QUALITY OF THE DMI FILE AS A BUSINESS SAMPLING FRAME

David A. Marker and W. Sherman Edwards, Westat, Inc.
David Marker, Westat, Inc., 1650 Research Blvd., Rockville, MD 20850

Key Words: Establishment surveys, enterprise surveys

1. Introduction

The U.S. Bureau of the Census and Bureau of Labor Statistics (BLS) each maintain business registers for use as sampling frames for their surveys. Due to confidentiality and data restrictions on Federal business registers, other government agencies needing to survey businesses are not able to access the registers maintained by the Census and BLS. The most complete privately-maintained register is the Duns Market Identifiers (DMI) register maintained by Dun & Bradstreet (D&B).

The Small Business Administration (SBA) was an early user of the DMI register as a sampling frame. While they determined that this register provided the best frame available to them, they did report a number of difficulties with frame coverage (Phillips, 1993). Among the difficulties were: not all firms reported every branch to D&B; total firm employment did not necessarily equal the sum of employment at all of the branches; there were lags in recording company births and new branches; and there were delays in cleaning out firms and establishments that were no longer in business, sometimes as long as four years in the late 1980s.

Westat has used the DMI register in the last five years to conduct two large-scale surveys of employer-provided health insurance. This has provided us with a great amount of detailed knowledge concerning the strengths and weaknesses of using the DMI register as a sampling frame. We first used the register as a sampling frame for surveys in eight states conducted for a private foundation. We then used the DMI register for the National Employer Health Insurance Survey (NEHIS) conducted for the National Center for Health Statistics, Health Care Finance Administration, and Agency for Health Care Policy Research. NEHIS involved the selection of over 100,000 private-sector establishments from 750 strata. This paper summarizes our findings of the suitability of the DMI register to serve as a sampling frame for future government surveys.

2. Description of the DMI Register

2.1 Frame Basics

The DMI register is compiled by D&B through a review of public records from bankruptcy and district courts, secretaries of state, departments of motor vehicles, and unemployment insurance agencies, as well as newspapers, yellow pages, credit reports, and records of businesses dealing with governments. All establishments are then requested (by telephone or in person) to update their information (including name, address, and size) annually. As of June 1997 there are approximately 10.8 million establishments on the DMI register. When the DMI register was used as a sampling frame for the NEHIS at the end of 1993, it contained 10.1 million establishments.

The DMI register contains extremely detailed information on the type of industry conducted by each establishment. Each establishment has associated with it up to six four-digit Standard Industrial Classification (SIC codes) codes. Each of these codes then has up to four four-digit extensions that provide even greater detail than the standard BLS coding scheme. For example, this allows one to separate elementary and secondary schools (SIC code 8211) into public, Catholic, or other private; and elementary, junior, or high school. This provides a potential for identifying 24 different types of business at each establishment.

No establishments are purposely excluded from the DMI register. Both the public sector and private sector are on the frame, as are some self-employed.

2.2 Measures of Size

The DMI register contains a large number of variables that might be of use as measures of size for stratifying and/or selecting samples with unequal probabilities. These variables include the number of employees at the establishment, the number that report to that headquarters, and the total in the firm; the percentage employee growth in the total firm over the last three or five years; and annual sales volume and percentage sales growth over the last three or five years. One limitation is that the number of employees

at the establishment is missing approximately 13 percent of the time. The authors did not use the sales data and therefore cannot attest to their response rates.

2.3 Definitions

DMI defines an establishment as one location of one firm. Multiple locations of a firm are theoretically reported separately, but are defined by respondents. Multiple establishments (of different firms) can be at the same location. Field employees (e.g., sales reps, interviewers) are treated as branch offices if they have an office, even if its in their home.

All intermediate headquarters and ultimate headquarters are identified by DMI, as are which branches and lower-level headquarters report to which higher-level headquarters. There are procedures for identifying these connections, although they must be asked for specially.

