

**The Accuracy of Public Mailing Lists:
The Middletown Community Study Experience**

Vaughn R.A. Call
Department of Sociology
Brigham Young University

Colter Mitchell
Department of Sociology, and
Institute for Social Research
University of Michigan

Cassandra Dorius
Department of Sociology
Pennsylvania State University

June 1, 2007

* Please direct correspondence to Vaughn R.A. Call, Department of Sociology, 2008C JFSB, Brigham Young University, Provo, Utah 84602; email: vaughn_call@byu.edu.

ABSTRACT

The Accuracy of Public Mailing Lists: The Middletown Community Study Experience

Sampling firms frequently claim a high level of accuracy in their lists and contend that their frequent updates of databases and the use of weighting make targeted lists a good option for any research study. This study compares a database list against school records from “Middletown” (Muncie, Indiana) to ascertain households with children. Comparisons of the household list data against census data indicates that the database likely included a good listing for most households. The check of household information (kids present) against school enrollment information indicated that the list service only matched 61 percent of the currently enrolled students and that the non-matches are not randomly distributed. Schools with higher levels of minorities and of students receiving free lunch have a higher percent of no matches.

The Accuracy of Public Mailing Lists: The Middletown Community Study Experience

The purpose of a sample survey is to allow researchers to make inferences to a specific population, without the cost of recording information for the entire population. One of the major issues in creating an inferential sample is the amount of overlap between the target population and the sampling frame. Groves et al (2004) describe a sampling frame as lists of procedures used to identify all the elements (or units) of a target population. We are examining the use of database lists as a possible sampling frame. While personal information listing companies may widely advertise the quality of their services by emphasizing how many millions of people are included in their lists and the ease of access to household address, home phone number and now email address lists, the quality of these lists is generally not known (Dillman, 2000). In addition, in recent years, a number of companies have begun to provide sample lists with the option to easily target numerous demographic and consumer characteristics. Given the expense of obtaining adequate criterion data that would allow quality assessments, questions about the accuracy of the information in the files are seldom addressed and coverage characteristics are largely unknown. In this paper we explore list accuracy by comparing a list of households with children obtained from a major public list source with a list of students from official school directories.

Our first goal in this paper is to examine how well a large, reputable database list might work as a sampling frame. We hope to shed some light on whether these lists can be used directly as sampling frames, or, in the least, in conjunction with other lists or procedures to reduce the amount of coverage error, thus producing the most accurate sampling frame possible. The second goal of this paper is to address the database list's ability to focus the sampling frame from a general target population of a specific geographic area to a demographically specific target population within the same geographic space. For example, how well does the list move from being a sampling frame of households to a being a sampling frame of households with children of school age?

Targeted Samples

Many sample suppliers advertise that their targeted phone, mail and internet lists are more accurate and cost effective (more completes in less time and for less money) than lists obtained using traditional sampling methods that require expensive screening interviews or mailings. These targeted lists are household or respondent specific and are created largely from self-reported sources, such as face-to-face interviews, questionnaires, warranty cards, subscription lists, online surveys, and a variety of respondent completed forms. Most larger list service companies provide a range of categories such as: age, gender, marital status, children in household, ethnic background, surname, height, weight, health, income, credit rating, congressional district, type of business, residential information, hobbies, interests, lifestyle characteristics, products purchased, and Census tract/block information. Within these categories, numerous sub-selections are also available providing literally hundreds of specific items and thousands of possible combinations of traits to select.

Sampling firms frequently claim a high level of accuracy in their lists and contend that their frequent updates of databases make targeted lists a good option for research studies. Some list service companies suggest that their targeted samples provide clients a 100% efficiency rate for the specified characteristic; meaning, in a study of people aged 65 and older, a targeted sample of people age 65+ should only include people within this age category. These targeted samples allow researchers the luxury of knowing they are contacting the population they are studying without the difficulties of screening. Some companies make their lists available on compact disk or the internet with search programs that facilitate the selection of numerous sample characteristics and allow the generation of targeted sample lists with minimal effort. Tight time lines, budgetary concerns, and the need to find rare populations may led research companies, organizations, and university researchers to buy their samples from sampling firms who specialize in creating targeted lists for almost any research need.

Although these lists have been around for decades, relatively few major academic studies have used them in projects that are later published. Some exceptions are the 1978 Muncie Community Survey (Caplow et. al., 1982) and the 18 year Oklahoma City Survey (St. John and Clark 1984; Welch, Tittle,

and Grasmick 2006). But both of those projects used the Polk Directory as a sampling frame and then, if necessary, provided additional procedures to oversample or screen. This begs the the question of just how accurate are targeted samples supplied by list providers? The “End of the Century” Middletown study provides a limited assessment of the accuracy of purchased target samples by comparing a basic target population characteristic, children in the home, with information from an actual sample frame provided by local schools.

