Clinical Trial Search Application Development  
The design and development of a Search Application of cleaned and curated Clinical Trial Data and all related information.

Nicole Koshy, NK, and Koshy  
School of Informatics and Computing, Indiana University - Purdue University Indianapolis, nickoshy@iu.edu  
Alden Jettpace, AJ, and Jettpace  
School of Informatics and Computing, Indiana University - Purdue University Indianapolis, ajettpac@iu.edu  
Austin Anderson, AA, and Anderson  
School of Informatics and Computing, Indiana University - Purdue University Indianapolis, [andersau@iu.edu](mailto:andersau@iu.edu)

**ABSTRACT:**  
Clinical Trials are research studies conducted on human volunteers to study the behavior and efficacy of treatments and interventional therapy to treat specific indications. Traditional repositories that catalog publicly available study data have search systems centered around trail information and only provide summary information of related data such as participating institutions, personnel involved, sites where trials are conducted etc. Our website offers a search suite that allows users to design a multivariate search that can combine inputs for 20-25 different criteria related to trial study, institutions, sites, and personnel involved as starting points to generate trial results that can be refined by adding additional parameters. This provides researchers and analysts with a way to generate tailored results that meet their requirements and do not require an inordinate number of repeated searches.

CCS CONCEPTS • Information systems → Applications → Digital libraries and Archives • Applied Computing → Life and Medical Sciences → Health Care Information Systems

**Additional Keywords and Phrases:** Clinical trials, phase, enrollment, NCT number, sponsor, collaborator, health care personnel, sites, trials, institutions, multivariate search, clinicaltrials.gov

**1 INTRODUCTION**

Publicly available clinical trial data is an information rich resource that can be used to study trends in the health industry, compare treatment modalities for indications, and collate information on study designs. This data alone is difficult to use for researchers who require collated data from a large number of trials, since they would have to download the information for all trials matching their requirements and sift through the data manually.

We have designed a database that stores curated and cataloged trial information that can be cross-referenced with detailed information about related entities such as participating institutions, researchers managing a trial, sites where a trial can be conducted, mesh keywords or associated publications so that researchers have analytically meaningful information to study clinical trials. This database can be accessed through a web interface that allows registered users to search trial information from a multitude of initiation points aside from trials so that even people who aren’t familiar with clinical trials can use the advanced search to look for data with ease.

**1.1 Aims**

The trial database stores trial study data that is connected to related entities through bridge tables to minimize redundancy and facilitate normalization. This structure also makes it easy to set up update trigger cascades that can be used to update the database.

The search suite is designed to make trial information accessible using criteria specific to the main entities in the database:

1. Trial Study Details
2. Participating Institutions
3. People involved in the study (Healthcare Personnel – HCP)
4. Trial Sites
5. Mesh Terms (medical keywords) associated with trials.
6. Associated Publications
7. Treatments used in the trial

Definitions for the individual search parameters can be found on the site homepage (homepage.php).

In the connecting layer, the multivariate search is set up to connect several simple SQL queries together to make the search faster and less complex. These queries can be modified to generate more refined results centered around entities other than trials.

Currently, the primary result of the search generates a list of trial summaries that match the input parameters. Each trial has dynamic links that allow users to lookup corollary data related to each trial on a separate page. This page can be used to look up data, one trial at a time.

