

SPECIAL REPORTS AND REVIEWS

The Burden of Selected Digestive Diseases in the United States

ROBERT S. SANDLER,* JAMES E. EVERHART,† MARK DONOWITZ,§ ELIZABETH ADAMS,||
 KELLY CRONIN,¶ CLIFFORD GOODMAN,¶ ERIC GEMMEN,¶ SHEFALI SHAH,¶
 AIDA AVDIC,¶ and ROBERT RUBIN¶,#

*University of North Carolina, Chapel Hill, North Carolina; †National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland; §Johns Hopkins University, Baltimore, Maryland; ||American Gastroenterological Association, Bethesda, Maryland; ¶The Lewin Group, Falls Church, Virginia; #Division of Nephrology and Hypertension, Georgetown University School of Medicine, Washington, DC

Background & Aims: Gastrointestinal (GI) and liver diseases inflict a heavy economic burden. Although the burden is considerable, current and accessible information on the prevalence, morbidity, and cost is sparse. This study was undertaken to estimate the economic burden of GI and liver disease in the United States for use by policy makers, health care providers, and the public. **Methods:** Data were extracted from a number of publicly available and proprietary national databases to determine the prevalence, direct costs, and indirect costs for 17 selected GI and liver diseases. Indirect cost calculations were purposefully very conservative. These costs were compared with National Institutes of Health (NIH) research expenditures for selected GI and liver diseases. **Results:** The most prevalent diseases were non-food-borne gastroenteritis (135 million cases/year), food-borne illness (76 million), gastroesophageal reflux disease (GERD; 19 million), and irritable bowel syndrome (IBS; 15 million). The disease with the highest annual direct costs in the United States was GERD (\$9.3 billion), followed by gallbladder disease (\$5.8 billion), colorectal cancer (\$4.8 billion), and peptic ulcer disease (\$3.1 billion). The estimated direct costs for these 17 diseases in 1998 dollars were \$36.0 billion, with estimated indirect costs of \$22.8 billion. The estimated direct costs for all digestive diseases were \$85.5 billion. Total NIH research expenditures were \$676 million in 2000. **Conclusions:** GI and liver diseases exact heavy economic and social costs in the United States. Understanding the prevalence and costs of these diseases is important to help set priorities to reduce the burden of illness.

Gastrointestinal and liver diseases inflict a heavy burden on the health and well being of Americans. The economic consequences for the nation are enormous. Although the burden of gastrointestinal (GI) disease is considerable, accessible information on the prevalence, morbidity, and costs is sparse.

Accurate information on the overall costs of digestive diseases could serve several purposes. The data could help to shape the research agenda of government, industry, and private foundations. The data could be used to evaluate or modify priorities and resource allocations of agencies that fund research or pay for health care services. Government and academia could use the data for assessing clinical services and manpower needs.

The most recent comprehensive summary of the prevalence and costs of digestive diseases was derived from information collected in the mid-1980s. To gather such information is difficult.¹ Although there are a number of publicly available national datasets and several proprietary databases, none of these sources is sufficiently comprehensive to capture costs across the full range of diseases in a consistent manner. Moreover, to completely capture the cost of a given GI disease requires estimates of both direct and indirect costs. Direct costs include costs related to goods or services associated with treatment, including physician visits, inpatient hospitalizations, outpatient hospital care, emergency (urgent) visits, and pharmaceutical costs. More difficult to measure are indirect costs, which include costs associated with losses in productivity caused by lost or impaired ability to work and by premature death, as well as the intangible costs of pain and suffering. In addition to economic costs, GI

Abbreviations used in this paper: GERD, gastroesophageal reflux disease; GHC, Group Health Cooperative of Puget Sound; GI, gastrointestinal; IBS, irritable bowel syndrome; ICD-9, International Classification of Diseases, 9th Revision; MEPS, Medical Expenditure Panel Survey; NAMCS, National Ambulatory Medical Care Survey; NHAMCS, National Hospital Ambulatory Medical Care Survey; NHDS, National Hospital Discharge Survey; NHIS, National Health Interview Survey; NIH, National Institutes of Health; NIS, nationwide inpatient sample; OPD, outpatient department; SAF, Medicare standard analytic files.

© 2002 by the American Gastroenterological Association

0016-5085/02/\$35.00

doi:10.1053/gast.2002.32978

diseases impose a significant impact on quality of life. This is particularly true for diseases that are chronic and debilitating in nature such as Crohn's disease and ulcerative colitis. Although there are useful data on mortality and costs, the data regarding impact on quality of life are limited and tend to be assessed in inconsistent ways.

To estimate the economic burden of GI disease, the American Gastroenterological Association convened a task force and commissioned a study to develop estimates that could be used by policymakers, physicians, and the public. This report summarizes current information on the burden to society from a selected group of common GI and liver diseases.

Materials and Methods

The goal of this study was to find information on the prevalence and cost of selected digestive and liver diseases. A select, rather than inclusive, set of GI and liver diseases was chosen owing to financial and time constraints. The diseases included in the report were selected based on the input of experts in gastroenterology and epidemiology, on the perceived high prevalence of the condition, on the availability of data, and the perceived research needs. As a result, we excluded some significant but uncommon or underdiagnosed diseases, such as celiac disease and hemochromatosis. The exclusion of any particular disease is no indication of its clinical or economic importance relative to others. The study was funded by an unrestricted grant from TAP Pharmaceutical Products, Inc., and was conducted by The Lewin Group, Falls Church, Virginia.² Table 1 lists the diseases included in the study in order of their corresponding codes according to the International Classification of Diseases, 9th Revision (ICD-9).

Table 1. Digestive Disease Categories by Diagnostic Codes

Disease	ICD-9 codes
Nonfoodborne gastroenteritis and other intestinal infections	001-009 (subset), 558.9
Foodborne illness	001-009 (subset), 558.9, 070.0, 070.1
Chronic hepatitis C	070.54
Colorectal cancer	153, 154, 197.5, 230.3, 230.4
Liver cancer	155.0, 155.2, 230.8
Pancreatic cancer	157, 197.8, 230.9
GERD	530.1, 530.2, 530.3, 530.81
Barrett's esophagus	530.2
Peptic ulcer disease	531, 532, 533, 534
Crohn's disease	555
Ulcerative colitis	556
Diverticular intestine	562
IBS	564.1, 564.5, 564.8, 564.9
Chronic liver disease and cirrhosis	571
Gallbladder disease	574, 575
Diseases of pancreas	577
Chronic diarrhea	787.91

Table 2. Cost Components Used to Determine Burden of GI Disease

Service/resources	Cost components
Physician office visits	Physician costs
Hospital facility	Facility costs associated with inpatient stays
Inpatient physician	Physician costs for inpatient physician services
Outpatient hospitalization	Facility costs associated with outpatient visits, physician costs for hospital outpatient physician services
Emergency care	Emergency room costs, physician costs for emergency room services
Pharmaceutical therapy	Pharmaceutical costs as determined in the Source Prescription Audit
Work loss	Indirect costs or valuation of work loss secondary to receiving health care

Usage was measured for each major type of service including physician care, inpatient hospital care, outpatient hospital care, emergency (urgent) care, and pharmaceutical therapy. For each variable, the relevant cost components were identified. Table 2 lists the health care resources and corresponding cost components included in the study.