BACKGROUND

In 1924, Robert and Helen Lynd selected Muncie, Indiana as the benchmark community for their classic “Middletown” community study (Lynd and Lynd, 1929). The 75-year interval between the initial survey and the 1999 replication represents one of the longest intervals of relatively continuous survey research in a single community. In the original Middletown study, the Lynds selected 164 working and business class families and interviewed the wives in each of these families. They specified that the families had to be native born, white Americans, in intact families, with school age children (ages 6-18), who resided within the Muncie city limits (Lynd & Lynd, 1929: 507). In 1978, researchers replicated this survey and used Polk City Directories to obtain a random sample of 2,192 married women. A combination of postcards and telephone calls to nonrespondents was used to determine whether these women had a school age child in their household (Caplow et. al., 1982). This two stage process identified 793 married women with school age children, but the process was very expensive and time consuming.

In 1999, researchers conducting the “ End of the Century” replication of the Lynds’ Community survey faced the problem of identifying households with school aged children but now had the advantage of significant advancements in computer and data management technology that was not available in the earlier Middletown surveys. For example, there was a substantial increase in the 1990s in the number of list service firms that provided researchers access to the addresses and phone numbers for millions of people. In addition, list service providers expanded the information in their data files to include numerous demographic and psychometric indicators on households. Similar to the 1978 study, the 1999 sample

design started with the selection of all Muncie residents using the 1998 Muncie City Directory CD. Using the data file extracted from the CD included we identified 7,265 Muncie households with children.

Once the households with children were identified, we conducted a small pre-test to verify that the targeted list of respondents with children was accurate. We randomly selected 20 names and attempted to call them. Of those we were able to contact, less than half said they had a child living with them. The large percentage of people without children in the household raised serious questions about the accuracy of our targeted list and the indicator for kids. Faced with the prospect of calling every person on the list to identify residents with children, as well as questions about the unknown number of people with children not on the targeted list, we dropped plans to use the Polk targeted list of residents with school age children. Rather than use the expensive 1978 mail and telephone procedures to screen the Polk list of addresses for mothers with school aged children in the home, we negotiated with the Muncie school system to obtain a complete directory listing of all school children and drew a random sample of mothers from this list.

The two different approaches to obtain a replicate Middletown sample of mothers with children in Muncie Schools resulted in access to two major data sources that comprise the basis for this study. The first is a well-established, Polk City Directory general public listing of residents in Muncie, Indiana. The *InfoTYME*[®] CD of Muncie listings was produced on November 10, 1998. The second was the official school directory listing of all students enrolled in Muncie public schools as of January 27, 1999.

DATA AND METHODS

Polk Directory Data

R.L. Polk is one of the oldest information services providers. Founded in 1870, Polk publishes over 1400 city directories and provides a wide range of information services. *InfoTYME*[®] combines city directory information with search software on a CD. The 1998 Muncie CD database provided names, addresses, and telephone information for the residents of Muncie, Indiana and surrounding communities. The innovative Target Marketing version of the program permitted targeted searches on the database with

the option to print out the specified sub-population sampling lists. Some of the possible searches included: census county number, resident's occupation, head of household or resident status, gender, marital status, presence of children, home ownership status, number of years in current home, dual earner family, and relative income (ex. high, average, below average, etc).

Polk's InfoTYME[®] provided individual level data for Muncie residents. We created a household listing from the individual listings by combining individual listings that matched on any two of three possible variables: last name, address, and phone number. We first identified households by combining all individual listings with the same information on all three variables: last name, address, and phone number. We also combined individuals if they had the same phone number and address, or the same phone number and last name, or the same last name and address. There were 6,335 individual listings with "current resident" as the last name. These "current resident" listings were not allowed to be combined when we searched for listings by phone number and last name or address and last name.

These matching procedures resulted in 40,102 households, a number that closely approximates the 41,513 households that Polk reported for Muncie (Polk City Directories, 2003). Of these 40,102 households, 4,772 (11.9%) were created by matching on last name, address and phone number, 1,484 (3.7%) households were created by matching on address and last name, 722 (1.8%) from phone number and address, 161 (0.4%) from phone number and last name, and 32,963 (82.2%) were individuals that did not match on any two of the three variables and were designated as individual households.

Using the individual information we were then able to determine which households had children present. Thus, from the Polk Directory we have three primary data sets: Polk Individual Data (N=48,050), Polk Household data (N=40,102), Polk Households with children (N=7,265). Respectively, each data set is a subset of the previous one (i.e. individuals to households, all households to households with children).

