

The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the authors.

**USING CENSUS BUSINESS DATA  
TO AUGMENT THE MEPS-IC**

by

**Kristin McCue \***  
**U.S. Bureau of the Census**

and

**Alice Zawacki \***  
**U.S. Bureau of the Census**

**CES 05-26**

**December, 2005**

All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, Discussion Papers, Center for Economic Studies, Washington Plaza II, Room 206, Bureau of the Census, Washington, DC 20233-6300, (301-763-1882) or INTERNET address [snguyen@ces.census.gov](mailto:snguyen@ces.census.gov).

## Abstract

This paper has two aims: first to describe methods, issues, and outcomes involved in matching data from the Insurance Component of the Medical Expenditure Panel Survey (MEPS-IC) to other business microdata collected by the U.S. Census Bureau, and second to present some simple results that illustrate the usefulness of such combined data. We present the results of linking the MEPS-IC with data from the 1997 Economic Censuses (EC), but also discuss other possible sources of business data. An issue in any linkage is whether the linked sample remains representative and large enough to be useful. The EC data are attractive because, given the survey's broad coverage and large sample, most of the MEPS-IC sample can be matched to it. We use the combined EC/MEPS-IC data to construct productivity measures that are useful auxiliary data in examining employers' health insurance offering decisions.

\* This paper is made available to encourage discussion and suggestions for revisions. It has not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed are those of the authors, and do not necessarily represent those of the U.S. Census Bureau. This paper has benefitted from helpful comments from seminar participants at CES and at the Agency for Healthcare Research and Quality.

## 1. Introduction

Rising health care costs and declining rates of insurance coverage have made employer-provided health insurance of great current policy interest. Understanding the factors that affect employers' decisions about whether or not to offer insurance to their employees, what type of insurance to offer, and what sort of cost-sharing arrangements to put in place all require detailed data on health plan features and costs, and characteristics of employers and their employees.

One response to this increased need for data has been development of the Medical Expenditure Panel Survey (MEPS) by the Agency for Healthcare Research and Quality. The Insurance Component (the MEPS-IC) is a part of the MEPS that was designed to support better understanding of the employer's role in determining whether and what sort of health insurance to offer to their employees. It collects information from employers on their health insurance offerings and on characteristics of their business and employees.

Because the survey collects a wealth of information on health plan offerings and cost sharing, other information collected is necessarily limited. Collecting detailed information on both health plans and a business's characteristics is also complicated by the fact that the respondent who is most knowledgeable about health insurance offerings will often be different from the respondent most knowledgeable about a business's output or receipts. However, the MEPS-IC draws its sample from a frame used for a wide variety of other business surveys, making it relatively straightforward to link to these other surveys for MEPS-IC sample members that are also included in other survey samples. The content of these other surveys can thus provide additional information about employer characteristics.

This paper presents results from carrying out such a linkage. It has two aims: first to describe methods, issues, and outcomes involved in matching data from the MEPS-IC to another set of business data collected by the U.S. Census Bureau, and second to present some simple results that illustrate the potential usefulness of the combined data.

Our primary focus here is on matching MEPS-IC data to the censuses of business, or the Economic Censuses (EC), which are an attractive source of data for a number of reasons. An issue in any linkage is whether the linked sample remains representative and large enough to be useful. Most businesses are included in the EC, so the matched sample is quite large. We will also argue that it is reasonably representative of the sectors covered by the EC.

The business censuses are collected every five years, for reference years ending in 2 and 7. Since the MEPS-IC data collection began with reference year 1996, we link data from the 1997 list sample of the MEPS-IC with data from the 1997 EC.<sup>1</sup> We use the EC

---

<sup>1</sup> Data for the 2002 Economic Censuses were collected in 2003, as were data for the 2002 MEPS-IC sample. The 2002 EC microdata files are not yet in final form at the time of writing, and so are not used here.

primarily as a source of data for measuring labor productivity, but draw on additional data that is available in the manufacturing sector.

We illustrate the usefulness of these combined data by examining productivity and earnings differences associated with health insurance offerings. Offering employer-sponsored health insurance is one business strategy to help attract and maintain a productive workforce. We expect more productive workers to have higher compensation, and for some of that increased compensation to take the form of health insurance benefits. After matching establishments found in the MEPS-IC with EC data, we calculate measures of earnings per employee and labor productivity for matched establishments. We can then compare differences in earnings and productivity between establishments that offer health insurance and those that do not.

The structure of health insurance benefits might also be related to productivity. To cope with rising premium costs, employers have restructured benefits to place more of the financial burden on employees, including raising employee contributions towards premium costs, and increasing deductibles and copayments. These changes may affect whether employees choose to take up health insurance, and how inclined they are to seek medical attention. If these changes affect employee health, a less healthy workforce might be less productive. The detailed benefits information collected by the MEPS-IC matched with the EC measures for productivity would allow researchers to look at these issues more carefully.

Section 2 of the paper discusses our data sources. Section 3 describes the matching methods and presents evidence on the success of the match—match rates and differences between the matched and unmatched parts of the samples. Section 4 discusses construction of the EC measures, section 5 presents some illustrative results from the match, section 6 discusses other data sources that might be usefully combined with the MEPS-IC, and section 7 concludes.

## **2. Data Sources**

The data we use here are drawn from two sources: the 1997 MEPS-IC and the 1997 EC. They can be linked together because sample members for both surveys are drawn from a common frame—the Business Register—so understanding the linkage requires some information about that data source as well.

