Measuring and Modeling Health Care Costs

Ana Aizcorbe, Colin Baker, Ernst Berndt, and David Cutler

February 2017

Medical care costs accounts for nearly 18% of Gross Domestic Product (GDP) and 20% of government spending. These numbers are so large that it is imperative to understand what we get for that spending.

As a country, we know a lot about where the medical dollar goes. Thirty-eight percent of medical care dollars are paid to hospitals, 31% is paid for professional services, 12% is for outpatient pharmaceuticals, and so forth. But this is not really what we value. The goal of medical care is not to poke, prod, or take pictures of our insides; rather, it is to improve our well-being. To really understand health care, we need to determine what it is doing for that most precious of commodities – our health.

Health accounting is not easy. Academics and statistical agencies have struggled with it for decades. Questions range from the mundane – how do colonoscopy prices vary across payers? – to the fundamental – to what extent is medical care improving the population's health? With this much uncertainty about the value of medical care, it is incumbent on public and private researchers alike to regularly survey the landscape. What do we know about medical care costs and output? Where can we make improvements in our measurement systems? What areas remain unexplored?

These issues were the subject of the Conference on Research in Income and Wealth in 2013, and they are the topic of this volume. The papers in this volume were presented and discussed at the conference. They were then revised, peer refereed, and revised again before this publication.

As if the topic itself were not controversial enough, there was nearly a government shutdown at the time of the conference. Such a shutdown would have prohibited government

employees from even attending. Fortunately, the shutdown failed to occur, and government employees were full participants in the conference.

The conference was held and the papers written at a time when the Affordable Care Act (ACA) was sure to be law for the next several years. As the book went to press, the new Trump Administration and Republican Congress were working on plans to repeal it. As the book is read in the future, some of the language surrounding the ACA may be out of date. Fortunately, the themes are not and the empirical analysis will be timely under any set of health care rules.

The Conference staff at the National Bureau of Economic Research, led by Carl Beck, Rob Sherman and Brett Maranjian, provided flawless logistical and related assistance. Equally fortunately, the papers were terrific. Befitting the difficulty of the issue, the papers are organized into several themes. We develop those themes briefly in this introduction, as they appear in the volume.

Section I: Methodological Issues in Measuring Health Care Costs and Outcomes

We begin with a survey chapter characterizing the current state of health care cost, outcome and productivity measurement by Paul Schreyer, Chief Economic Statistician at the Organization for Economic Cooperation and Development, and his coauthor, Matilde Mas, from the University of Valencia and Ivie. Schreyer and Mas lay out the type of challenges faced when measuring medical care costs and outcomes and provide a review of how this sector is currently measured in 30 OECD countries. Specifically, they discuss two types of issues that arise. Among the national accounting issues, difficulties in valuing non-market activity present challenges for properly measuring the dollar value of health care (nominal output). The second set of issues deals with how to decompose the growth in that value into price and volume

measures; that is, decomposing the change in nominal spending into changes attributable to things getting more expensive (inflation) vs changes in the quantity and quality of goods and services provided (growth in real output). In national statistics, this decomposition can be done either indirectly by using price indexes to deflate the growth in nominal output or directly by constructing volume indexes.

Valuing non-market activities—such as health care provided for free by the government—is difficult because there are no transaction prices with which to apply the usual methods. Thus, the universal practice in the countries under study is to value these activities at cost—using price indexes to calculate changes in the cost of providing inputs to treatments for conditions. As noted in Steve Landefeld's discussion of this paper (at the time of the conference, Landefeld was Director of the US Bureau of Economic Analysis), valuing these activities at cost implicitly assumes zero productivity gains in providing treatments, which in turn assumes away the possibility that innovations might allow more treatments to be provided at the same cost or the same treatments at a lower cost.

With regard to splitting out changes in spending into price and volume components, several issues arise. First, the authors note that one would, ideally, want to measure *the complete path* of treatments for a medical condition or episode of care. Doing so would properly account for shifts in treatment protocols that affect cost: for example, shifts from talk therapy to (lower cost) drug therapy in the treatment of depression. However, the organization of the available data accounts does not allow one to measure care using this definition. In particular, in the administrative data that is typically available, treatments at different venues (hospital, residential care, etc.) are reported separately and do not allow one to tie all of the spending to specific patients. For that reason, virtually all of the OECD countries use price or volume indexes for the

individual treatments. However, as noted by Schreyer and Mas, there is increasing interest in using disease-based price indexes that tie expenditures and activities to specific medical conditions. Construction of such disease-based treatment price indexes is becoming feasible in part because of the increasing availability of government and private sector medical claims data.

The increasingly available health care claims and outcomes data in electronic format covering millions of lives raises issues of how best to exploit such data statistically. One major problem with many of these observational claims data is that they are not generated from randomization, i.e. treatments and non-treatments are not randomly assigned to patients, but instead reflect the decisions of physicians, patients and payers resulting in data subject to selection biases. This is in contrast to experimental data emanating from randomized controlled trials, or from quasi-randomized data plausibly linked to a quasi-randomized data generating process. Can one use sophisticated statistical methods, such as propensity score procedures, with observational data to generate reliable estimates of causality that inform cost-effectiveness analyses? That is the focus of the second chapter in this section by Armando Franco of the University of California at Berkeley, Dana Goldman of the University of Southern California, Adam Leive of the University of Pennsylvania, and Daniel McFadden of the University of California at Berkeley.

Franco and colleagues start by noting the broad popularity of comparative effectiveness research. This research, which typically compares one drug to another, is generally based on randomized trials. However, randomized trials are expensive, often underpowered to detect rare outcomes, and typically focused on a homogeneous group of patients. Thus, it is natural to wonder if claims data can substitute for randomized trials.