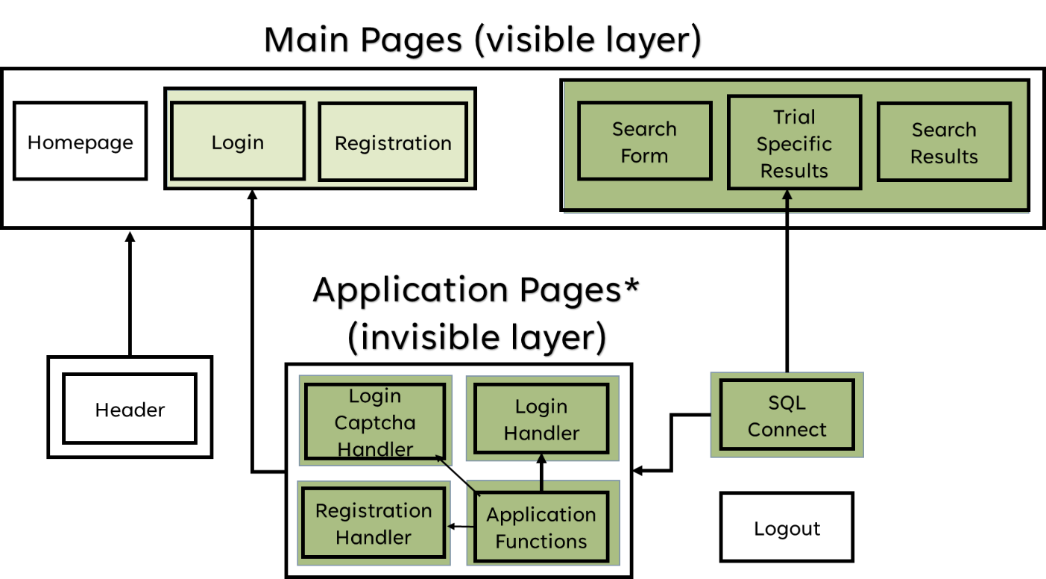
**1.2 Audience**

The website is designed to make detailed trial information accessible not only to healthcare researchers, physicians and analysts looking to study clinical trial data but also inexperienced users who do not have an expertise in the field. The application is designed to be scalable to increase the complexity of displayed search results and as such, can have commercial applications.

**2 APPLICATION DESCRIPTION**

**2.1 Application Summary**

The Clinical Trial Search is composed of four directly accessible pages, and a fifth page only accessible through the results of the search page. The four directly accessible pages are connected by a header-bar, which only shows Trial-Search access for registered and logged-in users. An outline of the architecture of the website is provided below.



**Figure 1:** Architecture of the website outlining how the pages are connected to each other and connecting application layer.

2.1.1 **Homepage**

The homepage is intended to be the first page any user will encounter, and provides a description of the website functionality, how to register and sign in for access to the trial search parameters and provides a description of the components of the trial information and the search parameters. The name and description of these entities is given below.

**Table 1:** Description of main Trial search entities

|  |  |
| --- | --- |
| **TRIAL ENTITY** | **DESCRIPTION** |
| Trial Details | Pertains to study set-up details and the status of a trial. |
| Treatments & Interventions | Any treatment method or substance that can be used in a trial. |
| Participating Institutions | Organizations involved in conducting the study, either financing the trial or collaborating with the sponsoring institution. |
| Mesh Terms  (Medical Subject Headings) | Associated medical keywords that are part of a hierarchical index that can be used to search for biomedical information. |
| Associated Publications | Publications which have referenced the trial. |
| Trial Sites | Disclosed locations where a Trial can be conducted. |
| Health Care Personnel (HCP) | Physicians and researchers who participate in conducting and managing a trial. |

2.1.2 **Registration**

The site itself is intended only for trusted users who want to search through trial information. They do not have access to edit any information in the database. Users are required to register and sign in to limit traffic to the site. This way, not too many connections to the database server can be made at any given time and helps maintain site security.

The registration page consists of a simple form which requires users to input a unique valid email, as well as a password with 5-part password criteria, as shown below:

* A minimum of 8 characters.
* At least one lowercase and uppercase letter.
* At least one numeric and one special character.

The user is required to repeat their password correctly. The page is equipped with preset alerts for any errors including invalid email, invalid-password, or empty inputs. If the user has successfully registered, their credentials will be added to the user-table in the database.

2.1.3 **Sign in**

Once registered, registered users can sign in and then get access to the search suite. The sign in page consists of a form to input user email and password and uses a captcha code system that generates a dynamic image which can be reset with a pressable button if the image is unclear to the user. Users can only sign in after entering correct credentials and pass the captcha test.

|  |  |
| --- | --- |
|  |  |

**Figure 2:** An image of the Signin page, and the same page with new captcha image after clicking.

2.1.4 **Trial Search**

The Trial Search page is one of the two main functionalities of the web application and allows users to dynamically search for Clinical Trials information based on several parameters related to the trial. The filters and their descriptions can be found on the homepage, and are described in Table 2 below:

