



History, Methodology and Findings

A. Overview and History of the Project

The purpose of this research project, which was jointly funded by the Health Section and the Society of Actuaries, was “to assemble and analyze a limited database which may serve as a model for expanded intercompany studies of health care benefits in the future.” The project focused on the types and incidence of large claims, defined for the research as claims totaling \$25,000 or more for any one year.

The research comprised two parts: (1) collection of data from participating insurers, followed by standardization and analysis of the data, and (2) presentation of the results of the analysis in a written document. The process involved the authors and individuals from the Society, actuarial firms, insurers, and providers.

In August 1992, the Society of Actuaries solicited input from a number of firms and institutions on their interest in the project and their comments and suggestions on the proposed data format and specifications. (See Appendix A.) Shortly thereafter, the Society released a Request for Proposal for a Research Project entitled, “Group Medical Insurance Large Claims Database Collection and Analysis,” with a due date of October 9, 1992. In November, letters from the Society requested the participation of insurers in the study and were accompanied by the data specifications for the study. (See Appendix B.) In April 1993, a follow-up letter was sent to insurers that did not respond to the November letter, requesting that they reconsider participation in the study. (See Appendix C.)

In February 1993, the authors were selected; receipt of data tapes from the Society began in July 1993, and

the final data tape was received by the research team in January 1994.

The Project Oversight Group (POG) comprised the Director of Research for the Society and practicing actuaries from the Health Section of the Society. The POG worked with the research team in designing the analytic plan and its presentation and met with the team periodically throughout the study. In March 1994, the authors met with two members of the Project Oversight Group to preview the format of and preliminary findings from the analysis. In April 1994 at the Orlando Spring Meeting of the Society, preliminary findings were presented as part of a panel on the activities of the Research Department of the Society (Tenenbein, A., Adams, W., Grazier, K.L., and Hickman, J., “Where Your SOA Research Dollar Goes,” *Record of the Society of Actuaries*, Vol. 20, No. 1 (1994), pp. 153–297).

The draft of the complete report was circulated to the POG and the Research Department in May 1994. Comments were received in writing and through a conference call. A proposal for incorporating the recommended changes was approved by the Society in November, and work was completed in December 1994. In early 1996, additional analyses were performed and incorporated into the report for public release.

B. Methodology

1. Data Collection

Over the course of the study, data for 1991 and/or 1992 in computer-readable format were received from 25 participants. One additional carrier submitted in

report format only the total charges incurred for 1,187 individuals in 1991 and 544 individuals in 1992; claimant-level data were not available. Formats for the computer-based files included half-inch nine-track tape reels of various densities, IBM 3480-compatible tape cartridges and microcomputer diskettes. Copies of the original data submissions were made by the Society and/or the research team prior to analysis. Copies of the tapes were housed at the Society of Actuaries offices in Schaumburg, Illinois. Information on the identities of the participants and claimants was not supplied to the research team. Unique numerical codes were assigned to each participant by the Society prior to receipt by the research team. Individual claimant identifiers (such as names or Social Security numbers) did not appear on any tapes received.

2. Data Editing

Tapes were read and files examined for completeness. Where provided by the participant, control totals for charges and numbers of claimants were checked against data supplied by the participants. Several stages of data editing were completed to ensure that valid, comparable data and field definitions existed within and across all data supplied. When questions about files or variable definitions arose, the Director of Research was contacted to intervene with the contributing carrier for clarification. In one case, a participant was asked to resubmit tapes so that year of claim could be determined from the data.

The Request for Participation, prepared by the Director of Research prior to the contract, specified the data elements required from the carriers. These elements appear in Appendix D. In most cases, not all data requested were received. For instance, unique identifiers for claimants were not always provided; in these instances, claims and demographic data were studied for logical patterns to create unique claimant identities and associated claims for each year.