Databases

A brief description of the datasets that provided the information included in this report can be found in the Appendix.

Measuring Usage in the Databases

The National Hospital Discharge Survey (NHDS), National Ambulatory Medical Care Survey (NAMCS), and National Hospital Ambulatory Medical Care Survey (NHAMCS) (emergency department and outpatient department files) datasets were used as the primary source for estimating usage figures for inpatient hospital stays, physician office visits, emergency room visits, and hospital outpatient visits, respectively. In addition to stays and visits, the average length of stay and total days of care by disease were obtained from these databases, which were used in calculating indirect costs (as described later).

To increase the statistical reliability in the estimates, the data were used only when there was an unweighted cell count (e.g., for the number of particular types of visits associated with a given disease) equal to or greater than 30 cases. A number less than 30 was judged to be too sparse to support reliable conclusions. The Group Health Cooperative of Puget Sound (GHC) claims database was used for usage estimates when the other databases lacked sufficient data or the American Gastroenterological Association advisory panel identified inconsistencies between the results and the published literature. This claims database was used as a secondary data source primarily because the enrollees are limited to the Pacific Northwest region of the United States.

Literature Review

The literature was the primary source for disease prevalence when the study diseases were clearly over- or underrepresented in the national databases, when database estimates were incompatible with expectations, and when there were plausible reasons for poor representation. A disease was underrepresented, for example, when the sample size of the population data captured in a national database was too small for deriving dependable estimates of prevalence or costs. A disease was overrepresented, for example, when it was apparent that patient data for the disease were aggregated with patient data for 1 or more other diseases that may have been similar. When the databases were judged to be inadequate, well-designed, peer-reviewed articles served as the main data source for descriptive epidemiology. Of the 17 diseases that we studied, 4 diseases were judged to be significantly underreported in the national databases, and prevalence was obtained from the literature (food-borne illness,³ hepatitis C,⁴ gastroesophageal reflux disease [GERD],⁵ and gallbladder disease⁶). Two diseases were judged to be overrepresented in the databases and the prevalence of these diseases was obtained from the literature (Crohn's disease⁷ and ulcerative colitis⁸).

Estimating Costs

Direct costs. For each type of service and for each disease, unit costs were applied to the total usage figures to calculate total direct costs. Where data allowed, both usage and average cost estimates were differentiated by whether the disease appeared in the source dataset as the first-listed (i.e., primary) diagnosis or as one of the second-listed (i.e., secondary, tertiary, and so forth) diagnoses. When the disease appeared as a first-listed diagnosis, 80% of the costs of the claim were attributed to this disease. When the disease appeared as a second-listed diagnosis, 20% of the costs of the claim were attributed to the disease. This method is somewhat arbitrary and has not been validated for digestive diseases. This method has been applied in past studies on the burdens of GI illness.¹ In the small number of cases in which the disease appeared as both a first- and second-listed diagnosis, 100% of the costs of the claim were attributed to the disease.

Inpatient hospital costs. The cost of inpatient hospitalization comprises 2 components: (1) hospital inpatient facility costs; and (2) physician visits and procedures for inpatients in the hospital. For most diseases, average hospital facility costs were obtained from the Nationwide Inpatient Sample (NIS), and inpatient physician costs were obtained from the GHC claims database.

The 1997 NIS was used to calculate hospital inpatient facility costs. The NIS reports total facility billed charges for each record (i.e., discharge) in the dataset. Charges include a hospital's mark-up over costs (including profit and bad debt) and, hence, overstate the economic resources used in patient care. To account for this, for each record (i.e., discharge) in NIS the study converted the monetary value of inpatient facility charges to inpatient facility costs. For each discharge record in the 1997 NIS, the total charge value was multiplied by a

hospital-specific cost-to-charge ratio determined by the Medicare Cost Reports by using the Health Care Financing Administration Form 2552-96, Worksheet S-2 for Cost Reporting Periods Beginning on or After October 1, 1996. For most diseases, this estimated daily cost figure from the 1997 NIS was used to calculate average daily inpatient facility charges, by primary diagnosis, secondary diagnosis, and both, respectively. For each disease, these 3 average daily facility cost figures were adjusted to 1998 dollars and then applied to the appropriate usage figure (i.e., total number of inpatient days by primary diagnosis, secondary diagnosis, and both) calculated from the 1998 NHDS. The sum of these 3 products represents the total facility costs for that disease in 1998 dollars.

The costs of physician services associated with inpatient hospital stays were calculated from the GHC claims database. This database lists physician visits and procedures in the form of Current Procedure Terminology/Healthcare Common Procedure Coding System, a classification that obviates the need for separating physician visits from physician procedures, as performed in previous burden of illness studies.¹ These average daily costs were then applied to the usage estimates (i.e., total number of inpatient days) from NHDS. All usage and cost data pertaining to the 3 cancers were acquired from the Medicare Standard Analytic File (SAF).

Outpatient hospital costs, emergency care, and physician office visits costs. The GHC claims database was used to estimate the average cost per physician office visit, average facility and physician cost per outpatient hospital visit, and average facility and physician cost per emergency room visit. These average costs were applied to the usage figures (i.e., visits by primary diagnosis, secondary diagnosis, and both) from NHAMCS (emergency department and outpatient department [OPD] files) and NAMCS, respectively, to estimate the total costs by setting of care for each disease. The Scott-Levin databases were used for information on prescription and over-the-counter medications.

Indirect costs. Indirect costs, sometimes referred to as *productivity costs*, typically include 3 categories of time costs: (1) costs related to the time required by the patient's family or other caretakers to receive medical care; (2) costs associated with lost or impaired ability to work or enjoy leisure activities because of morbidity; and (3) lost future earning potential owing to premature death. Given the scarcity of reliable data sources for indirect costs, this study addressed only productivity losses caused by time away from paid labor resulting from consumption of health care. It did not attempt to estimate the value of diminished productivity owing to morbidity in patients while in the workplace, time away from work but not receiving medical care (e.g., in recuperation or convalescence at home), foregone future earnings owing to premature death, or the value of foregone nonpaid labor or leisure time.