**Table 2:** A description of the search-filers of the Trial Search page

|  |  |  |
| --- | --- | --- |
| **Search Category** | **Filter Name** | **Description** |
| Trial Details | Trial ID | Unique Identifier of Trial |
| Trial Name | Name of Trial |
| Trial Start/Trial End | Start / Completion date of Trial |
| Recruitment Start/End | Start/End of patient recruitment |
| Trial Status | Status of trial (i.e.: completed, etc.) |
| Trial Phase | Scope of study |
| Trial Country | Country study was conducted in |
| Treatment | Treatment Name | Name of Treatment/Procedure |
| Treatment Mode Action | Mechanism of action |
| Small Molecule? | Is the treatment a low molecular weight compound |
| Participating Institutions | Institution Name | Name of Institution |
| County/State/City | Country/State/City of Institution |
| Sponsor or Collaborator? | Limits search to Institutions that were either main sponsor or merely collaborator |
| Mesh Terms | Mesh Term | Specific Mesh Terms |
| Indication | Associated Disease, if any |
| Cancer Stage | Cancer state, if any |
| Publications | Author | Name of Publication Author |
| Journal Name | Journal Article was published in |
| Publication Title | Title of Publication |
| Trail Sites | Name | Name of Site Location |
| Country/City/State | County/City/State of Trial Site |
| Healthcare Personnel | Name | Name of Health Care Personnel |
| Credentials | Professional Credentials of HCP |
| Investigator or collaborator? | Is the HCP a primary investigator or a collaborator |

The filter tabs of the Search Page are split by main entity and are collapsible to allow for easier navigation and appearance.

The Trial Search page dynamically generates a table of all the trials within the database that match the input search criteria. Each trial in the result has several links that open to a single dynamically generated page that provides all associated information related to a single trial.

2.1.5 **Primary Search Results**

The Trial Summary is only accessible through the results of the Trial Search page. Each Search Result link contains a URL that corresponds to the Trial ID and generates the complete associated information of the trial onto a single page through a series of SQL searches based on the Trial ID value. The page itself is divided into several sections, corresponding to the different filter-categories, which displays all the associated information in collapsible tables. These collapsing tables exist for better visual filtering of the information to the user, allowing them to focus on what they are interested in without crowding the page.

For example, if a user wanted to search for the Institutions which have participated in a specific trial related to Esophageal Cancer, they could simply search for that mesh term, go through the generated results and click on the [Institution Info] link for an associated trial. The link opens onto the complete full trial page which displays all associated results for a single trial. The URLs are designed to toggle to the specific part of the page that the link directs them to.

Calendar

Description automatically generated

**Figure 3:** A screenshot of the Search Page displaying the Toggleable sections of the Search tab on the left and the dynamically generated results for a Mesh term search for “esophageal cancer” on the right (<https://in-info-web4.informatics.iupui.edu/~nickoshy/Project_Final/main/ProjectPracticeForm.php>)



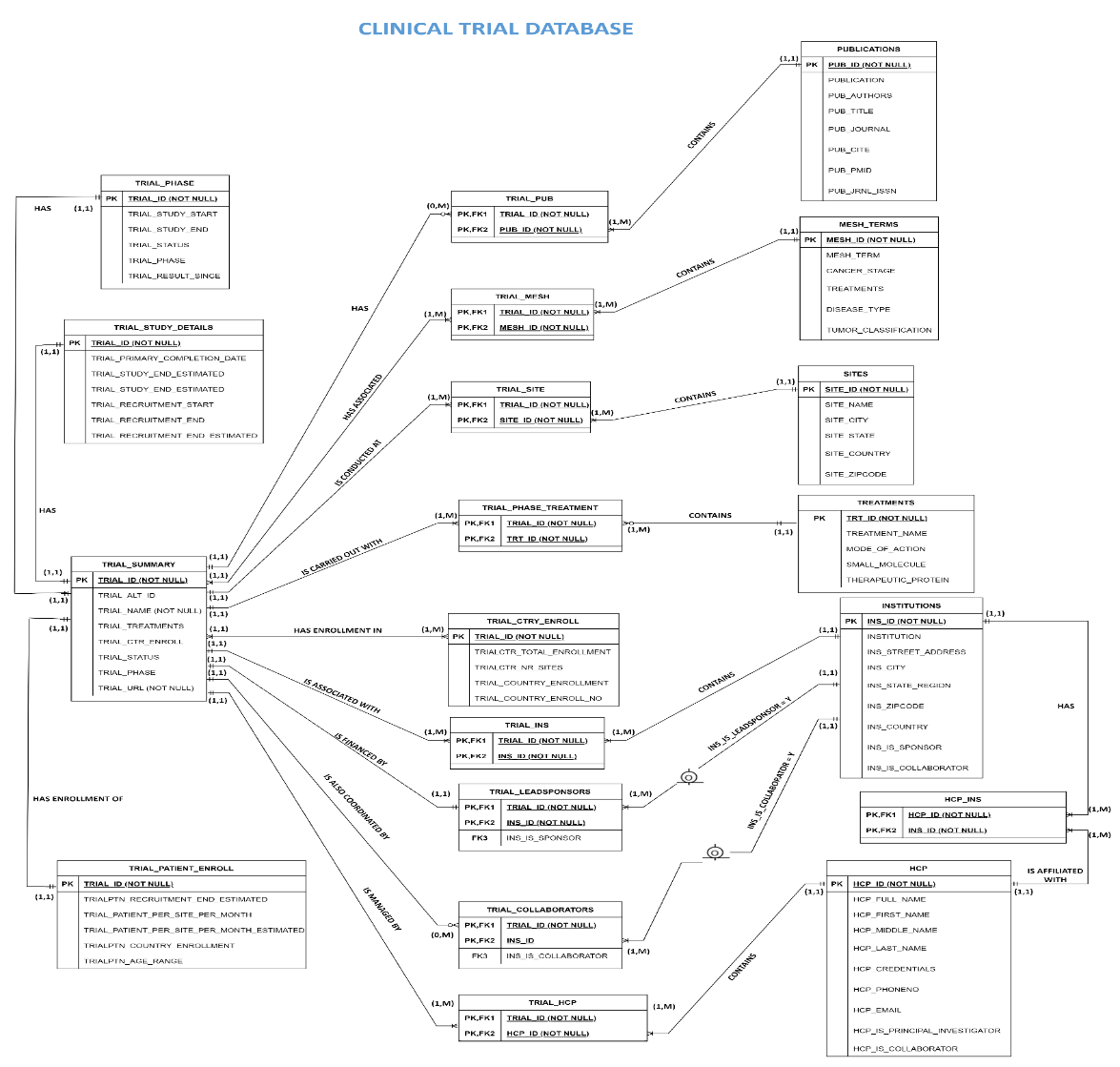
**Figure 4:** Screenshot of the Full Trial Page displaying results for a single trial, which opens specifically to the Institution details for the trial

