Longitudinal Employer - Household Dynamics

Technical paper No. TP-2003-05

Describing the Form 5500-Business Register Match

Date	:	January 2003
Prepared by	:	Anja Decressin, Kristin McCue, and Martha Stinson
Contact	:	U.S. Census Bureau, LEHD Program FB 2138-3 4700 Silver Hill Rd. Suitland, MD 20233 USA

This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics Program (LEHD), which is partially supported by the National Science Foundation Grant SES-9978093 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging, and the Alfred P. Sloan Foundation. The views expressed herein are attributable only to the author(s) and do not represent the views of the U.S. Census Bureau, its program sponsors or data providers. Some or all of the data used in this paper are confidential data from the LEHD Program. The U.S. Census Bureau is preparing to support external researchers' use of these data; please contact Ronald Prevost (Ronald.C.Prevost@census.gov), U.S. Census Bureau, LEHD Program, FB 2138-3, 4700 Silver Hill Rd., Suitland, MD 20233, USA.

Describing the Form 5500-Business Register Match

Anja Decressin, Kristin McCue,
and Martha Stinson

January 2,2003
(PREMLININARY)

1.Introduction

The accurate measurement of employer-provided health and pension plan coverage is critically important for policy makers seeking to make national decisions that will affect the well-being of the current and future infirm and elderly population. However, current information could be vastly improved by improving estimates of coverage and by bettering the firm level information currently available. We will exploit an underutilized resource –the 5500 files – to achieve this improvement. The 5500 file, which contains data from annual returns filed annually with the Internal Revenue Service by all employers covered by ERISA, provides information on employee benefit plans, including data on different types of pension and health insurance plans.

The 5500 file can improve coverage estimates because it helps reconcile inconsistencies in current data collection. Two major Census Bureau surveys -- the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) -- produce substantially different estimates of the uninsured.[1] Similar problems surround the collection of private pension information. A large part of the problem is due to that fact that respondents are typically not able to give accurate responses. For example, the health insurance information collected by the CPS is meant to represent coverage in the prior year, yet it is possible that the CPS actually provides a point-in-time estimate for some respondents. The 5500 file may help shed some insight on what the CPS actually seeks to collect. In addition, the CPS and SIPP each use a different set of questions to collect health insurance information, and the 5500 file may help to determine which set of questions produces a more accurate estimate of employer-provided health insurance. Finally, the 5500 file could be used to add information to the American Community Survey (ACS) to provide broader measures of health and pension coverage.

The 5500 file could additionally be used to enhance the Census Bureau's Business Register. Matching the 5500 file to the Business Register by EIN could add important new information on the types and geographic location of firms that offer pension and health benefit coverage. This operation could, in turn, substantially enhance the production of County Business Pattern statistics.

However, there are a number of important and difficult quality and coverage issues that need to be addressed before full use can be made of these data. The next sections provide background on employer-sponsored benefits and detail both the potential projects and the detailed data work that could be undertaken on a pilot basis to evaluate the possible contribution of the 5500 file to the statistical data infrastructure.

[1]Bennefield (1996).

2. Background

Employer-sponsored health insurance and pension plans play key roles in the economy. About two-thirds of non-elderly Americans received health insurance through employers in 1998 (Holahan and Kim 2000), and about four-fifths of full-time workers in medium and large firms participated in pension plans in 1997 (U.S. Bureau of Labor Statistics 1999). These pension plans provide retirement security for many workers in later life. In 1998, for example, 45 percent of Americans ages 65 and older received pension income from past employers or were married to someone who did. In addition, private pensions accounted for 26 percent of income for retirees with pensions (Social Security Administration 2000).

Health and pension benefits also have important effects on the labor market. The availability of employer-sponsored health benefits can discourage turnover by workers (Madrian 1994). Many defined benefit pensions also provide workers with incentives to remain with their employers until they are eligible to collect pension benefits and then encourage retirement once benefits become available. More generally, employee benefits are an important and growing component of compensation and an accurate measure is necessary to gauge the returns to work.