Franco and colleagues use data from 2006 to 2009 for Medicare Parts A, B, and D to examine these issues. The specific drug class they consider is angiotension II receptor blockers (ARBs), which are used to treat hypertension. The FDA has identified stroke and cancer as possible unintended consequences of using ARBs. They examine whether claims data confirm these results.

Even a cursory examination of claims data highlights the significant difficulties in comparative effectiveness research they entail. Some people discontinue treatment, while others switch from one treatment to another. Neither of these decisions is random. The paper considers two methods to control for non-random selection of people into treatments. First, they assume that physicians have a preferred drug to prescribe, and patients do not choose physicians on the basis of this unobserved propensity. Thus, they compare the outcomes for physicians that prescribe ARBs more frequently compared to physicians that prescribe them less frequently. Second, they instrument for the patient's choice of an ARB using the relative price of ARBs in comparison to other hypertension treatments. If patients do not choose plans on the basis of these price differences, this instrument can serve to randomize treatment to individuals.

Using each of these strategies, the authors find mixed evidence that ARBs lead to higher cancer rates, and some evidence that ARBs lead to higher stroke rates than other anti-hypertension medications. However, other signs are troubling. Use of ARBs presumed to be exogenous is associated with greater reports of pain. Since neither ARBs nor other anti-hypertensive medications would affect pain, these results suggest non-random assignment of people to treatments, even with the two methodologies. Overall, their conclusion is cautious in some parts, and optimistic in others.

Traditional measures of inflation and productivity published by government statistical agencies aim to provide metrics for the sector as a whole at the macro level. Recently, the arrival of new datasets and development of new methods have allowed further study into the methods underlying the official statistics, how they might be disaggregated from aggregate sectoral to disease-specific treatment metrics, and how those measures might be improved.

The final two chapters in this section provide examples of the kinds of decompositions that can be done using official statistics and discuss the potential frailties in the data and methods. Both papers exploit patient-centric data that allow them to define the treatment of diseases over the complete course of treatment. Specifically, both sets of authors generate indexes called Medical Care Expenditure indexes (MCEs) that allow one to decompose changes in expenditures into changes in price vs. changes in quantity. Notably, because these indexes do not account for changes in the quality of treatment (or patient outcomes), a National Academies Panel recommended using the label "medical care expenditure" indexes instead of price index – and that is what both these papers do.

The study by Abe Dunn of the Bureau of Economic Analysis, Eli Liebman of Duke and Adam Shapiro of the Federal Reserve Bank of San Francisco develops a decomposition of changes in medical care spending and applies that decomposition to a health claims database for commercially-insured patients. Their decomposition first breaks out changes in spending into two components: changes in per capita treated prevalence of disease and changes in the MCE index—that tracks changes in the cost of episodes of care. They then further break out changes in the MCE into changes in procedure prices—by constructing a procedure price index—and changes in utilization of procedures. In their analyses, episodes of care are measured using an episode grouping algorithm that uses the diagnoses reported in the claims data to allocate

spending into individual disease categories. The ability to drill down to the procedure level is made possible by their dataset, which is highly granular, and reports spending and diagnosis information for each procedure. As with many papers in the recent literature, given the complexity of the task, no attempt is made to account for changes in quality of care or patient outcomes.

Overall, Dunn, Liebman and Shapiro report that both prevalence and the cost of treating conditions contributed to the growth in spending from 2003 to 2007. Further breaking out the latter, they find that most of the growth in the MCEs comes from growth in the procedure price index; there is very little change in their index of the utilization of procedures. Given the similarity in their procedure price index and the official price indexes, their finding suggests no obvious bias in the official price indexes for health care spending. Although reassuring, as the authors note, their finding is not definitive owing to differences in the composition of patients in their data—only fee for service patients with commercial insurance and drug coverage—and the broader coverage of the official statistics—that include Medicare patients, for example.

Using the five top spending categories, Dunn, Liebman and Shapiro show that their finding of little change in the utilization of procedures is the net effect of two shifts that held down costs--shifts from inpatient to outpatient care and a shift from branded drugs to generic drugs—and a shift that works in the other direction--an increase in the utilization of procedures at physicians' offices.

The study by Anne Hall and Tina Highfill of the Bureau of Economic Analysis also focuses on MCE indexes. In particular, Hall and Highfill study the numerical importance for these indexes of different methods for allocating spending by disease—alternatives to the episode grouper used in Dunn et al—and different datasets—surveys vs. claims. They consider

two methods for allocating spending to disease categories: (1) the principal diagnosis method, which allocates all spending from an encounter to the first-listed diagnosis, and (2) a regression method, which scrolls up the encounter-level data to the patient level and uses fixed effects to indicate the conditions for which the patient was treated that year. They apply these methods to two different datasets that contain patient-centric data for Medicare beneficiaries: the Medical Expenditure Panel Survey (MEPS) and the survey and claims components of the Medicare Current Beneficiary Survey (MCBS).

Hall and Highfill find that when the different methods are applied to the same datasets, the primary diagnosis method produces higher average annual aggregate growth rates. They conclude the regression-based method should be employed with caution, given its sensitivity to outliers and propensity for producing volatile indexes. Regarding the different datasets, the MEPS is the only dataset with diagnoses attached to drug events, which significantly affects the resulting indexes. On balance, however, the MCBS is probably the preferable dataset for Medicare beneficiaries because of its greater sample size and its inclusion of nursing home residents. The optimal index may be a hybrid of the primary diagnosis method applied to Medicare claims and a regression-based index for pharmaceutical spending.

II. Analyses of Sub-populations and Market Segments

An alternative approach to disease-based measurement aggregated over all providers is to instead focus on costs and outcomes in a particular health care delivery sub-market, such as hospitals or physicians. The paper by Brian Chansky, Corby Garner and Ronjoy Raichoudhary of the Bureau of Labor Statistics examines different strategies for measuring output and productivity growth in private hospitals in the US from 1993-2010. Specifically, they consider

three methods: (1) the course of treatment model, where annual output is the number of inpatient hospital discharges and outpatient visits, (2) the procedures model, which counts individual services separately, and (3) and the revenue model, which measures output using the Producer Price Indexes (PPIs) from the Bureau of Labor Statistics to deflate hospital revenues. They link these utilization and cost measures to the treated conditions using Diagnosis-related Groups (DRGs)—for inpatient care—and 16 major disease categories reported in the American Hospital Association (AHA) survey—for outpatient care.

