTrials.gov* provides the public with a free access to clinical information on studies for a wide range of diseases and conditions. WHO maintains a universal portal called *International Clinical Trials Registry Platform (ICTRP). NIH-National Cancer Institute* also keeps a searchable database of federally and privately supported clinical trials conducted in the US and around the world. These databases have been well established and can be taken good advantage of while searching for multiple information base on users need. An ideal clinical trial database should include multiple medical-related entities to provide users (i.e. clinical practitioners, patients) with wide accesses to the up-to-date clinical information.

## 2 Requirements Analysis

### 2.1 User Identification

Clinical trial database is a clinical trial centric model focusing on the diverse needs from users. Its target users are people who concerned with real-world clinical or medicine information, i.e. researchers, clinical practitioners, healthcare company members, patients or any other stakeholders.

### 2.2 Possible Queries

Clinical trial database designed here is to answer users' diverse question about clinical trials, drugs and indications, pharmacy companies, hospital or organizations and contact informations. For example: What phase or status is that clinical trial? What are the organizations that are implementing the research? Which company develop a specific drug? What indication of the drug is approved by FDA?

### 2.3 Information Provided

In order to give answers to the questions, the clinical database should at least contain the information about the clinical trial, hospital or organizations, companies, drugs, indications and the contacts, each should have several descriptive attributes along with it.

### 2.4 Database Maintain and Update

As a clinical trial going on, some elements should be frequently updated, e.g., the status of the clinical trial (not yet recruiting, recruiting, active but not recruiting, completed, withdraw or terminated), the number of the enrollment for each trial. In addition, some other information could be modified if information changes, e.g., contact information, newly approved drugs, new indication of drugs etc.

### 2.5 User Story Examples

A

Julia from Arizona was diagnosed with breast cancer in 2018, she received line1 therapy before, but the disease recurred unfortunately. In order to seek for better therapies, he needs to search information of some breast cancer related ongoing clinical trials. It would better if the hospital which implement the clinical trial is close to where she lives.

*User requirement analysis:*

- *First, Julia needs to identify clinical trial ID that can meet her criteria: the trial should be recruiting, the tested drugs should be used for second line breast cancer treatment, also the organizations which conduct the clinical trial should locate at Arizona, so it would be more convenient for Julia to receive treatment and get surveillance near her place.*
- *After identify some IDs of clinical trial candidates, she needs to find out the contact information about these trials, for example, the names of the contact person, the contact information, the location of the organization and so on.*
- *Once Julia sign up for one of the clinical trials, the number of enrollment needs to be updated for that particular trial.*

B

Emma is a 5th year PhD student at a prestigious molecular biology institute, her research focus on developing biomarkers for human diseases. She wants to find a job in the US after graduation this year, but she doesn't really care about relocating. She is interested in a position at R&D department of a company, so she needs to find some basic information about related companies in advance.

*User requirement analysis:*

- *Emma needs to find some biotech companies that meet her criteria: the company should at least developing or registered drugs that is related to her research field, also she would be satisfied if the company locates in the US*
- *She needs to identify the drug type of her interest first, and then find a list of corresponding company which developed that drug.*

C

Robert is the principle of a pharmacy company, he is in charge of clinical trial design for new drugs. Before the company pushes forward their marketing strategy, he want to know some basic info about their potential competitors in the same area.

*User requirement analysis:*

- *In this case, Robert needs to use the clinical trial database to identify the clinical trials that have completed of cancer drugs, and also get more information about the trial.*
- *All he needs to do is to look at the clinical trial information, filter the status of the trial and check the NCT number of the trials of his interest.*
- *Also, he can update the status of his trial at his company once it's done, by switch the "status" from "recruiting" to "completed".*

## 3 Database Design

Take breast cancer clinical trial as an example, specifically ones that are associated with ADC (antibody-drug conjugate) drugs. Design a simplified clinical trial database with normalization in order to reduce redundancy, improve the integrity and meet the need of granularity.

