# *University of Maryland College Park*

****

Final Report



##### 

##### Authors: Joseph Bensen, Laraib Laubach, Tanachot Trachoo, Shashvat Vyas and Kevin Woo

##### INST 327-0203:

##### Instructor: Pamela Duffy

##### May 12, 2020

**Introduction**

Our team has created a database which tracks the spread of historically infectious diseases, analyzing their spread over a period of time. We have utilized various online and global resources to investigate where specific diseases originated, how quickly they have spread, and how effective treatments were, with our goal being to help people around the world react and prepare accordingly for future outbreaks. The database mainly focuses on the Swine Flu (H1N1), Ebola, and the Bird Flu (H2N2). Our other ideas were to track natural disasters around the world in real time, the change in their occurrences, and the cost of the ensuing catastrophes. In essence, we were interested in tracking world-wide occurrences and phenomena. The main theme for all of our ideas was to use information available to the public from the government and world agencies in order to produce a relevant and usable database. Ultimately, our group became most interested in analyzing the spread of historical diseases because of the current relevance to the ongoing COVID-19 pandemic.

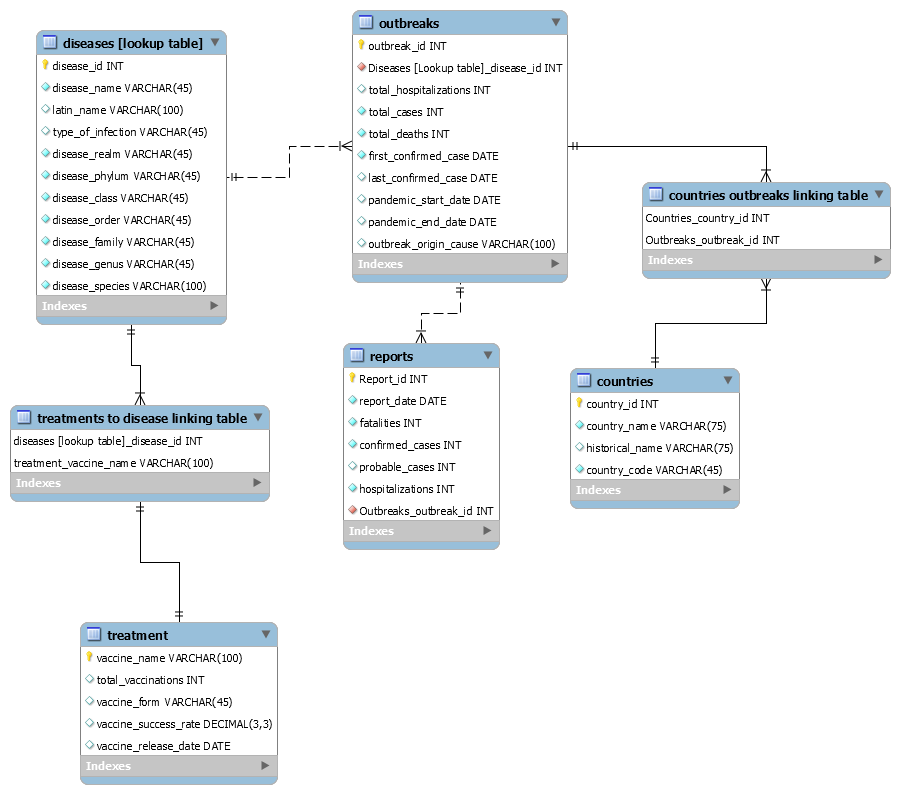
**Database Description**:

Tracking and analyzing the spread of historical diseases over time is the main function of our database. Timespan, proliferation, and the effects of these diseases are all based on crucial factors such as location and the severity of an initial outbreak. Sample data that we used included real historical data, as our topic is primarily focusing on historical diseases, and we wanted to be as accurate as possible. Our target audience that we are building the database for is everybody - this is because we believe that anyone should be able to use our database for their own interests. However, we also believe that this collection of data will attract the interest of researchers and analysts in the field of pathology. The entities we have focused our attention on are diseases that have already been tracked throughout history. The symptoms of various diseases, however, were not relevant for us to use in our database because either they are commonly known, or a simple web search would satisfy this inquiry.

We also understand that our database is going to raise questions that would call for various solutions and answers, but we are certain that within the restraints, we should be able to provide a solution for any relevant future issues. Some of the questions we anticipate to answer, for example, include “which disease spreads the fastest from the year X to Y?” and “which country has the highest rate of infected individuals for disease Y?”. The database will also be able to return queries that answer questions like “which disease was the deadliest from years X - Y, and how does it differ by different countries?” Essentially, our database provides functional ways to analyze diseases and will hopefully deliver insightful and meaningful information for researchers, students, the general public, and especially those seeking to prevent diseases from spreading.