Chansky, Garner and Raichoudhary find only minor differences in the resulting output measures and implied labor productivity measures: for 1993-2010, the three methods imply average annual labor productivity growth rates of 0.7%, 0.9% and 1.0%, with very similar results also for the underlying sub-periods. Perhaps it is not so surprising that the procedure (1.0%) and revenue (0.9%) models give very similar results: one uses a volume index based on procedures and the other uses a PPI (essentially a procedure price index) and, thus, obtains an indirect volume index.

The productivity growth implied by the course of treatment model is the slowest (0.7%) for the whole period and is surprisingly negative for the 2001-2010 period. This output measure—based on number of discharges by DRG—takes severity into account only imperfectly. The authors argue that the shift from inpatient to outpatient care meant that only the more-severe, labor-intensive patients still receive inpatient care, hence reducing output per labor hour in the inpatient setting.

Instead of looking just at costs in the hospital sector, an alternative disaggregation involves examining cost and outcome trends in a distinct sub-population. The paper by Allison B Rosen of the University of Massachusetts, Ana Aizcorbe and Tina Highfill of the BEA, Michael

Chernew of Harvard, Eli Liebman of Duke, Kaushik Ghosh of the NBER, and David Cutler of Harvard looks at decomposition methods using a large commercially insured population.

Rosen and colleagues consider three ways of partitioning medical spending to conditions. The first approach, which is typical in much of the literature, involves assigning each medical care claim to one or more diseases. For example, a visit to a primary care doctor that is coded as being for high cholesterol would be classified as spending for that condition. The difficulty with this approach is that many people have comorbid medical conditions. The claims based approach requires that physicians adequately solve the comorbidity problem – what factor is really contributing to the patient needing care? In practice, such an attribution is difficult to make, and may not even be possible for patients with particularly complex illnesses. The second approach is a regression approach. In this method, total spending for the year is regressed on the full set of conditions that a patient has. The resulting coefficients are used to back out spending for each condition. In practice, however, the regression approach is only as good as the underlying model of spending, which is itself problematic in a number of ways.

The third approach, which is developed by the authors in the chapter, is to use a propensity score methodology to cost diseases. The idea is to find people with a particular condition and compare their spending to a group of people who are otherwise similar but without the condition in question. The resulting spending difference is an estimate of the cost of treating that condition.

The data that Rosen and colleagues employ is from the Marketscan database, which has 2.3 million people under age 65 with both medical and pharmaceutical coverage. The authors note that the method chosen to allocate spending has a material impact on the findings. Broadly speaking, the claims-based approaches allocate more spending to acute conditions – a heart

attack for example – while the regression and propensity score approaches attribute some of that spending to comorbid conditions such as musculoskeletal problems and mental illness. In addition, the authors show significant problems with the claims method, where not all spending has a condition associated with it (for example, prescription drugs). Without a gold standard to which to compare, the authors do not choose a favorite methodology. They suggest that researchers should be very careful about methodology.

The final chapter in this section by Ralph Bradley of the Bureau of Labor Statistics and Colin Baker, then at the National Institutes of Health and now in the Office of the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services, focuses on a different sub-population, namely, the obese. The recent increase in the prevalence of obesity among Americans has received considerable press attention, not only for its possible adverse impact on the prevalence of chronic cardiovascular and metabolic diseases, but also for its impact on health care costs. A number of studies have examined the relationships among obesity and health care costs, with varying findings. Bradley and Baker begin by noting that most of the empirical studies examining obesity-health care cost relationships treat obesity, and for that matter, health insurance coverage, as exogenous variables. Unlike public campaigns to curb smoking that have been substantially successful, even with mounting evidence concerning the adverse health effects of obesity, obesity rates in the US have continued to increase. The principal contribution of the Bradley-Baker chapter is the construction, interpretation and empirical estimation of a microeconomic model where an individual's BMI is the outcome of a rational utility maximizing decision making process, i.e., BMI is endogenous rather than exogenous.

In their two-period *ex ante, ex post* micro model, Bradley and Baker specify that individuals trade off the disutility (psychic cost) of weight reduction (reduction in BMI) with the increased utility coming from better health. More specifically, both insurance status and BMI are simultaneously set *ex ante*, each depending both on observed and unobserved latent variables. After a draw of a random health status variable in the *ex post* period, the consumer chooses whether to visit a health service provider. If the consumer visits a health service provider, then based on the consumer's health status, the health service provider and the consumer jointly select a treatment intensity. Hence, as suggested by the chapter title, in the Bradley-Baker framework obesity, insurance choice and medical visit choice are endogenous, and together they simultaneously affect health care costs. The model predicts that *ex ante* moral hazard can occur as the presence of health insurance affects the BMI choice, and that adverse selection can occur where those with greater propensity to have higher BMIs will more likely purchase health insurance.