### **2.1 Business Register**

The Census Bureau maintains its Business Register (BR), a list of all private employers in the U.S., to serve as a frame for its surveys of businesses. The unit of observation is an establishment or business location. This is also the primary sampling unit for both the IC

list sample and the EC. While the BR is a list of establishments, it also includes information that identifies which establishments belong to the same enterprise or firm.<sup>2</sup>

The BR uses data from a number of sources to continually update its list of currently active establishments, but the starting point is data from business tax filings that IRS provides to the Census Bureau. These data include an extract from IRS's list of all known business, organizational, and agricultural taxpayers, information on employment and payroll from quarterly payroll tax filings (Form 941), and information on business receipts or revenue from business income tax returns.<sup>3</sup>

These filings do not directly identify establishments, as there is not a one-to-one relationship between EINs and establishments. Taxes filed under a particular EIN may include more than one establishment, and large businesses may file taxes under more than one EIN. To establish and update the business list requires identifying the relationships between businesses, EIN entities and establishments for all multi-establishment businesses. In non-EC years, the Company Organization Survey collects information to update these relationships, and in EC years it is collected as part of the EC. The Company Organization Survey surveys all large multi-units on an annual basis, and samples smaller multi-units.

The BR contains information from administrative records on industry, employment, and payroll for most establishments, along with receipts or sales for single-unit establishments. Administrative data on receipts/sales are also available for multi-unit businesses, but they are reported for a business's EIN(s) and the Census Bureau does not try to allocate them to establishments for multi-units. We use some of these data as an alternative source for establishments that do not respond to the EC, as we describe further below.

## **2.2 Economic Census data**

EC questionnaires are tailored to a particular industry, but also collect a core set of information from businesses in all industries. We include data from the Censuses of Manufacturing; Retail Trade; Wholesale Trade; Services; Finance, Insurance, and Real Estate (FIRE); Construction; Transportation, Communications, and Utilities; and Mining.

---

<sup>2</sup>Until recently the BR was called the Standard Statistical Establishment List (SSEL), but it was renamed after undergoing a redesign for the 2002 census. An establishment is a single physical location where business is conducted or where services or industrial operations are performed. An enterprise (or "company") is comprised of all the establishments that operate under the ownership or control of a single organization. An enterprise may be a business, service, or membership organization; consist of one or several establishments; and operate at one or several locations. It includes all subsidiary organizations, all establishments that are majority-owned by the enterprise or any subsidiary, and all the establishments that can be directed or managed by the enterprise or any subsidiary.

<sup>3</sup>Census also receives some information on industry classification for new Employer Identification Numbers (EIN) from the Social Security Administration (taken from applications for EINs), and from the Bureau of Labor Statistics' Business Establishment List. Walker (1997) provides a much more detailed discussion of the SSEL.

Despite the use of the term ‘census’, data are not collected from all businesses. In most sectors, the smallest businesses do not receive census forms, but are included in the estimates (and data files) as ‘administrative records’ cases. For example, in manufacturing, 40% of establishments (all of those with <5 employees, and some establishments with 5-19 employees) were not mailed forms in 1997. The cutoffs were chosen so that non-surveyed establishments were expected to account for less than 3% of the value of shipments. The BR does have limited information (generally payroll, employment, receipts, and industry) on all cases from the tax filings that are used in its construction. For administrative records cases, data items that are not available from tax filings are generally imputed using the values of items available from administrative records. For example, the value of shipments for small manufacturers are imputed using industry average ratios of value of shipments to receipts in combination with an establishment’s receipts from tax data.

The exact rules about which establishments are asked to fill out forms and which are not varies by sector. In some sectors we know how to identify which cases have reported data and which have imputed data while in others we do not. The variables that we need to measure labor productivity are generally those that are available directly from administrative records—employment, payroll, and sales. Thus, while there are likely to be quality differences between the reported and administrative data, the values we use are not imputed. In manufacturing, where we bring in measures such as capital that have no administrative source, imputations are more of a concern, but we are able to identify imputed and non-imputed cases. We have examined our results excluding imputed data where it is an issue, and concluded that our general findings hold with or without imputations, so we present results that include imputed data here.

The methodology for the Census of Construction differs from that used by the other ECs in ways that require some comment. Because very small establishments play an important economic role in that sector, the construction census does not have a size cutoff but rather samples single-unit establishments below a given size, and uses smaller sampling probabilities for the smallest establishments. All multi-unit establishments are included in the sample (so it is a census for known multi-units). In most non-construction sectors, Census estimates aggregates from the EC by taking sums across the population of establishments, some of which have imputed data and some of which have reported data, but all of which are included in the data files. In construction, sampling weights are used with data from reporting establishments to estimate aggregates, and cases that are not sampled are not included in the construction files. Thus, using the 1997 Census of Construction files as is leads to significantly lower match rates for the MEPS-IC list sample in that sector.

We have administrative records data for the basic variables we need to compute labor productivity for single-unit establishments from the BR; the construction sample was drawn from the BR, and the cases that are missing simply because they were not sampled are all single-units. This allows us to put together a supplementary data set consisting of all construction cases from the BR that are not in the Census of Construction files, which should closely approximate a Census of Construction with administrative records

included. This will not be an exact replica of such a file because of some differences in timing—the end of year BR files that we have would not exactly match the BR used at the time the Census of Construction sample was drawn because the BR is updated continuously. The sample includes all known multi-unit establishments, so we do not add any multi-unit establishments in doing this.