Despite the importance of employee benefits, there is still much we do not know about why firms offer health insurance and pension coverage and how they set plan parameters. Improving our understanding of these issues is critical to evaluating proposals to change the current system. Researchers have offered possible explanations for why employers offer pensions to their workers, but there is no consensus about which factors are most important to the sponsorship decision. For example, some have argued that pensions provide an inexpensive way for firms to compensate higher-wage workers because pensions are a tax-free fringe benefit (Munnell 1982). Ippolito (1985) has suggested that firms offer pensions in order to elicit cooperative behavior from labor unions. Other economists maintain that a primary function of pensions is to bond workers to their employers (Allen, Clark, and McDermed 1993; Ippolito 1991), or that they create retirement incentives that induce older, typically high wage workers to leave the firm (Kotlikoff and Wise 1989; Mitchell and Fields 1982). The retirement incentives created by DB pension plans may be particularly important because internal labor markets and age discrimination laws may limit the ability of employers to dismiss older workers.

Pensions may also be offered because they can enable employers to elicit effort from workers when productivity is difficult to monitor (Hutchens 1987, 1989; Lazear 1979, 1983). Workers who are dismissed after being caught shirking will incur large penalties if their career compensation packages were back-loaded, paying them less than their marginal products early in the career but more than their marginal products late in the career. The threat of severe penalties can deter workers with rising compensation profiles from shirking. Pension plans can also provide employers with an efficient means of deferring compensation.

Taxes obviously provide an incentive to offer compensation in the form of pensions and other non-taxed fringe benefits, but the way in which taxes affect the decision to offer pensions has been relatively unexplored in the economic literature. As a result, there is little evidence on the impact on pension coverage of changes in tax rules, such as limits on contributions to employer pensions and the availability of tax subsidized alternatives such as individual retirement accounts. Indirect evidence on the latter point might be gleaned from the findings of Poterba, Venti, and Wise (1996) that most IRA contributions represent new saving, rather than shifting from other accounts. Under this model, one might expect the availability of IRAs to have little or no effect on participation in and thus the demand for employer pension plans. But Engen, Gale, and Scholz (1996) reach the opposite conclusion with an equally plausible model and the issue remains unresolved.

The cost savings that employer-sponsored health insurance offer to workers are often cited to explain the provision of health benefits by firms. Health benefits received from an employer are tax free, while the cost of insurance purchased in the marketplace is generally subject to income taxes. In addition, employer-sponsored insurance provides workers with access to group rates, which are generally substantially lower than the rates available in non-group plans, especially for those with health problems (Chollet and Kirk 1998).

In addition, coverage data derived from 5500 data could be compared with coverage data collected through Census household surveys such as CPS and SIPP to aid in understanding why coverage counts differ. This would potentially aid in the development of improved survey methodologies.

3. Known Facts about Pension Plans

There have been many trends developing in employer sponsorship of plans, employee participation and pension design in the 1990's. It is imperative that we study these trends as well as their compounding effects with changes in the population demographics. This will lead to a better understanding of employment, personal savings, and firm benefits at the national level. It will also give us the opportunity to tackle other unanswered questions on the state and individual levels that will be able to inform the debate about pension policies.

Employer Sponsorship:

According to the Department of Labor, fifty-eight percent of private wage and salary workers participate in an employer-sponsored pension plan. For public-sponsored pension plans, ninety percent of the public wage and salary workers participate.

When we place the firms into different categories by employee size, we see a stark difference in participation between the small firms (less than 100 employees) and the medium and large firms (100-249 employees and 250 or more employees respectively). The medium firms and large firms are much more likely to offer a retirement plan. In

1997, seventy-nine percent of full-time employees were in a retirement plan sponsored by their firm. This is much higher than the participation rate in employer-sponsored pension plans in small firms. In 1996, only forty-two percent of employees in these small firms participated in a plan.