**2.2 Application Functionality**

The purpose of our application is to provide a secure, easy to read, and robust search engine for anyone interested in looking up information on clinical trials. The basic rationale is similar to that of other public repositories such as clinicaltrials.gov. However, our design has the added functionality of allowing a user to begin a search, using generic parameters such as Mesh terms and institutions, rather than only trial related information.

2.2.1 **Database and Datasources**

All the trial study details were scraped from clinicaltrails.gov. This includes the list of mesh terms, associated publication citations, summary information on institutions, investigators, and disclosed sites. Details of these specific entities were sourced from various non-aggregate sites where verified information on institutions, site addresses, and details of HCPs could be sourced. Publication information was collected primarily from Pubmed. The entity relationship diagram in fig. 5 shows the structure of the database.



**Figure 5:** The Entity relationship diagram of the tables in the Clinical Trial Search database

New data is cleaned and curated before input into the database. The table structures and relationships are designed to provide a 3NF normalized form that minimizes redundancy while connecting all relevant entities to each other through a series of bridge tables.

Users are not allowed to edit the data within the database as any information added to the database needs to first be thoroughly curated and validated, which can only be done by trained personnel working with site administrators who maintain the database. User registration information is stored in a user data table. This table has no relation to the clinical trial data and exists solely for ensuring that only verified users can access the application’s search functionalities.

2.2.3 **Security**

Because of the valuable nature of the data, and the amount of effort which goes into curating it, steps have been taken to ensure users cannot edit or delete any of the medical trial data within the search database. Every form on the application, including registration and sign in, has been set up using prepared statements to protect against SQL injection.

To protects against bot spamming, which would congest site traffic and slow the applications performance, aside from the standard login security measures we employ a randomly generated captcha image that must be entered successfully before a registered user can access the search suite.

Finally, to ensure users cannot edit source code and can only access certain pages when they are within the application, PHP header-statements have been prepared. This will send users back to the homepage if they try to access pages in the improper order. If a user is logged in, the header-statement ensures they cannot login again until they logout, nor can they register a new account until they logout. We also employ password hashing to ensure that user information is secure even within the database.

2.2.3 **Search Design**

The search suite is designed to provide a multivariate search tool by combining a series of smaller SQL queries that connect two or three related tables together on the PHP side. The advantage of having several interconnected smaller queries connected together, is that it makes the search faster than a single, long complicated query that can be difficult to troubleshoot and would take longer to return results as the database grows. This approach also makes it easier to add additional queries as we move towards providing different result views to the user. For example, a person searching for information on Institutions participating in certain trials may only be interested in looking up Institution information rather than trial data. In such cases, we would need to create different result views that make this information easier to navigate and analyze.

