

# Letters to the Editor

---

## Researcher's guide to health statistics

Sir

Health and medical statistics data seems ubiquitous. After all, the popular media in addition to the numerous health focused dot coms daily published, reports make it seem as if a wealth of data is readily available and in an easy to read format. Yet in reality, finding medical data can prove to be quite difficult. Before starting a search for health statistics, it is helpful to understand how health and medical data is collected and what types of data are most commonly available.

**What are health statistics?** Health and medical statistics incorporate a variety of data types. The most common statistics reported are vital (birth, death, marriage, divorce rates), morbidity (incidence of disease in a population) and mortality (the number of people who die of a certain disease compared with the total number of people). Other common statistical data reported are health care costs, the demographic distribution of disease based on geographic, ethnic, and gender variables, and data on the socioeconomic status and education of health care professionals.

**Challenges to finding data.** Seekers of health and medical statistics should be aware of the following challenges in their quest for data. Firstly, it is important to know that data collection the United States of America (USA) is a fairly recent phenomenon. For example, it was not until 1956 that Congress enacted legislation to establish the USA National Health Survey in order to collect statistics on disease, injury, impairment, disability and other health related topics. Locating historical data (pre 1956) will be more time consuming and consultation with primary resources will be necessary unless an analysis has already been published on your topic. Secondly, data collection is decentralized. Numerous federal, state and local agencies in addition to non-for-profit organizations are involved in the collection and dissemination of health related data. As a result, there may be little to no data in some areas and duplication of data in others. There may also be variations between how data is collected and described between these various agencies. For example, differences in time periods covered, geographic areas, and sample sizes used to tabulate the data will differ. In addition, pay special attention to the definitions and coverage used by each reporting agency as the terminology used may have different meanings. It is difficult to obtain a precise figure when it comes to national samples. In most cases the numbers reported are estimates. Also,

reporting standards and definitions change over time. If your figures seem higher or lower than expected, check to see if this change reflects actual trends or if the definition or reporting of your figure has changed. Finally, you may also see slight variations in the numbers reported between different agencies. Check to see whether the figures reported are compiled independently, or if they were taken from other sources. For example, commercial publishers often repackage government-compiled data. In most cases, the secondary reporting source has accurately reproduced the original data and conclusions, but when in doubt, track down the original report. Thirdly, data collection on a national level takes money and staff to compile. Only the federal government or a large organization has the resources to organize and manage large-scale data collection efforts. Finding data will be easier if you know who is responsible for collecting data in the area you are studying and the type of data they collect and publish. Government agencies collect data as part of a federal mandate; private organization's expend resources as part of their organizational mission or in response to the needs of the organization's membership. In either case the collecting agency has predetermined the need and collection scope for data in a specific area. It is possible that that data you are looking for has not been collected, or it has not been analyzed in the manner in which you need. You may need to obtain the original data and perform your own analysis. This will be easier if the data was gathered by a government agency, in which case the data may be available upon request, in a library, or for purchase. However, if the data was originally compiled by a private organization, it may not be possible to obtain the original data sets. Fourthly, data collection and analysis at a national level takes time to compile. Rarely will you find "real time" data except for estimates. Most statistics are out of date to some degree by the time they are published. If a report is published in 2000 it does not mean the data is from the most recent few years. Current health data is often based on extrapolations of older data, such as US Census data, compiled every 10 years, or on smaller sample sizes. Again, check to see when and how the data was collected. In most cases you will probably see at least a 3-5 year time lag between data collection and a published report. Fifthly, and compounding the difficulty of locating data at national or international level, there is a lack of adequate subject access through indexing sources (Meddling). Most health data is compiled and distributed by the federal government. It is notoriously difficult to track down government resources since federal reports for example, are as not well indexed. When possible see if the reporting

# Letters to the Editor

**Table 1** - Some useful publications for health and medical statistics.

Title	Author	Date
Statistics with applications to the biological and health sciences	Schork MA	2000
Statistics in medicine	Riffenburgh RH	1999
Probability without equations: concepts for clinicians	Holland BK	1998
How to report statistics in medicine: annotated guidelines for authors, editors, and reviewers	Lang TA	1997
Health statistics: an annotated bibliographic guide to information resources	Weise FO	1997
Biostatistics: a foundation for analysis in the health sciences	Wayne DW	1995

agency has the statistics you need. If not, then you will need to search the Journal literature in a bibliographic database like Medline<sup>1</sup> or CancerLit.<sup>2</sup> It may be difficult to always determine what type of statistical data is available from a citation or abstract. When in doubt, read the article to make sure that you are not missing potential data.

**International data.** Finally a note regarding obtaining international data. Locating international

health statistics is usually more problematic than locating USA data. Resources, particularly in developing countries, may not be available to collect data as extensively or comprehensively as in the USA. Data collecting efforts therefore vary considerably around the world and comparative data may not be available. In fact, there are really only 2 international agencies whose mission and budget allow for international data collecting. The World Health Organization<sup>3</sup> and the United Nations.<sup>4</sup> Consult the publications of these 2 agencies first when seeking health and medical statistics. Some useful books that can help to learn more with regards to health and medical statistics are listed in **Table 1**.

**Ibrahim Mansoor**  
*King AbdulAziz University Hospital*  
*Jeddah*  
*Kingdom of Saudi Arabia*

## References

1. US National Library of Medicine, Medline PLUS Service. Available from: URL: <http://www.ncbi.nlm.nih.gov/entrez>.
2. National Cancer Institute CancerLit Service. Available from: URL: <http://cnetdb.nci.nih.gov/cancerlit.htm>.
3. World Health Organization, international home page. Available from: URL: <http://www.who.int>.
4. United Nations 2001 home page. Available from: URL: <http://www.un.org>.