Although the logic of the Bradley-Baker micro model is relatively straightforward, measurement and econometric specification issues complicate the empirical implementation. In the *ex ante* period, the consumer makes expectations on her health status and medical spending in the *ex post* period. Based on these expectations, consumer i decides her insurance status I_i and her BMI level BMI_i. If individual i buys insurance, $I_i = 1$, else it is zero. Cost sharing respectively under insurance and no insurance is $c_{I,i}$ and $c_{N,i,}$ with $c_{I,i} < c_{N,i}$. Although the ideal BMI does not vary across the i individuals, there is a "natural" BMI denoted $B_{N,I}$, that occurs when the individual eats to satiation and pursues no other activity to manage weight. Hence, $B_{N,I}$ varies by individual. The lower the individuals ideal BMI goes below the satiated BMI, $B_{N,I}$, there is an increasing marginal disutility (i.e., hunger) of non-satiation. The econometrician

cannot observe $B_{N,I}$. When the ex post period begins, the consumer draws an unpredictable shock ε_i. After the draw of ε_i, the individual decides whether to visit a health service provider. Hence in the *ex ante* period the consumer simultaneously selects her BMI and insurance status (each of which depends on unobserved latent variables) to maximize her expected utility in the *ex post* period. Since $B_{N,I}$ is private, asymmetric information that only the individual knows, her health insurance premium cannot be risk adjusted for this private information, thereby generating adverse selection. The *ex ante* moral hazard occurs because the insured individual bears a smaller financial risk for her BMI decisions, and the BMI choice cannot be written into a health insurance contract. In the resulting Bradley-Baker Tobit cost equation, there are two selection effects – the insurance decision and the provider insurance effect, which is estimated as a multiple selection Heckman procedure. Notably, BMI decisions affect costs both directly and through health insurance decisions. To correct for the endogeneity of BMI, Bradley-Baker employ a control variable approach where a reduced form equation is estimated.

To implement the model empirically, Bradley-Baker use 2002-2010 annual data from the Medical Expenditure Panel Survey (MEPS). Between 2002 and 2010, the U.S. obesity prevalence rate increased from about 17.5% to 21.5%, with the mass BMI range migrating from the 21 to 26 range in 2002 to the 30 to 45 range in 2010, and with both distributions right skewed. To avoid a possible non-response bias, Bradley-Baker estimate a probit model for the probability of the respondent providing information on his/her BMI. Males and those with more education are more likely to respond, while the older and unemployed individuals are less likely to respond. Since corn syrup is an intermediate product for foods considered the major culprit behind obesity, Bradley-Baker construct a relative food price index as the Producer Price Index for corn syrup divided by the all items Consumer Price Index; its coefficient in the estimated *ex*

ante BMI equation is negative but not statistically significant. In the *ex ante* insurance choice equation, the coefficient estimate on the BMI variable is positive and significant, indicating that there is adverse selection with BMI. Young men have a lower propensity to purchase insurance, while individuals with children who are not beneficiaries from the State Childrens' Health Insurance program and where both spouses work in technical, professional or government occupations have a much higher propensity to purchase insurance. Within the structural BMI equation where private insurance is treated as an endogenous variable, the coefficient on the private insurance indicator variable is positive and significant, indicating the presence of *ex ante* moral hazard.

In summary, in the *ex ante* period, both insurance status and BMI are determined. If the individual purchases insurance, the financial consequences of illness are less severe, and the policy holder is not compensated by the plan for the savings generated by suffering additional disutility to get the BMI nearer to an ideal level – thereby generating *ex ante* moral hazard. Likewise, employer sponsored insurance premia do not appear to be risk adjusted for increases in BMI. As BMI increases, so does the risk of severe disease. This increases the expected utility of having health insurance, yielding adverse selection.

After having estimated the *ex post* cost equation, Bradley-Baker undertake several simulations. Of particular interest is a simulation of a 10% BMI reduction for all obese persons on costs. Bradley-Baker report a $45 per capita annual cost reduction were all obese people to reduce their BMI by 10% -- a rather modest amount. They conclude that while high BMI does increase costs, policies that are successful in reducing BMI will not generate the large cost savings previously estimated by other researchers. They conjecture that current intervention

programs to reduce obesity may underestimate the marginal disutility that obese individuals experience when they reduce an additional BMI.

III. Prescription Pharmaceutical Markets

Pharmaceutical markets present an important case for measuring and modeling health spending. Pharmaceuticals themselves account for more than 10% of medical spending. In addition, though, there are changes in the form of delivery and producer of the good (e.g., branded v. generic) that need to be accounted for. Indeed, the classic example of health price index adjustments that are thought to be essential is the lower price that results from substituting generic medications for branded ones. Despite the importance of pharmaceuticals in understanding medical care costs and prices, there has been relatively little work taking a close look at the pharmaceutical sector. The next section of this volume remedies this deficiency.

The first paper, by Murray Aitken of the IMS Institute for Healthcare Informatics, Ernst R. Berndt of MIT, Barry Bosworth of Brookings, Iain M. Cockburn of Boston University, Richard Frank of Harvard Medical School, Michael Kleinrock of IMS, and Bradley T. Shapiro of MIT, examines pricing and utilization trends around the time of patent expiration. The data they use are from IMS, one of the leading companies tracking pharmaceutical quantities and prices over time. Their analysis focuses on data from six molecules that lost patent exclusivity between June 2009 and May 2013, and which were among the 50 most prescribed medications in May 2013. Because the trends may differ across population groups, they divide the population by payer (Medicaid, Medicare Part D, commercial and other third party payers (TPPs), and cash customers) and age (above and below 65).

Their analysis reaches several conclusions. First, the major trends that occur with loss of exclusivity are on the price side. Many patients switch to generic versions of medications. This switch, at relative constant prices, lowers spending significantly. Because generics are less expensive than branded drugs, the total quantity of drugs consumed rises. Second, offsetting some of the lower spending from substitution is the fact that branded drug prices continue to raise prices after generics enter. The authors rationalize this as a result of an increasingly inelastic purchasing pool when price sensitive consumers have shifted to generic formulations, leaving brand-loyal consumers vulnerable to brand price increases.

Generic penetration rates differ across patient groups. They are generally highest for third party payers and lowest for Medicaid. Correspondingly, cash payers and seniors generally pay the highest prices for brands and generics, while third party payers (and those under age 65) pay the lowest prices. It is likely that third party payers can steer more patients to less expensive formulations, and they use this power to extract lower prices from pharmaceutical companies.