At the front end, the search parameters are segregated into smaller entities divided into togglable subsections that allow the users to open one or multiple tabs at the same time, as per their preference. Search fields are designed to receive text input or allow users to choose from dropdown lists. Several of the text search fields are equipped with “autocomplete” features that show recommendations by linking the search field to associated attributes in the relevant tables.

The dropdown lists are either hardcoded, if fewer options are present, or linked to a predefined list whose values are used to populate the list.

**3 APPLICATION COMPONENTS**

The search application is designed around flexibility of searches and immediate, repeatable results. It utilizes PHP to handle user inputs, query the application data within a SQL database, and display the results for the users using AJAX. This is accomplished through a mixture of displayed (main directory) and inclusion pages (inc folder), which work behind the scenes to help verify any user inputs and transform the pages.

**3.1 Files and Directories**

To organize the pages, directories were divided according to functionality of the resources included, as to whether to user was intended to directly see or interact with the results of a page and its functionality. The design is depicted in Figure 1. If the resource was purely for stylization, it was stored in the **/css/** directory.If the user was meant to see or interact with the page, then it was stored in the **/main/** directory. Pages which were meant to communicate with the SQL database, clean and validate input, or serve as part of a larger page, were placed within the **/inc/** directory. These pages are not intended to be viewed by users. Table 3 gives a description of each of the files within the application, including their directory, extension, and purpose.

**Table 3:** Files of Clinical Trial Search Application

|  |  |  |
| --- | --- | --- |
| **Directory** | **File Name** | **File Purpose and Description** |
| main | homepage.php | Explanatory homepage of application, location of successful signin. |
|  | register.php | User registration page |
|  | signin.php | User signin page |
|  | ProjectPracticeForm.php | Clinical Trial Search form |
|  | ProjectSearch\_2.php | Table of results of Trial Search Form |
|  | full\_trial.php | Page of all relevant information of trial selected from ProjectSearch\_2 |
|  | image\_fittedv2.jpg | Background image of register.php, signin.php, and full\_trail.php |
|  | CTsearch2.png | Background image of homepage.php |
|  | searchbcg.png | Background image of ProjectPracticeForm.php |
| inc | captcha.inc.php | Generates captcha image on signin.php |
|  | countries.txt | List of countries to generate country selection on ProjectPracticeForm.php search form |
|  | header.inc.php | Forms navigation bar and page title for homepage, signin, registration, trial search, and full trial pages |
|  | signin.inc.php | Confirms validity of user signin.php form, and signs in user if successful |
|  | register.inc.php | Confirms validity of user register.php form, and creates user if successful |
|  | sql\_connect.php | Establishes connection to database containing User table and Clinical Trial Data Tables |
|  | functions.inc.php | Creates helper functions to be used in other inclusion pages such as register.inc.php or signin.inc.php |
|  | logout.inc.php | Logs out user |
| css | header\_style.css | Stylizes Naviation Bar of header.inc.php and all pages which include it |
|  | search\_stylesheet.css | Styles ProjectPracticeForm.php and ProjectSearch\_2.php |
|  | trial\_style.css | Styles full\_trail.php page |

The inclusion files mainly served to help the login and registration functionality, by confirming that the account was valid and the user had successfully typed in the randomly generated captcha code. The inclusion files often reference each other, such as the pages [register.inc.php] or [signin.inc.php] including the pages [sql­\_connect.php] and [functions.inc.php].

If any page must confirm the inputs of a form or visualize clinical trial information, then it will include the [sql\_connecct.php] page. This includes [signin.inc.php], [register.inc.php], [ProjectPracticeForm.php], [ProjectSearch\_2.php], and [full\_trial.php].

**3.2 Technologies**

In order to dynamically generate search results with easily understandable presentation and as little distraction as possible, dynamic page generation was required along with the ability to show and hide elements as desired by the user. This required the primary use of PHP code to generate HTML rather than raw HTML itself, AJAX to enable toggle of search and result sections, as well as to bring up the results of the Search Results onto the Trial Search form page, and CSS and Bootstrap to better stylize the results.

3.2.1 PHP

Because the site allows accessibility to certain sections depending on whether a user is signed in or not, every accessible page and its primary support functionality pages in the [main] and [inc] directories are coded in PHP, with every accessible page except ProjectSearch\_2.php. initializing a session [session\_start], which was encoded within the included file [header.inc.php].