Record-level editing involved logical checks within and across fields. For instance, some diagnostic codes were inappropriate for certain age groups, resulting in identification of invalid records. As many data elements as possible were retained within a record to allow inclusion in as many analyses as possible. As a result, some case totals differ across analyses. A small proportion (less than 0.25%) of cases were eliminated due to illogical or missing fields.

Although the original RFP and the contract for the research included analysis of only 1991 data, the final study incorporates analysis of both 1991 and 1992 data submitted by the participants.

3. Data Standardization

All data were read into a microcomputer database, as specified by the Society. Data were organized by participating insurer and within insurer, by individual claimant. Individual records per claimant comprised fields for each of the elements requested in the original request, as well as fields for new variables constructed for the analysis.

4. Analysis of the Data

The analysis plan was constructed in consultation with the POG and Society actuaries. The unit of analysis for all analyses was the claimant. Most analyses are presented separately for each year for which data were submitted (1991, 1992).

Deductible amounts were defined as the minimum of each range in those analyses in which all ranges were presented. In other analyses, deductible levels were specified by the POG at \$25,000, \$50,000, \$100,000, \$150,000, and \$250,000. Age ranges were defined by the POG.

Carriers submitted diagnoses based primarily on the ICD-9 and ICD-9-CM classifications. Analyses utilized both the individual diagnoses as well as the diagnostic categories created by the POG. The project database retained both the actual diagnosis code and the defined diagnostic category.

C. Results

This section summarizes the findings of the analysis on all participants and all plans for both 1991 and 1992. The tabular results of the analysis follow in Sections II-VIII. For each level of analysis, tables present 1991 data separately from 1992 data. Records in which critical data elements were missing were removed for selected analyses; thus, subtotals and totals occasionally differ across certain tables.

The first levels of analysis are performed on all participants (subscribers and dependents) covered by any plan for each year and are reported in the tables in

Section II. These analyses of total, maximum, minimum, and average charges and numbers of claimants describe the gender, age and diagnostic characteristics of subscribers and dependents, combined and separately.

The second level of analysis (Section III) examines total, maximum, minimum, and average charges and claimant features by the different types of health benefit plans recorded in the claims. Indemnity and comprehensive plans, multiple-employer trusts, and self-funded plans are combined in the first category; managed care plans, health maintenance organizations (HMO), exclusive provider organizations (EPO), and point of service plans (POS) in a second category; and preferred provider organization (PPO) plans in a third category. As before, the analysis presents all claimants combined, followed by subscribers and dependents separately. Gender, age, and diagnostic characteristics are presented for each of these three groups for each year. Each table presents the number of claimants and percentages of totals and the maximum, minimum, and average charges for the participant group highlighted in the respective analyses.

The next major sections (Sections IV, V, and VI) present analyses by deductible level. As noted earlier, deductible levels were defined as \$25,000, \$50,000, \$100,000, \$150,000, and \$250,000. Findings are presented separately for 1991 and 1992. Section IV presents analysis by two-digit zip codes. Section V presents analysis of total charges by range of charges, excess and average charges by all plans, and by each category of plans. Section VI illustrates the mix of hospital, nonhospital, and total charges by deductible levels for those carriers submitting hospital-level data. This sample for analysis was reduced by the POG to maximize the validity of the hospital charge analysis; thus these analyses were performed on the sample of claims from 19 carriers with 1991 data and 20 carriers with 1992 data.

Section VII presents analysis by the 30 most frequent diagnoses for all participants for the two years and provides a summary of the relationships between charges and claim type.

Section VIII provides exposure analysis by subscriber status, gender, and age range for claims frequency and total charges at various deductible levels for each year. These analyses were performed using data only from the four insurers who, in addition to claims, submitted sufficient exposure data.

1. Incidence and Size of Large Claims

Large claims data were submitted by 23 participants for 75,789 claimants in 1991, and by 24 participants for 95,447 claimants in 1992.* Total charges were \$4,427,068,302 in 1991, and \$5,611,726,199 in 1992. Total charges per claimant ranged in 1991 from \$25,000 to \$4,518,420, and in 1992 from \$25,000 to \$7,104,081. Average charges per claimant were \$58,413 in 1991 and \$58,794 in 1992.

