

## Mortality and Prevalence Data on NIH's Categorical Spending Page

### Description of Data Sources

The disease statistics included in the “Estimates of Funding for Various Research, Condition, and Disease Categories (RCDC)” table are nationally-representative data gathered in 2019 by the National Center for Health Statistics (NCHS) at the Centers for Disease Control & Prevention (CDC). These data reflect two types of disease burdens: mortality and prevalence.

The mortality data are drawn from the [National Vital Statistics System](#) (NVSS), and show the number of deaths in which a particular disease or condition was cited as a cause of death on a deceased individual's death certificate including, but not limited to, the single underlying cause of death. In cases where more than one condition is mentioned, a particular death will be counted in more than one RCDC category. This method may under-estimate the number of deaths for some conditions. For example, chronic conditions such as obesity and atherosclerosis can greatly contribute to other conditions such as heart disease but may not be explicitly mentioned on the death certificate.

Data on prevalence are drawn from the [National Health Interview Survey](#) (NHIS). This survey produces estimates of the percentage of US residents who indicated that they were affected by a particular health condition. The survey collects different information for different conditions: some questions ask about occurrences in the last 12 months, while others ask about lifetime experiences; some questions ask about symptoms, while others ask if a health professional has ever made a particular diagnosis; and other questions may be asked regarding either adults or children under 18. Conditions for which prevalence is estimated among children are indicated as such in the table. Prevalence of conditions that are sex-specific (prostate cancer, breast cancer, etc.) are reported among that sex. In addition, the reported numbers are accompanied by a standard error, which estimates the uncertainty introduced by sampling. In cases where the standard error for a particular disease or condition is too large to meet NCHS standards of reliability or precision, those data are not reported.

In order to ensure that data for each RCDC category was collected in a consistent manner, the posted information comes from two broad data sources provided by NCHS. Other disease statistics may be publicly posted by CDC and other federal sources which are rigorously collected, but are not collected in the same way across a wide range of diseases. In addition, NCHS assembled the posted statistics specifically to match NIH's RCDC categories, and so some numbers on this page which were uniquely assembled to fit a category may not be publicly posted outside of this analysis.

### Frequently Asked Questions About This Analysis

For more information on how the RCDC categories are generated, please see the RCDC FAQ at <https://report.nih.gov/rcdc/faqs.aspx>.

#### Q. WHY IS NIH REPORTING THESE DATA?

A. NIH is reporting these data alongside NIH's categorical funding estimates in order to provide the public and policymakers with information that is helpful for understanding the NIH research portfolio and its relationship to public health needs. NIH chose to report estimates from these two sources because they provide relatively consistent and, within each source, comparable Federal data for a broad range of categories.

#### Q. ARE THESE DATA THE DEFINITIVE INDICATORS OF PUBLIC HEALTH NEED FOR EACH DISEASE AND CONDITION?

A. No, these data are not meant to be definitive measurements of disease burden, mortality, or prevalence for a given condition in the US. In general, NIH tries to understand disease burdens by examining patterns—both within the US and internationally—using multiple methods and measurements, chosen on a case-by-case basis as appropriate for each disease or condition.

#### Q. WHY DO THESE DATA COVER SOME CATEGORIES AND NOT OTHERS?

A. Some RCDC categories represent areas of research rather than disease (i.e., genetics, neurosciences, biomedical imaging), or population segments and their concerns (e.g., Minority Health, Health Disparities) and therefore do not map to diseases or conditions. Further, a number of RCDC categories representing diseases or conditions are not captured by the methods and questions of either data source. While estimates of prevalence or mortality may exist for these categories from other data sources, it is inadvisable to compare such estimates with the ones listed here, since substantive differences in estimates can result from differences in the methods by which they were collected.

#### Q. HOW WERE THESE DATA GENERATED?

A. NIH and NCHS collaborated to map relevant International Classification of Disease (ICD) codes, which allow for standardized reporting of diseases and conditions for epidemiology, health management and clinical purposes, to each RCDC disease or condition category, as appropriate. NCHS then used those codes to generate mortality and prevalence data. In some cases, the appropriate matching between ICD codes and RCDC categories was somewhat subjective, and in such cases these matches represent the best judgement of NIH and NCHS staff.

#### Q. ARE THE POSTED VALUES UNIQUE TO THIS ANALYSIS? DO THESE FIGURES EXIST ELSEWHERE?

A. All reported figures were derived from NVSS or NHIS, which provide mortality and prevalence data for particular ICD codes. Since these ICD-RCDC matches were created specifically for the purpose of this reporting, some of the combinations of ICD codes may be unique to this analysis, and may not have appeared previously.