Several of the pages included PHP purely as a programming language rather than to render HTML elements, and were used to support input validation and management of the user registration and login system. All these pages are found in the [inc] directory.

3.2.2 AJAX and jQuery

The goal of the Trial Search was for users to be able to search on increasingly stringent filtering parameters, to refine results as per their needs. Thus, AJAX was used to send the search form variables of [ProjectPracticeForm.php] to [ProjectSearch\_2.php] which would use the results of the search filters to create a custom SQL query and return the results.

Due to medicines complex nature, complicated and hard to remember terms, and many options involving states and cities, an auto-complete system was implemented into the search form for the treatment names, as well as for site and institution city and state input. This required opening a connection to SQL to extract all these unique values, before turning them into JSON-objects and implementing them into an autocomplete system via JavaScript [$(element).autocomplete] method for the specific PHP-constructed html elements. While this functionality is capable through JavaScript library, this had to be implemented through a jQuery framework in order to not overwrite the elements of the AJAX components used on the same page.

jQuery was also used in order to automatically open sections of the [full\_trial.php] page based on selecting of links generated in the [ProjectSearch\_2.php] search-results table. While each link ultimately lead to the same dynamically generated page from the specific trial-id, the link also included in the URL a variable associated with a specific id, one for each collapsible section of the [full\_trial] page. The specific id was extracted from the URL and used to open the specified section using the following code:

**Algorithm 1:** URL Variable Extractor

<script src = 'https://ajax.googleapis.com/ajax/libs/jquery/3.6.3/jquery.min.js'></script>

<script>

$(document).ready(function () {

$('.filter\_category').on('click', function() {

$(this).closest('.filter\_tab').find('.search\_crit').slideToggle(); });

var url = window.location.href.slice(window.location.href.indexOf('section='));

function GetParameterValues() {

var url = window.location.href.slice(window.location.href.indexOf('?') + 1).split('&');

for (var i = 1; i < url.length; i++) {

var urlparam = url[i].split('=');

return urlparam[1]; } }

console.log(GetParameterValues());

$('#' + GetParameterValues()).slideToggle();

});

</script>

3.2.3 CSS and Bootstrap

CSS and Bootstrap was used for a variety of reasons, with the two primary reasons being visual appeal and ease of use. With the plethora of data a user would have to sift through in this search, it is important to make things easy to manage.

All of the user-facing pages are wrapped within Bootstrap and CSS elements. Most of these are for visual appeal and ease of navigation, but there are certain functionalities that are also given by Bootstrap. The Bootstrap container is one of the most basic elements of Bootstrap, and it is used very frequently in the application. This allows not only for visual clarity with the multiple rows and columns set by the containers, but it allows for further augmentation that is discussed later.

The search page utilizes dynamic Bootstrap cards to contain the search and result forms as well as give them their collapsible and scrollable functionality. This is done by nesting the form within a Bootstrap card within the container set to house all the information on the search page.

This also applies to the consolidated trial result page that displays results for a single trial. The collapsible and side-scrolling capabilities are provided by CSS and Bootstrap to improve usability and data manageability. In conjunction with AJAX and jQuery that allows users to go right to the information they want to see. These functions help ensure that the user is only seeing the data that they want to see.

3.2.4 SQL

The main trial database was set up in MariaDB and consists of 20 tables connecting the main entities to each other as shown in Figure 5. Update cascades have been set up to make the database expandable and modifiable as new trial information becomes available.

**4.0 Future Work**

The database we have constructed for our Clinical trial search portal was designed to be updated with Stored Procedures or direct insertion. This means that we can continue to add new records as we collect it, as well as modify existing records as trial statuses and results get updated. We can also add more entities as we increase the granularity of the results we generate for users.

Similarly, we can add more search parameters to the web interface, as we add more entities to our existing structure. We can also add several different result views that center around other important entities other than clinical trials, so that users can broaden the scope of their analysis. All of this would require redesigning our current interface to improve the layout and shift to more advanced Bootstrap features.

One important feature that we would need to add is caching of the search parameters and primary search results even when users toggle to other pages or sections. Since we have such a complex search ecosystem, it is important that this feature is implemented to improve user experience and prevent frustration from having to continuously restart searches.

Finally, since trial information is such a vast and diverse big data resource, it is important for users to also be able to store the results of their search, for future reference. Hence this is another important feature that we should implement as we expand the functionality of our application.