Pension Design:

It is also interesting to note that many employers have shifted away from defined benefit (DB) plans toward defined contribution (DC) plan. According to the Congressional Research Service Report for Congress in 2001, fifty percent of employees in medium and large firms were covered by a DB plan, down from fifty-nine percent in 1991. At the same time the proportion of full-time workers in these firms who participated in DC plans increased from forty-eight percent to fifty-seven percent.

For small firms, only ten percent of employees participated in a DB plan in 1996, down from twelve percent in 1990. However, the proportion of employees who participate in a DC plan increased from twenty-eight percent to thirty-five percent.

When we look at industries, we find that some industries had higher participation rates than others. According to the Department of Labor, over seventy percent of employees who work in mining, manufacturing, communications and utilities, or finance, insurance and real estate industries participate in an employer-sponsored plan. This is striking compared to workers employed in agriculture, construction and retail trade. Less than fifty percent of these employees participate in an employer-sponsored plan.

Pension Coverage:

According to the Department of Labor, the overall pension coverage was forty-four percent in 1999 increasing from forty-one percent in 1995. Coverage in small firms, however, is much lower than coverage in the medium and large firms. In 1999, the pension coverage in small firms was only twenty-four percent, while the coverage rates in the medium and larger firms was fifty-five percent.

4. Tasks

As outlined previously in the document "New Uses of Health and Pension Information: The 5500 file at the Census Bureau: Statement of Work," there are two major tasks in which the Census research team is engaged: first, investigating coverage and quality issues, and second, documenting our findings. Below is a brief description of coverage and quality issues, followed by a summary of our progress on the issue, problems we have encountered, and future work to be done.

Also, we give a brief summary of the problems that we have encountered and that we are planning to address next. While we plan to extend the analysis to all the years between 1993 and 1999, we have focused for now on the years 1996 and 1997.

Task 4.1 What types of firms file the 5500?

The project team will document the coverage of the 5500 file. The team will also provide information about the characteristics of firms that do not offer pension coverage and how they differ from firms that sponsor plans.

Status of Task 4.1:

- For the years 1995, 1996, 1997, and 1998, we have converted the 5500 ASCII-data files (Form Annual Return, Form 5500 C, Form 5500 R) into SAS-data files.
- For all of those years we have identified which EINs have welfare plans, defined benefit plans, defined contribution plans, or "other" pension plans. (We have spent a considerable amount of time to create meaningful variables that indicate the type of plans an EIN has). See Table 1 and Table 2 below.
- Preliminary match of those EINs to the Census Business Register (BR)[2] for 1996, and 1998. See Table 1a and Table 1b below.
- Match of the EINs with a plan-end-year of 1997 to the BR (1996, 1997, and 1998). See Table 2a to 2c below.

Next Steps:

- Some technical issues regarding the layout of the raw 5500 file (See Task 4.4) and regarding the match to the BR have to be resolved.
- There are additional schedules for which data needs to be converted (Schedule A etc.). A preliminary version of the Form 5500 (1999) has just arrived. However, we will wait with processing that year until a final version is available.
- When the technical issues are resolved, we will document the characteristics of firms that do offer pension coverage.

Tables 1a and 1b: Statistics on Form 5500 Data

The following tables present information on plan types that have a plan-ending year of 1997. Since records with a plan-ending year of 1997 are found in filing year 1996 and 1997, those two files have been matched and the records with plan-ending year 1997 have been extracted.

Table 1a, column 1 and 2 show how many plans are classified (by us) to be pension defined benefit plans (PDB, 62,339), pension defined contribution plans (PDC, 660,702), fringe plans, health plans, mixed welfare or pension plans (2,837), plans where there is no information on the plan type (11,625), other pension plans (25,708), other welfare or fringe plans. Those categories are exclusive. Column 3 and 4 show the codes that have been entered with respect to question 6 of Form 5500 (not check boxes). There were a number of plans for which several codes were entered.

[2] Note: The Census Business Register (BR) is also referred to as the SSEL.

