
Matching Health Policy with Data: Data and Analytic Requirements for Federal Policymakers

Kenneth E. Thorpe, Ph.D.

The health data and statistical needs of our health care system continue to grow. Though we are expected to spend approximately \$1.4 trillion on health care next year, we know little about where the dollars are spent and what they are purchasing. Our national health statistics are currently collected through a patchwork of claims data and survey data. These data are collected periodically, are often out of date, and do not contain several key data elements critical for serious evaluation of the performance of our health care system. Failure to collect more timely and comprehensive data will undermine ongoing efforts for controlling the growth in costs and improving quality.

INTRODUCTION

Each year, HCFA publishes their estimates of national health expenditures. These data provide information concerning the sources of health care financing (i.e., private health insurance, Medicare, and Medicaid) and their uses. These data, known as the National Health Accounts (NHA), serve multiple purposes. From a macro perspective, they provide information regarding the changing role of the health sector in the economy. Moreover, the accounts provide a yearly benchmark for overall changes in national health care spending, as well as changes within sectors of the delivery system. In addition to the

“big picture” view of health care they provide, data from the accounts are used to address specific questions concerning health policy. For instance, each year, Federal policymakers introduce proposals for reforming the health care system. Though the nature of these proposals differ, they often seek to extend insurance coverage to the uninsured. Both the administration, as well as the Congress (in particular the U.S. Congressional Budget Office, and the U.S. General Accounting Office) are called on to provide analyses of the potential impact of various health care reform options.

The administration also uses these data in their efforts to develop health policy initiatives. Of course, the Federal role here is even broader. In addition to developing and analyzing health care reform proposals, the Federal Government also reports on access to health care, the health status of the population and the performance of the health care system. Though the NHA do not purport to serve all these functions, they generally serve as a basic, cross-cutting reference across a broad spectrum of reporting, analysis, and monitoring.

As currently collected, the NHA do not provide a timely or sufficient base for many of the health policy roles and responsibilities of the administration or the Congress. These shortcomings are traced largely to the piecemeal approach the Federal Government takes to collecting health care data. As no detailed nationally representative source of information on health care spending is available (though the Medical

Kenneth E. Thorpe is with the Rollins School of Public Health, Emory University. The views expressed in this article are those of the author and do not necessarily reflect those of Emory University or the Health Care Financing Administration (HCFA).

Expenditure Panel Survey [MEPS] is moving us in this direction), HCFA staff rely on multiple, often conflicting, sources of information. Of particular concern is our lack of systematic and timely information on private health insurance spending and benefits. As changes in this market have been dynamic, these omissions are especially troubling. This article outlines some of the ongoing analytic roles assumed by the administration, highlights the data required to meet these obligations, and provides some suggestions for improving the current Federal data collection strategy centering on the NHA.

CURRENT DATA AND ANALYTIC CAPABILITIES

The Federal Government collects several sources of data commonly used to provide insights into health care spending. These data also serve as the source of information for developing and evaluating health policy proposals. This section comments briefly on the government's current health care data capabilities and their uses. These data are used routinely to report on macro trends in the health care industry, across sectors, and payers. The data are also used to support policy development initiatives, assist in their implementation, and are used for program evaluation. The next section compares these capabilities with the ongoing requirements of Federal policymakers.

Perhaps the best known and most widely used source of data are the annual estimates of national health care spending as estimated by analysts at HCFA. The NHA provide a yearly snapshot of national health care expenditures. These yearly snapshots are periodically expanded to provide a similar, though less detailed, set of estimates for spending within each State. Data used to develop the accounts

are derived from multiple sources. To date, no single, coordinated, ongoing source of data is available for constructing the estimates. Thus, the estimates are developed from multi-sources. Some of these data sources are collected yearly, while others are collected once every 5 or 10 years. By the nature of the sources and uses construction of the accounts, these multiple sources of data are used to check and cross-reference both the allocation of spending across categories as well as program totals. Data for the accounts are derived from several agencies of the Federal Government, including the Census Bureau, the Bureau of Economic Analysis, the Bureau of Labor Statistics, HCFA, and the Agency for Health Care Policy and Research (AHCPR). Trade associations such as the American Medical Association, the American Hospital Association, Blue Cross and Blue Shield, and others also contribute. As a result, construction of the accounts is akin to a large jigsaw puzzle, with most, though not all, the pieces included. Moreover, even the dimensions of the component parts of the puzzle are a moving target. Periodic revisions in the data collection methods, and measures developed by each agency or trade association designed to advance the specific goals of each agency or association require periodic re-estimates and adjusts to the accounts. In this case, the size and dimension of the jigsaw pieces change frequently, forcing accommodating changes elsewhere in the puzzle.