For a majority of the diseases, indirect costs were calculated by the valuation of time away from work while receiving medical care. For example, 1 inpatient hospital day for a person of working age (16–64 years) was assumed to represent 1 day of missed work. The corresponding assumptions for ambulatory visits were 2 hours for physician office visit, 4 hours for

emergency room visits, and 8 hours for hospital OPD visits. These assumed hours were then applied to the total usage estimates (i.e., days in hospital, ambulatory visits) to estimate days lost from work in those settings. In valuing the work-loss time while seeking medical care, the National Statistical Abstract (2000) was used to obtain information on average daily wage, broken down by sex and age. These average daily wage figures were applied to sex- and age-specific usage estimates from the National Center for Health Statistics datasets.

When many persons miss work they do not actually lose their wages because a certain number of sick days are allowed for persons on salary. Nevertheless, there are costs of lost work time that accrue to employers and to society. Even when workers are replaced or other workers substitute their labor, there are friction costs of training and the inevitable diminished productivity associated with these events. For the purpose of this study, the equivalent of lost wages was used to value the lost productivity resulting from time away from work.

By limiting indirect costs to time away from work owing to doctor visits or hospitalizations, the present study substantially underestimates indirect costs. For example, costs associated with diminished capacity owing to morbidity while at work, and with work time lost during illness or recuperation at home, are not included or estimated. Costs to spouses or other family members are also not included. The decision to limit indirect cost determinations in this manner generates very conservative, although explicit and reproducible, estimates.

Assessment and Revision of Estimates

After completing all analyses from the individual datasets, the reliability and validity of the data were assessed. The unweighted cell count for each variable, the overall geographic and age distribution of the dataset, and the methods used in sampling the population were all considered in selecting the best available source for a given estimate. As previously noted, when the unweighted cell count for an individual disease decreased below 30 in a database, an alternative data source was used (e.g., instead of NAMCS for physician office usage, the GHC claims database was used).

Estimating Direct Costs for all Digestive Diseases

The current report is limited to 17 selected diseases. To estimate the cost of all digestive diseases, we extrapolated from published data from 1985.¹ The conditions in the present report mapped fairly well with conditions in the 1985 report with the following exceptions, which are not judged to be major: GERD was limited to esophagitis (ICD 530.1) in the 1985 report. Colorectal cancer was limited to ICD153-154. Food-borne illness and non-food-borne illness together were combined in a single category—gastroenteritis and other intestinal infections (ICD 01-09, 558.9). Chronic liver disease and cirrhosis was expanded to include ICD codes 570-573. Irritable bowel syndrome (IBS) was restricted to ICD codes 564.1, 564.9. Liver cancer was restricted to ICD 155. Pancreatic cancer was restricted to ICD 157. Chronic hepatitis C was

expanded to all types of viral hepatitis. Chronic diarrhea was not individually included in the 1985 report.

To estimate 1998 costs of all digestive diseases, we divided the 1998 total direct costs of the 17 diseases (minus diarrhea, which was absent in the 1985 report) by the direct costs for comparable diseases in 1985. We then multiplied this figure (2.08) by the 1985 total direct costs to estimate overall costs for digestive disease in 1998.

Estimating Total Indirect Costs

The present report limits indirect costs to time away from work owing to doctor visits or hospitalizations. The report does not include indirect costs owing to premature death or time away from job owing to chronic illness. To estimate total indirect costs we used published data¹ to determine the ratio of total indirect costs to total direct costs in 1985 for the selected GI diseases in this report (10,946 ÷ 17,200). We then applied that ratio (0.636) to the 1998 direct costs (minus chronic diarrhea, which was not present in the 1985 report).

Mortality

Mortality data were obtained from the National Center for Health Statistics. The numbers of deaths and mortality rate per 100,000 population were enumerated for 1998. The ICD codes used for the cost analyses were also used to determine mortality.

Results

The prevalence, direct costs, and indirect costs for the 17 GI diseases selected for this report are summarized in Table 3. The total direct and indirect costs in 1998 dollars of those diseases ranged from \$9.8 billion (GERD) to \$0.37 billion (Barrett's esophagus). Based on 1998 data, the most prevalent of these diseases is: non-food-borne gastroenteritis and other GI infections (135 million cases), food-borne illness (76 million), gallbladder disease (20.5 million), GERD (18.6 million), and IBS (15.4 million).

In terms of total direct and indirect costs, the 5 most costly diseases are: GERD (\$9.8 billion), gallbladder disease (\$6.0 billion), colorectal cancer (\$5.0 billion), peptic ulcer disease (\$3.3 billion), and diverticular disease (\$2.5 billion).

The top 5 diseases in terms of direct costs are: GERD (\$9.3 billion), gallbladder disease (\$5.8 billion), colorectal cancer (\$4.9 billion), peptic ulcer disease (\$3.1 billion), and diverticular disease (\$2.4 billion). The top 2 diseases in terms of indirect costs are non-food-borne gastroenteritis and other GI infections (\$505 million) and GERD (\$479 million), followed by gallbladder disease (\$294 million), food-borne illness (\$233 million), and chronic liver disease and cirrhosis (\$222 million). As emphasized previously, these indirect cost estimates in-

Table 3. Prevalence and Direct and Indirect Costs of Selected GI Diseases in Order of Year 2000 Total Costs

Disease	1998			2000 ^a	
	Prevalence ^b (\$ in thousands)	Direct cost (\$ in millions)	Indirect cost ^c (\$ in millions)	Total (\$ in millions)	Total (\$ in millions)
GERD	18,600	9326	479	9805	10,070
Gallbladder disease	20,500	5754	294	6048	6467
Colorectal cancer	422	4846	106	4952	5294
Peptic ulcer disease	6730	3058	201	3259	3441
Diverticular disease	2254	2358	141	2499	2667
Diseases of pancreas ^d	1151	2125	212	2337	2492
Nonfoodborne gastroenteritis and other intestinal infections	135,000	1602	505	2107	2238
Chronic liver disease and cirrhosis	5490	1421	222	1643	1752
IBS	15,396	1353	205	1558	1658
Liver cancer	10	978	10	988	1056
Pancreatic cancer	18	1225	30	1255	1370
Foodborne illness	76,000	886	233	1119	1192
Crohn's disease	359	708	75	783	826
Chronic hepatitis C	2530	694	51	745	758
Chronic diarrhea	3080	492	129	621	661
Ulcerative colitis	619	387	36	423	443
Barrett's esophagus	808	350	22	372	389
Total ^e	—	36,027	2749	38,776	42,749

^aInflated to year 2000 dollars by using Medical CPI.

^bPrevalence based on NHIS except for foodborne illness, hepatitis C, GERD, gallbladder disease, Crohn's disease, and ulcerative colitis (see text for details).