### 3.1 Table 1: ClinicalTrial

*Main table of the database. A subset of first 100 rows is visible to users*

- Trial_ID* (int): Each clinical trial is given an ID in this database, it's the primary key
- NCT_Num (varchar): Each clinical trial has an registered number, can be found at ClinicalTrial.gov
- Disease (varchar): The disease, disorder, syndrome, illness, or injury that is being studied

- Treatment (varchar): Specific usage of treatment, e.g., neo-adjuvant, adjuvant, line1 or line2 therapy for late stage patients
- Phase (int): Phase of the clinical trial, usually 1, 2, 3 sometimes 4
- Enroll_Num (int): Number of patients enrolled
- Status (varchar): Status of clinical trial, not yet recruiting, recruiting, active but not recruiting, completed, withdraw or terminated
- PostDate (date): Date of the clinical trial was first posted

### 3.2 Table 2: Organization

*In fact, most of the clinical trials can be conducted synchronously by multiple organizations. In addition, there may be the same organization name, so here each organization is encoded with a primary key Org_ID*

- Org_ID* (int): Primary key
- Organization (varchar): Name of the organization
- Country (varchar): Which country the organization is located in
- State (varchar): Which state the organization is located in
- PostCode (varchar): Post code information with limitation of 10 charaters. Because the format of the postcode in different counties are quite different, so we set the type of postcode to various character with a limtation of 10 digits. If all the organizations are in the US, then we can also set varchar(5) instead.

### 3.3 Table 3: Contacts

*Each clinical trial can have multiple contacts corresponding to the organization who conducts the trial. Trial_ID and Affiliation compose a composite primary key for this table, also create **index** for these two primary keys in order to speed up searches/queries*

- Trial_ID** (int): A foreign key, directing to Trial_ID at ClinicalTrial table
- Contact (varchar): Name of contact, who's responsible for the clinical trial of an organization
- Sex (varchar): F/M
- Phone (varchar): Phone number
- Email (varchar): Email information
- Affiliation** (int): A foreign key, directing to the id of his/her organization at the table of organization

### 3.4 Table 4: Drugs

*Each clinical trial can test multiple drugs, this table store information about those drugs*

- Drug *(varchar): Name of drug, primary key
- Company **(varchar): Company who developed the drug, a foreign key that directs to the table of companies
- Type (varchar): Biological or chemical etc.
- Target (varchar): Target of the drug, in BC for instance, HER2

### 3.5 Table 5: TrialDrugs

*Each trial may test multiple drugs while each drug may have more than one clinical trial that has been registered. In order to meet the normalization needs and reduce data redundancy, a bridge table named TrialDrugs is created. Both columns in this table are foreign keys, they are composite key, also create **index** for these two primary keys in order to speed up searches/queries*

- Drug **(varchar): Name of drug, foreign key
- Trial_ID **(int): Trial ID, foreign key

### 3.6 Table 6: Indication

*Some drugs through approved by FDA, still looking forward to expand their usage or entering the market of other countries. So they may either alternate the indication or disease of the clinical trial or implement their clinical trial at countries that can expand their lines of business. That's why one drug may have more than one indications, also one indication may have multiple effective drugs. In this table, these indications can either be approved by FDA or not*

- Drug **(varchar): Foreign key, directing to the drug name, create **index** on indication to speed up searches/queries
- Indication *(varchar): Together with drug column, form a composite primary key
- FDA (varchar): If an indication of the drug has got approved, this column is not null, otherwise it can be null
- MOD_ID (varchar): An oncology column for indications, Mondo Disease Ontology.