DMI's instructions are that owners/proprietors are to be included in the number of employees, as are part-time employees. Holding company's whose only employees are shared with another company's have its officers listed as its employees. These instructions are different from those used by BLS, for example, where owners/proprietors are not included in the count of employees.

2.4 Other Available Information

DMI attempts to list the names of the business principal, all corporate officers or partners, and any non-officer individuals responsible for major business segments (e.g., sales & marketing, data processing). They provide secondary names for many establishments. They also have indicators for when establishments are foreign owned, SBA small businesses, female owned, or minority owned, as well as an identifier of its census tract.

3. Survey Findings

3.1 Coverage

Estimates of the number of private establishments vary considerably due primarily to differences in definitions, including such factors as whether the self-employed are included and the identification of the establishment as a location or as an identifiable function, such as management of the

payroll. After deleting known government primary SIC codes from the October 1993 DMI register, the file contained 9,900,000 "location" records (while DMI provides a definition of "establishment" to all respondents, the frame is based on the list of establishments identified by each respondent). Some scattered government activities still remained in the DMI file, but their numbers were fairly small and could be further reduced by identifying government codes among secondary SIC codes.

The DMI count compares with 6,176,000 establishments included in the 1990 Census Bureau estimates (*Statistical Abstract, 1993*, Table 859). This Census Bureau estimate excludes self employed and farms. Both of these classes are covered by DMI, although their coverage is uncertain. Census also includes establishments in business any time during the calendar year; the impact of this definition is uncertain in validating a comparison with DMI. Because of the differences in definitions (e.g., differences in time period, as of a specific date versus any time during the year, temporary hiring of employees by otherwise self-employed with no employees, etc.), one cannot draw conclusions about the coverage of establishments (or locations) by the DMI compared to the Census coverage.

One facet of coverage is the extent to which new establishments are covered. This is a problem with government statistics as well as the DMI. Westat asked D&B to provide the age distribution of establishments in the eight states included in the private foundation study (Table 1). "Age" reflects either the year the firm (not necessarily the establishment) started or the year in which there was a reorganization of the firm.

The D&B report is dated October 27, 1993, and it may be presumed that several months elapse, typically, before a new establishment (location) is entered into the file. As a result, assuming that the number of new establishments is approximately constant, the 1993 data may represent only about one-third of the year-to-date's new establishments. It is also likely that some 1992 new locations may not have been entered at the time the tabulation was prepared. One may also presume that the records for (say) 1990, 1991, and 1992 include numerous new small businesses that failed or were consolidated or dissolved and whose records were still in the file in 1993. Taking these factors into account, the estimated number of new establishments not in the file as of October 1993 is approximately 1 or 2 percent of the

Table 1. Age of establishments on the DMI frame for eight states

Year started	Number of locations	Percent
0-1988	445,888	83.64
1989	23,031	4.32
1990	21,792	4.09
1991	20,526	3.85
1992	16,300	3.06
1993	5,569	1.04
Total	533,106	100.00

total file and may be approximately one-half of the businesses less than 1 year old as of that time.

Definitions of employees also vary, but not as much as definitions of establishments. BLS reported approximately 89,858,000 private sector employees in nonfarm establishments in 1992 (May issue of Employment and Earnings, reported in the Statistical Abstract, 1993, Table 661). BLS uses both a household survey and a survey of establishments to report on employment, and the latter survey is the basis for the data reported above. Number of employees was estimated from the DMI frame by summing (across a sample of 100,000 establishments selected for the NEHIS) the products of the sampling weight and the average number of employees in each employment size stratum for each state. This estimation process yielded approximately 103,486,000 private sector employees. The differences between this estimate and the estimates by the Census Bureau and BLS are due to DMI's inclusion of some farms and self-employed persons, differences in the time periods, and the fact that a nontrivial part of the DMI file covers out-of-business locations whose records have not been purged from the frame. Final NEHIS survey responses were post-stratified to adjusted BLS unpublished totals that attempted to take account of those structurally not included in BLS publications, in particular the more than 2,000,000 employees not covered by unemployment insurance. The post-stratified number of employees was approximately 98,000,000.

