DAY ONE PROJECT

Opening Up Mortality Data for Health Research

Sam Roosz
Michael Stebbins
Horatio Thomas

February 2021

Summary

Comprehensive and reliable mortality data is vital for public health research. Improving our infrastructure for managing these data will generate insights that promote longevity and healthy aging, as well as enable more effective response to rapidly evolving public health challenges like those posed by the COVID-19 pandemic. A modernized mortality data system will ultimately be self-sustaining through access fees, but will require federal investment to update state reporting infrastructure and data use agreements. The Biden-Harris administration should launch an effort to modernize our nation's infrastructure for aggregating, managing, and providing research access to mortality data.

Challenge and Opportunity

No metric for promoting the economy or health is more important than whether, when, and why Americans are dying. Yet the nation's system of record for death, the National Death Index (NDI) managed by the CDC, remains woefully unsuited to its task. Modernizing the NDI so it is fit for its purpose will improve the lives of Americans for generations to come.

The NDI was a triumph when it was established in the 80's. A researcher who wants to access the NDI must first endure a 2-3 month application review period, and, if recent data is desired, wait for the next annual batch refresh—constraints that are hardly suitable for mounting a response to a fast-moving public health challenge. Once researchers receive the data, they then must find a way to deal with uncertainty in cause-of-death reporting—a cause of death may be listed as pneumonia when in fact the underlying cause was lung cancer. Finally, the data is all but inaccessible to most researchers using big data analysis methods. The pricing of the NDI is still anchored to the costs of a single paper record pull, meaning that researchers looking to bring modern data science to their work may face direct list costs enumerated in the tens of millions of dollars; unsurprisingly, this has the effect of suppressing such research. Additionally, approval criteria for use of the NDI fail to articulate a clear process for how to incorporate the data into an interventional clinical trial. These collective barriers hinder efforts to make use of the data for valuable industry-sponsored studies (e.g., for cancer or COVID-19).

Modernizing our mortality data infrastructure will unlock an extensive array of benefits: more effective policy responses to public health threats like COVID-19 and opioid abuse, scientific insights that promote healthy aging and longevity, and the ability for industry to run more efficient clinical trials and promote learning health systems. It may also revolutionize individual clinical care. A modernized mortality database would revolutionize the study of rare malignancies, where prospective studies are challenging due to disease infrequency and outcomes monitoring (e.g., survival) is limited due to fragmented care and poor record sharing between healthcare providers. In addition, many treatment decisions for cancer patients are

driven by the medical community's best understanding of life expectancy (e.g., in prostate cancer). Better mortality data will help both cancer patients and physicians make more informed decisions by aiding the formulation of more personalized prognoses for patients.

Realizing this promise of better health and longevity requires strong coordination with state vital statistics offices as well as modest investment in modernizing the infrastructure and business model. For reference, the entire federal acquisition cost for the data in the last annual reporting period amounted to only $2.3M across all jurisdictions.

Plan of Action

There are two viable approaches to realizing the opportunities afforded by modernizing our infrastructure for mortality data:

First, the Biden-Harris administration could issue a presidential memorandum directing the CDC to work with state vital statistic offices on developing new data use agreements with a modern pricing structure that supports large cohort research. To facilitate its adoption, federal funding should guarantee a revenue share to states on research access to the NDI that, at a minimum, exceeds the annual revenue each state received under the current system. For instance, the federal government could guarantee states revenue of at least 125% of 2019 figures, with any further federal revenue surplus distributed back to states proportionally based on population. If further federal funds are appropriated for use by state public health departments, participating in this revised framework on mortality data should be a requirement for the receipt of those funds. The administration should also direct the CDC to modernize its management of the NDI to both move away from reliance on physical media to digital access for all use cases and to ensure timely review of research proposals. Similarly, the CDC should ensure that clear requirements and timelines are established for NDI use in industry-sponsored interventional research, and that priority review be granted to COVID-19 related studies.

Second, the Biden-Harris administration could empower an NGO to manage the data and its governance for use by state and federal entities, as well as by the research community. This can be done through, or in conjunction with, the National Association for Public Health Statistics and Information Systems (NAPHSIS), which represents state vital records and public health statistics offices in the US. As NAPHSIS maintains relationships with state vital statistics offices - and is indeed involved in setting the interoperability standards that enable the NDI - it may serve as an effective conduit to engaging states on a new business model. The administration can direct the CDC to grant funds to NAPHSIS (or a new NGO affiliate) to aggregate and manage access to these data, in addition to improving data quality through linking to external datasets.