^cIndirect cost estimates include only lost days of work owing to consumption of health care. The estimated indirect costs including costs from premature deaths and lost wages was \$22.8 billion.

^dExcludes diabetes mellitus.

^eTotal cost estimate does not include cost of Barrett's esophagus, chronic hepatitis C, and chronic diarrhea to avoid double counting (Barrett's esophagus is included in GERD, chronic liver disease and cirrhosis are often caused by hepatitis C, and chronic diarrhea is often a symptom of IBS).

clude only lost days of work owing to consumption of health care. By using information from the 1985 report on Digestive Diseases in the United States,¹ we estimate the total indirect costs of these selected digestive diseases in 1998 was \$22.8 billion after including indirect costs due to premature death and lost wages due to chronic illness.

The total direct and indirect costs for the 17 selected diseases in 1998 were \$38.8 billion. These diseases represent a significant portion, though not all, of GI diseases, hence, this estimate does not represent the cost of all GI diseases. The total cost estimate does not include the specific costs of Barrett's esophagus, chronic hepatitis C, and chronic diarrhea to avoid double counting (Barrett's esophagus is included in GERD, chronic liver disease and cirrhosis are often caused by hepatitis C, and chronic diarrhea is often a symptom of IBS). Excluding these diseases from the total cost estimate is conservative. Not all patients with chronic diarrhea have IBS; hepatitis C patients incur substantial costs even when they do not develop chronic liver disease.

Direct Costs

Table 4 provides detailed information on the sources of direct costs for the 17 selected conditions.

Data on costs for hospital facility costs come from the 1997 NIS. Costs for inpatient physician services, hospital outpatient visits, emergency room visits, and office visits come from the 1996 Medical Expenditure Panel Survey (MEPS) and 1997–1999 GHC databases. Information on drugs is derived from the Scott Levin databases. The direct costs in Table 4 are the same as those in column 3 of Table 3 with the exception of cancers of the colon, liver, and pancreas, for which Table 3 also includes costs for home health, nursing homes, and hospice care.

The highest direct costs were for GERD (\$9.3 billion), followed by gallbladder disease (\$5.8) billion, colorectal cancer (\$4.5 billion), and peptic ulcer disease (\$3.1 billion). For GERD, drug costs were responsible for 63% of the total direct costs. Pharmaceuticals were responsible for 20% of the total direct costs of peptic ulcer disease. Pharmaceutical costs were relatively low for IBS (\$80 million) because there are few effective drugs for this indication. In contrast, hospital inpatient costs were responsible for 75% of the cost of gallbladder disease.

For most diseases, hospital inpatient costs (hospital facility plus inpatient physician services) are dominant. This is even true for conditions that are largely managed in outpatient settings such as IBS, food-borne illness, and

Table 4. Direct Cost of GI and Liver Diseases (in Millions), 1998

	Hospital facility	Inpatient physician services	Hospital OPD	Hospital emergency room	Office visits	Drugs	Total
GERD	1991	549	213	78	603	5892	9326
Gallbladder disease	3510	827	940	241	234	2	5754
Colorectal cancer	3082	—	310 ^a	0	1093 ^b	20	4505 ^c
Peptic ulcer	1650	155	370	67	196	621	3059
Diverticular disease	1579	301	53	199	177	49	2358
Pancreatic disease	1687	256	29	40	64	49	2125
Nonfoodborne illness	729	188	82	251	275	77	1602
Chronic liver disease	1141	134	57	7	65	17	1421
IBS	848	149	35	13	228	80	1353
Pancreatic cancer	688	—	106 ^a	0	284 ^b	29	1107 ^c
Liver cancer	562	—	89 ^a	0	139 ^b	80	978 ^c
Foodborne illness	400	92	47	154	155	38	886
Crohn's disease	255	48	11	9	102	283	708
Hepatitis C	108	21	11	0	24	530	694
Chronic diarrhea	193	24	41	50	148	36	492
Ulcerative colitis	157	35	16	3	38	138	387
Barrett's esophagus	160	38	4	0	30	118	350

Data from NIS 1997 (hospital inpatient), 1996 MEPS, and 1997–1999 Puget Sound (hospital OPD, hospital emergency room, office visits), Scott Levin (drugs) (see text for details).

^aIncludes emergency room visits.

^bIncludes physician services across all settings.

^cDoes not include costs for home health care, nursing home care, or hospice.

non–food-borne gastroenteritis. For patients with these conditions who are admitted to the hospital, the direct costs can be considerable.

Hepatitis C is the condition other than GERD in which the total direct costs are dominated by pharmaceutical costs. More than 70% of the costs of hepatitis C treatment are drug costs, again owing to the expense of the medications to treat this condition. Conversely, inpatient hospitalization is responsible for the bulk of the costs of chronic liver disease (which includes hepatitis C).

For comparison purposes, Figure 1 collapses the 17 conditions into 6 major categories of conditions. The categories are somewhat arbitrary, and the cost for a

given category is partly due to the number of diseases in the category. Nonetheless, it is easier to compare 6 categories than 17. The acid-related disorders (GERD, peptic ulcer disease) are the largest single category, with \$12.4 billion in direct costs, caused largely by the cost of medications. Gallbladder and pancreatic disorders are responsible for \$7.8 billion. The GI cancers (colorectal, liver, pancreas) are next with \$7.0 billion in direct costs. The chronic GI disorders, including diverticular disease, IBS, Crohn's disease, and ulcerative colitis, account for \$4.8 billion in direct costs. The GI infections account for direct costs of \$7.3 billion. Chronic liver disease has direct costs of \$1.4 billion. As in previous analyses, these estimates exclude Barrett's esophagus, hepatitis C, and chronic diarrhea to avoid double counting. The chronic liver disease category does not include subjects with chronic hepatitis infection nor does it necessarily include nonalcoholic steatohepatitis unless the nonalcoholic steatohepatitis is associated with chronic liver disease. The liver category only includes 1 condition in contrast to the other categories that contain several conditions.

Hospital Outpatient Visits and Costs

Table 5 summarizes data on health care usage reflected by inpatient hospital stays, visits to hospital outpatient departments, emergency room visits, and physician office visits. In 1998 there were 5.4 million hospitalizations, 3.5 million outpatient hospital visits, 4.2 million emergency room visits, and 32.8 million physician visits for the 17 digestive diseases under study.