Table 2 shows, by state, the comparison of the estimates of number of employees from three data sources.[1] It should be noted that if the out-of-scope locations are in the neighborhood of 18 percent, the NEHIS sample is close to the other two estimates. Comparisons of the NEHIS estimates with the two government sources are shown in the second and third to last columns, and a comparison of the two government sources is given in the final column. The unweighted standard deviations of these last three columns are 7.1 percent, 8.1 percent, and 6.2 percent, respectively. Thus, the BLS and Census state estimates are almost as variable from each other as the NEHIS estimates are from either of them.

The coverage of smaller and newer establishments on DMI's register (or on either government register) is not as complete as it is for other establishments. However, on the basis of the above analysis, there is little reason to believe that there is serious undercoverage in the general DMI register other than self-employed with no employees (SENE). It was decided that an alternative frame was needed for the SENEs.

At the request of the government two potential register supplements were examined further. First, D&B has a separate file containing 2.6 million incomplete records. It was thought that this might contain records for new establishments that had not yet been introduced into the DMI frame. Second, how accurate is DMI's identification of the self-employed with no employees? These should be identified as having one employee on DMI (the instructions for DMI are to include the owner in the number of employees) and not be part of a multi-establishment firm. Samples of 304 and 198 records were selected from each of these types, respectively, and phone calls were made to determine the status of the establishments.

---

[1] Column totals in Table 2 differ from the sum of state estimates because of differences in the estimation procedures for the states and the national totals.

Table 2. NEHIS frame compared with BLS and Census frames (employment in thousands)

State	BLS Total*	All government*	BLS excluding government*	Estimates from Census**	Estimates from sample design	Sample minus BLS(%)	Sample minus Census(%)	BLS minus Census(%)
AK	247	73	174	158	166	-4.6	5.1	10.1
AL	1,673	338	1,335	1,341	1,530	14.6	14.1	-0.4
AR	963	168	795	751	890	11.9	18.5	5.9
AZ	1,520	278	1,242	1,236	1,436	15.6	16.2	0.5
CA	12,140	2,091	10,049	11,317	11,825	17.7	4.5	-11.2
CO	1,593	291	1,302	1,248	1,505	15.6	20.6	4.3
CT	1,522	205	1,317	1,482	1,675	27.2	13.0	-11.1
DC	677	287	390	427	518	32.8	21.3	-8.7
DE	344	49	295	311	364	23.4	17.0	-5.1
FL	5,339	869	4,470	4,608	4,869	8.9	5.7	-3.0
GA	2,982	535	2,447	2,499	2,609	6.6	4.4	-2.1
HI	541	111	430	433	429	-0.2	-0.9	-0.7
IA	1,251	222	1,029	1,008	1,258	22.3	24.8	2.1
ID	416	88	328	300	384	17.1	28.0	9.3
IL	5,205	768	4,437	4,647	5,274	18.9	13.5	-4.5
IN	2,538	387	2,151	2,150	2,332	8.4	8.5	0.0
KS	1,115	226	889	894	1,040	17.0	16.3	-0.6
KY	1,511	274	1,237	1,186	1,289	4.2	8.7	4.3
LA	1,625	338	1,287	1,271	1,449	12.6	14.0	1.3
MA	2,778	378	2,400	2,773	2,903	21.0	4.7	-13.5
MD	2,079	414	1,665	1,811	1,853	11.3	2.3	-8.1
ME	512	96	416	424	501	20.4	18.2	-1.9
MI	3,917	641	3,276	3,411	3,865	18.0	13.3	-4.0
MN	2,186	347	1,839	1,832	1,999	8.7	9.1	0.4
MO	2,320	372	1,948	2,013	2,285	17.3	13.5	-3.2
MS	962	209	753	725	856	13.7	18.1	3.9
MT	317	74	243	222	283	16.5	27.5	9.5
NC	3,133	510	2,623	2,675	2,826	7.7	5.6	-1.9
ND	277	67	210	196	242	15.2	23.5	7.1
NE	747	148	599	587	693	15.7	18.1	2.0
NH	485	72	413	441	469	13.6	6.3	-6.3
NJ	3,441	569	2,872	3,220	3,618	26.0	12.4	-10.8
NM	598	156	442	418	560	26.7	34.0	5.7
NV	641	86	555	537	609	9.7	13.4	3.4
NY	7,728	1,428	6,300	7,074	7,811	24.0	10.4	-10.9
OH	4,842	734	4,108	4,246	4,515	9.9	6.3	-3.3
OK	1,210	270	940	941	1,185	26.1	25.9	-0.1
OR	1,271	231	1,040	1,017	1,188	14.2	16.8	2.3
PA	5,071	698	4,373	4,599	4,868	11.3	5.8	-4.9
RI	421	61	360	393	424	17.8	7.9	-8.4
SC	1,529	292	1,237	1,266	1,337	8.1	5.6	-2.3
SD	307	65	242	215	273	12.8	27.0	12.6
TN	2,232	355	1,877	1,869	2,061	9.8	10.3	0.4
TX	7,271	1,334	5,937	5,865	7,182	21.0	22.5	1.2
UT	768	157	611	571	690	12.9	20.8	7.0
VA	2,840	589	2,251	2,321	2,507	11.4	8.0	-3.0
VT	249	43	206	215	235	14.1	9.3	-4.2
WA	2,216	423	1,793	1,761	1,875	4.6	6.5	1.8
WI	2,349	356	1,993	1,949	2,193	10.0	12.5	2.3
WV	639	132	507	483	566	11.6	17.2	5.0
WY	205	57	148	132	172	16.2	30.3	12.1
Totals†	108,743	18,962	89,858	93,476	103,486	15.2	10.7	-3.9

