

Development of a Multi-Year Population-Based Data Warehouse of Standardized Medical Costs: Methods and Results

Priscilla Van Grevenhof, Mayo Clinic, Rochester, MN
 Judith L. Wagner, Congressional Budget Office, Washington, DC
 James M. Naessens, Mayo Clinic, Rochester, MN
 Charles H. Darby, Mayo Clinic, Rochester, MN
 Jill M. Killian, Mayo Clinic, Rochester, MN
 Eva R. Helgeson, Mayo Clinic, Rochester, MN
 and Roger W. Evans, Mayo Clinic, Rochester, MN

INTRODUCTION

A collaboration of researchers at the Mayo Medical Center and Olmsted Medical Center created a SAS® data warehouse of detailed billing data. This unique warehouse contains billing data for most of Olmsted County from 1987 to present. The challenge for the developers was to create a warehouse of all historical data from five different institutions that was stored on a variety of different platforms and in a variety of forms. The goal is to give researchers easily accessible, consistent data for analysis. SAS tools were used in the development of the warehouse and are used for accessing, manipulating and analyzing data.

RESEARCH OBJECTIVE(S)

As an adjunct to the Rochester Epidemiology Project, researchers from the Mayo Clinic and the Olmsted Medical Center have collaborated to create a SAS data warehouse for population-based studies of health care utilization and costs. Detailed billing data, representing the vast majority of medical care services for residents of Olmsted County, Minnesota, are available since 1987. Because most research questions have relevance beyond the boundaries of Olmsted County, the costing methodology, when applied to the detailed utilization, provides a standard unit cost measure that is adjusted for inflation to a base year and for geographic variation. The methodology used protects proprietary cost data from disclosure and will provide consistent unit cost estimates for ongoing research studies of health care costs. (See Appendix A for overview of cost algorithms.)

STUDY DESIGN

The Cost/Utilization Data Warehouse development in SAS consisted of the following phases: (1) mapping of all data sources to a common set of data elements with consistent definitions across all years; (2) editing, cleansing, verifying, removing duplicates, and determining gaps in the data; (3) developing reference tables containing coding system descriptions, coding translations, and costing procedures;

(4) applying a standardized, inflation-adjusted estimate of the cost of each service and procedure, reflecting the national average of providing the service in constant 1995 dollars, using Medicare hospital cost reports and fee schedules as the basis for costing; (5) providing a method for updating with new data periodically; (6) documenting methods used. Services provided from 1987 to 1995 currently populate the warehouse. Future updates including adding costs in 1996 dollars have been planned.

PRINCIPAL FINDINGS

Mapping different systems into a common format is a laborious task. Experts in each of the systems were used to assist in re-mapping. When re-mapping of procedural codes was not possible, an interpolation method was used to cost out services. Less than 3% of the services required interpolation. (See Appendix B for bar chart of hospital/non-hospital costs.) Verification of complete capture of services was done by nurse abstractors comparing samples of patient services with chart review. Additional manual verification of the cost methodology was done on random samples of services in each year of service. Documentation is essential to the success of this system.

CONCLUSIONS

The initial projects using the SAS data warehouse examine patterns of care for disease-specific populations. The warehouse concept offers researchers a robust resource delivering the data in an easy-to-use, consistent, and accessible format. Any provider/payor with an adequate historical billing archive and processing system could potentially replicate our process. Our community-wide population-based medical care costing resource is a valuable tool for investigation of many issues relevant to health services research.

IMPLICATIONS FOR POLICY, DELIVERY, OR PRACTICE

Measuring differences in health care costs attributable to patient characteristics, interventions, or temporal trends has become an integral part of health care outcomes research. Studies of relative costs or cost-effectiveness are increasingly relevant to clinical decision making. In all such studies, the objective is either to compare the value of health care resources used or estimate the effect of patient or intervention characteristics on the cost of health care. With the standard costing methodology used by this system researchers now have the ability to make these comparisons.

Priscilla Van Grevenhof, Mayo Clinic,
200 First Street, SW, Rochester, MN 55905
phone: (507)284-5585 e-mail: vangreve@mayo.edu

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

REFERENCES

- Anderson, Scott and Morton, Steve. *Data Modeling for SAS Data Warehouse*. A SAS Institute White Paper.
- Emmerich, Thomas. *The Rapid Warehousing Methodology*. A SAS Institute White Paper.
- Inmon, Bill, Imhoff, Claudia, and Gosselin, Larry. *Managing The Data Warehousing Effort: Advanced Topics*. Barnett Data Systems Course Notes.
- SAS Institute Inc. (1996), *The SAS Data Warehouse*. A SAS Institute White Paper.
- Tidename, Sue and Chu, Robert. *Building Efficient Data Warehouses: Understanding the Issues of Data Summarization and Partitioning*. Proceedings of the 21st Annual SAS Users Group International Conference.