### **2.3 List sample of the MEPS-IC**

The MEPS-IC is an annual collection of information on employer-provided health insurance. The list sample is a stratified random sample of approximately 25,000 private establishments per year that have one or more paid employees.<sup>4</sup> Stratification variables include state and establishment and enterprise employment. From 1996 to 2002, sample sizes within strata were set to support national estimates and state-level estimates for most states. Larger samples were rotated among the smallest states and the District of Columbia during these years to allow for state-level estimates from these areas at least once every four years. Details on the strata and sample allocations can be found in Sommers (1999).<sup>5</sup>

The MEPS-IC collects data on organizational characteristics of the establishment, its workforce, and fringe benefit offerings.<sup>6</sup> Some questions, including the number of years in business and retiree health insurance offerings, are asked about the enterprise rather than the establishment. Organizational characteristics include size, type of ownership, and business activity. Workforce characteristics include the percent of the workforce that is female, 50 years of age and older, unionized, and earning low/medium/high wages. Non-health fringe benefit information includes yes/no questions about vacation, sick leave, and pensions. Much of this data is also collected from employers who do not offer health insurance benefits.

Establishments that do offer health insurance to their employees complete survey questionnaires for up to four health plans.<sup>7</sup> Data is collected on a plan's provider arrangement (e.g., fee-for service, health maintenance organization) and requirements for referrals. For single and family coverage (and employee-plus-one coverage in some years), respondents are asked to report on premiums, and employer and employee contributions. Information on deductibles, copayments, and coinsurance is also collected.

---

<sup>4</sup>There is also a list sample of governments, which we do not use here. The MEPS-IC data collection has included employers of members of the MEPS-Household Component sample in some years. Those employers are not included in our sample either.

<sup>5</sup> In 2003, the sampling design was changed to take into account other factors found to be closely associated with health insurance offers, enrollment, and insurance contributions. See Sommers (2004) for details.

<sup>6</sup> Copies of the MEPS-IC survey instruments are available online at <http://meps.ahrq.gov/survey.htm#icsurveyinstrument>.

<sup>7</sup> If an establishment offers five or more plans, the employer is asked to report on the three largest plans and a randomly selected fourth plan.

## 2.4 Scope of Data Sources

The scope of the BR and of the MEPS-IC is somewhat larger than that of the EC. The EC excludes the following activities that are included on the BR:

- agriculture, forestry, and fishing
- primary and secondary schools, colleges and universities
- some specialized financial activities
- government activities (though business activities carried out by government-owned enterprises, such as liquor stores or hospitals, are included)
- rail transportation, airlines, and postal service activities
- religious, labor, political, and private household activities.

The MEPS-IC includes most of these sectors in its sampling. However, government activities are sampled from a list of governments rather than from the SSEL, and are not included in the sample of private employers that we use here.

## 3. Matching

### 3.1 Methods

Our primary method for matching cases between the 1997 MEPS-IC list sample and the 1997 economic censuses is to simply use the Census Bureau identification number that comes from the BR and is included in both sets of files. Where there were duplicates, we eliminated one of the duplicates by dropping cases with zero reported payroll or missing SIC codes, and by dropping cases with SIC codes that were at odds with the file that they were in.<sup>8</sup>

The primary Census Bureau business identifier changes when an establishment changes ownership or when it is reclassified from single-unit to multi-unit or vice-versa, which complicates finding matches. For example, an establishment that appeared on the BR as a single-unit when the MEPS-IC sample was drawn might later return a census form that indicated that it was part of a multi-unit.<sup>9</sup> It would appear with a new Census ID in the

---

<sup>8</sup> Matching to each of the EC files separately, we end up with a small number of cases that match to more than one EC file. Generally this happens because of the switch in industry coding systems from SIC to NAICS that was part of the 1997 economic census. While individual establishments were only asked to fill out one EC form, aggregate estimates were created under both coding systems, and the files we have sometimes include all cases that would be in a sector under either classification system. For example, all retail bakeries were included in retail trade under the SIC system. Bakeries that both make and sell baked goods for immediate consumption (e.g. doughnut and cookie shops) were moved to services under NAICS. Those that make and sell baked goods at the same location (but not necessarily for immediate consumption) were moved to manufacturing, while those that do not produce baked goods on location remained under retail trade. Establishments in the second category of retail bakeries contributed data to SIC estimates for retail trade, but NAICS manufacturing estimates. As a result, data for the establishments show up in both the retail trade and manufacturing files—we kept the data from the retail trade file.

<sup>9</sup> This might occur because of a real change (a single-unit opens a second location, or is acquired by a multi-unit) or because the original information was incorrect. Businesses that are too small to be sampled by the Company Organization Survey do not have an opportunity to be classified as a multi-unit until they



EC files, but would probably retain the single-unit identifier in the MEPS-IC files, and so would not be identified as a match. The BR files also contain an identifier that links physical locations over time even when the primary identifier changes. We handle unmatched cases by pulling all records from the BR that match the MEPS-IC physical location identifier, and using an alternative identifier, if found, to match to the EC files. This secondary match accounts for about two percent of matched establishments.

### 3.2 Match rates

Table 1 presents the basic results of the match. We succeed in matching about 88 percent of the 1997 MEPS-IC list sample to an establishment record in the EC. As mentioned above, the scope of the EC is somewhat narrower than that of the MEPS-IC, and our matched sample is necessarily limited to the narrower scope. Differences in scope account for about 70 percent of the non-matches, with the majority of out-of-scope cases accounted for by agriculture; administrative or auxiliary establishments;<sup>10</sup> and private schools. We successfully match about 96 percent of MEPS-IC establishments that are in-scope to the EC.

