Disease Incidence and Prevalence – Summary of Findings

Click here to read a bibliography of this project.

Introduction

The Health Care Delivery Policy Program at Harvard University’s Kennedy School of Government’s Mossavar-Rahmani Center for Business and Government tracks the incidence and prevalence of fifteen chronic diseases identified in the Institute of Medicine’s *Crossing the Quality Chasm*. Statistical sources include newspapers, magazines, journals, reports, governmental sources (CDC, UNAIDS, etc.), non-profit agencies (American Diabetes Association, American Cancer Society, etc.), foundations and consultants (Robert Wood Johnson Foundation, RAND, etc) and original articles and presentations by the Health Care Delivery Policy Program.

Measurement of Disease Incidence and Prevalence

Difficulty in measuring disease incidence and prevalence stems from lack of standardized methods to obtain accurate data at a national level. The most recent Census (2000) did not ask participants about diseases and there is no national data registry for all medical conditions. Current measurement systems and surveys include, but are not limited to:

**National and State Registries**
- **National Notifiable Diseases Surveillance System** (NNDSS) - The Centers for Disease Control and Prevention and the Council of State and Territorial Epidemiologists operate the NNDSS. The NNDSS tracks selected diseases and conditions, mainly communicable, such as tuberculosis and AIDS. AIDS statistics are published semi-annually in the *HIV/AIDS Surveillance Report* and reported in the *Morbidity and Mortality Weekly Report* (MMWR).
- **Surveillance, Epidemiology and End Results Program** (SEER) – SEER is administered by the National Cancer Institute, which contracts with population based state registries to track data on residents diagnosed with cancer on an annual basis.

**Population Based Surveys**
- **Behavioral Research Factor Surveillance System** (BRFSS) – BRFSS is a monthly telephone survey administrated by the Centers for Disease Control and Prevention (CDC) of adults over 18 years of age. The BRFSS tracks current and lifetime prevalence of diseases. Of the fifteen chronic conditions this project studies, the BRFSS tracks arthritis, asthma and diabetes.
- **National Health and Nutrition Examination Survey** (NHANES) - Conducted by the National Center for Health Statistics (NCHS), NHANES is a continuous, annual, cross-sectional, nationally representative survey of civilian, noninstitutionalized people of all ages, providing estimates of prevalence of selected diseases. In addition to in-home interviews, medical professionals examine participants. About 5,000 individuals are surveyed in NHANES. The NHANES tracks asthma, diabetes, high cholesterol and hypertension.
- **National Health Interview Survey** (NHIS) - Administered by the National Center for Health Statistics, the NHIS is a continuous household interview survey of the civilian noninstitutionalized population. The 2007 NHIS interviewed 59,755 people. Proxy responses are collected for individuals (such as children) who cannot answer the questions. Data are collected on the respondent’s illnesses and chronic conditions within the past year or in his or her lifetime. The NHIS tracks asthma, diabetes and HIV testing.

**Epidemiological Studies** - Medical institutions, voluntary and health insurance organizations, researchers and epidemiologists conduct studies to estimate prevalence and incidence of diseases. For example, the National Bureau of Economic Research (NBER) used the Framingham Stroke Risk Profile, the Cardiovascular Health Study and the ARIC Study to predict incident stroke. Health insurance organizations often use claims data to estimate disease incidence and prevalence.

**Estimates** – Physicians in primary practice may have ballpark figures for the percent of asthmatics, diabetics, etc., based on years of their own experience in care. Pharmaceutical companies can estimate the number of people with high cholesterol based on sales of their cholesterol-lowering medications. Many estimates have no data source and are basically someone’s best guess at a number.
Figure 1. Prevalence of Chronic Diseases: 2004-2007

