**Project Title:** American Cancer Databases

**Names & Email Addresses of Investigators/Team Members:**

Bethel Kumsa | [bkumsa@smu.edu](mailto:bkumsa@smu.edu)

Ben Goodwin | [bgoodwin@mail.smu.edu](mailto:bgoodwin@mail.smu.edu)

**Introduction:**

*Introduction to be written once the final draft is complete.*

**Background/Research:**

The National Cancer Database (NCDB) was created as a tool for documenting, researching, and improving cancer diagnoses, treatments, and outcomes in the United States. In the decades since its 1989 establishment through a partnership between the Commission on Cancer and the American College of Surgeons, the NCDB has undergone various changes in support of its foundational goals.

The database’s first structural changes took place in 1996 and 2001. Prior to 1996, hospitals were able to voluntarily submit clinical cancer data to the NCDB. However, beginning in 1996 submission of cancer data to the NCBD became a requirement of members of the Commission on Cancer (CoC). Following 2001, database submissions and privileges were limited exclusively to member institution of the CoC. NCDB’s next largest systemic changes took place in 2002 and 2005, during which the quality constraints on data submissions were greatly increased.

NCDB has honed other various notable strengths. Amongst these strengths is geographically diverse data representing 49 states and Puerto Rico. Further, collected data is highly standardized. To meet its high-quality constraints, the NCDB employs rigorous internal annual quality testing as well as trained, Certified Tumor Registrars (CTRs) who maintain and correct data in real-time.

Despite these strengths, there are concerns that NCDB’s sole reliance on CoC member institutions for data collection may bias the data’s representativeness of the larger American population. Fortunately, amongst NCDB’s growing strengths are large and growing sample sizes, which may effectively offset such bias. Recent research found that the NCDB records 72% of new cancer diagnoses in the United States. This is an increase from previous estimates. The percent of cancer diagnoses represented by the database vary by cancer type, geographical location and demographics.

**Case Study:**

A possible major area of improvement is the NCDB’s accessibility. NCDB limits both the submission and accessibility of its data to members of its governing institutions. To more clearly understand how NCDB’s accessibility compares to related national databases with similar goals, we attempt querying and analyzing breast cancer data through the National Cancer Database, Surveillance, Epidemiology, and End Results Program, and the National Cancer Institute’s Genomic Data Commons.

National Cancer Database

* Access approval process
* Need for specialized software
* Granularity of publicly accessible data
* Institutional goals & governing bodies – Specifically, how does this impact the availability of the data?
* Data analysis of selected cancers or cancer variables

Surveillance, Epidemiology, and End Results

* Access approval process
* Need for specialized software
* Granularity of publicly accessible data
* Institutional goals & governing bodies – Specifically, how does this impact the availability of the data?
* Data analysis of selected cancers or cancer variables

National Cancer Institute’s Genomic Data Commons

* Access approval process
* Need for specialized software
* Granularity of publicly accessible data
* Institutional goals & governing bodies – Specifically, how does this impact the availability of the data?
* Data analysis of selected cancers or cancer variables

**Conclusion:**

*Conclusion to be written once the final draft is complete.*

**References:**

*References will be converted to appropriate format for final draft. References will also be incorporated throughout the paper above.*

1. [Link](https://link-springer-com.proxy.libraries.smu.edu/article/10.1245/s10434-007-9747-3) |2008 | The National Cancer Data Base: A Powerful Initiative to Improve Cancer Care in the United States
2. [Link](https://smu.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi_proquest_miscellaneous_934268616&context=PC&vid=01SMU_INST:01SMU&lang=en&search_scope=CentralIndex&adaptor=Primo%20Central&tab=CentralIndex&query=any,contains,database%20management&sortby=rank&offset=0) |2011 | Biocurators and Biocuration: surveying the 21st century challenges
3. [Link](https://link.springer.com/article/10.1245%2Fs10434-013-2901-1) | 2013.03 | Comparison of Cases Captured in the National Cancer Data Base with Those in Population-based Central Cancer Registries
4. [Link](https://search-proquest-com.proxy.libraries.smu.edu/docview/2121471806/fulltextPDF/DB17E44A6A944D61PQ/1?accountid=6667) | 2013.08 | Domain knowledge and data quality perceptions in genome curation work
5. [Link](https://ascopubs.org/doi/pdf/10.1200/PO.17.00175) |2018 | Future of Evidence Synthesis in Precision Oncology: Between Systematic Reviews and Biocuration
6. [Link](https://link-springer-com.proxy.libraries.smu.edu/article/10.1245/s10434-019-07215-z) | 2019.03 | Thirty Years of Counting Cancer Cases: The Collective Effort and Efficacy of the National Cancer Database
7. [Link](https://link.springer.com/article/10.1245%2Fs10434-019-07213-1) | 2019.02 | Incident Cases Captured in the National Cancer Database Compared with Those in U.S. Population Based Central Cancer Registries in 2012–2014