TABLE 1 ABOUT HERE

Table 2 presents estimates of the fraction of the aggregate employment represented in the matched and unmatched portions of the MEPS-IC sample. Out of scope establishments represent about 10 percent of employment, while unmatched in-scope establishments represent about 4 percent of total employment. If we consider only in-scope employment, un-matched cases represent about 5 percent of employment. Both out-of-scope and unmatched establishments have above average employment, as their shares of employment exceed their shares of establishment counts.<sup>11</sup>

TABLE 2 ABOUT HERE

One reason we might not find a match in the EC files is non-response to the survey, though imputation of missing data based on administrative records means that non-response has a bigger effect on the quality of the data we have than on its presence in our database. Some cases in the MEPS-IC sample do not have an industry code from the BR. Even very small establishments with missing industry codes are asked to complete a

---

complete a census form, so the timing of when they make the transition to being a small multi-unit is really only captured in 5 year intervals.

<sup>10</sup> Auxiliary establishments are those primarily providing services to other establishments of the same business, rather than to outside customers. There is a Census of Auxiliaries, but the sales measure excludes the value of transactions with other parts of the company and so would not be appropriate for measuring labor productivity. Of the 308 cases that have flags indicating that they are auxiliaries in the MEPS-IC files, 72% match to the 1997 Census of Auxiliaries, and another 18% match to other EC files.

<sup>11</sup> The estimates in Table 2 are constructed using the product of the survey weight and establishment employment as weights. Thus differences between Tables 1 and 2 are due to both differences in employment and in survey weights (which depend on sampling rates and response rates). Using only the survey weights to construct an analogous table yields estimates quite close to those in Table 1, so the differences between Tables 1 and 2 are primarily due to employment differences between matching and nonmatching establishments.

classification questionnaire, and their response would be used to assign a code on the BR, so the absence of an industry code probably indicates that the classification questionnaire was never returned.

Even after trying to fill in the construction sample in the EC files with data from the BR, construction establishments account for almost one-third of non-matches. These cases all appear on the BR but have missing data on employment or sales, so their administrative records do not provide any useful information and we treat them as non-matches. Thus one weakness of the combined data is that construction will be under-represented in the matched file.

### **3.3 Comparison of matched and unmatched samples**

An important question is whether the matched sample is reasonably representative of the population described by the full sample of MEPS-IC respondents. As a starting point, Table 3 presents mean characteristics for the matched and unmatched samples. We know that the matched sample cannot represent the parts of the economy that are not in-scope to the EC, so we exclude those from the comparison. There are only small differences in insurance characteristics between the two samples, which is somewhat surprising given how much some of the other characteristics differ by match status. Matched cases are from older firms, are less likely to be part of a multi-unit, and include fewer very small and very large establishments. The industry distributions are also quite different: manufacturing, retail trade, and services are over-represented in the matched sample, while construction is dramatically under-represented. Matched establishments have larger shares of part-time and female workers than do unmatched establishments, while they have much lower shares of temporary or seasonal employees.

TABLE 3 ABOUT HERE

To look more closely at how these differences interact, we run a series of logistic regressions to identify the characteristics associated with a high probability of matching. Ideally, the non-matches would be randomly selected from the in-scope cases, in which case the logistics should have primarily coefficients that did not differ significantly from zero.

TABLE 4 ABOUT HERE

Table 4 presents our logistic results. All three columns include controls for industry, state, firm age, multi-unit status, establishment size, and whether or not the establishment offers health insurance to its employees. The same sample is used in each specification. The second and third columns include a control for percent enrolled which is set to zero for establishments without health insurance. We use imputed values for this variable when it was not reported. The third column adds controls for worker characteristics. Because we do not have imputed versions of most of those variables, we handle missing values by adding dummy variables for missing values and set the variables themselves to

zero where missing. The estimates presented are average predicted marginal effects,<sup>12</sup> along with bootstrapped standard errors for those effects. Marginal effects for dummy variables were computed as the difference between the prediction with the dummy set to one and with it set to zero.

Given that the matched data is of interest mostly for answering questions about health insurance, selectivity on health variables is of primary interest. The coefficient on the health insurance dummy is significantly different from zero in the first two columns but in all specifications is very small—controlling for employer characteristics, the match rate for establishments that offer health insurance is estimated to be about 1 percent higher than that for establishments that do not. Results in the second column show that there is no significant relationship between enrollment rates and match rates among establishments that offer health insurance.

When we include worker characteristics in the third column, the health insurance dummy becomes insignificant, but this is primarily because the standard error increases rather than because the coefficient changes much. None of the added workforce characteristics appears to have an important relationship to the match rate, but their addition generally makes the estimates less precise. Most of the variables that are included in both columns (2) and (3) have very similar coefficients whether or not the workforce characteristics are included.<sup>13</sup>

We do find evidence of selection effects when looking at establishment size, multi-unit status, and industry. Our prior thinking was that we would have the most trouble matching small, new, single-unit establishments because they would be the most likely to turnover or change ownership in a short period of time. We do find that smaller and younger establishments are less likely to match, but controlling for size, multi-unit establishments are significantly less likely to be matched than single units. In fact, if we drop the multi-unit dummy from the specification, the health insurance dummy coefficient becomes negative and significant because of the strong correlation between health insurance offering and multi-unit status.