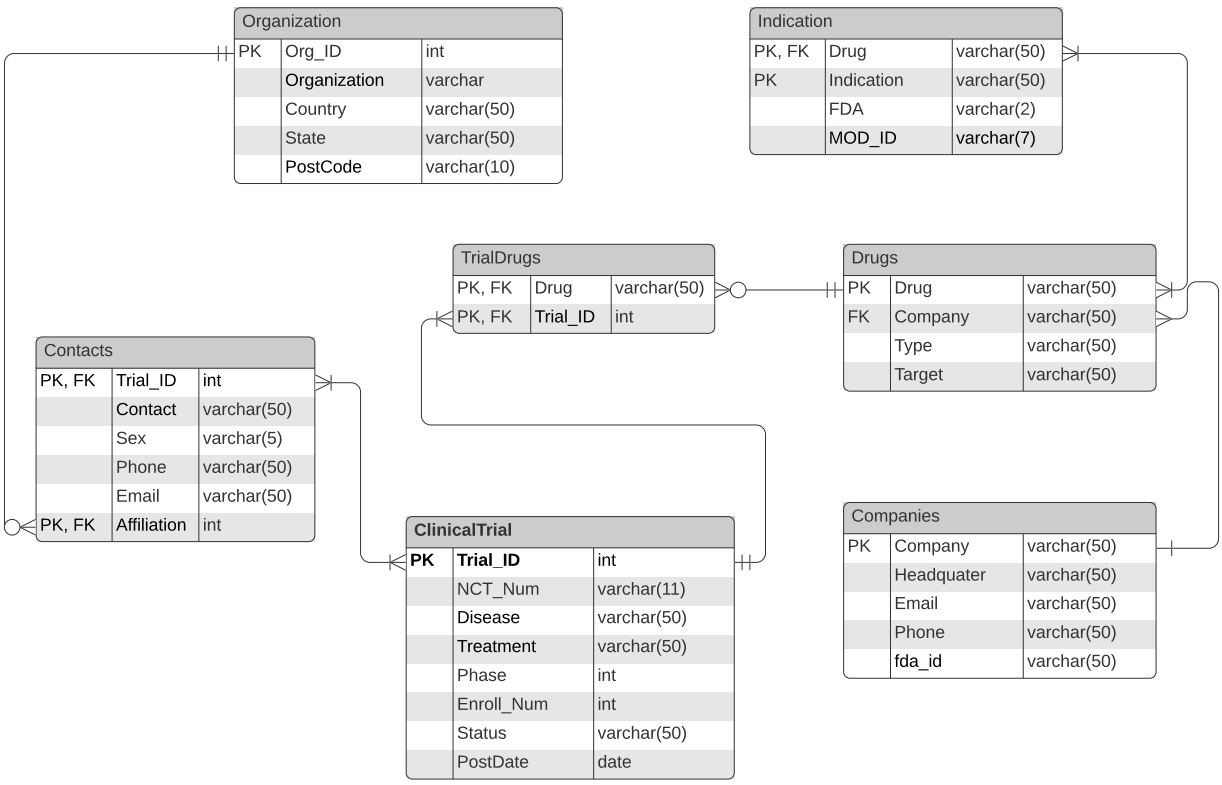
### 3.7 Table 7: Companies

*Company developed the drugs, so it is important to include the information about the companies*

- Company *(varchar): Name of the company, primary key
- Headquarter(varchar): The location of head quarter
- Email (varchar): Official email
- Phone (varchar): Official phone number
- fda_id (varchar): each company has an ID issued by FDA for market entry approval

### 3.8 Entity Relationship Diagram

*The clinical trial database is designed in a snowflake schema: meet the requirement of normalization, reduce data redundancy, good scalability, though it has some disadvantages that the query speed is slower than the star schema, and may need to do more join when writing a query.*



## 4 Mondo Disease Ontology

In this clinical trial database, several places can be populated by terms from an ontology or terminology, e.g., "Disease" column in ClinicalTrial table, "Target" in the Drugs table.

### 4.1 Usage

Specifically, here choose "Indication" column in "Indication" table to encode with an ontology called *Mondo Disease Ontology*, in order to give structural information about indications in the database.

### 4.2 Developer

Mondo Disease Ontology (MonDO) is a product of Monarch Initiative, and is developed by members of the Monarch Initiative.

### 4.3 Maintenance

A GitHub tracker is established by Monarch Initiative team to provide the public to contribute to the ontology, medical expertise can make a term request to the developers, and developers will response to the request and update the ontology.

### 4.4 Cimino's Desiderata

- ✓ *Content*

  Developers or users can add terms as needed, and they can combine as needed.