<table>
<thead>
<tr>
<th>Condition</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer Disease/Dementia</td>
<td>Alzheimer’s Association, National Institute on Aging, NIH</td>
</tr>
<tr>
<td>Arthritis</td>
<td>National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Center for Disease Prevention and Health Promotion, CDC, NIH</td>
</tr>
<tr>
<td>Asthma</td>
<td>Asthma and Allergy Foundation, CDC, MEPS, NCHS, NHIS, Phoenix Healthcare Intelligence</td>
</tr>
<tr>
<td>Back Problems</td>
<td>American Academy of Orthopedic Surgeons, MEPS, NIH, North American Spine Society</td>
</tr>
<tr>
<td>Cancer</td>
<td>American Cancer Society, NHIS, Lance Armstrong Foundation, Research America, SEER</td>
</tr>
<tr>
<td>Condition</td>
<td>Data Sources</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Depression/Anxiety</td>
<td>National Institute of Mental Health 67, Partners in Behavioral Health Sciences 37, Screening for Mental Health 80, Surgeon General 13</td>
</tr>
<tr>
<td>Diabetes</td>
<td>BRFSS 49, MEPS 76, NCHS 53, NHANES 18, RHIS 20, NIDDK 30</td>
</tr>
<tr>
<td>Emphysema</td>
<td>Mayo Clinic 48, NHIS 53</td>
</tr>
<tr>
<td>Gallbladder Disease</td>
<td>MEPS 76</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>CDC 77, Kaiser Family Foundation 1, NCHS 52, U.S. Preventive Services Task Force 10, UNAIDS 70</td>
</tr>
<tr>
<td>Heart Disease</td>
<td>BRFSS 62, NHANES 18, 30, NCHS 53</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>NHANES 18, 30</td>
</tr>
<tr>
<td>Hypertension</td>
<td>New York Academy of Sciences 34, NHANES 18, 30, 56</td>
</tr>
<tr>
<td>Stomach Ulcers</td>
<td>NIDDK 82, NCHS 53, NHIS 53, 69</td>
</tr>
<tr>
<td>Stroke</td>
<td>American Heart Association 54, NHANES 30, NHIS 53, 69</td>
</tr>
</tbody>
</table>

**Data Collection Challenges**

Ambiguity surrounding what is being tracked is an obvious problem - take the case of HIV/AIDS. Patients testing positive for Human Immunodeficiency Virus (HIV) infection may or may not develop Acquired Immunodeficiency Disease (AIDS), but could serve as carriers for the disease. Therefore, it is important to distinguish whether studies track HIV or AIDS (or both). Oftentimes, the distinction is unclear in the literature. The third study below probably provides the clearest picture of HIV/AIDS epidemiology:

- According to the Joint United Nations Program on HIV/AIDS (UNAIDS), in 2004, an estimated 950,000 people were living with AIDS in the US. 70
- HIV infection affected 850,000 to 950,000 persons in the United States in 2005, according to a Review of the Evidence for the U.S. Preventive Services Task Force. 10
- The Kaiser Family Foundation reported that in 2004, there were 944,000 cumulative cases of AIDS and 1.6 million HIV infections, at a rate of 40,000 new cases a year. In 2006, there were more than 1 million people living with HIV/AIDS and about 1/2 million living with AIDS. 1

In addition, many studies look at clusters of conditions, resulting in large (but vague) prevalence figures:

- The CDC testified that arthritis and chronic joint symptoms affected nearly 70 million Americans in 2005. 72
- The American Heart Association estimated that a total of 64.4 million Americans had one or more types of cardiovascular disease in 2004. 73
- In 2005/2006, NHANES asked 3,056 adults over age 40 if they have ever had pain or discomfort in their chest. 30% answered yes. 18
- The 2004 Medical Expenditure Panel Survey found the following number of people accounting for expenditures by disease: asthma/COPD - 46,718,000. 76
- A 2005 study sponsored by the National Institute of Mental Health found that there was a lifetime prevalence of 20.8% for mood disorders, which included depression. 9

Tracking chronic disease by age, gender, ethnicity or geographic location can also yield dramatically diverse results. Alzheimer’s Disease and dementia are examples of conditions that increase in likelihood as the studied population ages. For example, in 2006, the Alzheimer’s Disease Education and Referral Center estimated that about 5% of men and women ages 65 to 74 and nearly 50% of those age 85 and older had Alzheimer’s Disease. 38

Another layer of confusion stems from difficulties in determining manifestation of a disease or condition. Data may be collected on:

1. Whether an individual was diagnosed with, experienced symptoms of or were treated for a disease or condition
   - During their lifetime
   - During the year of the survey
   - Another specified time period
2. If they “have” a disease or condition at this particular point in time (a somewhat subjective measure at best)

The following examples illustrate these differences and their effects on prevalence:

- According to the National Comorbidity Survey Replication, more than 10% of noninstitutionalized Americans were estimated to have major depressive disorder in their lifetime, with 6.6% having an MDD during the last 12 months. 29
- Health United States listed that 10% of adults over age 20 had diabetes and 7.2% had diagnosed diabetes in 2004. 53
- Roughly two-thirds of Americans will suffer back pain at some point in their lives and 25% of Americans suffer from severe, chronic back pain. 5
Would an individual be considered asthmatic if he was diagnosed with asthma ten years ago, henceforth successfully treated and without an asthmatic episode since that time? What about someone who has not had an asthma attack during the past year? Keep in mind, however, that the individual could hypothetically have an asthma attack the day after he was surveyed.