Conclusion

The Biden-Harris administration should launch an effort to modernize our nation's infrastructure for aggregating, managing, and providing research access to mortality data for researchers working to understand health, aging, and contributions to all causes of death from accidents to disease. Doing so will enable improved understanding of the causes of death and insights into healthy living.

DAY ONE PROJECT

Frequently Asked Questions

Why hasn't this been done already?

The National Death Index has no federal law or appropriation provisioning its operation. Instead it grew out of an effort initiated in the 1970s that (impressively) built a fragile voluntary coalition of all 57 reporting jurisdictions. The voluntary nature of this agreement meant that the scope was defined by the conditions of the most restrictive member. There are reasonable fears that attempts to revise the NDI and its contracts will expose the existing arrangement to scrutiny in a time when mortality data itself is politicized, potentially leading to a worse outcome than the status quo. Similarly, there is concern about how expanded use of the NDI might come into conflict with state vital records and privacy statutes. This fragility only increases the importance of establishing a stable solution for monitoring our nation's mortality data through the use of appropriate federal incentives for states to overcome these obstacles.

What does current Federal / State coordination look like?

State vital statistics offices aggregate mortality data from coroners' offices. Most states receive this data through electronic death registration systems (EDRS), though 8 jurisdictions currently do not use them (including NC, WV, CT, RI), introducing delay in the receipt of records. The states then report the data to the National Center for Health Statistics at CDC, where it is assembled into the NDI and released annually (an "early release file" is also available quarterly after 90% of expected deaths are reported). The state vital records offices are represented by the National Association for Public Health Statistics and Information Systems (NAPHSIS) which supports national coordination of its members.

Don't we already release death data?

The National Death Index is the only complete resource for death data, though there are others that hold partial death information. The Social Security Administration offers the Death Master File, though following changes in reporting rules in 2011 it only covers a fraction of overall deaths (<20%). NAPHSIS offers a service called EVVE Fact of Death, which allows users to query whether a person has died with low latency, but only roughly half of states participate for healthcare use cases, and it doesn't include cause-of-death information. Private industry has also created partial solutions from scraping data from obituaries and other similar sources.

Doesn't this having identities stolen?

No more than published obituaries do. Making these data useful doesn't require disclosing social security numbers or other similar privileged personally identifiable information from the NDI (identifiers are submitted to NDI and fact and cause of death are returned). If anything, the lack of an effective national repository for mortality data increases the risk of stolen identities by making it harder for Medicare / Medicaid, state voter rolls, etc. to identify individuals registering as a known decedent.

Will this cost a lot of money?

As the NDI was never supported by legislative appropriation, it has been entirely self-funding through fees. These fees are less than they could be given the many restrictions on data use that make the data less fit for modern use, and as a result the total funding for NDI is remarkably small (as noted above, the total amount paid out to all jurisdictions was only $2.3M in the last reporting period). A relatively small incentive for states to refresh these contracts will go a long way in promoting public health.

How is the death data most likely to be used?

There are many ways this data can be used to promote health and longevity. Some examples include monitoring long-term outcomes of clinical trial populations (e.g., this would be tremendously useful for COVID vaccines), identifying effects of cancer treatments on overall survival of patients, and identifying nutritional and environmental correlates to longevity (made possible through linking mortality data at scale to other records). Currently, studies that track death as an outcome either need to do so in a very short time period or with rigorous long-term patient follow-up. Making the NDI more accessible will dramatically increase the number of studies that can ask questions about why Americans die and what levers we can pull to promote longevity—there will undoubtedly be many creative approaches by researchers not contemplated here.

How do we get states to come along?

We suggest a close collaboration with NAPHSIS (which understands the needs and barriers of state vital statistics offices) and tying these revised methods and contracts to dollars appropriated for state public health activities through potential upcoming legislation surrounding the pandemic.

Why is death data important for policy?

It's hard to improve what you don't measure. To assess the impact of interventions (for instance in response to COVID or the opioid epidemic) we need to have a reliable measure of the outcome we're looking to reduce, and right now we just don't. Likewise, with the integrity of our elections being cast into the public doubt, a national death registry could serve to blunt myths about large volumes of dead people voting.

Why NAPHSIS?