Table 1a (1997):	Exclusive plan types (created through combinations of codes and checkboxes)		Codes of Form 5500, question 6 (not checkboxes)	
Plan type	Frequency	Percent	Frequency	Percent
PDB	62,339	5.73	62,452	5.74
PDC	660,702	60.73	663,028	60.95
Fringe	216,102	19.86		
Health	66,066	6.07	68,082	6.26
Mixed welfare or pension	2,837	0.26		
No info on type	11,625	1.07	11,625	1.07
Other pensions	25,708	2.36	17,632	1.62
Other welfare or fringe plans	42,515	3.91	78,356	7.2
Total	1,087,894	100		

Table 1b show the marked checkboxes of Form 5500. For a substantial number of plans several checkboxes were marked.

Table 1b (1997):	Checkboxes of Form	
	Frequency	Percent
Fringe plan	241,906	22.24
Welfare plan	978,735	89.97
Pension plan	747,870	68.74
Total	1,968,511	

Table 2a, 2b, and 2c: Statistics on the match between Form 5500 Data and the Census Business Register

The following two tables present statistics from matching Form 5500 Data with plan-end-year of 1997 to BR data of 1997.

Table 2a: Unique EINs from Form 5500 1996 and Form 5500 1997 with plan-end-year 1997
737,966

Table 2b: In the first match we took the unique EINs of the Form 5500 file and pulled out the corresponding alphas from the multi-unit file of the BR (1997). With the resulting list of alphas, we went back to the BR and, in the second match, pulled out the corresponding EINS. That led to a total of 236,258 unique EINs (those are associated with multi-unit firms). 608,472 EINs did not match to the multi-unit BR file.

Table 2b (1997)	First match to multi-units in BR 1997	Second match to multi-units in BR 1997	No match to multi-units in BR 1997
Unique EINs	129,494	236,258	608,472
Unique EIN-alpha combinations	133,674	249,716	N/A
Unique alphas	114,644	114,644	N/A

Table2c: In the final match to the BR, 817,660 EINs matched to the BR (either to the single-unit or multi-unit file). 27,070 EINs did not match to the BR (the 27,070 EINs are 3.67 % of the 737,966 Form 5500 EINs from Table 2a).

Table 2c (1997)	Final match to BR 1997	No match to BR 1997
Unique EINs	817,660	27,070
Unique alphas	114,644	N/A

Task 4.2 What are the gaps in reporting?

Determine companies in the Census Business Register that do not have benefits plans versus those companies who have plans and do not file Form 5500

Status of Task 4.2:

- Completed match of EINs in the Form 5500 data (with plan-end-year of 1997) to the Business Register years of 1996 to 1998. See Table 2c, Tables 3, Tables 4a to 4e, and Tables 5a to 5c for the match to the 1997 BR.

Next Steps:

- MEWAs
- Investigate accuracy of ALPHA match
- Use MEPS survey to provide comparison health insurance offer rates by firm size (for Census purposes)
- Taft-Hartley plans

Table 3: Statistics for EINs from BR that DO NOT match to Form 5500 Data

One of the most useful aspects of the Census Business Register is that it allows us to determine what types of firms do and what types do not file Form 5500 with the IRS. This section describes the characteristics of firms in 1997 Business Register that do match as well as the ones that do not match in any way to 5500 filings for plans ending in 1997. This overview gives us an idea of what sectors of the economy might not offer benefit plans as well as what sectors simply might not file forms. It also examines the importance of firm size in terms of likelihood of filing.

Table 3 shows the total number of EINs from the 1997 Business Register that do not match to 5500 data described in previous sections. However, many of these non-matching EINs are not actually valid companies. The second row shows how many of these records are ghosts or deletes (old records that are flagged to be purged in later years of the Business Register), are missing SIC codes, or have FIPS codes for places outside the 50 states plus the District of Columbia. The total number of EINs deemed to be valid companies is listed in the last row.