Exposure data submitted by four of the insurers allowed an assessment of the rates of high-cost claimants among the 5,916,025 exposed individuals. In this sample, the incidence of high-cost claims was 1%. Analysis of charge data from these claims reveals an average charge per claim exceeding the \$25,000 deductible of \$193.20 in 1991 and \$226.90 in 1992.

2. Characteristics of the Claimant

Age

The mean age of claimants was 43.7 years in 1991 and 44.5 years in 1992, with a range of 0 (birth) to 107. The majority (68% in 1991 and 70% in 1992) of claimants were between the ages of 30 and 64. This age group accounted for similar proportions of charges each year: 65% in 1991, and 68% of charges in 1992. In both years, 50- to 59-year-old claimants accounted for the highest percentage of claims (22% in 1991, 23% in 1992), and the highest percentage of total charges (21% in 1991, 23% in 1992). The highest average charges per claimant were for claimants from birth to age 1, averaging roughly \$91,000 per claimant. The second-highest average total charges per claimant by age range was in the over-75 age group, at \$65,000 per claimant. These relationships and charge levels held in both years for which data were available.

* As noted earlier, although twenty six insurers submitted data for the study, one insurer submitted no claims-level data; instead this participant submitted in paper form only the grand totals for all claimants for each year. These data are not included in any of the analyses in this report. In addition, not all insurers submitted data for both years.

Gender and Age

Among females in both years, the largest proportion of claims was filed for the 40–49 age group (approximately 20%), accounting for an equivalent proportion of the total charges for all females.

For males, the 50–59 age group was responsible for 23.5% of all claims in 1991 and 24.6% in 1992, and 21% and 23.3%, respectively, of all charges in each year.

In both years, the highest average charges per case for both males and females were seen in claimants under the age of 1. High-cost cases among infants averaged \$89,951 for females in 1991 and \$88,766 in 1992; male infants' average charges were \$91,497 in 1991 and \$94,841 in 1992.

The average charge for females of all ages was \$56,086 in 1991 and \$55,683 in 1992; for males, it was \$60,534 in 1991 and \$61,706 in 1992.

Gender and Diagnosis

For all female participants, malignant neoplasms and circulatory system disorders accounted for the largest proportion of claims and charges. In 1991, 17.2% of all female high-cost cases had a diagnosis of malignant neoplasm, and 13.6% recorded a circulatory system disorder. In 1992, 19% of claims were for neoplasms and 13.5% were for circulatory disorders. These two categories of diagnoses alone accounted for \$600 million in charges in 1991 and \$900 million in charges in 1992.

For males, the predominant diagnoses, as measured by numbers of claims, were associated with circulatory system disorders (28.6% in 1991 and 29.9% in 1992), accounting for comparable portions of the total charges for males each year.

The highest *average* charges for both females and males, however, were for congenital disorders and perinatal conditions, including prematurity, at approximately \$80,000 per case each year.

Age and Diagnosis

Among infants, those under age 1, congenital abnormalities and premature births account in both years for the vast majority of charges and claims. About 68% of claims and 68% of charges are associated with these conditions for those ages.

Among children aged 1–9, congenital abnormalities (27% of claims) persist but are joined by neoplasms (11%), and injuries and poisonings (10%) in their increasing frequency and contribution to charges.

Among those aged 10–18, across both years, mental disorders and drug and alcohol abuse diagnoses rise in their preponderance (44.8% of all claims) and in their contribution to charges (36%). Injuries and poisonings still contribute roughly 13% of the cases and 16% of the charges for this age group.

Among those from 19 to 39, mental disorders, injuries, and neoplasms account for approximately half of all cases and charges.

From age 40 on, circulatory system diseases and malignant neoplasms account for an increasingly dominant share of the cases and charges. Between 35% and 55% of all cases and charges in both years are from these categories of illness.