The reported industry effects are relative to manufacturing, which has one of the highest overall match rates—over 96 percent of cases were matched. The industry measure here is based on the categories collected for the MEPS-IC because that is available even for unmatched cases. In excluding out-of-scope cases, we use SIC codes from the BR that would determine which Economic Census a case would be routed to. While we excluded cases that would not have been sent to the EC because they were classified as in agriculture on the BR, we still have a few cases that classify themselves as agriculture on their MEPS-IC form. Those cases, unsurprisingly, have very low match rates, but we do not think of that as an important concern.

---

<sup>12</sup>That is, we compute marginal effects at  $\bar{X}_i$  for each establishment  $i$ , and then report an average across establishments. The standard errors are computed by carrying this out for each bootstrapped sample.

<sup>13</sup> Adding the workforce characteristics to the specification in column (1) (rather than that in column (2), as is presented in the table), we reach essentially the same conclusion about the effects of the workforce characteristics.

The other industries with large negative effects are mining and construction. Mining is a very small industry to begin with, so the low match rate is of little practical importance, but construction is more of a concern. These results include our supplemental sample of construction single-units (described in section 2) as matched cases—without them the construction effect is much larger—but we still find a much lower rate for construction. While we can find all of the original cases on the BR, these ‘unmatched’ cases represent establishments that have zero values for administrative data fields on the BR and so have no useful data to match in.

While we do not report state effects here, they are included in the estimation. There do not appear to be important differences in match rates across states. With California as the omitted category, at most two states have significant marginal effects in these specifications and even those differences are fairly small.

We find the results of the match encouraging. We match most cases, and do not find any evidence of large selectivity effects associated with health insurance offerings among the cases that we do not match. A caveat is that the small health insurance effects are found when controlling for size and industry, which do show evidence of selectivity. Overall, establishments that offer insurance have a two percent higher match rate. Controlling for size accounts for about half of that gap because larger establishments are more likely to match and more likely to offer health insurance.

#### **4. Auxiliary measures from the Economic Censuses**

The match to the EC files provides us with several new measures that are useful in considering employers’ decisions about health insurance offers. Here we describe how we construct these measures, and in the next section we illustrate their usefulness.

##### **4.1 Labor productivity**

The measures that we have available to us from all of the Economic Censuses are limited, so our general measure of labor productivity is simply sales (or, in some sectors, receipts) per employee. In creating per-employee measures, we use the figure reported for the number of employees at the establishment for the pay period including March 12<sup>th</sup>, 1997, because that is what is available for all of the ECs.

In manufacturing we can also measure value added, derived by subtracting the cost of materials, supplies, containers, fuel, purchased electricity, and contract work from the value of shipments. That value is then adjusted by the addition of value added by merchandising operations plus the net change between the beginning and end of year in inventories of finished goods and work-in-process. For manufacturing, we present both value added per employee and sales per employee as measures of labor productivity.

## **4.2 Workforce characteristics**

For the matched sample as a whole, we can use the EC data to construct average earnings per worker by simply dividing payroll by the number of employees. In manufacturing we have some additional measures. The number of employees and total payroll are reported separately for production and non-production workers, and for production workers we also have total number of hours of work. In addition to average earnings per worker by production status, we use these variables to create the percentage of the workforce involved in production and average hours per production worker.

## **4.3 Fringe benefit spending**

In manufacturing, we also have measures of total employer spending on fringe benefits. This includes the cost of both voluntarily provided benefits like health insurance and pensions and the cost of taxes that finance required benefits such as workers' compensation and Social Security. Using this measure we calculate average per-worker spending on fringe benefits and the ratio of fringe benefit spending to payroll.

## **5. Employer health insurance offers and labor productivity**

In Table 5 we present mean values of these measures for employers that offer health insurance and those that do not. We do this both for the overall sample and for manufacturing alone. The table also includes means for a set of establishment characteristics from the MEPS-IC that are useful in understanding what the EC measures might mean.

The estimates for the overall sample make clear that establishments that offer health insurance also have employees with substantially higher earnings. The productivity differences between the two groups are even larger than the differences in pay, which helps to explain why we see a positive relationship between earnings and benefits. Average earnings per employee are 57 percent higher for the insurance group, while productivity is about 73 percent higher. The means for the first several MEPS-IC measures make clear that this is probably due to no-insurance employers having a combination of more part-time employment and more low wage employees than employers with insurance. Differences in temporary or seasonal work could also potentially account for lower average earnings, but those rates do not differ much by insurance status.

Our findings appear to support the idea that offering employer-sponsored health insurance is a business strategy that helps to attract and maintain a more productive workforce. Employees at establishments that offer health insurance earn more and are more productive. Observable worker characteristics that are correlated with productivity—such as full-time status—undoubtedly explain part of this relationship, but effects of other characteristics probably also play a role.

The manufacturing estimates give more detail on the differences between employers that offer insurance and those that do not. Earnings per employee are quite similar for employers with insurance and those without. Production and non-production workers both earn more on average in establishments that offer insurance, while the fraction of the workforce accounted for by production workers does not differ much with insurance status. The gap in labor productivity is even larger in manufacturing, and the gap in productivity holds up when we use our preferred measure of value added per employee. Employers that offer health insurance have much higher levels of capital intensity, which helps to account for the higher levels of labor productivity. While there are substantial average differences between the manufacturing and non-manufacturing samples in the prevalence of part-time work and in earnings distributions, the differences within sample associated with offering insurance are quite similar for the two sets of estimates.

