

## Data Appendix

Here we provide some additional details of how we selected our analytic sample and outline some differences between the public and private samples.

### *Variable Construction*

Broadly speaking, there are three categories of variables that we use in different parts of our analysis: health care spending and utilization, demographics and diseases.

*Health care spending and utilization:* For spending, we compute the total expenditures in the year for inpatient care, outpatient care and prescription drugs. We sum the inpatient and outpatient spending to construct total medical spending. For utilization, we compare the number of inpatient hospital admissions (stays), the total number of days in the hospital across all admissions, the number of outpatient visits (including hospital visits for outpatient care and physician visits) and the number of prescriptions. For the private sample, these variables are constructed entirely from the medical claims data. In the MCBS, the medical (inpatient and outpatient) data are also claims-based, but supplemented and validated with survey responses. The exception to this procedure in the MCBS is prescription drug utilization: Because Medicare did not cover prescription drugs for most of our sampling window, drug utilization is self-reported in MCBS.

*Demographics:* The key demographic variables we use are age, gender and income, which are available in both the private claims data and the MCBS. In the MCBS the income data are self-reported for each household responding. We do not directly observe income in the claims data. We proxy it using the median household income for the 3-digit zip code of the respondent's residence, taking the median value from the 2000 Census. While this variable is measured with more error in the private sample, it is not clear that the level of measurement error should systematically differ across regions.

*Diseases:* To condition on the observable health of individuals, we construct indicator variables for 30 different chronic health conditions: essential hypertension, congestive heart failure (CHF), diabetes, asthma, hypercholesterolemia, ulcer, depression, chronic obstructive pulmonary disease (COPD), allergic rhinitis, migraine, arthritis, chronic sinusitis, anxiety disorder, cardiac disease, vascular disease, epilepsy, gastric acid disorder, glaucoma, gout, hyperlipidemia, irritable bowel syndrome (IBS), malignancies, psychotic illness, thyroid disorder, rheumatoid arthritis, tuberculosis, angina, human immune-deficiency virus (HIV), anemia or stroke. While these represent the most common diseases identified in medical claims data, most of them are relatively uncommon in the general population (heart disease is the most common). These variables are identified using ICD-9 diagnosis codes in the year for both samples, so they are directly comparable.

We use the MCBS claims-based definitions of diseases to make the samples most comparable. Unlike the private sample, the MCBS contains a number of additional, self-reported health measures—including self-reported health status, questions about whether the respondent had ever been diagnosed with a number of specific conditions and limitations to the activities of daily living. We experimented

with including the self-reported health measures in the MCBS in our analysis for the public sample, either in addition to or in place of the claims-based measures, and it did not substantially alter our results.

### ***Sample Selection Criteria***

The data in our private sample come from a diverse set of employer-provided plans, including point-of-service (POS) plans, preferred provider (PPO) plans, managed care organizations (MCO) and a very small number of indemnity plans. The data in our public sample all come from Medicare.

To enter our sample, patients must have a diagnosis of ischemic heart disease (IHD). The heart disease patients are identified differently in the two samples. In the public sample, these patients are identified using the self-reported survey question of whether respondents were “ever told” by a physician that they had heart disease. In the private sector sample, patients are identified directly from the claims data. Specifically, we select patients with a primary diagnosis of 410.xx, 411.xx, 412.xx, 413.xx, or 414.xx, on any claim (outpatient or inpatient) in any year of our sample (2000-2006).

Once we identify a private sector patient with heart disease, we take all available years of data over this time period for those patients, including years prior to the first observed diagnosis. We include the years prior to diagnosis to move us closer to the “ever-diagnosed” criterion in the MCBS. Empirically, keeping the pre-diagnosis years appears to provide us with samples that are more comparable in terms of health. In our data, the fraction of patients that have a diagnosis of heart disease (identified off the medical claims) in a given year is 37% in the public sample and 32% in the private sample. If we drop the years prior to first diagnosis for the private patients, this climbs to 50%.

The ideal comparison would be to select patients with the same objectively measured health condition in the same year of diagnosis. In principle we could do that with our analysis; for example, we could restrict the public and private samples to just those patients with a diagnosis of heart disease in the year. The key limitation of this is that it substantially reduces the sample sizes (to 77,882 in the private sample and 9,208 in the public sample). These smaller samples turned out to be unworkable in terms of power. We do note, however, that the point-estimates for the analysis using the restriction based solely on an observed diagnosis in the year are extremely similar to those presented in the paper, and the qualitative results are the same.

Finally, to eliminate potential changes due to the introduction of Part D, we exclude 2006 in both samples for all analyses involving pharmaceutical spending and utilization.