Finally, they explore the impact of an 'authorized generic' during the 180-day exclusivity period – a molecule which has been authorized as an official generic version, and has a 180-day exclusivity period as a generic drug. They find that having an authorized generic has a significant impact on prices and volume of prescriptions, but this varies across molecules. In two of the cases studied, the brand and its licensee collectively retained almost two-thirds share of the market by volume, in the others they captured less than half. Price discounts off the brand prevailing during the "triopoly" period (the period with a branded medication and a two authorized generic medications) also showed substantial variation. In some cases, the price of the authorized generic product was between the brand and the independent generic, in others it was

significantly below the independent generic. All told, these dynamics have important implications for price and quantity of pharmaceuticals.

A particularly important pharmaceutical market is for so-called 'specialty drugs' – drugs which are administered by physicians to patients through a non-oral route (e.g., injected, infused or inhaled) or taken directly by patients after requiring very exacting production processes. Many drugs with prices exceeding $10,000 per annual treatment are specialty drugs, whose prices have become controversial. Insulin is a classic example of a specialty drug. On the supply side, because of the difficulty of production, there are often few suppliers of any particular medication. As a result, shortages can (and do) occur.

Rena Conti of the University of Chicago and Ernst Berndt of MIT examine how the loss of patent exclusivity affects the prices and utilization of specialty drugs. To do this, they utilize a unique set of information on drug prices and sales from IMS health. They focus on cancer medications because specialty drugs are particularly important for the treatment of cancer and the side effects associated with their use.

Loss of patent exclusivity allows generic firms to enter a market; Conti and Berndt show that they do so. After a patent expires, between two and five generic firms enter the market. However, true competition is somewhat lower than this, since many of the drugs are made by the same company and marketed by different intermediaries. Thus, the manufacturer likely has more market power than it appears.

Even so, loss of exclusivity results in significant generic price declines. Conti and Berndt estimate that generic drug prices fall by 25-50 percent after exclusivity is lost. The prices of specialty drugs administered by physicians through infusion or injection fall by more than the price of orally formulated drugs, but each declines greatly. Prices of the branded product

increase, however, a result consistent with prior studies. For people who continue taking the branded drug, even when a generic drug is available, there is little reason for the brand manufacturer not to increase the price substantially. The combined volume of the generic plus brand medications taken rises after loss of exclusivity; it is clear that some patients and physicians are put off by the high cost of the patented medication.

Welfare results are difficult in any market, particularly one for life-saving goods. But Conti and Berndt note one summary to the welfare analysis. With generic entry, there are effectively two prices for cancer medications: the (now higher) branded price, and the lower generic price. There is also greater use of the medications after patent expiry. Greater price dispersion and higher overall utilization are hallmarks of increased consumer welfare. Thus, Conti and Berndt tentatively conclude that loss of patent protection is associated with increases in consumer welfare.

In contrast to the Conti-Berndt chapter that focuses attention on high-profile generic injectable drugs that have experienced drug shortages, the analysis of Christopher Stomberg of Bates White Economic Consulting reveals that shortages of injectable and non-injectable drugs have very similar time trends: the correlation between the number of ongoing injectable and non-injectable drug shortages is 0.94, while the average length of ongoing drug shortages for injectable and non-injectable drugs is also highly correlated at 0.89. This suggests that whatever are the factors explaining drug shortages, they apply equally and with roughly the same timing in both the injectable and non-injectable markets. It also suggests that shortage theories relying on distinguishing features of injectable drugs (e.g., changing reimbursement of Medicare Part B drugs) are incomplete, and that broader causes such as changes in competition, market structure

and quality monitoring – affecting both markets – merit further scrutiny. Stomberg examines each of these three broader potential causes.

Although there may be no single "cause" of drug shortages, Stomberg notes that the overwhelming majority of shortages affect generic drugs. A key difference between brand and generic drugs is the low margin available to manufacturers on generic drugs, particularly for those drugs that have been on the market for some time. Given the US Food and Drug Administration's AB-rating of generic drugs, generic versions are not only essentially perfectly substitutable with the same-molecule brand, but also with each other. While both quality/purity of product and reliability of supply are costly attributes for the manufacturer to provide, they are generally invisible to buyers and patients. In non-pharmaceutical markets where the quality/purity and reliability of supply attributes are observable, a premium is paid for them. An important consequence of this institutionalized substitutability among generic drugs is that when competition takes the form of near-Bertrand auctions where suppliers are asked to meet or beat the price of the competition to win a supply contract, the firms surviving the intense price competition with any sort of profit margin will need to implement relentless cost-cutting. Given that many dominant modern generic manufacturers are multi-product firms with dozens if not hundreds of products on the market at any one time, once price competition has had its relentless effect on prices for more mature generic products, revenues and profits for individual products may not make a large contribution to the bottom line of the company. As a result, when faced with supply disruptions of any magnitude on older mature low-margin products, generic manufacturers may not find it worthwhile to address manufacturing quality issues, instead reducing their investments in maintenance and product quality. The nature of market competition in U.S. generic pharmaceutical markets thus leads to a "race to the bottom" in both

price and quality. If the current costs of plant maintenance and product quality investment exceed the discounted expected value of lost profits due to a shutdown, then the investments are not worth undertaking. Note that this market competition affects both injectable and non-injectable drugs.