Subscriber Status and Gender

Subscribers accounted for the majority of claimants in both years in the database (60%) and for a comparable proportion of total charges (58%). In both 1991 and 1992, 52% of the claimants were male; this group accounted for 54% of the total charges in both years. In both years, the largest proportions of claims and the percentage of total charges per claim were filed by male subscribers (37% in 1991, 36% in 1992), followed by female dependents (25% in each year), female subscribers (23% in 1991, 24% in 1992), and male dependents (16% in each year.)

Subscribers Only

The study sample of 45,245 subscribers incurred total charges of nearly \$2.6 billion, an average of \$57,018 per subscriber claimant in 1991. In 1992, the sample of 56,945 subscriber claimants incurred \$3.3 billion in total charges, an average of \$57,348 per subscriber claimant.

The average age of subscriber claimants was 49.9 years in 1991 and 50 years in 1992. In 1991, average total charges among subscribers in the oldest age group (over 75) incurred the highest average charge per claimant (\$65,465), followed by the 30–39 age group (\$57,482). In 1992, although the sample of subscribers under age 18 was small, the highest average charges were incurred by this group (\$77,610), followed by the over-75 age group (\$71,752).

Males accounted for over 61% of the subscriber claimant group in 1991 and 60% in 1992. Female subscribers incurred, on average, lower total charges than male subscribers in both years of study. In 1991, the female subscriber claimant averaged \$54,432 in claim charges, while the male subscriber claimant averaged \$58,652. In 1992, the female subscriber average total charge was \$53,262 versus \$60,072 for the males.

There is a greater percentage of male subscribers, versus female subscribers, in every age group except the under-18 age group in 1991 and the 19–29 age group in 1992. The largest difference between the charges reported by male and female subscribers occurred in the oldest age group (75 and older), where an average difference in claim amount of \$27,603 was seen.

For both male and female subscribers, circulatory system conditions and cancers were responsible for the greatest number of claims in both 1991 and 1992. Of male subscriber claims, 33% was for these conditions and of female subscriber claims, 17%, at average charges of \$55,000–\$60,000. Malignant neoplasms accounted for approximately 21% of female subscriber claims and 17% of male subscriber claims.

In 1991, 48% of subscribers for whom plan type was reported filed claims under an indemnity, comprehensive, or self-funded plan or a multiple-employer trust. Of these claims, 33% were filed under a PPO plan and 19% under an HMO, an EPO, a POS or other type of managed care plan. In 1992, traditional plans accounted for 42% of claims, PPOs for 37% of claims, and other managed care plans for 21% of claims.

Average total charges in both years were highest for the subscribers covered by managed care plans (HMO, EPO, POS, or other managed care plan). Subscribers enrolled in managed care plans, on average, incurred total charges of \$64,214 in 1991 and \$59,715 in 1992. This compares with average charges of \$55,255 in 1991 and \$55,150 in 1992 among subscribers enrolled in traditional insurance plans, and \$54,567 (1991) and \$55,748 (1992) among subscribers in PPO plans. In 1991, the average charges for claims among subscribers in managed care plans was 16% to 18% higher than the averages among subscribers in the other two groups of plans. In 1992, this charge difference narrowed somewhat to 7% to 8% higher than traditional plans.

Dependents Only

The study sample of 30,498 dependents incurred total charges of more than \$1.8 billion, an average of

\$60,481 per dependent claimant in 1991, almost \$3,500 more than the average subscriber claimant. In 1992, the sample of 38,442 dependent claimants incurred more than \$2.3 billion in total charges, an average of \$60,921 per dependent claimant, nearly \$3,600 more than the average subscriber claimant.

The average age of dependent claimants was 34.5 years in 1991 and 36.3 years in 1992. In both years, in contrast to subscribers, average total charges among the youngest dependents was the highest (roughly \$91,000).

Females accounted for more than 61% of the dependent claimant group in both years. As with female subscribers, female dependents incurred total charges that were 11% less, on average, than those incurred by male dependents (\$57,633 for females versus \$64,916 for males in 1991 and \$58,054 versus \$65,375 in 1992).