In addition to the NHA, the U.S. Department of Health and Human Services (DHHS) periodically conducts national surveys providing information on health care use, health insurance coverage, total spending, and spending by employers and workers. These surveys include the National Health Interview Survey (HIS), the MEPS, and the National

Employer Health Insurance Survey (NEHIS). The HIS routinely collects information regarding health care use and health insurance.

Despite these limitations, the NHA provide a timely and useful snapshot of total health care spending and their components. Along with the MEPS, the NEHIS will provide information concerning employer and employee spending on health insurance spending. Perhaps the most important recent innovation in DHHS' data collection strategy is the MEPS. Prior to the MEPS, DHHS launched two previous national surveys, the National Medical Expenditure Survey (NMES-1987) and the National Medical Care Utilization and Expenditure Survey (NMCUES-1977). These two national surveys provided the most detailed, nationally representative surveys concerning health care spending and health status. Though comprehensive, they were undertaken once a decade. The MEPS, in contrast, now builds from the HIS sampling structure (the NMES and its predecessor did not), and will be collected on an ongoing basis. The longitudinal feature of the MEPS, and the financial economies associated with using the HIS sampling structure are notable improvements. The MEPS will provide nationally representative data concerning household spending (both sources and uses), data on health status, and health insurance coverage.

Data and Analytic Requirements

Despite the abundance of health-related data collected by various Federal agencies, Federal policymakers, and policy analysts face substantial gaps between their data and analytic needs and available data. This section outlines briefly the nature of these data and analytic needs, and compares them with the type and frequency of data currently collected.

Health Data Needs—Macro View

This sections starts with the broadest data and analytic needs, our macro views of the health care system. From a big picture view, estimates of national health care spending by source and use are important annual sources of information. Though the national totals are important, and cited frequently, the components of the accounts are also important. It is here where ongoing improvements would be helpful. During 1984 and 1989, HCFA convened a national advisory group to review the national health accounting methodology. The panel presented several recommendations (I counted 22) from their meetings. The recommendations were published in Haber and Newhouse, 1991. Many of these proposals remain valid concerns and suggestions 19 years later. I focus on two areas of concern, largely mirroring those expressed at the 1989 meeting, and focusing primarily on providing more detail in both the sources and uses measures. The major issue concerns an area of special interest among many researchers-detailed information on private health insurance spending. Though the accounts tabulate private insurance benefit payments, out-of-pocket spending, and administrative costs, the lack of detail is a particular problem. Over the past 14 years, substantial changes have been transpired within the private insurance market, most notably the shift from fee-for-service to managed care. Tracking these changes in the industry are important for the accounts. At the very least, policymakers and researchers would benefit from tabulations of benefit payments and out-of-pocket spending by type of managed care plan.

The accounts currently display total benefit payments across all private plans. Moreover, out-of-pocket payments are totals across all individuals, both insured

and uninsured. These categories are far too aggregated to allow for simple comparisons of changes in benefit design and out-of-pocket burdens facing consumers.

In addition to displaying more disaggregated data within the private insurance industry, it would also help to include a measure of covered lives. Matching covered lives to health plans would provide an opportunity for more meaningful comparisons of changes in per capita costs within the private sector.

Health Data Needs—Micro View

The major health care data and analysis shortcomings are those needed for supporting the development, analysis, implementation, and monitoring of major changes in health care policy. Each year, dozens of proposals are advanced that would change significantly the terms under which insurance is purchased, and under which coverage is expanded to new populations, among other reforms. Analysis of these options requires more disaggregated data than generally collected through the NHA. In general, analysis of health policy options often require the use of some form of microsimulation modeling. Ideally, the key data elements of the microsimulation modeling would link directly to the NHA “baseline” totals. However, developing microsimulation models in light of the nature of the data used to create the accounts, is problematic. In its current form, this requires a “top down” strategy—one in which the elements of the model are derived from the aggregate estimates. Alternatively, the accounts could be developed from ongoing and comprehensive surveys (or other approaches) that would sum up to the account totals, yet provide sufficient detail to build the models for estimating the impacts of policy ini-

tiatives. In either case, data currently available within the Federal Government to support these activities are seriously deficient.