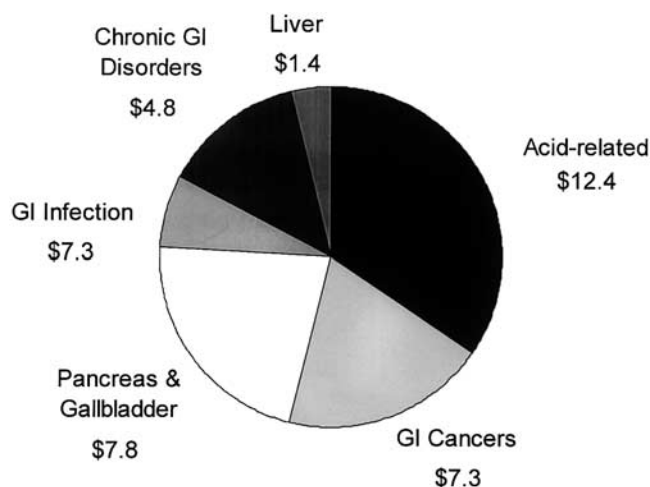


Figure 1. Direct cost (in billions) of gastrointestinal diseases by category, 1998.

Table 5. Health Care Usage for GI and Liver Diseases, 1998

Disease	Inpatient hospital stays	Inpatient hospital stays (primary diagnosis only)	Outpatient hospital visits	Emergency room visits	Physician office visits
GERD	968,341	175,725	829,848	459,719	9,114,361
Gallbladder disease	673,762	434,127	669,700	564,364	1,682,581
Colorectal cancer	267,104	172,938	350,975	50,049	1,787,879
Peptic ulcer disease	469,857	189,465	255,991	188,363	2,651,884
Diverticular disease	455,914	230,058	147,785	165,343	2,216,519
Diseases of pancreas	327,474	211,802	78,460	195,341	531,792
Nonfoodborne	465,254	—	240,191	1,159,149	3,638,729
Chronic liver disease cirrhosis	356,667	89,781	186,237	55,082	757,793
IBS	475,409	194,269	144,925	86,963	3,657,623
Liver cancer	22,588	11,521	12,476	16,899	38,085
Pancreatic cancer	71,054	34,031	31,036	14,937	221,883
Foodborne illness	256,626	—	142,384	787,977	2,022,860
Crohn's disease	84,622	37,196	51,265	69,768	1,180,354
Chronic hepatitis C	60,638	6160	46,311	—	316,920
Chronic diarrhea	165,298	863	210,390	455,328	2,210,094
Ulcerative colitis	46,630	23,447	64,952	25,672	469,692
Barrett's esophagus	60,906	18,863	9756	—	230,770
Total	5,228,144	1,830,246	3,472,68	4,194,954	32,729,81

Data from NHCS (hospital inpatient), NHAMCS (hospital OPD), NHAMCS (hospital emergency room), NAMCS (office visits) (see text for details).

For most conditions, physician visits represent the largest component of health care usage. The total number of physician office visits reflects, among other things, disease prevalence—the highest numbers of visits were GERD (9.1 million visits), IBS (3.7 million), non–food-borne gastroenteritis (3.6 million), and peptic ulcer disease (2.7 million). Among the GI conditions studied, non–food-borne gastroenteritis and food-borne illness are the most common reasons for emergency room visits, followed by gallbladder disease and GERD.

At the time of discharge, several diagnoses may be coded. For inpatient hospital stays, GERD, gallbladder disease, and IBS account for the greatest numbers of diagnoses. However, these highly prevalent conditions may not have been the primary reason that the patient

was hospitalized. Based on primary diagnoses for inpatient hospital stays, the most common conditions in this study were gallbladder disease (434,000), diverticular disease (230,000), and diseases of the pancreas (212,000). Although diseases of the pancreas are less common, they often require management in the hospital.

Mortality

In addressing the burden of GI disease, this report focused largely on incidence and costs. However, these digestive conditions can be fatal. The number of deaths and the mortality rate per 100,000 population for selected conditions are shown in Table 6. Colorectal cancer, because of its high incidence and mortality, leads the list with nearly 50,000 deaths in the United States each year.

Table 6. Total Deaths and Mortality Rate by Sex for Selected Digestive Diseases, 1998

Disease	Women		Men		Total	
	Deaths	Mortality rate (per 100,000)	Deaths	Mortality rate (per 100,000)	Deaths	Mortality rate (per 100,000)
Colorectal cancer	24,842	17.96	23,357	17.68	48,199	35.64
Pancreatic cancer	14,529	10.5	13,806	10.45	28,335	20.95
Chronic liver disease and cirrhosis	8849	6.39	16,343	12.27	25,192	18.66
Liver cancer	3336	2.41	6575	4.98	9911	7.39
Peptic ulcer disease	2460	1.78	2235	1.69	4695	3.47
Diverticular disease	2404	1.74	1010	0.76	3414	2.5
Pancreatic disease	1547	1.12	1648	1.15	3195	2.27
GERD	760	0.55	718	0.54	1478	1.09
Gallbladder disease	678	0.49	465	0.35	1143	0.84
Crohn's disease	268	0.19	201	0.14	469	0.33
Ulcerative colitis	245	0.18	220	0.17	465	0.35
IBS	43	0.03	25	0.02	68	0.05

Data from Centers for Disease Control. Data based on underlying cause of death, classified in accordance with the ICD-9 codes.

Pancreatic cancer and chronic liver disease are each responsible for more than 25,000 deaths. Liver cancer, though highly fatal, is responsible for a smaller number of deaths (about 10,000) because the disease is less common. Nearly 5000 people die each year in the United States from peptic ulcer disease. Although overall hospitalizations for peptic ulcer disease have decreased, hospitalizations for complicated peptic ulcer (which can be fatal) have remained largely unchanged.⁹ Both GERD and gallbladder disease are occasionally fatal. These deaths most likely represent complications of surgery or sepsis.

Estimated Costs for all Digestive Diseases

Excluding chronic diarrhea, the direct costs of the 17 selected diseases were approximately \$17.2 billion in 1985. The direct cost of all digestive disease in 1985 was \$41.5 billion.¹ The use of the ratio of 17 selected diseases in 1998 (minus diarrhea) divided by the same diseases in 1985 gives $(36,027 - 492)/17,212 = 2.06$. Applying this ratio to the 1985 total direct costs gives a 1998 direct cost total of \$86.5 billion in direct costs for all digestive diseases. This \$85.5 billion figure is likely to be an underestimate, and excludes indirect costs, which can be considerable.