Appendix A

Overview of Cost Algorithms

Hospital

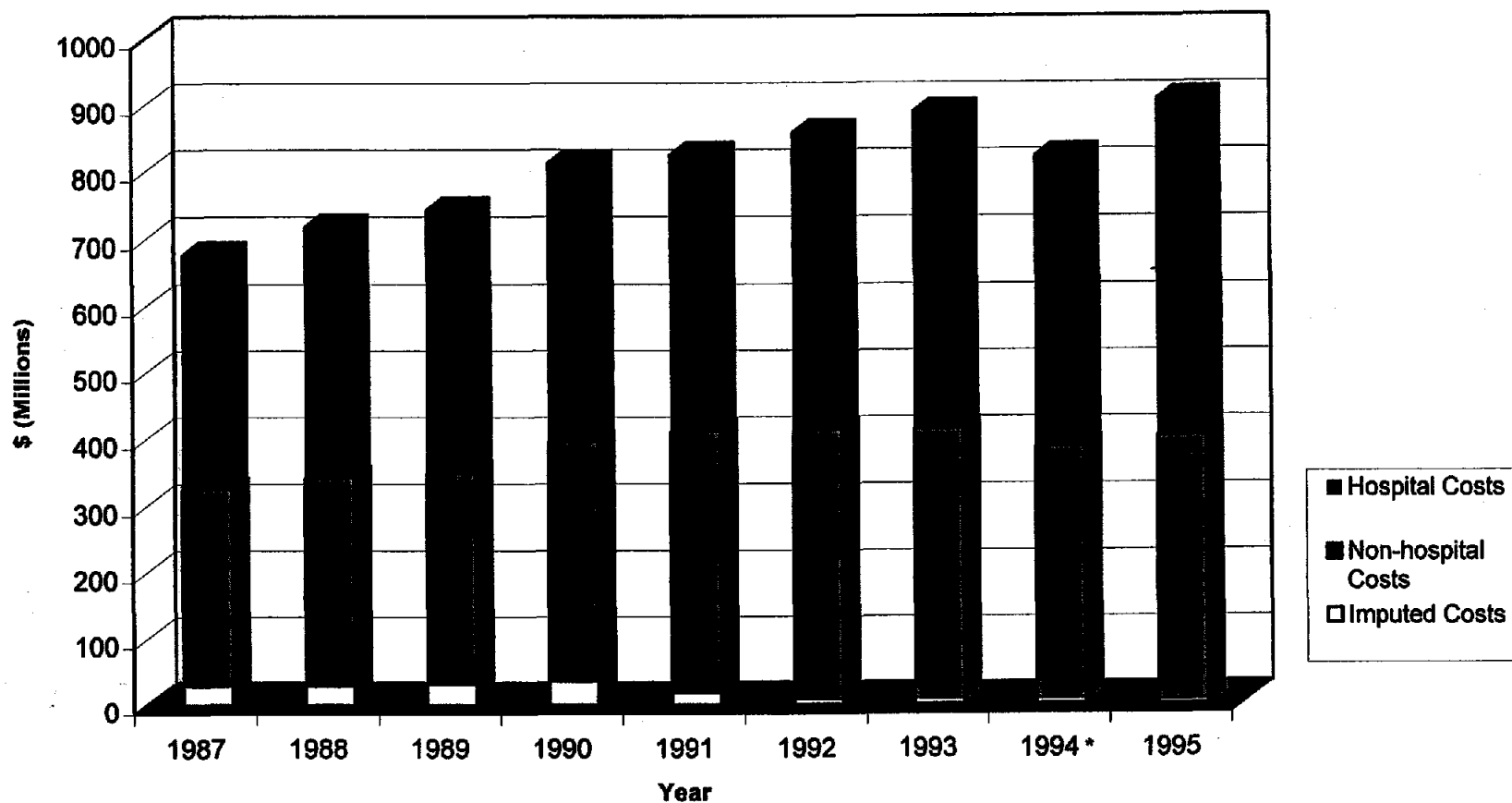
- Estimated costs were calculated based on the departmental cost to charge ratio in HCFA's Medicare Cost Reports for each hospital line item service. An inflation factor and geographic adjustment factor were then applied.
- Imputed costs were calculated using a year specific, institutional cost to charge ratio. Imputation was done when the hospital line item charges could not be linked to a specific department. An inflation factor and geographic adjustment factor were then applied.

Non-Hospital

- Estimated costs are based on mapping each non-hospital line item service into HCFA's 1995 National Fee Schedules by CPT-4 code. Manual translation of historic CPT-4 codes was limited to the most frequent and most costly services.
- Imputed Costs were calculated using the average HCFA fee across all departments in the year of service. Imputation was done when non-hospital line items could not be linked to the fee schedule.

Appendix B

Imputed and Total 1995 Estimated Costs



Total Cost Across all Years Millions = \$7088.68

Standardized and inflation-adjusted (1995 dollars) unit costs for all medical care services provided by institutions within Olmsted County.

*One hospital incomplete for one quarter