Table 3: EINs from 1997 BR	
	Frequency
EINs that don't match	9,856,162
EINs that are "Ghost" records, EINs that have geocodes outside of US, EINs that have missing or '000000' sic	1,713,959
Total "good" EINs that don't match	8,142,203

Table 4: Statistics for 5500-BR matches and for BR records that do not match.
Tables 4a and 4b show that the vast majority of companies that do not match to the Form 5500 data are in fact very small companies. Of the non-matches, 83% have 5 or fewer employees, and an additional 16% have between 6 and 100 employees. Large firms (>=100 employees) account for only .4% of all non-matches but they do account for 12.5 % of all matches. These results are no surprise. Larger firms are more likely to offer benefits and hence be required to file Form 5500. Even more encouraging, it appears that the majority of large firms in the Business Register can actually be matched to a 5500 Form filing.

Table 4a: Firm size – Total Employment at EIN – 11 categories

	Match	No Match	**Total**
5 or fewer employees	**215,438**	**6,744,339**	**6,959,777**
(Row Percentage)	3.10%	96.90%	100.00%
(Column Percentage)	*28.21%*	*82.83%*	
6-25 employees	**275,348**	**1,169,926**	**1,445,274**
	19.05%	#NAME?	#NAME?
	36.05%	*14.37%*	
26-50 employees	**104,629**	**141,526**	**246,155**
	42.51%	57.49%	100.00%
	13.70%	*1.74%*	
51-100 employees	**72,841**	**54,511**	**127,352**
	57.20%	42.80%	100.00%
	9.54%	*0.67%*	
101-250 employees	**54,803**	**22,574**	**77,377**
	70.83%	29.17%	100.00%
	7.18%	*0.28%*	
251-500 employees	**19,825**	**5,550**	**25,375**
	78.13%	21.87%	100.00%
	2.60%	*0.07%*	
501-750 employees	**6,963**	**1,533**	**8,496**
	81.96%	18.04%	100.00%
	0.91%	*0.02%*	
751-1000 employees	**3,528**	**678**	**4,206**
	83.88%	16.12%	100.00%
	0.46%	*0.01%*	
1001-2500 employees	**6,510**	**1,062**	**7,572**
	85.97%	14.03%	100.00%
	0.85%	*0.01%*	
2501-5000 employees	**2,099**	**265**	**2,364**
	88.79%	11.21%	100.00%
	0.27%	*0.00%*	
5001+ employees	**1,707**	**239**	**1,946**
	87.72%	12.28%	100.00%
	0.22%	*0.00%*	
Total	**763,691**	**8,142,203**	**8,905,894**
	8.58%	91.42%	100.00%
	100.00%	*100.00%*	

Table 4b: Firm size – Total Employment at EIN – 3 categories			
	Match	No Match	**Total**
5 or fewer employees	**215,438**	**6,744,339**	**6,959,777**
(Row Percentage)	3.10%	96.90%	100.00%
(Column Percentage)	*28.21%*	*82.83%*	
6-100 employees	**452,818**	**1,365,963**	**1,818,781**
	24.90%	75.10%	100.00%
	59.29%	*16.78%*	
100+ employees	**95,435**	**31,901**	**127,336**
	74.95%	25.05%	100.00%
	12.50%	*0.39%*	
Total	**763,691**	**8,142,203**	**8,905,894**
	8.58%	91.42%	100.00%
	100.00%	*100.00%*	

Table 4c presents the breakout of non-matching EINs by major SIC division. As expected, the majority of EINs that cannot be matched to the Form 5500 data are in construction, retail trade, and services, exactly the industries least likely to offer benefits. The majority of EINs that can be matched are services, manufacturing and wholesale trade.