- ✓ *Concept orientation*

  MonDO is developed to solve the problems like guarantee the mapping between different ontology to be one-to-one, especially in areas with evolving disease concepts such as rare disease.

- ✓ *Concept permanence*

  MonDO uses a logic-based structure for unifying multiple disease resources, concepts meanings are not changed and will not be deleted from the library.

- ✓ *Non-semantic identifiers*

  Concepts typically have unique identifiers, two disease names or identifiers are equivalent or one-to-one.

- ✓ *Polyhierarchy*

  Like the HPO, Mondo provides a hierarchical structure which can be used for classification or “rolling up” diseases to higher level groupings, by unifying multiple disease resources, multiple classifications are created.

- ✓ *Formal definitions*

  Mondo precisely annotate each mapping using strict semantics, so that we know when two disease names or identifiers are equivalent or one-to-one, in contrast to simply being closely related.

- ✓ *Rejection of NOC*

  MonDO will catch all terms are attractive from multiple disease resources.

- ✓ *Multiple granularities*

  Different levels of expressivity is all included in this ontology, due to its aims to harmonize disease definitions across the world.

- ✓ *Multiple consistent views*

  Multiple views of the hierarchy required to support different functional requirements and levels of detail

✓ *Context representation*

It uses concepts to define each other, combine concepts if needed and use the concept to deal with the users' needs.

✓ *Graceful evolution*

Any update of the ontology can be requested by sending issue on GitHub.

✓ *Recognized redundancy*

Mechanism for recognizing equivalence is made. For instance, in .obo version if the ID is one of Orphanet, OMIM, DOID or EFO then the xref precisely shadows the equivalence axiom.

### 4.5 Alternatives

MonDO is chosen to code the indications because it meets the need of granularity when encoding the diseases, it terms different expression of diseases more precisely, compared to some other ontologies. For example, disease *classic Hodgkin lymphoma* is termed as "LYMPHOMA, HODGKIN, CLASSIC" in OMIM, "Hodgkin lymphoma" in SNOMET CT. Both cases fail the expectation of coding the precise expression. But in MonDO, it is coded exactly the expression from the table in the database, *classic Hodgkin lymphoma* is coded 0009348 in MonDO.

However, there still be some alternative ontologies for this *Indication* column, e.g., National Cancer Institute Thesaurus (NCIT), Orphanet Rare Disease Ontology (ORDO). Both can encode diseases at different levels.

## 5 Analysis

### 5.1 User Story A

Julia was diagnosed with breast cancer in 2018, she received line1 therapy before, but the disease recurred unfortunately. Now she is looking for trials that are recruiting.

*Julia lives in Arizona, so it's better to find a hospital in the state.*

SQL query:

SELECT * FROM ClinicalTrial

LEFT JOIN Contacts

ON ClinicalTrial.Trial_ID = Contacts.Trial_ID

LEFT JOIN Organizations

ON Organizations.Org_ID = Contacts.Affiliation

WHERE Status = 'Recruiting'

AND Treatment = 'line2'

AND State = 'Arizona'

```
SELECT * FROM "ClinicalTrial"
LEFT JOIN "Contacts"
ON "ClinicalTrial"."Trial_ID" = "Contacts"."Trial_ID"
LEFT JOIN "Organization"
ON "Organization"."Org_ID" = "Contacts"."Affiliation"
WHERE "Status" = 'Recruiting'
AND "Treatment" = 'line2'
AND "State" = 'Arizona'
```

Data Output | Explain | Messages | Notifications

| Contact character varying (50) | Sex character varying (5) | Phone character varying (50) | Email character varying (50) | Affiliation integer | Org_ID integer | Organization character varying | Country character varying (50) | State character varying (50) | PostCode character varying (10) |
|---|---|---|---|---|---|---|---|---|---|
| Abbigail Hibbert | M | 5452126299 | jonas@optonline.net | 1 | 1 | Ironwood Cancer & Re... | US | Arizona | 85224 |
| Lorelei Puckett | F | 9826504197 | heidrich@verizon.net | 2 | 2 | Banner MD Anderson ... | US | Arizona | 85234 |
| Haaris Glover | M | 2827947275 | evilopie@comcast.net | 3 | 3 | Cancer Treatment Cent... | US | Arizona | 85338 |
| Mitchel Trevino | F | 8982563718 | tfinniga@yahoo.com | 3 | 3 | Cancer Treatment Cent... | US | Arizona | 85338 |