Looking at the fraction of establishments offering pensions makes clear that it is fairly rare for an employer to offer a pension without also offering health insurance but health insurance without a pension is not so unusual. Overall employer costs for fringe benefits are substantially larger for insurance-offering establishments, which is unsurprising given that the overall measure would include spending on the employer portion of health insurance contributions. What is surprising is how little difference there is in the ratio of fringe costs to payroll. Given the low rates of pension provision among no-health-insurance employers, it must be that most of the cost of fringes for these employers is for government programs rather than voluntarily provided benefits.

## **6. Alternative sources of supplemental data**

We have illustrated that the EC can provide several useful auxiliary measures for the MEPS-IC, but this is not the only potential source. Census collects many other business surveys that have narrower scopes than the Economic Censuses. With a few exceptions, these other surveys also are based on samples drawn from the BR, so it is relatively straightforward to match data across files where the survey samples overlap. Here we discuss other sources of data that might be useful in combination with the MEPS-IC, including the likely composition and sample size of the matched data.

Census conducts separate annual surveys of businesses in manufacturing, retail, wholesale, and services. While the content of these surveys varies across sectors, they have several common features that are worth noting. Each is designed to provide timely measures of economic activity, and so each disproportionately samples large businesses that account for the majority of economic activity. Each selects a sample of businesses shortly after information from the EC is incorporated in the BR and then follows that sample until a new panel is selected five years later.

### **6.1 Annual Survey of Manufactures (ASM)**

The *Annual Survey of Manufactures (ASM)* surveys approximately 55,000 manufacturing establishments. Of the 366,000 establishments in manufacturing in the 1997 EC, about

200,000 are potentially eligible for inclusion in the ASM, with the remaining 166,000 excluded from data collection based on minimum size cutoffs.<sup>14</sup> About 16,600 of the establishments in the survey are included with certainty because of their large size (or the large size of the firm that they belong to.) Another 40,000 establishments are sampled with probability roughly proportionate to size. The ASM provides annual measures similar to those drawn from the Census of Manufactures in Table 5—for example, value added, capital expenditures, and overall costs of fringe benefits.

The MEPS-IC averages about 2,400 cases per year in manufacturing, of which about 43 percent also appear in the ASM sample, giving a matched sample of roughly 1,000 employers per year.<sup>15</sup> Matched establishments are on average much larger than those not matched: those in both the ASM and MEPS-IC average close to 500 employees per establishment, while those in the MEPS-IC only have about 60 employees on average. Large establishments are more likely to offer health insurance, and manufacturers are both larger and more likely to offer insurance given their size than employers in most other industries. Because of this, the matched sample is of limited use for considering questions about whether or not employers offer insurance—over 95 percent of employers in the matched sample do. However, the sample may be of more interest for considering differences across employers in health plan features such as cost sharing or retiree coverage.

## 6.2 Annual Trade and Service Surveys

Linkage of the MEPS – IC establishments with the *Service Annual Survey*, *Annual Retail Trade Survey*, and *Annual Trade Survey* could similarly provide additional information for service and trade industries. For these surveys, there is the added complication that the unit of analysis is an industry segment within a company, so for multi-units measures are aggregated above the establishment level. For example, if a multi-unit has ten establishments—five each in two different industries—in these surveys the company would have at most two records, one for each of the two industries. Either record could match to up to five establishments in the MEPS-IC (though in most cases would match to fewer than five as it would be unlikely that all five were in the MEPS-IC sample). Thus, researchers must consider the proper level of aggregation in combining information from these surveys with the MEPS-IC establishment data. Many decisions about health insurance are made at the firm level in multi-units, so the more aggregated unit of observation may be quite suitable to many analyses.

The *Service Annual Survey* collects information on approximately 45,000 businesses in the personal, business, automotive, amusement and recreation, social, health, and other professional services. Linkage with this data would provide information on operating revenues and operating expenses for service establishments in the MEPS – IC. The

---

<sup>14</sup> The size cutoffs vary by detailed industry. Administrative records are used for these cases to impute values for survey variables for use in producing aggregate estimates. These cases generally account for less than 3 percent of the estimated aggregate value of shipments.

<sup>15</sup> These figures vary by year. These are rounded numbers based on the 1996 and 1998-2001 MEPS-IC and ASM samples. (In census years, the ASM is wrapped into the Census of Manufactures, so we omit 1997.)

*Annual Retail Trade Survey* includes about 22,000 retail businesses and provides measures of the dollar value of retail sales, sales taxes collected, inventories, and account receivables' balances. Data on inventories, dollar value of annual sales, purchases, and gross margins is available from the *Annual Trade Survey* for approximately 6,500 wholesale businesses. We discuss the likely size of samples matched to the MEPS-IC below.

### **6.3 Business Expenses Survey (BES)**

The *Business Expenses Survey (BES)*, called the *Business Expenditures Survey* in 1997) collects information every five years (i.e., years ending in the digit "2" or "7") from the combined samples of the *Service Annual Survey*, *Annual Retail Trade Survey*, and *Annual Trade Survey*. It collects data on revenues and operating expenses for service and trade businesses, including payroll and fringe benefits costs similar to those collected in the *Census of Manufactures*. We carried out a preliminary match of the 1997 BES to the 1997 MEPS-IC to get a feel for approximate matched sample sizes for both the BES and the annual services and trade surveys. Different decisions about how to handle aggregation issues would yield somewhat different numbers, but when we match data from the two surveys at the company/2-digit SIC level of aggregation, about ten percent of establishments in the MEPS-IC sample find a match in the BES sample. The matched sample includes a disproportionate number of large establishments, and so accounts for about 20 percent of employment of MEPS-IC sample establishments. The distribution of the sample across industrial sectors should also approximate what one would obtain by matching the annual service and trade surveys to the MEPS-IC. About half of the matched sample establishments are in retail trade, about 40 percent in services, and the remaining 10 percent in wholesale trade. While retail trade accounts for the largest number of establishments, services accounts for about three-quarters of matched employment.