Table 4c: Industry – Major SIC Division

	Match	No Match	**Total**
Agriculture	**12,214**	**471,837**	**484,051**
(Row Percentage)	2.52%	97.48%	100.00%
(Column Percentage)	*1.62%*	*5.79%*	
Mining	**3,696**	**32,251**	**35,947**
	10.28%	89.72%	100.00%
	0.49%	*0.40%*	
Construction	**52,480**	**1,019,589**	**1,072,069**
	4.90%	95.10%	100.00%
	6.95%	*12.52%*	
Manufacturing	**88,225**	**401,881**	**490,106**
	18.00%	82.00%	100.00%
	11.68%	*4.94%*	
Transportation, , Communications, Electric Gas, and Sanitary	**26,387**	**353,897**	**380,284**
	6.94%	93.06%	100.00%
	3.49%	*4.35%*	
Wholesale Trade	**83,822**	**519,241**	**603,063**
	13.90%	86.10%	100.00%
	11.10%	*6.38%*	
Retail Trade	**68,736**	**1,646,715**	**1,715,451**
	4.01%	95.99%	100.00%
	9.10%	*20.22%*	
Finance, Insurances, Real Estate	**78,025**	**712,568**	**790,593**
	9.87%	90.13%	100.00%
	10.33%	*8.75%*	
Services	**335,986**	**2,948,143**	**3,284,129**
	10.23%	89.77%	100.00%
	44.50%	*36.21%*	
Public Administration (non-missing SIC)	**5,527**	**36,081**	**41,608**
	13.28%	86.72%	100.00%
	0.73%	*0.44%*	
Total	**755,098**	**8,142,203**	**8,897,301**
	8.49%	91.51%	100.00%
	100.00%	*100.00%*	

*Number of EINs with SIC=blank, '999999', '000000' is 8,593 for the matched group and 1,521,778 for the non-matched group.

Table 4d shows the type of firm organization of the matching and non-matching EINs. The overwhelming majority of the non-matching EINs are single-unit companies, whereas 17 % of the matching EINs are multi-unit companies. Only 1% of the non-matching EINS are part of a multi-unit company. The EINs that are part of a multi-unit company can be grouped together using an enterprise identifier called Alpha. The 104,734 non-matching EINs associated with multi-unit companies represent 90,905 different enterprises. The 196,007 matching EINs associated with multi-unit companies represent 108,824 different enterprises. This table again confirms that successfulness of the match to the 5500 data. Most large enterprises are matched.

Table 4d: Firm Organization

	Match	No Match	**Total**
Single-Unit Company	567,684	8,037,469	**8,605,153**
Multi-Unit Company: Number of EINs (units within Enterprise)	196,007	104,734	**300,741**
Total	**763,691**	**8,142,203**	**8,905,894**
Number of Alphas (Enterprise, unique Multi-unit companies)	108,824	90,905	199,729

* The numbers here exclude EINs with a missing employment variable.

Table 4e shows the states where EIN is located. In the case of a single-unit firm, this state is the location of business activity. In the case of a multi-unit firm, this state is from the address of the EIN, meaning that it is possibly the headquarters of a particular division of the multi-unit enterprise. Companies with multiple EINs may be associated with multiple states and within EINs, there may also be plants in several different states.

Table 4e:	Firm Geography (Headquarters)			
	Match	Percent of Total matched EINs 761,582	No Match	Percent of Total non-matched EINs 8,068,648
AL	8,846	1.16%	114,791	1.41%
AK	1,674	0.22%	21,461	0.26%
AZ	10,585	1.39%	128,671	1.58%
AR	5,404	0.71%	81,975	1.01%
CA	84,752	11.13%	910,579	11.18%
CO	12,670	1.66%	156,363	1.92%
CT	13,940	1.83%	101,120	1.24%
DE	3,406	0.45%	25,602	0.31%
DC	3,546	0.47%	19,720	0.24%
FL	32,855	4.31%	543,850	6.68%
GA	17,840	2.34%	224,266	2.75%
HI	4,383	0.58%	33,104	0.41%
ID	3,179	0.42%	48,796	0.60%
IL	39,398	5.17%	349,456	4.29%
IN	15,690	2.06%	163,219	2.00%
IA	10,307	1.35%	99,579	1.22%
KS	8,226	1.08%	92,265	1.13%
KY	8,855	1.16%	102,925	1.26%
LA	8,936	1.17%	118,717	1.46%
ME	3,768	0.49%	45,458	0.56%
MD	15,252	2.00%	145,678	1.79%
MA	25,285	3.32%	179,254	2.20%
MI	28,694	3.77%	265,190	3.26%
MN	18,786	2.47%	150,923	1.85%
MS	4,814	0.63%	73,907	0.91%
MO	15,629	2.05%	169,421	2.08%
MT	3,075	0.40%	43,339	0.53%
NE	5,878	0.77%	64,766	0.80%
NV	3,783	0.50%	54,144	0.66%
NH	4,427	0.58%	41,119	0.51%
NJ	30,691	4.03%	284,951	3.50%
NM	3,467	0.46%	51,252	0.63%
NY	59,834	7.86%	587,750	7.22%
NC	19,234	2.53%	219,885	2.70%
ND	2,843	0.37%	29,168	0.36%
OH	36,128	4.74%	285,852	3.51%
OK	7,602	1.00%	106,600	1.31%
OR	10,754	1.41%	118,458	1.45%
PA	37,336	4.90%	311,190	3.82%
RI	4,716	0.62%	32,560	0.40%
SC	7,555	0.99%	107,132	1.32%
SD	2,547	0.33%	32,742	0.40%
TN	12,480	1.64%	147,624	1.81%
TX	38,708	5.08%	559,843	6.88%
UT	4,961	0.65%	61,822	0.76%
VT	2,072	0.27%	26,069	0.32%
VA	17,685	2.32%	187,781	2.31%
WA	14,899	1.96%	198,408	2.44%
WV	3,732	0.49%	45,797	0.56%
WI	19,268	2.53%	154,784	1.90%
WY	1,187	0.16%	22,875	0.28%
Total	761,582	100%	8,142,201	100%