National Institutes of Health Expenditures

Table 7 shows expenditures by the National Institutes of Health (NIH) for 10 selected GI and liver diseases in the fiscal year 2000 based on data from the Office of the Director. The most heavily funded areas

Table 7. NIH Research Expenditures and Total Costs for Selected Diseases in Fiscal Year 2000

Disease	Research \$ (in millions)	Total direct and indirect costs (in millions)	Standardized research dollars/ disease costs ^a
Chronic liver disease and cirrhosis	218.6	1752	55.6
Chronic hepatitis C	66.5	758	39.1
Ulcerative colitis and Crohn's disease	27	443	27.2
Foodborne illness	57.5	1192	21.5
Colorectal cancer	205.2	5294	17.3
Liver cancer	46.2	1056	13.6
Pancreatic cancer	20	1370	6.5
IBS	8.2	1658	2.2
Peptic ulcer disease	12.1	3441	1.6
Gallbladder disease	14.5	6467	1.0
Total	675.8	23,431	

Data from NIH Office of the Director; NIDDK, NCI, and NIAID.
NOTE. Indirect costs included in this analysis are very restricted. If full indirect costs were included, the ratios would be very different.
^aResearch dollars/disease costs compared with gallbladder disease.

were chronic liver disease and cirrhosis (\$218.6 million), colorectal cancer (\$205.2 million), hepatitis C (\$66.5 million), and food-borne illness (\$57.5 million). The least amount of research dollars was allocated to IBS (\$8.2 million). The listed conditions received measurable funding from the NIH. Omission from Table 7 does not indicate that there was no research funding from the NIH, simply that specific estimates were not available. However, no more than a few total million dollars in research were provided for the omitted conditions. The last column of Table 7 presents a standardized ratio of NIH research expenditures/total costs. Each ratio was divided by the ratio for gallbladder disease, which had the lowest ratio. As shown in Table 7, the ratio of research dollars to disease burden for chronic liver disease and cirrhosis was 55.6 times greater than the ratio for gallbladder disease. It is important to note that the indirect costs included in Table 7 (from Table 3) are likely to significantly underestimate total indirect costs. Therefore, the ratios shown in the last column of Table 7 might change substantially if indirect costs due to disability or premature death were included.

Discussion

The present report estimates that a group of 17 GI and liver disorders was responsible for \$36.0 billion in direct costs and, based on a very conservative determination, \$2.8 billion in indirect costs in 1998. Adjusting for inflation, this represents approximately \$43 billion in the year 2000. GERD was the most costly disease in 2000 with total costs of \$10 billion, followed by gallbladder disease (\$6.5 billion), and colorectal cancer (\$5.3 billion). The most prevalent disorders were non-food-borne gastroenteritis (135 million cases) and food-borne illness (76 million). In the aggregate, acid-related disorders accounted for the highest direct costs (\$12.4 billion), followed by gallbladder and pancreatic disorders (\$7.8 billion), and GI cancers (\$7.3 billion).

Although the methods used to measure the economic burden of these diseases were as rigorous as the data allowed, these cost figures represent conservative estimates for several main reasons. First, estimates exclude important categories of direct or indirect costs owing to lack of data in the literature and in datasets. For many diseases, the scope of direct costs does not include costs of long-term care, such as nursing home care or home health care. The datasets available for this study did not allow for estimates from federal facilities (e.g., Armed Forces hospitals, Veterans Health Administration facilities, and Indian Health Service). These sources represent about 10% of health care in the United States. Second,

some data sources are older than others, and though adjustments can be made for inflation, they cannot be made for trends in prevalence, costs, or other relevant parameters over time. Third, the methods used to measure indirect costs, a function of work loss while consuming health care, greatly underestimate total indirect costs of many of these diseases. If more reliable data sources were available to adequately capture other segments of indirect cost (e.g., reduced productivity while working, lost unpaid labor, lost leisure time), it is probable that overall cost estimates would be significantly greater, with a greater share attributable to indirect costs. There are insufficient data for productivity losses owing to mortality, illness, and family members and caretakers. Losses due to pain, suffering, and diminished quality of life are considerable, yet also unmeasured for many of these diseases.

Because of the cost-assignment method used, hospital costs may be somewhat overstated for certain diseases. For all diseases, 80% of the cost of a hospital claim is attributed to the primary diagnosis, whereas 20% is applied to a secondary diagnosis. To the extent that a secondary diagnosis does account for 20% of inpatient resource use (e.g., more physician time, prolonged hospitalization), this method results in a proportionate reflection of hospital costs. For certain prevalent diseases that are predominantly symptomatic and often appear as a secondary diagnosis (e.g., chronic diarrhea or GERD), this methodology might result in an overestimation of related inpatient costs if, in fact, these diseases contribute less than 20% of total inpatient resource use when listed as a secondary diagnosis.

People who suffer from these diseases, but do not have the health insurance coverage or other resources at their disposal to access health care services, are underdiagnosed and undertreated. Because some of these diseases are inversely correlated with socioeconomic status, there is likely to be underreporting.

The choice of specific ICD codes for each disease was deliberate, but sometimes arbitrary. For example, we did not include ICD 553.3 hiatus hernia among GERD diagnoses because hernia is not always associated with GERD. We excluded acute hepatitis C from the hepatitis C category to focus on chronic hepatitis C. For Barrett's esophagus we used 530.2, which is the ICD-9-CM code for Barrett's esophagus. From the ICD index: "Barrett's syndrome or ulcer (chronic peptic ulcer of esophagus)." We excluded ICD 155.1, cancer of the intrahepatic bile duct, from liver cancer to focus on hepatocellular cancer. It is also important to recognize the potential for miscoding at the time of discharge. There is no guarantee

that discharge codes are always completely valid or reliable. Because hospitals are paid based on diagnosis codes, there is the potential for up-coding.

The last comprehensive look at the economic burden of GI disease was a study by Brown and Everhart¹ that estimated the cost of digestive diseases in the United States for the year 1985. In that report, direct costs were estimated by summing all resource payments for patient services. Indirect costs represented forgone earnings owing to hospitalization, premature death, and work loss associated with acute GI diseases. Brown and Everhart¹ found that the total cost of GI diseases was \$56.1 billion in 1985. Of the total, \$41.5 billion was attributable to the direct costs and \$14.6 billion to indirect costs. The study by Brown and Everhart¹ was more comprehensive than the present study because it included all GI diseases and it based its calculation on charges, a measure considerably larger than costs.

Other investigators have explored the relationship between burden of disease and NIH funding. Gross et al.¹⁰ found that research funding for human immunodeficiency virus/acquired immune deficiency syndrome, breast cancer, and dementia was much greater than other diseases with similar disability-adjusted life-years, with 1 disability-adjusted life-year defined as the loss of 1 year of healthy life to disease. That study suggested that the health impact of disease does not coincide with the level of public research funding. However, that study used hospitalization as a surrogate for the economic impact of disease and therefore systematically underestimated the true economic burden of diseases studied. It is not possible to make a direct comparison between the data presented by Gross et al.¹⁰ and the data in the present report. However, the 1994 NIH expenditures for certain diseases such as diabetes (\$298.9 million), acquired immune deficiency syndrome (\$1.4 billion), and breast cancer (\$381 million) were considerably higher than any of the GI diseases in the present report.