* U. S. Bureau of Labor Statistics, Bulletin 2320: and Employment and Earnings, monthly, May Issue; Statistical Abstract, 1993.

** U.S. Bureau of the Census employees "on-board March 12, 1990.

† States do not add to totals due to differences in methods of estimation.

Of the 304 incomplete records, only 64 were identified as current business establishments. A similar number (68) were clearly out–of–scope, and the remaining were not successfully contacted despite repeated attempts. Forty-three of the current business establishments stated that they had more than one employee. These were recontacted to try and determine whether this frame contained many establishments that had not yet made it into the DMI frame (since the 2.6 million records do not have DUNS numbers they cannot be directly matched against the DMI frame).

Of the ten establishments that had stated that they were part of multi-establishment firms six participated in the call back. These included locations of well-known national corporations. All six had existed for at least five years, offer health insurance, and have from 5 to 300 employees.

Twenty-seven of the 33 single location firms participated. One had only been in business six months and two others between one and two years. The majority (at least 19) had been in business more than five years. Only two had more than ten employees and eight offered health insurance.

After discussions with the government it was agreed that this 2.6 million record frame did not contain a high number of new establishments. Most of these records that did correspond to establishments may also be found on the regular DMI frame. Thus it was decided not to include any sample from the 2.6 million incomplete D&B record frame in the NEHIS.

Sixty percent (120) of the sample of records on DMI reporting one employee responded that they were indeed establishments, but only 74 of the 198 were self-employed with no employees. The SENE population of inference was therefore not covered by the sample from the DMI frame. Sampled DMI cases

determined to be SENEs were considered out-of-scope for the NEHIS. It was estimated that between one-fourth and one-third of the 1.2 million DMI establishments with one employee on the frame are likely to actually have more than one employee at the time of the survey (in addition to the 46 such establishments that were identified out of the 198, there were 38 nonrespondents, some of whom also are likely to have more than one employee). Therefore, it was decided to include a sample of such establishments in NEHIS. If they responded that they were either part of a multi-establishment firm or had more than one employee they were eligible for the survey.