Regulatory actions regarding quality/purity monitoring are a second potential broad cause of shortages. In particular, in a market where product quality is not generally observable but the actions of the regulator are, the FDA's actions may play an important role in setting expectations for both buyers and sellers. The profit-maximizing decisions of producers may be to undertake only those expenses required to pass the FDA's threshold and no more – leading to a generally consistent low level of quality. Were product quality an observable attribute, manufacturers might find it optimal to differentiate themselves by optimizing around different levels of observable quality. Manufacturers facing the uncertainty of whether they will be subject to an FDA inspection may well pick a level of quality that is below the public regulatory threshold if the probability of future inspection is less than one. Manufacturers may assign different probabilities to the possibility of detection, and/or may be risk averse to varying degrees, which could lead them to choose heterogeneous levels of quality. To the extent such heterogeneity exists and it translates into differences in marginal production costs, an adverse selection problem could arise. With Bertrand-like competition, the producers most likely to survive in the market are those that are most willing to take a risk with low spending on quality, giving them a low marginal cost and an advantage in price competition. Moreover, even if the relatively risk-loving low-cost firms were eventually inspected and shut down, the consequences could be long-lasting if they have already edged out high quality competition, leaving no alternative higher-quality supply available. Thus, in Stomberg's scenarios, a key ingredient is the FDA's setting

clear expectations and time-consistent quality monitoring policies. If the FDA sets expectations both about the probability of inspection and the quality threshold in one time period, but then changes one or the other of these subsequently, it could potentially cause either disruption or time-inconsistent issues. Stomberg conjectures that altered FDA inspection rates, to the extent they reflect exogenous regime changes, are a plausible factor that could contribute to increased shortage rates (at least in the short run), and this would be an effect likely to cut across both injectable and non-injectable drugs. Later in the chapter Stomberg analyzes this possibility empirically.

The third broad possible cause of shortages put forward by Stomberg is limited price responsiveness on both the demand and supply side, at least in the short run. For suppliers, short-run price inelasticity generally stems from FDA regulatory requirements for approval of new manufacturing facilities and/or abbreviated new drug applications required to market generic drugs in the US, as well as technological obstacles to adding new capacity. On the demand side, patients' medical necessity for prescription drugs and the fact that neither they, nor their physicians, generally pay market prices for generic drugs argues for low responsiveness of demand to changes in price. Absent price responsiveness in the market, endogenous incentives for manufacturers to address supply issues are likely to be attenuated.

Stomberg then implements an empirical analysis of one of the three possible broad factors causing shortages – changes in the FDA's regulatory activity. Using FDA data on the number of inspections of manufacturing facilities and the number of citations issued, separately for US and ex-US manufacturing sites, Stomberg regresses the number of monthly shortages (both newly reported, and ongoing) on current and lagged values of the inspection and citation measures. He identifies a consistent and statistically significant predictive relationship between

FDA regulatory activity in drug quality inspections and citations and the incidence of new drug shortages, with the relationship being similar across both injectable and non-injectable drugs. He concludes that changes in regulatory activity may be one of the cross-cutting factors contributing to the ongoing wave of drug shortages, and that supply interruptions resulting from changes in regulatory activity can be viewed as a necessary step on the road to a different quality equilibrium. He cautions, however, that the predictive power of his empirical model is modest, leaving a substantial amount of variation in new drug shortage starts remaining unexplained by the regulatory activity factor. Pricing and market structure (such as changes in the number of generic manufacturers for a molecule due to mergers and acquisitions) could be additional important factors to consider in future research on drug shortages.

IV. Issues in Industrial Organization and Market Design

Many of the previous chapters refer to issues of how the industrial organization (IO) of medical care affects costs and outcomes. Several of the papers address this topic directly. Laurence Baker and Kate Bundorf of Stanford University, along with Anne Royalty of Indiana University, start with a central issue in physician markets: how to measure the concentration of physician ownership.

As Baker, Bundorf, and Royalty note, measuring concentration is important for several reasons. In the hospital industry, hospitals with greater market shares have higher prices for both inpatient and outpatient care. Some data suggest that this is true for physicians as well, although measures of physician concentration are scarce. Concentration may also influence quality, with some authors suggesting that more concentrated markets have higher quality, and others suggesting lower quality.

Measuring physician concentration is difficult because ownership patterns are difficult to follow. A small physician practice may be owned by a larger group, which itself might be owned by a big health system. Is the physician practice small, or part of a large system? Baker, Bundorf, and Royalty propose to use Medicare data to measure concentration. Specifically, they investigate the use of Tax Identification Numbers (TINs) to measure physician firms. The TIN is the organization that receives the payment from Medicare for physician services. For a measure of financial integration (their aim), this is a natural measure of concentration.

Baker, Bundorf, and Royalty use the TINs to characterize physician practices in the period 1998-2010. They reach several conclusions. First, they conclude that TINs provide a reasonable way to group practices. They tend to be consistent over time and identify groups of physicians that are known to be large. Second, many physician markets are highly concentrated. For many specialties in many areas, physician HHI indices are well above 2,500, the standard measure that triggers antitrust worry. Third, these concentration measures have been increasing over time. The increase is particularly pronounced in areas such as surgeries, while concentration has fallen over time in some medical specialties. Fourth, they do not find a large advantage to incorporating data on ownership of physicians by hospitals or other systems. Most physicians still practice independent of institutional providers. Finally, they note that other data will need to be added to Medicare claims, since data on pediatricians, obstetricians, and some other specialists are not always prevalent in Medicare data. Even still, they conclude that they have identified a promising way to measure market concentration.

A particularly important market in which to analyze competition is the health insurance market. Many countries rely on insurance market competition to promote high quality, low cost access to medical care. For example, the Affordable Care Act in the United States provides

subsidies to individuals to purchase insurance in state-based insurance exchanges. Medicare also has a private insurance option, as do national health care systems in the Netherlands, Germany, Switzerland, and other countries.

Competition in health insurance is different from competition in other markets, however. In most markets, the cost of serving people is independent of who buys the product; the cost of producing a pill, for example, depends only on manufacturing and distribution costs, not how sick the patient is. In health insurance, that is not the case. Insurers that attract less healthy enrollees will have higher costs than those that attract healthier enrollees, even with the same coverage network and prices paid.

For this reason, payments to health plans in choice-based system are often 'risk-adjusted'. The goal behind risk adjustment is to pay more for less healthy enrollees, so that such individuals do not raise the price to all enrollees. Typical risk adjustment formulae base payments on demographics along with clinical conditions.