The distribution of costs is helpful in understanding the burden of these diseases. For example, GERD is the condition with the highest costs, with direct costs in excess of \$9 billion. The largest component is the cost of drugs—\$5.8 billion. Although GERD is a highly prevalent disease, the total direct costs for GERD are heavily influenced by the cost of medications. The high cost of acid-reducing medication is also reflected in the costs for peptic ulcer disease, which are also heavily influenced by drug costs.

Gallbladder disease has the second highest direct costs, and the distribution of those costs is quite different. Hospital inpatient costs dominate the overall expen-

ditures for gallbladder disease. The high inpatient costs for gallbladder disease are caused by high rates of cholecystectomy. Drugs account for a very small proportion of total direct costs for the treatment of gallbladder disease.

Combined public and private research expenditures on GI disease were approximately \$1.4 billion in 2000. However, this magnitude of effort is significantly less than the research commitment to other diseases with lower health and economic burdens in the United States based on the Parexel's Pharmaceutical R&D Statistical Sourcebook 2000¹¹ and the NIH Office of Budget Information. The data shown in Table 7 suggest poor correlation between NIH research expenditures and cost of disease.

The present report has quantified hospitalization as well as outpatient care, pharmaceuticals, and losses to productivity for the 17 GI conditions in economic terms. As such, this provides a broad picture of the health and economic impact of GI disease, and helps to highlight the relative economic burden to the nation of direct and indirect costs of these conditions. For example, there is a large economic impact of lost productivity caused by certain diseases, including IBS, GERD, various types of GI infections, and peptic ulcer disease. As the federal government and industry consider research priorities, it is important to account for the health and economic impact of disease, including both direct and indirect costs.

The present report substantially underestimates indirect costs. For example, Sonnenberg and Everhart¹² used data from the 1989 National Health Interview Survey (NHIS) to obtain data on peptic ulcer disease. They found that 10%–15% of subjects with a recent ulcer reported that they had been in poor health, incapable of major activity, or unable to work for some time during the 12 months preceding the interview. A total of 20%–25% of the subjects with recent ulcers complained about restricted activity and had spent 7 or more days per year in bed. The investigators estimated expenditures of \$5.65 billion in 1993, which is about 70% more than the estimate in the present report. To develop an estimate of more complete indirect costs we applied the ratio of indirect to direct costs from the 1985 report on the cost of digestive diseases¹ to the direct costs in the present report. The estimate, \$22.8 billion, probably underestimates indirect costs.

The purpose of this report was to assemble available information on a number of digestive and liver diseases, and to attempt to estimate the economic costs to the nation. The report has clearly shown high costs—approx-

imately \$36.0 billion in direct costs for the selected 17 diseases and an estimated \$85.5 billion in direct costs for all digestive diseases. At the same time the report has shown the difficulty in determining the economic burden, particularly indirect costs, and highlights some of the serious shortcomings in using the available data to estimate cost and usage. This study used analysis of administrative databases that were assembled for other reasons. The advantage of administrative databases is their large size, complete coverage (e.g., Medicare), and convenience. The data have already been collected. However, because the data were not originally collected to support research, they suffer from potential problems of accuracy and reliability.

Hopefully, this report will motivate future investigators to rigorously determine the costs of these diseases by using both administrative databases and primary data collection methods. The report should challenge policy makers, clinicians, researchers, and the public to develop strategies to reduce the huge economic and social burden of these diseases.

References

1. Brown D, Everhart JE. Costs of digestive diseases in the United States. In: Everhart JE, ed. Digestive diseases in the United States: epidemiology and impact. US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institutes of Diabetes and Digestive and Kidney Diseases. Washington, DC: US Government Printing Office, 1994:55–82.
2. The burden of gastrointestinal diseases. Bethesda, MD: American Gastroenterological Association, 2001.
3. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV. Food-related illness and death in the United States. *Emerg Infect Dis* 1999;5:607–625.
4. Alter MJ, Kruszon-Moran D, Nainan OV, McQuillan GM, Gao F, Moyer LA, Kaslow RA, Margolis HS. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N Engl J Med* 1999;341:556–562.
5. Frank L, Kleinman L, Ganoczy D, McQuaid K, Sloan S, Eggleston A, Tougas G, Farup C. Upper gastrointestinal symptoms in North America: prevalence and relationship to healthcare utilization and quality of life. *Dig Dis Sci* 2000;45:809–818.
6. Everhart JE, Khare M, Hill M, Maurer KR. Prevalence and ethnic differences in gallbladder disease in the United States. *Gastroenterology* 1999;117:632–639.
7. Loftus EV Jr, Silverstein MD, Sandborn WJ, Tremaine WJ, Harnsen WS, Zinsmeister AR. Crohn's disease in Olmsted County, Minnesota, 1940–1993: incidence, prevalence, and survival. *Gastroenterology* 1998;114:1161–1168.
8. Loftus EV Jr, Silverstein MD, Sandborn WJ, Tremaine WJ, Harnsen WS, Zinsmeister AR. Ulcerative colitis in Olmsted County, Minnesota, 1940–1993: incidence, prevalence, and survival. *Gut* 2000;46:336–343.
9. Kurata JH, Corboy ED. Current peptic ulcer time trends. An epidemiological profile. *J Clin Gastroenterol* 1988;10:259–268.
10. Gross CP, Anderson GF, Powe NR. The relation between funding by the National Institutes of Health and the burden of disease. *N Engl J Med* 1999;340:1881–1887.
11. Matthieu M. Parexel's pharmaceutical research and development

statistical sourcebook. Waltham, MA: Parexel International Corp, 2000.

12. Sonnenberg A, Everhart JE. Health impact of peptic ulcer in the United States. *Am J Gastroenterol* 1997;92:614-620.
13. LeClere FB, Moss AJ, Everhart JE, Roth HP. Prevalence of major digestive disorders and bowel symptoms, 1989. Advance data from vital and health statistics; no 212. Hyattsville, MD: National Center for Health Statistics, 1992.

Received September 21, 2001. Accepted January 3, 2002.

Address requests for reprints to: Robert S. Sandler, M.D., M.P.H., CB# 7080, 719 Burnett-Womack Building, University of North Carolina, Chapel Hill, North Carolina 27599-7080. e-mail rsandler@med.unc.edu; fax (919) 966-2478.

Supported by an unrestricted grant from TAP Pharmaceutical Products, Inc. to the American Gastroenterological Association.