### **6.4 Annual Capital Expenditures Survey**

The *Annual Capital Expenditures Survey* collects data at the firm level, asking for total assets, sales/revenue, and capital investments in structures and equipment by nonfarm businesses for each industry in which a sampled firm had activity. In 1997, the sample included about 35,000 employers, while more recent samples have been larger. Slightly less than 20 percent of the 1997 MEPS-IC establishments belong to firms that are also in the ACES sample for that year. Like the surveys discussed above, the sample disproportionately includes large firms, and as a result the matched establishments account for almost three-quarters of employment among MEPS-IC establishments. The ACES data may be of particular use to researchers interested in examining the relationship between productivity differences and health insurance offerings, as it provides annual estimates of capital expenditures much like the information from the *Census of Manufactures* that we used in our analysis.



## 7. Conclusions

Our match of the 1997 MEPS-IC with the 1997 EC was successful in terms of both match rates and the representativeness of the matched sample. Ninety-six percent of the establishments in the MEPS-IC that were in scope to the EC were matched, while 95 percent of total in scope employment was accounted for by matched establishments.

We run a series of logistic regressions to look at the relationship between the probability of matching and establishment characteristics such as industry, size, and health insurance characteristics. We also examine specifications that include the percent enrolled and workforce measures. The results indicate that the unmatched sample does not have large selection biases with respect to insurance offerings. Establishments that offer insurance do have a two percent higher match rate than establishments that do not offer insurance, but this is largely explained by the fact that larger establishments are more likely to match and more likely to offer insurance.

Given the high match rate and general representativeness of the matched MEPS-IC /EC sample, this data is a useful source of information for a variety of analyses of employer-sponsored health insurance. Based on simple comparisons of means by insurance status, we find that earnings and productivity are substantially higher in establishments that offer health insurance. This in part reflects more part-time and low wage employment at establishments that do not offer insurance. While we do not pursue this here, linking measures of labor productivity, compensation, and payroll with the MEPS-IC data on health plan characteristics would also allow examination of relationships between the structure of health insurance benefits (e.g., premiums, contributions, enrollment) and productivity.

One weakness of the matched sample is that establishments in the construction industry are underrepresented, which may be a concern for some analyses of the health insurance offering decision. However, given that construction has lower rates of health insurance coverage than most other industries, this may be less of a concern in studying characteristics of health plans among establishments that offer insurance such as eligibility, enrollment, or plan types and premiums.

Matching the 1997 MEPS-IC with the 1997 EC is just one possible combination. It will also be possible to match in the 2002 EC when those data become available. As this paper has discussed, matches are possible with a number of other Census Bureau datasets depending on the focus of the analysis. By adding more detailed information on employers and their workforces, matching economic data from other surveys can be valuable in expanding the ability of the MEPS-IC to address concerns about the access and affordability of health insurance.

## References

- Sommers, John P. 1999. "List Sample Design of the 1996 Medical Expenditure Panel Survey Insurance Component." Rockville (MD): Agency for Health Care Policy and Research. [MEPS Methodology Report No. 6](#). AHCPR Pub. No. 99-0037.
- \_\_\_\_\_. 2004. "Updates to the Medical Expenditure Panel Survey Insurance Component List Sample Design, 2004." Agency for Healthcare Research and Quality [Working Paper No. 04008](#).
- Walker, Ed. 1997. "The Census Bureau's Business Register: Basic Features and Quality Issues." *1997 Proceedings of the American Statistical Association*.

**Table 1: Outcome of Match to Economic Census for MEPS-IC Cases**

	Number	Percent of Total	Percent of In-Scope
<b>Out-of-Scope Establishments</b>			
Agriculture	853	3.3%	
Administrative/Auxiliaries	308	1.2%	
Schools	177	0.7%	
Other	<u>761</u>	<u>3.0%</u>	
Total Out-of-Scope	2,099	8.1%	
<b>In-Scope Establishments</b>			
Matched	22,782	88.5%	96.3%
Unmatched			
Construction	251	1.0%	1.1%
Missing industry code	158	0.6%	0.7%
Other unmatched	<u>465</u>	<u>1.8%</u>	2.0%
Total Unmatched	874	3.4%	3.7%
Total In-Scope	23,656	91.9%	100.0%
<b>Total MEPS-IC List Sample</b>	25,755	100.0%	

**Table 2: Outcome of Match to Economic Census for MEPS-IC Cases: Employment**

	Number	Percent of Total	Percent of In-Scope
<b>Out-of-Scope Establishments</b>			
Employment	10,861,562	10.2%	
<b>In-Scope Establishments</b>			
Matched Employment	91,447,653	85.6%	95.3%
Unmatched Employment	<u>4,546,811</u>	4.3%	4.7%
Total In-Scope Employment	95,994,464	89.8%	100.0%
<b>Total MEPS-IC List Sample</b>	106,856,026	100.0%	