Though the nature of health policy proposals often differ with respect to several key design features, the proposals do require estimates of their Federal costs (as is discussed later, the Federal budget impacts of the proposals involve both changes in Federal outlays as well as revenues). For example, several recent proposals have attempted to extend health insurance coverage to selected populations (i.e., children, those age 55-64 years, the temporarily unemployed, and others). The policy development process requires information regarding the cost of providing the set of benefits, to newly eligible populations, the number expected to participate, and impacts on employers, employees, and individuals that already have coverage. These data are also needed to develop estimates for each State. Increasingly, many of the Federal proposals involve significant involvement in policy design and implementation by the States. At the very least, analysts charged with developing these estimates need detailed information regarding the costs of alternative sets of private insurance benefits (specific to the newly covered population in question), the distribution of insurance coverage in the population, and the type of cost of insurance currently provided (and how much is paid by employers and employees).

Data requirements to support this type of analysis require at least the following elements:

- National and State-by-State estimates of health insurance spending, plan generosity (or alternatively the actuarial value of the plan), population demographics, health status, and financial information.

- Data linking employment information (firm size, industry, nature of employer insurance contributions, plan choices provided) to the individual.

Ideally, these data elements concerning source and type of spending would also serve as the basis for developing the more aggregated NHA.

Though the Federal Government often collects various data elements as previously noted, they are not generally linked across the surveys, nor are they completed on a timely basis. For instance, during the health reform debate of 1993 and 1994, the most recent national survey linking the largest number of these data elements together was the NMES-1987. These data were not only woefully outdated for estimation purposes, but did not match the estimates of national health care spending as produced by HCFA (the estimates of national spending in 1987 differed by approximately 10 percent).

It is also essential that the NHA totals and the data used to support policy analytic work through microsimulation models match. In addition to estimating changes in the Federal costs associated with legislative proposals, policymakers want to know how the proposals will impact overall health care spending, spending at the State level (let alone at the congressional district), as well as changes in spending by employers and households. This requires the need for sophisticated linking of data used in policy analytic micro models to the more aggregated totals reported in the NHA.

In addition to the obvious need to align the process of developing estimates of the Federal costs of health policy reform initiatives, it is also important to link the data with estimators calculating revenue impacts of Federal proposals. This adds another layer of complication, for it not only requires intradepartmental data linking and reconciliation, but must include

interagency alignment. For purposes of estimating the budgetary impacts of Federal health policy proposals, both the U.S. Treasury Department and the Joint Committee on Taxation assume the baseline projections of gross domestic product are fixed. Thus, increases in health insurance benefits stemming from new policy initiatives paid by employers will increase the cost of fringe benefits and therefore, lower cash wages. At the same time, proposals reducing the cost of private health insurance will lead to increases (relative to the baseline projections of payroll) in taxable payroll. Relative to baseline projections, the latter would increase income and social insurance tax receipts while the former would reduce them. These “indirect” tax effects of various reform proposals are often quantitatively very important. For instance, over time, President Clinton’s Health Security Act would have reduced aggregate spending on private health insurance. The resulting increases in cash wages were estimated to increase tax receipts. In fact, the scoring of these increased receipts represented (by far) the largest increase in revenues associated with the plan. Indeed, between fiscal years 1998 and 2004, higher tax receipts were estimated at \$124 billion. This compares with the substantially more visible increase in the excise tax on tobacco which “only” increased Federal receipts by \$72 billion over the same period (Congressional Budget Office, 1994). These figures both have the same impact on calculations of changes in the net budget deficit associated with policy proposals.

MAJOR GAPS: HOW TO CLOSE THEM

The observations presented previously highlight two key uses for the NHA. The first involves more macro comparisons of

changes in total health care spending, and changes in total spending by source of funds and their use. Using the account data as part of health policy formation and cost estimates represent a second critical use. Ideally, data used to model health reform options should match the more aggregate measures used in the accounts. To date, however, significant gaps in the Federal data collection strategy remain.