### 5.2 User Story B

Emma is a 5th-year PhD student at a prestigious molecular biology institute, her research focus on the biomarker of breast cancer. She wants to find a job in the US after graduation, a position at R&D department of a company would be of her interest.

*List the organizations that meet her needs:*

SQL query:

```
SELECT Companies.Company FROM Indications
LEFT JOIN Drugs
ON Indications.Drug = Drugs.Drug
LEFT JOIN Companies
ON Companies.Company = Drugs.Company
WHERE Indications = 'Breast Cancer'
AND Headquarter = 'US'
```

```
SELECT "Companies"."Company" FROM "Indication"
LEFT JOIN "Drugs"
ON "Indication"."Drug" = "Drugs"."Drug"
LEFT JOIN "Companies"
ON "Companies"."Company" = "Drugs"."Company"
WHERE "Indication" = 'Breast Cancer'
AND "Headquater" = 'US'
```

Data Output | Explain | Messages | Notifications

| | Company [PK] character varying (50) |
|---|---|
| 1 | BMS |
| 2 | Athenex |
| 3 | Pfizer |
| 4 | MSD |
| 5 | Genentech |
| 6 | Spectrum |
| 7 | Seattle Genetics |

She gets 7 companies: BMS, Athenex, fizer, MSD,Genentech, Spectrum, Seattle Genetics

Discussion: Though it's a clinical trial database, the queries do not have to be centered around the clinical trial. Different users can get the information they need in this database.

### 5.3 User Story C

Robert is the principle of a pharmacy company, his team is implementing a phase 3 clinical trial of their new drug for metastatic breast cancer, and the NCT number is *NCT01120184*. Before his team pushing forward their marketing strategy, he want to know who are their potential competitors in the same area.

In this case, he needs to use the database to find the NCT number of the completed phase 3 trials of breast cancer drugs, and also get the name of the drugs.

SQL query:

```
SELECT DISTINCT NCT_Num FROM ClinicalTrial
WHERE Phase = '3'
AND Status = 'Completed'
AND NOT NCT_Num = 'NCT01120184'
```

ORDER BY NCT_Num

```
SELECT * FROM "ClinicalTrial"
WHERE "Phase" = '3'
AND "Status" = 'Completed'
AND NOT "NCT_Num" = 'NCT01120184'
ORDER BY "NCT_Num"
```

Data Output | Explain | Messages | Notifications

| | Trial_ID [PK] numeric | NCT_Num character varying (11) | Disease character varying (50) | Treatment character varying (50) | Phase integer | Enroll_Num integer | Status character varying (50) | PostDate date |
|---|---|---|---|---|---|---|---|---|
| 1 | 27 | NCT00829166 | Breast Cancer | line2 | 3 | 495 | Completed | 2018-04-27 |
| 2 | 23 | NCT01419197 | Breast Cancer | line2 | 3 | 404 | Completed | 2018-03-24 |
| 3 | 18 | NCT02131064 | Breast Cancer | neoadjuvant | 3 | 444 | Completed | 2019-12-31 |

### 5.4 Extra Question

What's the highest status of T-DMI's clinical trial? Is it completed or not?