**Table 3: MEPS-IC Sample Means by Match Status**

	Matched	Unmatched
Offers insurance	.54	.52
Percent enrolled	.37	.38
Pension Offered	.30	.28
Firm less than 5 years old	.30	.38
Multi-unit	.23	.41
Establishment size		
Less than 5 employees	.48	.55
5-9 employees	.22	.17
10-24 employees	.18	.13
25-99 employees	.09	.11
100-499 employees	.03	.02
500 or more employees	.003	.005
Industry		
Manufacturing	.07	.02
Construction	.09	.34
Retail trade	.25	.11
Wholesale trade	.08	.07
Personal services	.04	.02
Business services	.07	.04
Other services	.25	.13
Finance, insurance, real estate	.11	.11
Transportation, communication, utilities	.04	.07
Mining	.01	.01
Agriculture, fishing, and forestry	.003	.07
Employee characteristics		
Percent part-time	.25	.19
Percent temporary or seasonal	.13	.22
Percent low wage (<\$6.50/hr)	.19	.19
Percent high wage (>\$15.00/hr)	.21	.22
Percent women	.45	.31
Percent 50 or older	.20	.22
Union (=1 if >25% of employees belong to union)	.03	.06
Region		
West	.22	.20
South	.34	.35
Midwest	.24	.29
Northeast	.21	.16

Notes: Excludes out-of-scope cases. N=22,782 matched cases and 874 unmatched cases. Means for employee characteristics exclude cases with missing values.

**Table 4: Explaining Probability of a Match**

(Max likelihood logit estimation, dependent variable=1 if matched, marginal effects reported)

	(1)	(2)	(3)
Offers insurance	.010** (.004)	.019** (.007)	.016 (.011)
Percent enrolled		-.011 (.006)	-.010 (.010)
Firm less than 5 years old	-.010** (.003)	-.010** (.003)	-.009 (.032)
Multi-unit	-.085** (.007)	-.085** (.007)	-.082* (.033)
5-9 employees	.009** (.003)	.008** (.003)	.008* (.004)
10-24 employees	.013** (.003)	.013** (.003)	.013** (.003)
25-99 employees	.009** (.003)	.008* (.003)	.008* (.004)
100-499 employees	.014** (.004)	.014** (.003)	.014** (.004)
500 or more employees	.008 (.006)	.008 (.006)	.009 (.006)
Construction	-.186** (.022)	-.192** (.023)	-.186** (.045)
Retail trade	.008 (.005)	.005 (.005)	.007 (.009)
Wholesale trade	-.011 (.008)	-.012 (.009)	-.011 (.010)
Personal services	-.010 (.011)	-.013 (.011)	-.012 (.012)
Business services	-.012 (.009)	-.014 (.010)	-.013 (.142)
Other services	-.011 (.007)	-.013 (.007)	-.013 (.008)
Finance, insurance, real estate	-.018* (.008)	-.020* (.009)	-.019* (.009)
Transportation, communication, utilities	-.023* (.010)	-.023 (.012)	-.019 (.067)
Agriculture, forestry, fishing	-.452** (.059)	-.461** (.057)	-.442** (.155)
Mining	-.109* (.047)	-.108* (.043)	-.106 (.058)

Note: Percent variables all specified as decimals (e.g., 50%=.50). Industry measures based on MEPS C060, hence the presence of agriculture. Also includes state dummies and dummies for missing values of employee characteristics. \* Significant at .05 level, \*\* Significant at the .01 level

**Table 4 continued: Explaining Probability of a Match**

	(1)	(2)	(3)
Percent part-time employees			.001 (.004)
Percent low wage (<\$6.50/hr)			-.007 (.008)
Percent high wage (>\$15/hr)			-.004 (.008)
Percent women			.006 (.005)
Percent 50 or older			-.003 (.005)
Union dummy: 25% or more union employees			.008 (.011)

Note: Percent variables all specified as decimals (e.g., 50%=.50). Industry measures based on MEPS C060, hence the presence of agriculture. Also includes state dummies and dummies for missing values of employee characteristics. \* Significant at .05 level, \*\* Significant at the .01 level

**Table 5: Sample Means by Insurance Status**

	Total Matched Sample		Manufacturing	
	Offer Insurance	No Insurance	Offer Insurance	No Insurance
<b>EC measures</b>				
Earnings per employee (\$)	30,371	19,390	30,298	19,826
Average earnings, production workers (\$)	—	—	25,860	16,766
Average earnings, non-productions workers (\$)	—	—	45,852	27,919
Production workers as a percentage of workforce	—	—	.70	.69
Labor productivity (\$ sales or receipts per employee)	207,211	120,061	184,834	85,213
Value added per employee (\$)	—	—	91,968	46,138
Capital intensity (\$ book value per employee)	—	—	74,542	30,400
Average hours worked per year by production workers	—	—	2,108	1,480
Employer cost of fringe benefits per employee (\$)	—	—	7,384	4,451
Fringe costs divided by payroll	—	—	.24	.23
<b>MEPS-IC measures</b>				
Percent of employees part-time	.16	.36	.06	.21
Percent of employees temporary or seasonal	.15	.16	.15	.16
Percent of employees earning wage <\$6.50/hr	.13	.28	.08	.24
Percent of employees earning wage >\$15.00/hr	.28	.12	.24	.11
Percent of employees female	.45	.48	.31	.35
Percent of employees over 50	.18	.22	.18	.21
Firm less than 5 years old	.16	.40	.09	.41
Multi-unit	.46	.04	.46	.02
Establishment size				
Less than 5 employees	.18	.63	.06	.54
5-9 employees	.19	.21	.08	.22
10-24 employees	.22	.12	.16	.14
25-99 employees	.25	.04	.33	.09
100 or more employees	.15	.01	.37	.01
Union (=1 if >25% of employees belong to union)	.06	.01	.13	.01
Pension offered	.49	.11	.55	.08