One positive change since the end of the 1993-1994 national debate over health care reform has been the change in the MEPS strategy. Changes in the MEPS strategy were adopted as part of an effort by DHHS to rationalize its data collection strategy, provide more timely data, and provide a better match between data requirements and data collection. At least two important changes were made as part of the MEPS that differ from the NMES. First, the MEPS household sampling strategy relies on the HIS sampling structure, allowing a more explicit link between the two surveys as well as saving money (previously the NMES has its own independent sampling approach). Second, rather than reporting results once every 10 years, the MEPS will provide ongoing data and information each year. Both of these changes are welcome improvements. However, this should be viewed as the first, of several, substantial changes required to improve both the macro and micro roles played by the accounts.

Several departments within the Federal Government collect data useful for the accounts, though it is often difficult to combine these data in a timely and useful fashion. The primary purposes of these surveys are generally not to provide a comprehensive picture of the health care industry, with the detail needed for policy analysis and formation. Thus, while we spend hundreds of thousands of dollars per year on multiple surveys, conducted by several agencies

within the Federal Government, this fragmented approach to health data collection does not yield timely and sufficiently useful data for policymakers and researchers. Collecting health data by relying on several surveys conducted outside DHHS is problematic. In particular, each of these surveys fundamentally addresses different data needs and constituencies. In addition, DHHS also conducts numerous surveys, which in many cases collect overlapping data and information.

There are at least two approaches for addressing these problems. One option is for DHHS to coordinate and expand the collection of health data and surveys conducted by other departments. Given the conflicting agendas of these surveys, this would be the most difficult option to pursue. Alternatively, DHHS could reduce fragmentation and duplication of its own survey and data collection instruments, and develop its own unified approach for collecting health care data. Though several groups within and outside DHHS clearly focus on these issues (i.e., The Department Data Council, researchers at HCFA, AHCPR, the Centers for Disease Control and Prevention, The National Center for Health Statistics, and others), fragmentation and duplication persists. To collect the requisite data needed for health policymaking, requires appointing a new departmental wide data council empowered to closely examine all data collection instruments. This group could develop specific approaches for collecting the data needed with DHHS. Absent this focus, several remaining data and analytic gaps will persist. A sampling of these are listed later, along with some thoughts for addressing these issues.

Though data collected from the MEPS panel structure are a welcomed improvement both in the scope of data it will pro-

vide as well as its timing, significant data gaps still remain. Action steps would include:

- Developing a unified framework for estimating spending by source of payment (Medicare, Medicaid, private health insurance, etc.), utilization data, economic, demographic, health insurance, and data on health status. Ideally, estimates for each of these elements are required both nationally as well as for each State. The latter data requirement simply reflects the ongoing fiscal shifts occurring between the Federal Government and the States. Moreover, these data are necessary for monitoring the implementation of the growing volume of state-based changes in health policy.
- DHHS should continue to focus on coordinating and refining the several population-based sources of health care it collects. Moreover, and most importantly, efforts should continue to work with the Departments of Commerce and Labor to coordinate data collection instruments and measures. Ideally, the MEPS would contain all the data elements it currently collects, augmented with the detailed income and employment data collected through the March supplements to the Current Population Survey, detailed information concerning Medicare and Medicaid spending, use and demographics, and the detailed data concerning employer and employee health insurance information envisioned through the NEHIS. Whether and how the NEHIS survey, or more directly the data it is attempting to collect, is merged or coordinated with the MEPS over time remains an issue.
- The coordination of data both within DHHS and across Federal agencies should develop explicit approaches for reconciling, linking, or bridging several

population-based surveys conducted by several departments that ostensibly collect similar information. This would include, but is not limited to, the U.S. Bureau of Labor Statistics' Consumer Expenditure Survey, Medicare's Current Beneficiary Survey, the Current Population Surveys, the MEPS, the NEHIS, and ongoing surveys of long-term care.

- Questions concerning health insurance, use, expenditures, and source of spending collected from the surveys should match the desired display in the more aggregated NHA data estimated by HCFA.
- Though the MEPS is designed to collect ongoing national information, State-level estimates concerning the broad-range of health spending, use, health status and insurance data will not be available. While regional data will periodically be collected through the MEPS, Federal and State policymakers need these data collected at the State level. Moving in this direction could be facilitated through some additional Federal appropriations, continued consolidation of other population-based surveys within the Government, or both. Moreover, States could be encouraged to "buy-in" to the ongoing MEPS survey structure, and through this marginal cost-sharing arrangement receive a State-based MEPS. Without Federal coordination, States will continue to collect these data using their own instruments and sampling structures. These approaches may differ across the States, making cross-State comparisons more difficult.
- DHHS should develop an explicit strategy and approach for developing microsimulation models to address issues of health policy and reconciling these data and models with the

broader aggregates used for the NHA. Discussions among the professional staff in the administration and the Congress for improving such models should also continue.