Table 3 shows the national distribution of private sector establishments on the DMI frame by firm and establishment sizes.

An abstract file should be purchased from D&B to improve the accuracy of the resulting sample. The abstract file contained size, corporate structure, and other information for every establishment on the DMI frame but does not include establishment names and street addresses. Using the abstract file, Westat was able to examine alternative designs for producing a sample that would result in more accurate estimates than would have been possible under the original plan. This approach offers the advantages outlined below.

- It facilitated the definition of "firm. "

- It provided a fixed reference frame that remained constant during the survey, thus facilitating comparison between frame and sample results.

- It facilitated the comparison of sample allocations and the selection of more nearly optimal designs than could have been achieved otherwise.

Table 3. DMI sample frame by firm and establishment size

Firm size	Establishment size								
	Unknown	1 no other	1-5	6-24	25-49	50-249	250-999	1000+	Total
1-49	1,197,959	1,105,384	4,884,932	1,539,707	210,621	0	0	0	8,938,603
50-999	26,094	22	105,387	147,541	61,946	190,392	16,991	0	548,373
1000+	29,504	29	83,353	147,134	51,041	77,550	18,825	7,478	414,914
Total	1,253,557	1,105,435	5,073,672	1,834,382	323,608	267,942	35,816	7,478	9,901,890

- It permitted the examination of sample sizes per detailed stratum and facilitated the collapsing of detailed strata to increase the likelihood of obtaining sufficient complete interviews for variance computations.

- It allowed for improved quality control over the selection of the sample of establishments, because the selection was performed at Westat rather than by D&B.

3.2 In-scope and Eligibility

One weakness of the DMI register is that it contains a lot of listings that are not current establishments. Such listings can be as frequent as 40 percent for very small establishments of small firms (see first few lines of Table 4). Many of these represent temporary business locations, e.g., construction offices, that are no longer used. The effect of this when using the DMI register as a sampling frame is that it increases screening costs, especially for surveys that emphasize very small establishments. It does not, however, introduce any biases into the estimators.

Table 5 shows similar results by establishment size from the earlier survey conducted in eight states. That survey did not take firm size into consideration. Establishments with only one employee were excluded

and those of unknown size were sampled with the 5-9 size class. The eligibility rules were also somewhat different from that used for the NEHIS.

A separate issue is the accuracy of the DMI register with respect to the number of employees in establishments. These estimates were used to create one of the principal stratifications for sample allocation and selection for both health insurance studies. Table 6 compares the classification of number of employees in the DMI abstract frame with the same classification of number of employees found during interviewing.

The frequencies are heavily loaded along the principal diagonal (boxed), which indicates that, for the most part, the DMI estimates of the number of employees were reasonably accurate. In some cases, however, the estimates of size were quite different from what was reported during data collection. Weight trimming in some cases can avoid the domination of the estimates by a few establishments with extremely large weights. Such cases are infrequent enough that they do not destroy the effectiveness of the DMI estimates for stratification purposes. For the NEHIS the "unknowns" were classified into the 1-5 employee size stratum for firms with less than 50 employees and into the 6-24 size stratum for larger firms. This proved to be the correct classification for capturing the modal group, as shown in the table. Unfortunately, the "unknowns" ranged widely in size and are as variable as the rest of the data. In summary, it appears that the DMI data were effective in classifying the frame into size strata.