* Four EINs of the matched file and two EINs of the non-matched file have FIPS code of '00'.

The next 2 tables attempt to isolate truly problem cases, where one would expect the EIN to match to the Form 5500 but it does not. Hence companies with 100 or fewer employees (8,042,306 EINs) or major SIC group of public administration (additional 6,266 EINs) are dropped. Furthermore, we recognize that EINs in the 1997 Business Register may not have a Form 5500 filing for a plan ending in 1997, but may have a filing for a plan ending in 1996 or 1998. Hence these tables look for matches in the Form 5500 data for plans ending in 1996 and 1998 and drop EINs which match to one of these years (1,640 EINs). This leaves a total of only 18,436 EINs that are of some concern. The tables below show the industries and firm sizes of the remaining EINs. There are very few large EINs left and most EINs are in the services sector.

Table 5a: Single-Unit problem non-matches: Industry – Major SIC Division

	Frequency
Agriculture	219
Mining	63
Construction	1,098
Manufacturing	1,653
Transportation, Communications, Electric, Gas, and Sanitary Services	778
Wholesale Trade	452
Retail Trade	1,854
Finance, Insurances, Real Estate	682
Services	11,637
Total	**18,436**

Table 5b: Single-Unit problem non-matches: Firm size – Total Employment at EIN –

	Frequency
101-250 employees	13,537
251-500 employees	3,011
501-750 employees	793
751-1000 employees	335
1001-2500 employees	540
2501-5000 employees	121
5001+ employees	99
Total	**18,436**

Table 5c looks at multi-unit EINs that are not coded as belonging to the public administration sector (drop 126 EINs) and that also did not match to Form 5500 data for plan years ending in 1996 or 1998 (drop 4,895 EINs). There are 99,713 EINs that meet these criteria, representing 86,955 unique enterprises. Again, most of the non-matches are in the service and retail trade sectors.

Table 5c: Multi-Unit problem non-matches: Industry–Major SIC Division	
	Frequency
Agriculture	566
Mining	498
Construction	1,136
Manufacturing	3,213
Transportation, Communications, Electric, Gas, and Sanitary Services	3,994
Wholesale Trade	9,511
Retail Trade	41,318
Finance, Insurances, Real Estate	9,767
Services	29,710
Total	**99,713**

Tables 5d and 5e:
There are some two-digit SIC groups with major industry group that seem to have more non-matches than others. Inspection of these groups in some cases raises the possibility that perhaps some sectors are simply not required to file the Form 5500. In other cases, it seems likely that benefits are not widely offered in these SIC groups.