SQL query:

SELECT DISTINCT Drug, Phase FROM ClinicalTrial

LEFT JOIN TrialDrugs

ON ClinicalTrial.Trial_ID = TrialDrugs.Trial_ID

WHERE Drug = 'T-DM1'

ORDER BY Phase DESC

Limit 1

```
SELECT "NCT_Num", "Phase", "Status" FROM "ClinicalTrial"
LEFT JOIN "TrialDrugs"
ON "ClinicalTrial"."Trial_ID" = "TrialDrugs"."Trial_ID"
WHERE "Drug" = 'T-DM1'
ORDER BY "Phase" DESC
Limit 1
```

Data Output | Explain | Messages | Notifications

| | NCT_Num character varying (11) | Phase integer | Status character varying (50) |
|---|---|---|---|
| 1 | NCT03975647 | 3 | Recruiting |

## 6 Build Database System

Use PostgreSQL as relational database management system, and build the clinical trial database with 31 clinical trials of breast cancer drugs. Some clinical trial information is originate from ClinicalTrials.gov, the data has been reorganized, and some data from tables is fabricated. The implementation of the design of the clinical trial database can be versatile, DBMS developers can build their own needs.

Export the entity relationship diagram into DBMS using the commands below:

CREATE TABLE "ClinicalTrial" (

"Trial_ID" int,

"NCT_Num" varchar(11),

```
  "Disease" varchar(50),
  "Treatment" varchar(50),
  "Phase" int,
  "Enroll_Num" int,
  "Status" varchar(50),
  "PostDate" date,
  PRIMARY KEY ("Trial_ID")
);

CREATE TABLE "Contacts" (
  "Trial_ID" int,
  "Contact" varchar(50),
  "Sex" varchar(5),
  "Phone" varchar(50),
  "Email" varchar(50),
  "Affiliation" int
);

CREATE INDEX "PK, FK" ON  "Contacts" ("Trial_ID", "Affiliation");

CREATE TABLE "Indication" (
  "Drug" varchar(50),
  "Indication" varchar(50),
  "FDA" varchar(2),
  "MOD_ID" varchar(7),
  PRIMARY KEY ("Indication")
);

CREATE INDEX "PK, FK" ON  "Indication" ("Drug");

CREATE TABLE "TrialDrugs" (
  "Drug" varchar(50),
  "Trial_ID" int
);

CREATE INDEX "PK, FK" ON  "TrialDrugs" ("Drug", "Trial_ID");
```

```
CREATE TABLE "Companies" (
  "Company" varchar(50),
  "Headquarter" varchar(50),
  "Email" varchar(50),
  "Phone" varchar(50),
  "fda_id" varchar(50),
  PRIMARY KEY ("Company")
);

CREATE TABLE "Organization" (
  "Org_ID" int,
  "Organization" varchar,
  "Country" varchar(50),
  "State" varchar(50),
  "PostCode" varchar(10),
  PRIMARY KEY ("Org_ID")
);

CREATE TABLE "Drugs" (
  "Drug" varchar(50),
  "Company" varchar(50),
  "Type" varchar(50),
  "Target" varchar(50),
  PRIMARY KEY ("Drug")
);
```

## Reference:

[1] Anhøj, Jacob. 2003. “Generic Design of Web-Based Clinical Databases.” *Journal of Medical Internet Research* 5 (4): e27.