Table 4. In-scope and eligibility rates from NEHIS

Firm size	Establishment size	Located by telephone interviewers	In scope rate (is in business)	Screener response rate	Eligibility rate (private sector with employees)
<50	unknown	65.0%	82.0%	75.0%	57.0%
<50	1 no other location	80.0%	81.0%	84.0%	22.0%
<50	1 - 5	80.0%	86.0%	80.0%	66.0%
<50	6 - 24	92.0%	94.0%	80.0%	91.0%
<50	25 - 49	96.0%	96.0%	79.0%	89.0%
50 - 999	unknown	98.0%	90.0%	80.0%	86.0%
50 - 999	1 - 5	97.0%	92.0%	80.0%	89.0%
50 - 999	6 - 24	98.0%	95.0%	76.0%	94.0%
50 - 999	25 - 49	99.0%	97.0%	80.0%	84.0%
50 - 999	50 - 249	98.0%	96.0%	79.0%	88.0%
50 - 999	250 - 999	99.0%	97.0%	78.0%	90.0%
1,000+	unknown	100.0%	90.0%	70.0%	79.0%
1,000+	1 - 5	100.0%	90.0%	77.0%	89.0%
1,000+	6 - 24	100.0%	93.0%	72.0%	92.0%
1,000+	25 - 49	100.0%	94.0%	72.0%	86.0%
1,000+	50 - 249	100.0%	96.0%	72.0%	84.0%
1,000+	250 - 999	100.0%	96.0%	75.0%	85.0%
1,000+	1,000+	100.0%	95.0%	76.0%	83.0%
	Total	88.0%	90.8%	78.1%	76.2%

Table 5. In-scope and eligibility rates from earlier survey

Establishment size	Located by telephone interviewers	In scope rate (is in business)	Eligibility rate (private sector with employees)
2--4	78.6%	95.4%	72.8%
5--9	88.2%	97.6%	89.5%
10--24	93.1%	98.1%	94.7%
25+	96.5%	97.8%	90.5%

Table 6.   Effectiveness of the DMI employment data for stratification

Firm size stratum	DMI size stratum	Employee size class found during data collection						
		1-5	6-24	25-49	50-249	250-999	1,000+	Total
<50	Unknown	943	548	105	84	23	11	1,714
	1, No other	735	75	8	2	2	0	822
	1-5	6,923	1,470	75	50	19	3	8,540
	6-24	1,471	4,667	281	62	2	1	6,484
	25-49	60	329	803	138	6	0	1,336
50-999	Unknown	50	56	29	34	5	0	174
	1-5	387	141	20	28	5	2	583
	6-24	122	902	113	68	9	3	1,217
	25-49	18	150	418	171	8	4	769
	50-249	130	289	708	3,861	146	9	5,143
	250-999	16	22	23	178	716	30	985
1,000+	Unknown	32	61	29	52	17	14	205
	1-5	382	121	16	22	6	1	548
	6-24	183	965	122	46	10	6	1,332
	25-49	23	139	370	117	16	8	673
	50-249	27	87	257	1,687	121	32	2,211
	250-999	10	14	16	244	922	67	1,273
	1,000+	4	14	3	38	150	698	907
All	Unknown	1,025	665	163	170	45	25	2,093
	1, No other	735	75	8	2	2	0	822
	1-5	7,692	1,732	111	100	30	6	9,671
	6-24	1,776	6,534	516	176	21	10	9,033
	25-49	101	618	1,591	426	30	12	2,778
	50-249	157	376	965	5,548	267	41	7,354
	250-999	26	36	39	422	1,638	97	2,258
	1,000+	4	14	3	38	150	698	907
	Totals	11,516	10,050	3,396	6,882	2,183	889	34,916

3.3   Linkages and Firm Size Comparison Between Frame and Survey

If explicitly requested, DMI can provide what they refer to as DIAS and hierarchy codes with their sample abstract file.   These codes can be used to identify the establishment structure within firms. The DIAS code provides a sort order that groups together all establishments in a firm, from ultimate headquarters down to each branch.   The hierarchy code then identifies how many levels of headquarters are between the establishment and its ultimate firm headquarters. Combining   this   information   with   other   DMI information such as the field that reports the DUNS

number of the headquarters to which each establishment reports, allows users to understand the structure of the entire firm. It can also allow users to subdivide firms to suit their analytic needs, for example to separate General Motors' major subsidiaries from its automotive core.