Table 5d: The most problematic two-digit SIC groups for the single units are the following:

Table 5d: Common Sub-industries among single-unit non-matches with 100+ employees

SIC 2-digit code	Name of sub-industry group	% of Non-Matches
17	Construction Special Trade Contractors	3.74
23	Apparel and finished products made from Fabrics	1.97
41	Local and Suburban Transit and Interurban Highway Passenger Transportation (includes local Transit	1.25
42	Motor Freight Transportation and Warehousing	1.18
51	Wholesale Trade-non-durable goods	1.27
58	Eating and Drinking Places	6.42
70	Hotels, Rooming Houses, Camps	3.12
73	Business Services	16.26
79	Amusement and Recreation Services	2.92
80	Health Services (hospitals)	11.21
82	Educational Services (universities)	18.34
83	Social Services	2.95
86	Membership Organizations	3.54
87	Engineering, Accounting, Research, Management	3.34
	Total	**77.51**

Table 5e: The two-digit SIC groups with the highest concentrations of non-matches for multi-unit firms are

Table 5e: Common Sub-industries among multi-unit non-matches		
SIC 2-digit code	Name of sub-industry group	% of Non-Matches
1	Agricultural Production Crops	0.36
42	Motor Freight Transportation and Warehousing	1.35
50	Wholesale Trade-durable goods	5.91
51	Wholesale Trade-non-durable goods	3.63
52	Building Materials, Hardware, Garden supply, Mobile Home dealers	1.89
54	Food Stores	4.57
55	Automotive Dealers and Gasoline Service Stations	4.86
56	Apparel and Accessory Stores	2.35
57	Home Furniture, furnishings, and Equipment Stores	3.23
58	Eating and Drinking Places	14.05
59	Miscellaneous Retail	9.95
65	Real Estate	4.73
67	Holding and other Investment Offices	0.95
70	Hotels, Rooming Houses, Camps	1.66
72	Personal Services	4.69
73	Business Services	4.36
75	Automotive Repair, Services, Parking	2.62
79	Amusement and Recreation Services	1.46
80	Health Services (hospitals)	5.27
82	Educational Services (universities)	0.66
83	Social Services	3.34
86	Membership Organizations	0.88
87	Engineering, Accounting, Research, Management	2.4
	Total	**85.17**

In conclusion, when one looks at the very largest firms (5000+ employees), either multi-unit or single-unit and excludes a few key industry sub-groups such as 41, 80, and 82, there are few EINs that do not match to the Form 5500 data.

Task 4.3 Identifying problems associated with matching to the Business Register

Some companies from the Business Register may offer benefit plans but the EIN will not be found in the Form 5500 data. Try to identify these companies and if possible, attach these EINs to plans filed.

Status of Task 4.3:

- The match of our created 5500 plan-end-year file (Task 4.1) to the BR added 12% new EINs using the ALPHA match. See Table 2b and 2c.

Next Steps:

- Compare combined employee counts to participant counts
- Investigate examples of how multi-units structure their plans

Task 4.4 Identifying data quality issues

We will devote considerable effort to understanding the editing process and potentially applying them to all the files.

Technical issues encountered:

- When converting the 5500 ASCII data (Form 5500 Annual Return, Form 5500 C, Form 5500 R), we encountered problems with the 1995 and 1996 layouts available to us (the layouts for 1997 and 1998 seem to be fine).
  Note: We have discovered that there were some return keys entered erroneously. There are only a few (below 100) observations involved.

- When examining the yearly 5500 file, we have noticed that there are multiple reports/records for the same plan (e.g. for 1997 this affects around 40,000 records). In those instances there are multiple records for the same EIN and the same Plan Number (PLNNUM). In some of these, there has been a name change of the plan, and/or the plan-end date has changed. However, in others no obvious other change could be detected by us
  Note: We have identified those multiple records for 1997 and kept the plans (within the plan-end-year of 1997) that reported the highest employee count.

- More research needs to be done with respect to the plan participant count and eligible employee count, as those variables can be used to verify coverage of the 5500 file (see Task 4.3).