*Additional notes:* (Our last\_confirmed\_case column in our ‘outbreaks’ table accepts null values to signify an ongoing outbreak; for example, COVID-19 would have a last\_confirmed\_case of NULL)

**Logical Design**



**Physical Design**

Please see our attached .sql file for our complete backup with the required sample data and tables.

**Sample Data**

Our sample data is almost completely derived from historical data from sources like the CDC, the WHO, and Kaggle. We were able to fulfill our initial desire of keeping the data authentic in order to provide a relevant and usable database.

**Views/Queries**

We were able to develop five queries that can answer specific questions while fulfilling all query requirements:

1. How many deaths and confirmed cases were there for diseases with outbreaks after the year 2000?
2. How many total deaths and confirmed cases were there for each disease with deaths greater than 50,000?
3. How many outbreaks were tracked per country?
4. What was the total number of deaths from H1N1 before the TIV injection vaccine was released?
5. Which of the diseases covered are classified as viruses?

**Changes From Original Design**

Initially, our first draft ERD was not approved by Snehal and Isabeau, the class’ Teaching Assistant and our Academic Peer Advisor & team mentor. Isabeau suggested that we should change certain aspects of the ERD as we were not exactly fulfilling the ‘seven tables’ portion of the criteria for creating the database. We met with Isabeau and fixed the ERD in an attempt to enhance it in time for the final report submission. A progress report was also included detailing the changes and improvements we had made to our database. Furthermore, we met with Professor Duffy who suggested practical ideas which we could incorporate into our tables to expand upon our initial plan. One of these ideas was an ‘Outbreaks’ table which would track the origin of a disease and length of time an outbreak occurred. Professor Duffy was extremely helpful in guiding our team to improve the logical design aspects of this project. After we completed collecting our data for the diseases, we stored all of these values in an excel sheet and forward engineered the data, so we could import the files into MySQL.

**Lessons Learned**

Building our database was where we ran into our most challenging problems. The process was not as smooth as we had initially imagined, and there were many corrections and alterations along the way. Our group encountered several problems that needed to be addressed as we approached the deadline to turn in our final report and deliverables. As mentioned earlier, our initial ERD drafts did not satisfy many of the project’s requirements, and it took us multiple attempts to finally work through the issues. The database we had planned to create with our first ERD was a non-relational database - which defeated the purpose of the project (creating a relational database). Thankfully, we had come to realize this mistake after consulting our TA, Isabeau, and Professor Duffy. With their help we completely reorganized our ERD to correctly exemplify concepts of relational databases. Practically scrapping an entire ERD and building a new one was difficult, but we had to acknowledge our lack of understanding and the mistakes we made to deliver a better project that satisfied everyone.

While gathering our sample data, we learned that organizing through the copious amounts of unrelated data and reformatting them into excel sheets was an arduous task. The steps it would take to import the necessary files with relevant data would take some time. Each team member had found their own datasets to use for the project, but only two members contributed to 90% the data gathered. The other members were then allocated different tasks, and they accepted the distribution of work. Such a mishap was due to the lack of communication that our team occasionally suffered due to the nature of virtual learning. Eventually, we developed our communication with weekly meetings and became a much more effective team as we learned from this experience that communication is very much key to success. Although we believed that everything was under control and overseeing the work of all members was not necessary, we needed to know exactly what each of us needed to do, and what we needed from the other members. By developing better communication skills, our team dynamic and management significantly improved.

**Potential Future Work:**

This database has an enormous amount of potential for real-world applications, especially with what is currently occuring around the world. As stated above, our initial goal of our database was to track the spread of historical diseases over time to be able to more-effectively predict how new diseases would spread around the world. Although the final database differed slightly with the amount of data that was easily available, our general theme remained consistent with our goals. We plan on uploading our database code to Stack Exchange which would allow other researchers to utilize our database theme and input additional data they have found to effectively predict how future diseases would impact the world. This database could be used to upload the enormous amount of data collected from the CoronaVirus when reliable data becomes available and the multiple waves have occured.