Jacob Glazer of Tel Aviv, along with Tom McGuire and Julie Shih of Harvard University explore optimal risk adjustment in their paper. Glazer, McGuire, and Shih begin by noting a fundamental anomaly with risk adjustment based on conditions. The procedures used for risk adjustment make the weights used a function of the data on enrollees, but the enrollees are a function of the risk adjustment formula. Thus, the formula builds in adverse selection.

The question that Glazer, McGuire, and Shih ask is how to account for this adverse selection in designing risk adjustment formulae. Their analysis has both a theoretical and an empirical component. Theoretically, they design the optimal second best risk adjustment formula – second best because there is always sorting based on the risk adjustment formula itself. They show that optimal risk adjustment can be determined by constrained regression, where the

constraints (on the risk adjustment weights) require that risk adjustment transfer sufficient funds to the premium group to achieve the desired subsidy in equilibrium. Intuitively, the second best risk adjustment trades off several features, including the degree of adverse selection, which itself is based on peoples' (possibly incorrect) forecasts of their own future spending.

Empirically, they use data from seven years of the Medical Expenditure Panel Study (MEPS) to estimate the optimal risk adjustment formula. The sample is selected to be representative of people in the insurance exchanges. They consider choices between a typical Gold and Silver plan, using data on spending to sort people to plans. Not surprisingly, the market fares poorly when there is no risk adjustment; the Gold plan attracts sick people, and the Silver plan enrolls healthy people. Conventional risk adjustment improves the situation significantly. But the optimal risk adjustment is even better. Glazer, McGuire, and Shih show that the optimal risk adjustment formula has significantly lower welfare cost than the conventional risk adjustment. They also show how to incorporate other constraints on pricing that may be desired, for example limiting cost differentials between older and younger people.

Rather than focusing on risk adjustment characteristics of aspects of the Affordable Care Act (ACA) as in Glazer, McGuire and Shi, Pinar Karaca-Mandic, Jean Abraham, and Roger Feldman of the University of Minnesota, along with Kosali Simon of Indiana University attempt to establish a pre-ACA implementation baseline of data from which to compare post-ACA changes. The Affordable Care Act (ACA) of 2010 is likely the most significant new health care legislation passed by the US Congress during the Obama Administration. One important provision of the ACA is that all low-income Americans above the poverty line who lack access to affordable employer-sponsored insurance will be eligible for subsidies to purchase individual insurance in state-based or federally operated Insurance Exchanges. Since in 2012 only about

5% of the non-elderly population has coverage in the individual market and by 2016 this proportion is projected to increase to about 17%, this provision of the ACA may greatly expand the size and importance of the individual market. Another provision in the original ACA legislation sought to simplify the health insurance shopping experience for small employers with 50 or fewer full-time equivalent employees, and allow their employees to choose among options in an analogous Health Insurance Exchange, though without similar access to exchange tax credits; in 2012, only 35.2% of private sector establishments with fewer than 50 employees offered health insurance to their employees, compared with 95.9% of establishments with 50 or more employees. Other important provisions of the ACA legislation seek to control insurance premium increases through rate review regulation, and by regulating insurers' medical loss ratios (MLRs – which generally represents the proportion of health insurance premium revenues that are paid out by the insurer in medical claims).

What will be the effects of these and other provisions of the ACA legislation? To answer such important questions (and undoubtedly, there will be differences of opinion), it will be necessary to establish a pre-ACA, or at least a pre-ACA implementation baseline of data, from which to compare post-ACA changes. In this chapter, Karaca-Mandic and her coauthors discuss challenges in describing and measuring the size, structure, and performance of the individual and small group markets. Along the way they discuss improvements in data availability beginning in 2010 that could in principle address some of these issues. Finally, using data from the National Association of Insurance Commissioners (NAIC), they evaluate insurance market structure and performance during the 2010-2012 immediate post-ACA time period, focusing on enrollment, the number of participating insurers, premiums, claims spending, MLR, and administrative expenses.

Regarding the size of the individual market, earlier work by one of the co-authors and collaborators found that federal survey estimates of the individual market varied widely, from 9.5 million non-elderly in the Medical Expenditure Panel Survey Household Component, to 25 million in the American Community Survey; they attributed the wide range to variability in the precision of the survey questions, as well as differences in the reference period of the insurance questions (a particular point in time vs. any time during the previous calendar year), which generate significant measurement issues since enrollment patterns in the individual market are typically dynamic throughout a given year. Using their best judgment to narrow the range of individuals with health insurance coverage, the current authors still find a 2-1 ratio, from 8 to 16 million. This is disappointing, for it suggests we may never know to what to compare the post-ACA individual enrollment. Since most household surveys do not ask working individuals about the size of their employer, obtaining baseline enrollment data for the small group market may be even more elusive, although estimates based on employer surveys linked to administrative data appear more reliable.

Regarding the structure of the individual and small group markets for health insurance, an obvious issue is whether these markets are "competitive" and how market structure interacts with premiums/prices. In this context, the authors document very substantial heterogeneity in market competition across states and regions. Counting the number of competitors in a state is not a trivial issue, for health insurance is sold by life insurance firms, fraternal, and property/casualty insurers, as well as by health insurance firms. The authors report on data from "credible" firms, defined as having a minimum number of member years (e.g., at least 1000 in 2010 and 2011). The authors compute Herfindahl-Hirschman Indexes (HHIs) for the individual and small group markets in 2010 and 2012. Fourteen states had an individual market HHI less than 2,500 (a

minimum threshold for highly concentrated) in both years, while in the small group market the number of states with an HHI < 2,500 was relatively stable at 18 in 2010 and 20 in 2012. However, in 2012 13 states had an HHI > 5,000 in the individual market (indicating very highly concentrated), while only six states had that large an HHI in the small group market. Not only is the individual market highly concentrated in many states, but the average HHI in all states increased from 3,680 in 2010 to 3,920 in 2012. Overall, the small group market was slightly less concentrated relative to the individual market; the average HHI across all states was 3,252 in 2010 and 3,353 in 2012.