Based upon both the NEHIS pre-screening, in which alphabetic matches with DMI-based multi-establishment firms were checked for inclusion in corporate families, and on comparison of survey responses and frame linkage information of whether establishments were part of multi-establishment firms or not, the corporate linkage information in the DMI file appears to be reasonably reliable. Only four percent of alphabetic matches (that were not otherwise identified by DMI) were found to be actually in the corporate structure with which they were matched, and 13.8 percent of establishments were classified differently as to inclusion in a larger firm between the frame and the survey. Particularly considering the difference in time between the frame data and the survey data and the likelihood of some response error in each source, the Kappa value of 0.71 indicates very good agreement between the DMI frame and respondent reports of whether or not an establishment was part of a multi-establishment firm.

To summarize, very few establishments that DMI says are unattached responded that they were really part of a larger firm. The reverse is hard to measure since in large firms with subfirms it is not always clear to respondents which "firm" they belong to. Thus, it appears that the DMI frame may be used with some confidence in drawing samples for firm-level estimates.

The agreement between survey and frame information on firm size was more problematic. While more than 50 percent of establishments in the survey sample had relative agreement (reports with ratios between 2/3 and 3/2) in reports of firm size, there were many cases with large proportional disagreements. While some of these disagreements may be trivial (the difference between firms of size two and of size four, for example), the fact that the distribution of ratios was skewed as much or more for multi-establishment firms as for single-establishment firms indicates that the discrepancies may not easily be ignored.

There are many different definitions of "firm," and the definition chosen should be driven by the analysis being conducted. In the NEHIS, the concern was with the operational organization of health benefits, which varied considerably among large firms. Some were very centrally organized, while others had separate benefits administration for subsidiaries or for regions or divisions. Generally, the firm size reported in the survey seemed to fall somewhere in between the entire-firm and headquarters sizes shown on the DMI register. However, the NEHIS data do not support determination of, for example, how closely any of the firm size figures match with the operational definition of "health benefit groups." Further, in modeling policy changes using NEHIS data (or other survey data including firm size), one should consider the effects of policy changes on how corporations define their own structures. For example, the Clinton health plan would have exempted firms over a certain size from some regulations. If firms felt that it was in their interest to be exempt, they would perhaps organize to maximize their firm size.

4. Conclusions

In general the DMI register works well as a sampling frame for high quality establishment-based surveys. The coverage of establishments appears to be near 98 or 99 percent (based on a study in eight states). Family farms and self-employed coverage is less, although most of these with employees are probably covered. Other missing establishments are likely to be new small establishments. Approximately one-half of new establishments get on the list within the first year (from the same study of eight states). Coverage of all employees is probably higher than the coverage of establishments since it is much more likely for large establishments to be included in the register.

These weaknesses are similar to those of the Census and BLS frames. The Census frame is probably at least as likely to miss small new establishments, especially nonmanufacturing establishments in small firms. Self-employed are problematic for all business registers but are excluded from the Census frame. The BLS frame also misses many new establishments in new firms since it depends on establishments beginning to pay unemployment insurance to be identified for their register. The BLS frame also does not include employees who are not eligible for unemployment insurance, such as employees of many religious institutions and railroads. BLS estimates these employees to number around 2,000,000.

The main weakness of the DMI register is its inclusion of many small establishment listings that are

no longer in business. This requires extra screening, especially in surveys that emphasize small businesses. This weakness, however, does not cause bias, only an increase in costs over what would result from a cleaner list. Use of the DMI abstract file allows for higher quality control on sample selection and for the implementation of more complicated sample designs. Overall, careful use of the DMI register can result in surveys that meet the standards expected of high quality government data collection.

Reference

Phillips, Bruce D. (1993). *Perspectives on Small Business Sampling Frames.* Proceedings of the International Conference on Establishment Surveys, American Statistical Association, Alexandria, VA. pp. 177-184.