In other words, when a disease is successfully treated, is someone still considered sick? Health United States 2006 estimated that 16.5% of the US population had high serum total cholesterol in 2004. The authors went on to clarify that, “individuals who take medicine to lower their serum cholesterol levels and whose measured total serum cholesterol levels are below the cut-offs for high and borderline high cholesterol are not defined as having high or borderline high cholesterol, respectively.” 29 From this perspective, those who were successfully treated were not counted as “having” the disease.

Even with the most carefully crafted surveys, participants may not understand the question or respond in a way consistent with the surveyor’s intent. There is also the possibility that survey participants may in fact “have” a disease, but were never tested for it and do not know they have it. For example, over 12 million Americans were diagnosed with some form of diabetes in 2004, while another 6 million were considered to be diabetic without their knowledge. 28

Some diseases are by their nature difficult to track. People suffering from depression may be reluctant to seek care because of the stigma associated with mental health disorders, or because the disease may be compromising their self-care or communication abilities. They may not even be aware that they are depressed, not know what depression constitutes, or be embarrassed by or deny the diagnoses. 58

Often, statistics on disease prevalence vary by who is reporting the data and how the numbers are interpreted. Two newspaper articles published in the same week summarized a Lancet study about a diabetes drug trial that included an International Diabetes Federation worldwide adult diabetes prevalence estimate of 5%: 21

- New York Times - About 20 million people in the United States had diabetes in 2006. 66
- Boston Globe - More than 21 million people in the United States had diabetes in 2006. 15

Numbers without a data source are a common problem in the press and online – “Some doctors believe that diabetes affects 20 million Americans,” 81 “Experts estimate that 360,000 new cases of Alzheimer's disease are diagnosed each year in the United States,” 27 and “Scientists think that up to 4.5 million Americans suffer from Alzheimer’s Disease” 38 provide little or no concrete evidence of how the “doctors”, “experts” or “scientists” arrived at their thoughts.

Changing Clinical Guidelines and the Emergence of “Pre-Diseases”

A popular reporting trend has been moving away from describing populations as “at risk” of a disease to labeling them as “prediabetic”, “prehypertensive”, “preasthmatic”, etc. Alarming new numbers have been touted for these “prediseases”:

- An estimated 58 million Americans (29% of adults) had hypertension and 45 million were prehypertensive (20% of adults) in 2004, and the lifetime risk of developing hypertension was 90%. 66
- At least 21 million people had diabetes in 2006 and 54 million had pre-diabetes 35

It is possible that anyone is likely to develop a disease at any point in his or her lifetime, but the relevance of these statistics has yet to be determined.

National clinical guidelines and standards for classification of diseases are reviewed and revised on a regular basis. For example, “based on new guidelines for diagnosis issued by an expert panel, about 41 million people had prediabetes in 2004.” 59 When changes occur, it is debatable how studies published under the older guidelines should be viewed. An algorithm to extract meaningful comparisons of data over time would be helpful to include in standards and guidelines when changes occur.

Disease Incidence

The incidence of disease is defined as the number of new cases of a disease or medical condition that began during a set period of time (usually one year). Incidence is harder to track than prevalence, especially for diseases that may take years to diagnose. There is no centralized registry that collects data on the onset of all diseases and conditions at the national level; HIV/AIDS and cancer are the only diseases researched by the HCDP that are tracked and reported through national surveillance systems. The Division of HIV/AIDS Prevention, Surveillance and Epidemiology, National Center for HIV, STD & TB Prevention is required to report new cases of HIV/AIDS on an annual and provisional basis. All 50 U.S. states and territories report AIDS cases to the CDC using a standard case report form. The Surveillance, Epidemiology and End Results Program (SEER) is affiliated with the National Cancer Institute (NCI). SEER cancer registries collect data on cancers diagnosed in hospitals, medical facilities and geographic areas (approximately 26% of the U.S. population). SEER cancer incidence rates are calculated based on Census Bureau population rates.
Various surveys, such as the BRFFS and NHIS, ask whether participants have had a certain disease during the course of the year, but that does not answer the question of whether the disease “began” that year. The disease could have been undiagnosed for many years before, and the patient could have waited until symptoms became apparent to seek treatment. Many times, people receive medical care for an unrelated issue, and their doctor finds that the patient had suffered a stroke or developed a tumor. Many people go through their lives unaware that they have high cholesterol, hypertension, etc., and may be at risk for a fatal event. If a person in this scenario has a heart attack and does not survive, there is no accurate way of determining when he or she first developed heart disease.