# Appendix 1: An overview of database tables

**Table 1 ClinicalTrial**

| Trial_ID | NCT_Num | Disease | Treatment | Phase | Enroll_Num | Status | PostDate |
|---|---|---|---|---|---|---|---|
| **1** | NCT03734029 | Breast Cancer | line2 | 3 | 540 | Recruiting | 2020-06-12 |
| **2** | NCT04553770 | Breast Cancer | neoadjuvant | 2 | 88 | Not yet recruiting | 2020-06-28 |
| **3** | NCT04494425 | Breast Cancer | line2 | 3 | 850 | Recruiting | 2019-04-26 |
| **4** | NCT04420598 | Breast Cancer | line2 | 2 | 39 | Recruiting | 2019-12-04 |
| **5** | NCT04042701 | Breast Cancer | line2 | 1 | 115 | Recruiting | 2017-09-22 |
| **6** | NCT03975647 | Breast Cancer | line2 | 3 | 460 | Recruiting | 2019-01-29 |
| **7** | NCT03587740 | Breast Cancer | adjuvant | 2 | 82 | Active, not recruiting | 2019-01-02 |
| **8** | NCT03529110 | Breast Cancer | line2 | 3 | 500 | Active, not recruiting | 2018-03-07 |
| **9** | NCT03523585 | Breast Cancer | line2 | 3 | 600 | Recruiting | 2015-02-16 |
| **10** | NCT03368196 | Breast Cancer | line2 | 1 | 12 | Active, not recruiting | 2014-09-21 |
| **11** | NCT03366428 | Breast Cancer | line2 | 1 | 51 | Active, not recruiting | 2016-12-14 |
| **12** | NCT03364348 | Breast Cancer | line2 | 1 | 52 | Recruiting | 2015-04-27 |
| **13** | NCT03153163 | Breast Cancer | line2 | 1 | 11 | Completed | 2019-07-03 |
| **14** | NCT03032107 | Breast Cancer | line2 | 1 | 27 | Active, not recruiting | 2019-09-15 |
| **15** | NCT02924883 | Breast Cancer | line2 | 2 | 202 | Completed | 2020-08-21 |
| **16** | NCT02562378 | Breast Cancer | line2 | 1 | 15 | Completed | 2020-01-08 |
| **17** | NCT02144012 | Breast Cancer | line1 | 3 | 34 | Terminated | 2020-01-06 |
| **18** | NCT02131064 | Breast Cancer | neoadjuvant | 3 | 444 | Completed | 2019-12-31 |
| **19** | NCT02073916 | Breast Cancer | line1 | 1 | 24 | Completed | 2019-12-14 |
| **20** | NCT01976169 | Breast Cancer | line2 | 1 | 28 | Active, not recruiting | 2019-12-08 |
| **21** | NCT01816035 | Breast Cancer | line2 | 1 | 13 | Completed | 2020-02-29 |
| **22** | NCT01772472 | Breast Cancer | adjuvant | 3 | 1486 | Active, not recruiting | 2018-04-04 |
| **23** | NCT01419197 | Breast Cancer | line2 | 3 | 404 | Completed | 2018-03-24 |
| **24** | NCT01120184 | Breast Cancer | line1 | 3 | 299 | Completed | 2018-03-07 |
| **25** | NCT00951665 | Breast Cancer | line2 | 1 | 107 | Completed | 2020-01-22 |
| **26** | NCT00943670 | Breast Cancer | line2 | 2 | 51 | Completed | 2019-10-10 |
| **27** | NCT00829166 | Breast Cancer | line2 | 3 | 495 | Completed | 2018-04-27 |
| **28** | NCT00679341 | Breast Cancer | line1 | 2 | 67 | Completed | 2018-04-22 |
| **29** | NCT03429101 | Breast Cancer | line2 | 1 | 6 | Terminated | 2020-02-28 |
| **30** | NCT02725541 | Breast Cancer | neoadjuvant | 2 | 0 | Withdrawn | 2017-10-30 |
| **31** | NCT02144012 | Breast Cancer | line1 | 3 | 34 | Terminated | 2017-08-11 |

**Table 2: Organization**

| Org_ID | Organization | Country | State | Postcode |
|---|---|---|---|---|
| **1** | Ironwood Cancer & Research Centers | US | Arizona | 85224 |
| **2** | Banner MD Anderson Cancer Center | US | Arizona | 85234 |
| **3** | Cancer Treatment Centers of America at Western Regional Medical Center | US | Arizona | 85338 |
| **4** | UCLA School of Medicine | US | California | 90095 |
| **5** | Stanford Cancer Institute | US | California | 94305 |
| **6** | Medizinische Universität Innsbruck | Austria | Innsbruck | 6020 |
| **7** | Kepler Universitätsklinikum | Austria | Linz | 4020 |
| **8** | LKH - Universitätsklinikum der PMU Salzburg | Austria | LKH | 5020 |
| **9** | Institut Jules Bordet | Belgium | Bruxelles | 1000 |
| **10** | Tom Baker Cancer Center | Canada | Alberta | T2N4N2 |
| **11** | Chinese PLA General Hospital | China | Beijing | 100853 |
| **12** | Institut Paoli Calmettes | France | Bouches-du-Rhône | 13009 |