In 1991, there are proportionately more females than males in all but the three youngest age groups, which included dependents 18 and under, and in all but the two youngest groups in 1992. Male dependents, on average, incurred higher total charges than female dependents across all age groups except 60–64, in which charges were similar, and 65–69, in which the average among females was considerable higher than the average among males. The differences among male and female dependents in 1991 were widest among the 19–29 year group (\$8,651), ages 40–49 (\$7,849), and over 75 (\$7,451). The year 1992 revealed a somewhat different pattern. Average total charges were higher for male dependents than female dependents across all but four age groups. In these four (1–9, 60–64, 65–69, and 70–74), female dependents incurred higher average total charges than male dependents, ranging from \$1,156 to \$4,141.

Among dependents for both years, the highest frequency of reported diagnoses occurred for circulatory conditions and neoplasms (roughly 14% for each in each year); mental disorders, drug and alcohol abuse (12% in 1991 and 9% in 1992), and congenital disorders and premature births (10% in each year). In 1991 and 1992, although the order of the diagnoses was not entirely the same, the most expensive claims for dependents were associated with diagnoses of congenital abnormalities and premature births (\$86,027, on average), nervous system conditions (\$77,678), health-service-associated illnesses (\$76,464), and infectious and parasitic disorders (\$73,899).

In 1991, 48% of dependents for whom plan type was reported filed claims under an indemnity, comprehensive, or self-funded plan or a multiple-employer trust.

Of these claims, 33% were filed under a PPO plan and 19% under an HMO, an EPO, a POS, or other type of managed care plan. In 1992, 41% were enrolled in the more traditional plans, while 36% were in PPO plans and 23% in managed care plans.

As with subscriber claimants, average total charges in both years were highest for the managed care group (HMO, EPO, POS, and other managed care) in both years. Dependents enrolled in managed care plans, on average, incurred total charges of \$69,086 in 1991, and \$65,431 in 1992. This compares with average charges of \$57,954 (1991) and \$59,173 (1992) among dependents enrolled in traditional insurance plans, and \$58,649 (1991) and \$59,882 (1992) among dependents in PPO plans. In 1991, the average charges among dependents in managed care plans were 19% higher than the averages among dependents in either traditional plans or PPOs. In 1992, the gap narrowed somewhat to 11% higher than charges for cases under the more traditional plans or PPOs.

3. Geographic Distribution

Of all claims, 42% recorded sites of care in the southern states; 21% in the midwest states; 20% in the western states; and 14% in northeastern states. These rates were consistent across both years. Zip code analysis revealed significant differences in average claimant charges by region. When aggregating zip codes into four regions, averages were highest in claims submitted from the western states (roughly \$64,000 in each year); followed by the southern states (roughly \$58,000 each year); the northeastern states (approximately \$56,500 each year); and the midwest states (\$56,000).

4. Hospital Charges

For insurers providing hospital charges, the percentage of total charges represented by hospital charges ranged from 14.6% to 98.6% (1991 and 1992); the weighted average of hospital charges for all insurers was 61.4% of total charges in 1991 and 61.7% in 1992. Based on this analysis, a subset of insurers was selected by the POG to reflect the weighted average, rather than the extremes in range. This resulted in a subset of 19 carriers in 1991 and 20 carriers in 1992, whose hospital charges as a percentage of total charges ranged from 59% to 87%.

5. Highest Cost and Most Frequent Diagnoses

For this series of analyses, the POG categorized diagnoses (ICD-9 and ICD-9-CM) into diagnostic groups and age into slightly different age categories from the earlier analysis. Separate analyses are presented on the aggregated diagnoses and on the 30 most frequently occurring individual diagnostic codes.