**WORKS CITED**

Anthrax. (2001, October 25). Retrieved from <https://en.wikipedia.org/wiki/Anthrax>

Bubonic plague. (2008, March 19). Retrieved from https://en.wikipedia.org/wiki/Bubonic\_plagueBubonic plague. (2008, March 19). Retrieved from <https://en.wikipedia.org/wiki/Bubonic_plague>

BCG vaccine. (2001, September 15). Retrieved from <https://en.wikipedia.org/wiki/BCG_vaccine>

CDC H1N1 flu | Questions & answers novel H1N1 influenza vaccine. (n.d.). Retrieved from <https://www.cdc.gov/h1n1flu/vaccination/public/vaccination_qa_pub.htm>

Cholera. (2002, February 25). Retrieved from <https://en.wikipedia.org/wiki/Cholera>

Cipargamin. (2010, September 3). Retrieved from <https://en.wikipedia.org/wiki/Cipargamin>

Coronavirus disease 2019. (2020, February 5). Retrieved from <https://en.wikipedia.org/wiki/Coronavirus_disease_2019>

Corynebacterium diphtheriae. (2020, April 16). Retrieved from <https://en.wikipedia.org/wiki/Corynebacterium_diphtheriae>

Dengue fever. (2002, February 18). Retrieved from <https://en.wikipedia.org/wiki/Dengue_fever>

Ebola 2014-2016 outbreak dataset. (n.d.). Retrieved from <https://www.kaggle.com/imdevskp/ebola-outbreak-20142016-complete-dataset>

Ebolavirus. (2020, May 10). Retrieved from <https://en.wikipedia.org/wiki/Ebolavirus>

Ebola vaccine. (2014, October 2). Retrieved from <https://en.wikipedia.org/wiki/Ebola_vaccine>

Hiv. (2001, November 21). Retrieved from <https://en.wikipedia.org/wiki/HIV>

Influenza A virus subtype H2N2. (2005, October 20). Retrieved from <https://en.wikipedia.org/wiki/Influenza_A_virus_subtype_H2N2>

Influenza pandemics. (n.d.). Retrieved from <https://www.historyofvaccines.org/content/articles/influenza-pandemics>

Kivu Ebola epidemic. (2018, July 27). Retrieved from <https://en.wikipedia.org/wiki/Kivu_Ebola_epidemic>

List of country codes by Alpha-2, Alpha-3 code (ISO 3166). (n.d.). Retrieved from <https://www.iban.com/country-codes>

Malaria. (2001, October 30). Retrieved from <https://en.wikipedia.org/wiki/Malaria>

Measles. (2002, June 22). Retrieved from <https://en.wikipedia.org/wiki/Measles>

Middle East respiratory syndrome. (2013, May 30). Retrieved from <https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome>

MMR vaccine. (2002, August 8). Retrieved from <https://en.wikipedia.org/wiki/MMR_vaccine>

Polio. (2002, January 17). Retrieved from <https://en.wikipedia.org/wiki/Polio>

Preventing tetanus, diphtheria, and pertussis among adults: Use of tetanus toxoid, reduced diphtheria toxoid and acellular pertussis vaccine recommendations of the Advisory Committee on Immunization Practices (ACIP) and recommendation of ACIP, supported by the healthcare infection control practices advisory committee (HICPAC), for use of Tdap among health-care personnel. (n.d.). Retrieved from <https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5517a1.htm>

Rabies. (2002, June 15). Retrieved from <https://en.wikipedia.org/wiki/Rabies>

Russian scientist dies of Ebola after lab accident. (n.d.). Retrieved from <https://www.cidrap.umn.edu/news-perspective/2004/05/russian-scientist-dies-ebola-after-lab-accident>

Smallpox. (2001, October 10). Retrieved from <https://en.wikipedia.org/wiki/Smallpox>

Smallpox vaccine. (2002, July 8). Retrieved from <https://en.wikipedia.org/wiki/Smallpox_vaccine#Types>

Study supports overall effectiveness of 2009 H1N1 vaccine. (n.d.). Retrieved from <https://www.cidrap.umn.edu/news-perspective/2011/08/study-supports-overall-effectiveness-2009-h1n1-vaccine>

Tuberculosis. (2001, September 29). Retrieved from <https://en.wikipedia.org/wiki/Tuberculosis>

Vaccines. (2018, December 13). Retrieved from <https://www.cdc.gov/cholera/vaccines.html>

Vaccine-preventable diseases. (2009, May 16). Retrieved from <https://en.wikipedia.org/wiki/Vaccine-preventable_diseases>

Viboud C , et al. (n.d.). Global mortality impact of the 1957-1959 influenza pandemic. - PubMed - NCBI. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/26908781>

Years of Ebola virus disease outbreaks error processing SSI file. (2019, December 4). Retrieved from <https://www.cdc.gov/vhf/ebola/history/chronology.html>

Zika fever. (2007, June 28). Retrieved from <https://en.wikipedia.org/wiki/Zika_fever>

1968 flu pandemic. (n.d.). Retrieved from <https://www.britannica.com/event/Hong-Kong-flu-of-1968>

2009 H1N1 pandemic. (2020, March 13). Retrieved from <https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html>

2009 swine flu pandemic vaccine. (2009, June 15). Retrieved from <https://en.wikipedia.org/wiki/2009_swine_flu_pandemic_vaccine>