Figure 2 displays published disease incidence estimates for six out of the fifteen chronic diseases tracked by this project. Diagnosis, admissions, emergency care and claims data are collected by every medical organization to some degree, so the possibility of utilizing these forms of data for determining and better understanding the incidence and onset of other chronic disease exists.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s /Dementia</td>
<td>Florida Senate</td>
</tr>
<tr>
<td>Cancer</td>
<td>CDC 63, NCI 39, Research America 24, SEER 69</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>CDC 52, 53, 69</td>
</tr>
<tr>
<td>Heart Disease</td>
<td>AHRQ 78, NHANES 30</td>
</tr>
<tr>
<td>Stomach Ulcer</td>
<td>American Gastroenterological Association 57, Current Medical Diagnosis and Treatment 75, BRFFS 63</td>
</tr>
<tr>
<td>Stroke</td>
<td>American Heart Association 25, 30, 54, 71</td>
</tr>
</tbody>
</table>

### Relationship of Disease Prevalence to Disease Costs

Looking at chronic diseases both in terms of prevalence and cost has the potential to help planners and policymakers model disease management and healthcare delivery systems that allocate funds in more directed ways to achieve better health outcomes while reducing expenditures. The HCDP Deconstructing the Costs project examines the costs and expenditures on medical care at the national and individual levels. We have searched the literature for small and large-scale studies that result in cost estimates, particularly focusing on those studies that examine chronic diseases both in terms of their prevalence and cost.

Figure 3 compares disease prevalence to two metrics of cost – cost per person and total (direct and indirect) cost to the US. Some interesting observations can be made based on the chart:

- Some diseases, while high-ranked in prevalence, can be managed relatively inexpensively, through medication and lifestyle modification (like high cholesterol). Others, like Alzheimer’s Disease, often require expensive nursing home stays and have higher mortality (thus lowering prevalence).
Some conditions, like stroke, may have high initial costs (hospitalization, rehabilitation, etc.), with the possibility of lower continuing costs if the condition is well managed.

Some conditions, like back problems and diabetes, are ranked higher in their cost to the US than their cost per person. This may be partially due to national cost studies accounting for “indirect costs”, such as lost workplace productivity.

There are few major individual-level cost studies. Much of the available cost data are based on small sample sizes. For example, the Predictors Study, a multicenter cohort of 236 patients with probable Alzheimer’s Disease in 3 university-based AD centers found that total direct cost increased from approximately $9,239 per patient per year when patients were at the early stages of the disease, to $19,925 by year 4.  