In terms of new regulations, the MLR regulations were among the first ACA provisions to be implemented. Beginning January 2011, insurers in the individual and small group markets must spend at least 80 percent of their premium revenue on medical care and quality improvement activities, while insurers in the large group have a minimum threshold of 85 percent, with those insurers not meeting those thresholds being required to provide equivalent rebates to their policyholders beginning in 2012.

Since 2010 passage of the ACA, the National Association of Insurance Commissioners (NAIC) has actively collaborated with the US Department of Health and Human Services to design standard measures, definitions and methodologies related to the regulatory targets such as MLR. Although this may make pre- and post-ACA comparisons problematic, in principle it allows for evaluation of post-ACA trends. However, one study by three of the current coauthors examining NAIC 2001-2009 data found that markets with only one credible insurer (defined as having at least 1,000 member-years of enrollment) have lower MLRs, controlling for insurer characteristics, health care provider market structure and other market attributes, and population-level demographics and health status. Although a number of definitional changes for

measurement of MLRs took place in 2009-2012, the current authors report on a previous study examining 2010 and 2011 NAIC *Supplemental Health Care Exhibit* filings, which found that the average MLR increased from 80.8% to 84.1% in the individual market, while it remained unchanged at 83.6% in the small group market. Distinguishing for-profit from not-for-profit insurers, they also report that nonprofit insurers already had high MLRs in 2010 relative to for-profit insurers (88.1% vs. 71.8).

In terms of early responses to the MLR regulation, in a study by three of the four current co-authors, it is found that individual market insurers with 2010 MLRs that are more than 10 percentage points under the 80% threshold experienced a 10.94% percentage point increase in MLR from 2010 to 2011 (controlling for a wide variety of factors), while those within five points under the threshold experienced only a 2.91 percentage point increase in MLR. Individual market insurers with MLRs more than ten points above the threshold in 2010 reported a decrease, on average, relative to insurers that were only slightly above the 80% threshold. A similar pattern of changes in insurers' MLRs was found in the small group market.

In summary, while there is some post-2010 data available on various performance metrics of insurers in the individual, small group and larger employer insurance markets indicating potential improvements in meeting MLR targets, the authors conclude that even after various plausible data curating procedures are implemented, Federal household surveys give widely different estimates of how many individuals were covered in the individual market prior to passage of the ACA. Hence, it may be difficult to track changes in enrollment and to conduct studies based on a pre/post ACA design using Federal household surveys because of the limitations in properly estimating the size of the individual market at the baseline. However, unlike in the individual market, the authors conclude that better estimates of the small group

market enrollment can be obtained from the Insurance Component of the Medical Expenditure Panel Survey. Moreover, since major improvements were made in the NAIC's Supplemental Health Care Exhibit filings in 2010, at least some empirical evaluations on certain regulatory developments such as those involving MLRs, can be reasonably assessed for the single pre-ACA year (2010) and for early post-ACA years beginning in 2011.

V. Potpourri

The two final chapters in this volume deal with somewhat different topics than those presented by the author conference participants. Nonetheless, they are equally important in addressing these issues. Didem Bernard and Thomas Selden of the Agency for Health Care Research and Quality (AHRQ), and and Yuriy Pylypchuk of Social and Scientific Systems, Inc., examine the total amount of public spending on medical care and its 'benefit incidence' in 2010. That year was important in part because it laid the foundation for modeling of the Affordable Care Act. And the effort here is particularly important in supporting the modeling that AHRQ and other agencies do to understand the likely impact of health care reforms.

The data that are used are primarily from the Medical Expenditure Panel Study. However, the MEPS is known to understate certain categories of spending and certain categories of people (for example, high spenders). Thus, the first challenge for the research is to adjust MEPS spending to national totals. The authors follow previous methodology that they and others developed to do this. In addition, the authors use data from the NBER's TAXSIM model to attribute tax expenditures to relevant groups.

The results show large amounts of government spending for medical care, directed primarily at the elderly. For example, total government spending on medical care is

approximately three times higher for the elderly than the adult population, and five times higher for the elderly than for children. Much of this spending is for the very poor, but not all of it. Medicaid benefits are predominately for the poor, but tax expenditures for employer-provided health insurance reach much higher up in the income distribution – both because higher income people are more likely to have employer-provided insurance and because the value of the tax exclusion is higher at higher incomes. Because medical costs have increased over time, the value of this spending has risen as well.

One of the fundamental issues in the measurement of health costs is determining how such costs relate to health benefits. Nominal prices count only what is spent. Real prices – and corresponding real output – require a quality adjustment. Frank Lichtenberg of Columbia University explores a novel way to measure the health benefits of medical innovation. Lichtenberg's methodology is to measure how much medical knowledge is learned about diseases, measured as the number of publications referring to the disease. He then relates this to mortality reductions for the disease.

The clinical setting Lichtenberg considers is cancer. Cancer is natural to study because there are about 45 well-identified sites, the National Cancer Institute calculates consistent incidence and mortality data by cancer site since 1975, and research innovation can be measured through MedLine searches. For each cancer site, Lichtenberg calculates the number of articles published pre-1975, and the number of articles published between 1975 and various later years.

Lichtenberg shows a clear relationship between recent research findings and mortality declines. The number of articles published in the last 5-10 years has a large and significant effect in lowering mortality. The effects are such that many cancers with declining mortality would have increasing mortality were it not for new research findings.

Taken together, the papers in this volume present a compelling case that we have made significant advances in understanding the cost of medical care, and that we can continue to make such improvements in the future. Current and future analyses will have much to learn from the studies reported here.