**Table 3: Contacts**

| Trial_ID | Contact | Sex | Phone | Email | Affiliation |
|---|---|---|---|---|---|
| **1** | Abbigail Hibbert | M | 5452126299 | jonas@optonline.net | 1 |
| **1** | Lorelei Puckett | F | 9826504197 | heidrich@verizon.net | 2 |
| **1** | Haaris Glover | M | 2827947275 | evilopie@comcast.net | 3 |
| **1** | Jaydon Busby | M | 2068525896 | mcsporran@mac.com | 4 |
| **1** | Theia Parks | F | 6957294838 | andrei@comcast.net | 5 |

**Table 4: Drugs-1**

| Drug | Company | Type | Target |
|---|---|---|---|
| **Abraxane** | BMS | chemical | HER2 |
| **DS-8201a** | AstraZeneca | biological | HER2 |
| **Lapatinib** | GSK | biological | HER2 |
| **non-pegylated liposomal doxorubicin** | SJZZY | chemical | |
| **Paclitaxel** | Athenex | chemical | |
| **Palbociclib** | Pfizer | chemical | CDK4/6 |
| **Pembrolizumab** | MSD | biological | |
| **Pertuzumab** | Genentech | biological | HER2 |
| **Poziotinib** | Spectrum | biological | |
| **T-DM1** | Roche | biological | HER2 |
| **Tucatinib** | Seattle Genetics | biological | HER2 |

**Table 5: TrialDrugs**

| Trial_ID | Drug |
|---|---|
| **1** | DS-8201a |
| **2** | DS-8201a |
| **3** | DS-8201a |
| **4** | DS-8201a |
| **5** | DS-8201a |
| **5** | Pembrolizumab |
| **6** | Tucatinib |
| **6** | T-DM1 |
| **7** | T-DM1 |
| **8** | T-DM1 |
| **9** | DS-8201a |
| **10** | DS-8201a |

**Table 6: Indication**

| Drug | Indication | FDA | MDO_ID |
|---|---|---|---|
| **Abraxane** | Lung Cancer | 1 | 8903 |
| **Abraxane** | Pancreatic Cancer | 0 | 11739 |
| **Abraxane** | Heart Disease | 0 | 5010 |
| **Abraxane** | Breast Cancer | 0 | 7254 |
| **DS-8201a** | Breast Cancer | 0 | 7254 |
| **Lapatinib** | Breast Cancer | 1 | 7254 |
| **non-pegylated liposomal doxorubicin** | Ovarian Cancer | 0 | 8170 |
| **non-pegylated liposomal doxorubicin** | Breast Cancer | 0 | 7254 |
| **Paclitaxel** | Ovarian Cancer | 1 | 8170 |
| **Paclitaxel** | Pancreatic Cancer | 1 | 11739 |

**Table 7: Companies**

| Company | Headquater | Email | Phone | FDA_ID |
|---|---|---|---|---|
| **BMS** | US | telbij@aol.com | 6454547724 | 18225662 |
| **AstraZeneca** | UK | jane.doe@astrazeneca.com | 18002369933 | 38722592 |
| **GSK** | UK | jmgomez@att.net | 2283741652 | 12341426 |
| **SJZZY** | China | maneesh@att.net | 2057451487 | 67457235 |
| **Athenex** | US | sbmrjbr@mac.com | 2262788773 | 68772611 |
| **Pfizer** | US | bhtower@me.com | 6912133624 | 75467444 |
| **MSD** | US | monkeydo@gmail.com | 9358311541 | 34654657 |
| **Genentech** | US | sjmuir@verizon.net | 6363886567 | 34263271 |
| **Spectrum** | US | budinger@msn.com | 4669947123 | 53734675 |
| **Roche** | Switzerland | dkeeler@icloud.com | 3287947494 | 41648547 |
| **Seattle Genetics** | US | cantu@live.com | 4787348131 | 32414765 |

# Appendix 2: PgAdmin 4 window (Tables and Views)