We suggest NAPHSIS as it has established relationships and trust with all of the national reporting jurisdictions, and has its own partial fact-of-death reporting system in the form of EVVE. These suggest it could be a capable and effective honest broker for this data. The maintenance of this data by a non-governmental honest broker may also increase trust in the data in periods when a federal administration may be seen as untrustworthy by segments of the public.

Didn't the National Institutes of Health (NIH) announce that NDI would be free for all researchers?

NIH, realizing the high costs of NDI access deterred much mortality research, decided to carve out these costs for its researchers and pay them directly (up to $100,000 without approval). This is frankly draining the NIH budget to put a band-aid on a broken system. NIH still pays for the data, and large studies are still deterred by the $100,000 limit, not to mention the challenges still facing non-NIH researchers.

Doesn't the NDI provide discounts for higher record volumes?

The NDI suggests users inquire about potential discounts for record requests of greater than 100,000 records, but fails to provide any transparent pricing or predictability for such requests. As the baseline pricing would lead to prohibitive fees for multi-year large cohort studies, this lack of clear pricing guidelines has the effect of deterring many valuable large cohort research studies.

Where do we go from here / what more can be done?

The first priority should be revising the access structure and use cases for this data. Following that, enriching the mortality data would be incredibly valuable for research. There are reporting biases in death data (e.g., one coroner will report a death as pneumonia while another will say lung cancer for the same death) that can be normalized through linking in other available data. Through linking in Healthcare Cost and Utilization Project (HCUP) or state all-payer claims database (APCD) data, there is potential to assign a probability score that a given death has an additional cause not listed on the coroner's report. This can be quite valuable in programmatic and policy activities to track the burden of deaths for areas of public interest like deaths of despair, opioid overdose, COVID, or similar.

About the Authors

Sam Roosz is co-founder and CEO of Crescendo Health, a healthcare technology company focused on empowering patients to contribute their data to clinical trials. Most recently, Sam co-founded Datavant, the leading provider of de-identification and linking solutions for health data. He is currently launching a data-driven healthcare non-profit and planning his next venture. Sam received a degree in Molecular and Cellular Biology from Harvard and holds an MBA from the Stanford Graduate School of Business.

Michael Stebbins, Ph.D. is a geneticist and public policy expert who served as the Assistant Director for Biotechnology in the Obama White House Office of Science and Technology Policy. He is currently the President of Science Advisors, a science and health consulting firm he founded in 2018 to provide science, technology, and public policy guidance to private companies, philanthropies, and non-profit organizations. He is also a Senior Fellow of the Federation of American Scientists. While at the White House, Dr. Stebbins' work led to large initiatives across the Federal government to address antibiotic resistance, protect pollinators, improve veterans' mental health, increase access to federally funded scientific research publications and data, promote the preferential purchasing of antibiotic free meats, reform the regulatory system for biotechnology products, drive Federal purchasing of bio-based products, and improve the management of scientific collections. Dr. Stebbins previously served as the Vice President of Science and Technology for the Laura and John Arnold Foundation, science advisor to the Obama Presidential Campaign, and on the Obama White House Transition Team. He is the former director of biology policy for the Federation of American Scientists and worked for U.S. Senator Harry Reid and at the National Human Genome Research Institute. Before coming to Washington, he was a senior editor at Nature Genetics. Dr. Stebbins is on the Board of the Value in Cancer Care Consortium and chair of the Board for Vivli. He serves on the scientific advisory boards for Datavant and Amida Technology Solutions.

Horatio Thomas, M.D., M.M.Sc. is a resident physician in the Department of Radiation Oncology at the University of California, San Francisco (UCSF). He completed his undergraduate and medical school training at Harvard University. His academic interests include medical education and clinical and translational outcomes in oncology. With regards to medical education, his research explores personal and professional factors that affect residents' professional pursuits, resource-efficiency in residence, and the element of training programs that shape residents' professional development. In oncology, he investigates social, clinical, pathological, and molecular features of patients with

malignancy that correlate with their clinical outcomes. He continues to build on this research with emphasis on how current clinical practices affect the health of underserved populations based on socioeconomic status, race, sexual orientation, and/or gender identity.

About the Day One Project

The Day One Project is dedicated to democratizing the policymaking process by working with new and expert voices across the science and technology community, helping to develop actionable policies that can improve the lives of all Americans. For more about the Day One Project, visit dayoneproject.org