Total charges by diagnostic grouping are largest for both years for three diagnostic categories: circulatory system diseases, malignant neoplasms, and injuries and poisonings. Analysis of average total charge by diagnostic grouping by gender shows highest average charges for males for congenital conditions (\$85,000), infectious and parasitic diseases (\$82,000), and blood-related diseases (\$81,000). For females, highest average charge per claimant for the top three diagnostic categories were for premature births and congenital conditions (\$80,000), nervous system disorders (\$75,000), and infectious and parasitic diseases (\$70,000).

As deductibles increase, however, the level of excess charges represented by circulatory disease and cancers remains stable at about 18% to 20%. However, the percentage of excess charges for claims for congenital disorders, and for injuries and poisonings, increases with increasing deductible levels. This phenomenon is also reflected in the age, status, and gender analysis by deductible level.

Analysis of the 30 most frequently occurring individual diagnoses supported the findings of the aggregated analysis, with coronary atherosclerosis, acute myocardial infarction, and post-myocardial infarction syndrome accounting for the three most frequent diagnoses recorded. Interestingly, affective psychosis ranked fourth and osteoarthritis, fifth.

Analysis of the highest cost diagnoses by gender and age revealed the preponderance of congenital conditions in male and female infants. The highest average charges per claimant, however, were found in those cases with diagnoses of digestive systems disorders in males and in the childbirth diagnostic codes and sensory organ disorders in females. In male and female children and adolescents, the most common diagnoses were in the mental and substance abuse categories; however, the highest average charges were for infectious diseases and cancers for males, and for cancers and nervous systems disorders for females.

Of the 18- to 65-year-old male claimants, more than 30% showed circulatory system disorders, by far the most common diagnosis in this age group for males. For females, almost 19% of the cases recorded cancers, and roughly 15% noted circulatory disorders. The highest average charges for males in this age group were in cases with diagnoses of blood-related disorders, and for females, infectious and parasitic diseases. In this age group, females showed significantly lower average charges than did males.

In the over-65 age group, almost 40% of males and 30% of females reported circulatory disease as the primary diagnosis of the high-cost claim. The highest average charges for both genders in this age range were for diagnoses associated with health services use (V codes) or symptoms and ill-defined conditions (ICD codes 780–799).

D. Recommendations

The study conceived by the Health Section of the Society of Actuaries resulted in a database of more than 171,000 large claims for two years, 1991 and 1992. The findings reveal characteristics of large claims that should be of use to actuaries and other researchers. The findings also generate as many new and interesting questions as they do answers. Questions on the trends in large claims, relationships among variables in explaining the incidence of large claims, and the types of plans most likely to influence the occurrence of these claims are but a few. The availability of the database will allow a wide variety of specialty questions to be asked on reasonably timely data.

Although the study has and will continue to benefit its users, the database itself could be updated over time to make its value even greater. Trends beyond two years could be studied if the current database were

extended. Substitutions of one therapy over another by diagnosis could be analyzed for their impacts on cost or prevalence of catastrophic claims.

Additional data elements that could be collected could expand the utility of the database. Contributions of more extensive exposure data from each contributor would allow for more intensive actuarial analysis. Adding actual dates of claims to the datasets submitted would allow for analysis of patterns of treatment within and across years. Retaining multiple diagnoses and procedures per claim would improve the analysis of the clinical picture of the large claimant, and permit analysis of, for instance, risk adjustment techniques. Adding sites of care to the individual claims would permit accurate identification of sources of care and the opportunity to explore offset effects of outpatient care for inpatient services, or medical care services for mental health services. More information on the plan coverage, including details on deductibles and copayments for individual employer groups, would add greatly to the ability to associate plan design with utilization and costs. Finally, careful specification of the definitions of all variables provided by the contributors would decrease the time required to interpret, assemble, and standardize the data received. Standardizing the medium on which the data are prepared and submitted for study would also decrease the time and costs of compilation and update.

The widespread participation of so many carriers may indicate that the request for data, in spite of potential shortcomings, met with the reality of the data systems from which the large claims were extracted. Complicating the request for data or increasing its customization substantially may result in reduced participation. A follow-up of participants in this study for potential changes in future requests for data may help clarify these decisions.