

The Importance of Data Ethics in Data Interpretation

Tonja Cunnigham

August 3, 2024

As a newly minted MLS graduate, I embarked on the Spring 2024 National Center for Data Services (NCDS) / Data Curation Network (DCN) internship with a quaint perspective on data. I believed that data, particularly research data, existed as an objective collection of facts and figures, shielded from human biases. However, my internship experience has dispelled this notion. Data and research do not exist in isolation; they are shaped by the biases, assumptions, and limitations of those who generate, collect, analyze, and, most importantly, interpret it.

Following a presentation and discussion about human participants in research with Jen Darragh, Senior Research Data Management Consultant at Duke University, I gained a deeper appreciation for ethical considerations in research and data management. It led me to consider how ethical lapses along the path of a data lifecycle may have consequential negative effects on end users and society at large. Our presentation included well-known examples of unethical research in the health sciences field, such as the story of [Henrietta Lacks](#), as well as the documents and guidelines established to safeguard research participants, such as [the Nuremberg Code](#), [Declaration of Helsinki](#), [the Belmont Principles](#), [the Common Rule](#). While these examples focused attention on the violation of human rights of those participating in research studies, the subject led me to consider the importance of extending concern about

data ethics beyond the conclusion of the research. What happens when data and research are used to justify cruel and unfair practices in society?

Consider the eugenics program that existed in North Carolina for almost 50 years. From 1933 to 1977, the North Carolina Eugenics Board—established by the state legislature—approved the sterilization of citizens deemed to be somehow mentally unfit (Kickler, 2024). Over time, criteria expanded to include criminality, promiscuity and even poverty. Misguided by crude interpretations of scientific theories (such as Darwin’s natural selection and Mendel’s laws of inheritance), the board authorized this cruel treatment for over 7,000 people (Stockton, 2016). Orders for sterilizations were often reviewed in just 10-15 minutes, accompanied by minimal medical history information. Institutions, legal or charitable, could request sterilization for inmates, patients, and even welfare recipients. In some instances, coercive tactics were employed, threatening denial of welfare benefits to vulnerable women if they did not comply (Rose, 2022).

The ethical violations of the program were numerous, often including a lack of informed consent. Initially, both men and women, primarily white citizens, underwent sterilization. However, by the 1960s, the program shifted its focus. Black citizens, the poor, women, and individuals with disabilities became the primary targets. Post-1964, over 60% of those sterilized were Black, and 6 out of 7 were women. The youngest known victim was a 14-year-old girl who went to a hospital to give birth after being raped, only to be deemed promiscuous and recommended for sterilization by a hospital social worker (N.C., 2012). The sociological effects of this program may not have been studied, but I would imagine that allowing medical

professions and social workers, supposed trusted members of a community, to have the power to request that such draconian actions be taken against citizens created a long-lasting wariness and breach of trust. Although many states had similar programs during their history, the North Carolina program was the only one that allowed sterilization orders for citizens who were not institutionalized.

Perhaps this program could have ended much sooner, or have never begun, if stringent ethical data interpretation guidelines were applied to the use of eugenics research as the basis for public policy. There are several data ethics principles that policymakers should consider in this era to protect people from similar assaults to their rights and freedoms, reproductive and otherwise, including:

1. Bias awareness in the steps of data collection, processing, and analysis (Commerce Data Ethics Framework, 2022).
2. Fairness in considering how different demographic groups are represented in data and how the resulting interpretation could impact various populations. (Ferrara, 2024).
3. Algorithmic accountability to blunt the effects of programmed biases in artificial intelligence driven data analysis, leading to issues such as discriminatory policing practices against certain demographic groups (There's More, 2022).
4. Transparency about the known assumptions, limitations and potential biases in the steps of data collection, analysis, and final interpretation (Weiss et al, 2023).

My experiences during the NCDS/DCN internship have significantly impacted my understanding of the data lifecycle, highlighting the critical importance of ethical considerations

in every aspect of data curation, management, and interpretation. Data ethics and interpretation are inseparable. The main purpose of creating and sharing research is to add to the body of knowledge that can be used one day. The question then becomes who uses it and to what end. As curators, managers, and stewards of data, we must recognize our biases, uphold ethical standards, and ensure that data reflects reality without distortion. And we should not shy away from challenging the misuse of data and research when we become aware of it.

References

Commerce data ethics framework. *United States Department of Commerce*. (2022).

<https://www.commerce.gov/sites/default/files/2023-02/DOC-Data-Ethics-Framework.pdf>

Ferrara, E. The Butterfly Effect in artificial intelligence systems: Implications for AI bias and fairness. *Science Direct*. (2024).

<https://www.sciencedirect.com/science/article/pii/S266682702400001X?via%3Dihub>

Kickler, T. Eugenics Board. (n.d.). *North Carolina History Project*.

<https://northcarolinahistory.org/encyclopedia/eugenics-board/>

N.C. to compensate victims of sterilization in 20th century eugenics program. *ABC News*, (2012).

<https://abcnews.go.com/Health/WomensHealth/north-carolina-compensate-victims-eugenics-program-sterilized/story?id=15328707>

Rose, J. (2022). A brutal chapter In North Carolina's eugenics past. WFAE.

<https://www.wfae.org/united-states-world/2022-02-22/a-brutal-chapter-in-north-carolin-as-eugenics-past>

Stockon, R. (2016). How American eugenics programs inspired The Nazis. *All That is Interesting*.

<https://allthatsinteresting.com/american-eugenics>

There's More to AI Bias Than Biased Data, NIST Report Highlights. (2022). *National Institute of Standards and Technology (NIST)*.

<https://www.nist.gov/news-events/news/2022/03/theres-more-ai-bias-biased-data-nist-report-highlights>

Weiss, M., Nair, L. B., Hoorani, B. H., Gibbert, M., & Hoegl, M. (2023). Transparency of reporting practices in quantitative field studies: *Journal of Informetrics*.

<https://www.sciencedirect.com/science/article/pii/S1751157723000214>