It is important that estimates of the costs and coverage implications and of health care reform proposals developed through microsimulation models are reconciled with our “best” estimates of aggregate health care spending. This reconciliation will involve discussions concerning the most appropriate methods for developing the aggregate NHA (as well as State) estimates. At issue is the extent and mix of administrative program data and population-based survey data (for instance the MEPS) used to develop the NHA. Though HCFA has traditionally used both sources, the accounts have traditionally relied more heavily on administrative data (from census, Medicare, and Medicaid) and industry data) generally supplemented with data from the NMES (and now the MEPS). The baseline aggregate spending totals developed through the population-based survey approaches—such as the MEPS—should be easily reconciled with the aggregate estimates developed through for the NHA.

Though reconciling data used to create microsimulation models with the macro health account data have received relatively little attention, they are actually critically important for developing consistent estimates of health care reform proposals. The Federal Government and the Congress have developed several simulation models for developing program spending baselines, and for estimating the costs and coverage implications of health policy proposals. DHHS used at least three such models as it developed estimates for President Clinton’s Health Security Act. In addition, the U.S. Congressional Budget Office has its own model as does the Joint Committee on Taxation. Each model will generally pro-

duce different estimates concerning costs and populations covered under various reform proposals. Some of these differences are irreducible (and desirable) while other sources of variation are less desirable and should be coordinated. Assumptions concerning behavioral changes represent welcome differences across models used to estimate program impacts. These differences should be explicit, and policymakers should be aware of how assumptions impact estimates. However, each of these models also produced different estimates of program impacts since each incorporated a different starting point concerning aggregate spending and its distribution by source of payment (in particular baseline estimates of private health insurance). For instance, if the model employed by the U.S. Treasury Department has a higher or lower starting point (baseline) estimate of private health insurance, and the distribution of coverage across the wage-earning population differs from models used to generate premium estimates generated by DHHS, then estimates of direct Federal outlays and changes in the indirect tax receipts associated with reform will not be internally consistent.

CONCLUSIONS

Changes made within DHHS since 1994 have improved greatly the scope and timing of information available to policymakers and researchers. For instance, some of the data coordination strategies within DHHS have come to fruition, though in many respects these changes simply represent a first step, albeit an important one. Policymakers and researchers need more timely, broader and detailed aggregate and distributional data to estimate the impact of policy options, monitor their implementation, and report on their impact. We need to continue to push for a convergence (or at least a reconciliation) in the categories

of measurement and displays used in the NHA data with micro models used in the Federal Government to estimate program impacts. A substantial gap still remains both in the development of State-level accounts, as well as the development of micro modeling capabilities designed to understand and measure reform impacts at the State level.

Expanding the scope and timing of data will be expensive. As the debate over the future of Medicare and Medicaid as well as the private insurance market continues, the need for timely, comprehensive health care data is essential. Our current system for collecting such data suffers in several dimensions. These include lack of timely data, lack of appropriate links across data we collect, as well as continued duplication of data collection efforts. DHHS has taken a leadership role in identifying these issues within the Department and has made important strides in solving many of these issues. However, to be successful, these discussion and efforts at coordination need to cut across all departments in the administration. Given the relatively low amount the Federal Government currently spends to support data collection activities, a strong case can be made to support such

efforts. Some of the additional costs can be financed through additional savings and efficiencies in sampling design, and reduction in duplication of effort within executive departments and across them. Financially partnering with the States to collect more timely and detailed State-based estimates may also represent a prudent course of action.

ACKNOWLEDGMENT

The author wishes to thank Katharine Levit, HCFA, for comments on an earlier draft.

REFERENCES

Haber, S., and Newhouse, J.P.: Recent Revisions to and Recommendations for National Health Expenditures Accounting. *Health Care Financing Review* 13 (1):111-116, Fall 1991.

U.S. Congressional Budget Office: *An Analysis of the Administration's Health Proposal*. Washington. U. S. Government Printing Office, February 1994.

Reprint Requests: Kenneth E. Thorpe, Ph.D., Robert W. Woodruff Professor and Chair, Department of Health Policy and Management, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322. E-mail: Kthorpe@sph.emory.edu