Appendix

1995 National Health Interview Survey

Unless noted, estimates of prevalence for each disease were obtained from the NHIS. The NHIS is an annual nationwide survey of civilian households administered by the National Center for Health Statistics, a component of the Centers for Disease Control and Prevention. Approximately 100,000 persons from 40,000 households were interviewed. All information collected in the survey was from face-to-face interviews of responsible family members residing in the household. Proxy responses were accepted for family members who were not at home and were required for all children and for all family members who were physically or mentally incapable of responding for them. Illnesses were coded by using a slight modification of the ICD-9.

The 1995 NHIS was the primary data source for the period prevalence of the majority of diseases represented in this report. NHIS was also used to calculate estimates of indirect costs via work loss days; however, low, unweighted cell counts prohibited the use of these estimates for some diseases.

The 1995 NHIS included the most recent questionnaire on digestive system conditions. Respondents were asked if anyone in the family had any of 22 different digestive conditions during the 12 months before the interview. For each disease condition reported, respondents were asked to recall, over the prior 2 weeks, the number of restricted activity days, bed days, work-loss days, and school-loss days owing to the condition.

The 1995 prevalence rate was applied to 1998 U.S. population figures to estimate the 1998 prevalence. Information from both proxy and self-respondents may be underreported because some respondents were unaware of relevant information, forgot it, or did not wish to reveal it to an interviewer. As such, the self-reported nature of

the NHIS likely results in underestimates of prevalence.¹³

1998 National Hospital Discharge Survey

The NHDS dataset was used to identify inpatient hospitalization information, including number of hospitalizations, average length of stay per disease, and the total number of days of care by each disease. The NHDS, conducted by the National Center for Health Statistics, is a national survey of inpatient usage in nonfederal short-stay hospitals, children's hospitals, and general hospitals. The database comprises medical records from a nationally representative sample of 270,000 inpatient records drawn from the 500 nonfederal hospitals covered in the NHDS.

1998 National Ambulatory Medical Care Survey

The NAMCS dataset provided information on physician office visits, including number of visits and demographic data. The NAMCS, conducted by the National Center for Health Statistics, is a survey of office visits made by ambulatory patients to a nationally representative sample of approximately 1500 nonfederally employed physicians. The settings included in this survey are: free-standing private, solo, or group office; free-standing clinic; neighborhood health center; privately operated clinic; local government clinic; and health maintenance organization or other prepaid practice. The total number of records (i.e., visits) is about 23,400.

1998 National Hospital Ambulatory Medical Care Survey

The NHAMCS, conducted by the National Center for Health Statistics, includes 2 separate files: a hospital emergency department file and an OPD file. Each file has a sample size of 29,400 records (i.e., visits). From these files, data were obtained on the number of emergency department and OPD visits for each disease, as well as patient demographic data.

The hospital OPD is defined as outpatient clinics within a hospital. Clinics are determined to be eligible for NHAMCS if ambulatory medical care is provided under the supervision of a physician, under the auspices of the hospital. Clinics where only ancillary services are provided, or where physician services are not provided, are not included in the survey. Furthermore, ambulatory surgery centers, both those in hospitals as well as free-standing centers, are not included in NAMCS or NHAMCS.

1997 Nationwide Inpatient Sample

The NIS Release 6 is part of the Healthcare Cost and Utilization Project, administered by the Agency for Healthcare Research and Quality. This database was used to estimate the average facility cost per hospital day for each given disease. This value, along with the total days of hospital care per disease (from the 1998 NHDS), was used to calculate the total facility costs associated with hospital inpatient stays.

The NIS approximates a 20% sample of community hospitals in the United States, defined by the American Hospital Association as all non-federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions for 1997. Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric hospitals. Excluded are long-term hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities. NIS Release 6 is based on a stratified probability sample of hospitals, with sampling probabilities proportional to the number of community hospitals in the United States in each stratum. NIS Release 6 is drawn from 22 states and contains information on all inpatient stays from over 1000 hospitals, totaling about 7.1 million records in 1997.

1996 Medical Expenditure Panel Survey

The Agency for Healthcare Research and Quality's longitudinal household dataset, the Medical Expenditure Panel Survey (MEPS), provides information, collected in 3 rounds, on a nationally representative sample of the civilian noninstitutionalized population of the United States. The 1996 household survey contains information on 23,200 persons, whose responses were weighted to a nationally representative total of 268.9 million persons. MEPS collects self-reported information on respondents' health care usage and expenditures, and then links these data with additional information collected from the respondents' medical providers, employers, and insurance providers. MEPS contains specific information about direct medical resource usage (the number of physician visits, inpatient hospital stays, emergency room visits, and prescribed medications) and the corresponding costs (amount paid by insurer type, amount paid out-of-pocket, and so forth) for given conditions. MEPS details diagnoses to only the 3-digit level. For the GI diseases fully defined by 3-digit ICD-9 diagnosis codes (see Table 1) and when sufficient disease counts were found, the 1996 MEPS file was used to

calculate average costs for all physician office visits and for hospital outpatient visits, including physician services.

1998 Medicare Standard Analytic Files

The SAF contains all medical claims of 5% of Medicare beneficiaries. The 1998 Medicare SAF served as the primary source for all usage and cost data for the 3 cancers included in the study. The SAF was required for the cancers because it most accurately measured all types and settings of care among this patient population. The SAF files contained final action and adjustment-resolved claims, and were updated each July with data from the previous calendar service year. SAF included the following files: Inpatient (Part A), Hospital Outpatient, Home Health Agency, Hospice, Skilled Nursing Facility, and Physician/Supplier (Part B).

1997–1999 Group Health Cooperative of Puget Sound Claims Database

The GHC claims database includes approximately 450,000 covered lives from a mixed-model health maintenance organization. These claims data contain information on fully allocated costs per claim. For many diseases, this database was used to estimate average cost per physician office visit, average physician cost per day in the hospital, average facility and physician cost per outpatient hospital visit, and average facility and physician cost per emergency room visit.

2000 Scott-Levin Source Prescription Audit and the Physician Drug and Diagnosis Audit

The Scott-Levin Source Prescription Audit, drawn from a nationwide sample of pharmacies, served as the source for drug expenditures for each disease. Approximately 35,400 retail pharmacies, 63% of the entire retail pharmacy universe in the United States, are surveyed on a monthly basis, and 72% of all dispensed prescriptions are included. A shortcoming of this source was that it did not include drugs administered in physician offices. Disease-specific usage of drugs, used to determine the proportion of drug spending that is attributable to a given disease, was measured by using the Scott-Levin Physician Drug and Diagnosis Audit. This audit monitors patient visits and associated treatment regimens among approximately 365,000 office-based physicians in 29 specialties across the 9 census regions in the United States. It tracks patient drug requests, drug switching, and prescribing volume. Diagnosis activity is organized by ICD-9 codes.