---

**Figure 3. Comparing Prevalence and Cost Data: 2004-2007**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Prevalence</th>
<th>Cost per Person</th>
<th>Cost to US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High Cholesterol</td>
<td>Alzheimer/Dementia</td>
<td>Heart Disease</td>
</tr>
<tr>
<td>2</td>
<td>Back Problems</td>
<td>HIV/AIDS</td>
<td>Arthritis</td>
</tr>
<tr>
<td>3</td>
<td>Arthritis</td>
<td>Arthritis</td>
<td>Cancer</td>
</tr>
<tr>
<td>4</td>
<td>Hypertension</td>
<td>Stroke</td>
<td>Hypertension</td>
</tr>
<tr>
<td>5</td>
<td>Heart Disease</td>
<td>Emphysema</td>
<td>Alzheimer/Dementia</td>
</tr>
<tr>
<td>6</td>
<td>Asthma</td>
<td>Cancer</td>
<td>Diabetes</td>
</tr>
<tr>
<td>7</td>
<td>Diabetes</td>
<td>Asthma</td>
<td>Stroke</td>
</tr>
<tr>
<td>8</td>
<td>Depression/Anxiety</td>
<td>Gallbladder Disease</td>
<td>Back Problems</td>
</tr>
<tr>
<td>9</td>
<td>Cancer</td>
<td>Diabetes</td>
<td>Depression/Anxiety</td>
</tr>
<tr>
<td>10</td>
<td>Stomach Ulcer</td>
<td>Heart Disease</td>
<td>High Cholesterol</td>
</tr>
<tr>
<td>11</td>
<td>Alzheimer/Dementia</td>
<td>Depression/Anxiety</td>
<td>Emphysema</td>
</tr>
<tr>
<td>12</td>
<td>Stroke</td>
<td>Hypertension</td>
<td>Asthma</td>
</tr>
<tr>
<td>13</td>
<td>Gallbladder Disease</td>
<td>Back Problems</td>
<td>HIV/AIDS</td>
</tr>
<tr>
<td>14</td>
<td>Emphysema</td>
<td>Stomach Ulcer</td>
<td>Gallbladder Disease</td>
</tr>
<tr>
<td>15</td>
<td>HIV/AIDS</td>
<td>High Cholesterol</td>
<td>Stomach Ulcer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prevalence</th>
<th>Cost per Person</th>
<th>Cost to US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer/Dementia</td>
<td>4.5 million</td>
<td>$31,000</td>
<td>$92 billion</td>
</tr>
<tr>
<td>Arthritis</td>
<td>58 million</td>
<td>$17,500</td>
<td>$138.5 billion</td>
</tr>
<tr>
<td>Asthma</td>
<td>29 million</td>
<td>$5,336</td>
<td>$22 billion</td>
</tr>
<tr>
<td>Back Problems</td>
<td>60 million</td>
<td>$527</td>
<td>$56 billion</td>
</tr>
<tr>
<td>Cancer</td>
<td>13.5 million</td>
<td>$9,333</td>
<td>$126 billion</td>
</tr>
<tr>
<td>Depression/Anxiety</td>
<td>14.5 million</td>
<td>$2,655</td>
<td>$38.5 billion</td>
</tr>
<tr>
<td>Diabetes</td>
<td>19 million</td>
<td>$4,211</td>
<td>$80 billion</td>
</tr>
<tr>
<td>Emphysema</td>
<td>3 million</td>
<td>$11,333</td>
<td>$34 billion</td>
</tr>
<tr>
<td>Gallbladder Disease</td>
<td>3 million</td>
<td>$4,500</td>
<td>$13.5 billion</td>
</tr>
<tr>
<td>HIV/AIDS</td>
<td>0.9 million</td>
<td>$18,000</td>
<td>$19 billion</td>
</tr>
<tr>
<td>Heart Disease</td>
<td>46 million</td>
<td>$3,848</td>
<td>$177 billion</td>
</tr>
<tr>
<td>High Cholesterol</td>
<td>105 million</td>
<td>$352</td>
<td>$37 billion</td>
</tr>
<tr>
<td>Hypertension</td>
<td>58 million</td>
<td>$1,810</td>
<td>$105 billion</td>
</tr>
<tr>
<td>Stomach Ulcer</td>
<td>10 million</td>
<td>$400</td>
<td>$4 billion</td>
</tr>
<tr>
<td>Stroke</td>
<td>5 million</td>
<td>$11,600</td>
<td>$58 billion</td>
</tr>
</tbody>
</table>

* Average prevalence data from Figure 1. For sources of average cost data, see the Deconstructing the Costs bibliography.  
*Italicized costs per person are based on national cost divided by prevalence, as no published studies at the individual level were available.
Conclusion

While this report offers a solid background to help individuals understand concepts in disease prevalence and incidence, until we have an accurate way to collect and report the data, the numbers themselves are somewhat subjective. Even the CDC notes in *Health United States*, “Estimates are considered unreliable. Data preceded by an asterisk have a relative standard error of 20%–30%.”

To effectively improve the health of the nation, it is crucial that researchers, policymakers and healthcare providers have accurate, clear data on disease incidence, prevalence and medical care. Research should be conducted on best practices to achieve these objectives, and to determine which organizations need to take leading roles in data collection, dissemination and oversight. Collaboration between the government, healthcare and research communities is imperative for preparing clinical guidelines and improving methods of care. The Healthcare Delivery Policy Program will continue to develop recommendations on coordination of healthcare information to provide more complete and accurate epidemiological data for healthcare decision making.

Bibliography

Click here to read an annotated bibliography of over 140 sources of data.


37. **How Common is Depression?** Partners in Behavioral Health Sciences, University of Arkansas for Medical Sciences, Accessed December 29, 2007.


72. Testimony of the CDC Coalition Concerning the Public Health Budget for Fiscal Year 2006 Submitted for the Record to the House Appropriations Subcommittee on Labor, Health and Human Services and Education, CDC Coalition, 2005.

Updated 1/27/08