Muncie School Directory Data

We contacted the superintendent of the Muncie school district and requested permission to obtain a directory listing of all students in Muncie public schools. Our request was reviewed by school officials and approved. The school district provided a directory listing (name, address, and telephone number) for all students enrolled in Muncie elementary and high schools as of January 27, 1999 (n=8,523). While the exact count of dropouts and residential relocations was not available for the entire school district, efforts to obtain parental permission forms for the High School Student Survey (n=2235) identified only 38 students who had dropped out of school or moved out of the city. This list was ideal for drawing a sample of mothers of school age children.

It is important to note three major differences between the school directory and the city directory list of households with children. First, the school list provided the student's name, address, and phone number but not the parents' names. Which, for most students, wouldn't be much of an issue, but students living with family with a different name last name (e.g. step-families, extended family, etc) would take further processing to match correctly. Second, the school directory was a listing of school age children (5-19) in Muncie public schools. The city directory included all listings with kids (presence of children) in the home (R.L. Polk & Co. 1998: 30). According to the 2000 census there were 3,943 children under the age of five in Muncie. If these children did not have an older brother or sister in school, their family would not be listed on the school district list, but should be on the Polk list. Third, the school list only contains students enrolled in public schools and about 1,330 students were enrolled in 10 private schools with about 490 of these students enrolled in a single school (K-12) associated with Ball State University (National Center for Education Statistics and Indiana Department of Education, 2000). However, most of these schools were pre-kindergarten/kindergarten schools for four and five year olds and some of these students resided outside of Muncie. The overlap between age and nonpublic pre-kindergarten/kindergarten schools and the lack of information on how many have older siblings make it

difficult to calculate how many families, particularly young families, were excluded from the public school list.

Matching

This study compares listings in the Muncie school directory against the Polk City Directory for Muncie, Indiana. While there are recognized omissions and small errors in the school directory list, we use the school directory as the criterion. We searched for each of the 8,523 student phone numbers in the Polk household data. If we located the student phone number we examined to see if the address and/or last name also matched the information in the Polk City Directory. If the telephone number was not matched, we entered the student's address. If neither the phone number nor address search produced a match, the name was coded as a no match.

We specified the type of match as follows: 1) student's last name, address, and phone number all match (perfect match); 2) student's last name and phone number match; 3) address and phone number match; 4) student's last name and address match but no phone number listed in Polk database; 5) student's last name and address match but phone number does not match; 6) student's last name and address match but no phone number in school directory; 7) only the student's telephone number, 8) only the student's address matched, 9) no match on address or telephone, and 10) Youth Opportunity Center (a school where a small number of students live and attend school). We accounted for several street name changes and a second coder verified each "no match" entry did not have any matching information.

To address how well the listing performed as a sampling frame we use the codes above to estimate the undercoverage depending on different information requirements. We removed the 85 students attending the Youth Opportunity Center from the analysis because their address on the school directory was the address of the Center. By adding different codes together we get a number of households that would have been found depending on the information required by the researcher to conduct survey (i.e. telephone, mail, mixed-mode, etc). We argue that this indicates how much undercoverage existed in the Polk directory. That is, the smaller percent of "no matches" the lower the undercoverage, and the more accurate the sampling frame.

To examine the accuracy of the screening or filtering abilities of the Polk database, we rely on two separate analyses. First, we compare the Muncie City Polk Directory data to the 2000 Census data for Muncie for three measures: sex ratio, marital status and fertility. Our second analysis compares the Muncie school data to the Polk's screened data for children. Practically speaking, we attempt to answer the question: if a researcher had initially screened for children in the home, how well would the Polk directory have preformed? To do this we follow the steps indicated above, but search for the student information in just the sub sample of the Polk Directory where addresses are recorded as having children. Unlike the previous search, however, we require a match to have information on at least two pieces of information (i.e groups 1-6) instead of any information (i.e. groups 1-9).

RESULTS

Coverage

Table 1 about here

Table 1 presents the percentage of the student directory entries matched in the Polk directory. About 95 percent of the students in the Muncie Public Schools did have some location information in the Polk Directory. This suggests that the nearly all of people may be expected to have some information within the directory. Almost 1/3 of the students matched on all three variables (last name, address, and telephone number). This is the largest group and probably primarily consists of students who share their last name and household with their guardians. Approximately 16% of the students had a phone number that matched an entry in the Polk directory and also matched on either the last name (3%) or an address (13%), but not both. The 3 percent with the same last name and phone number could indicate either the number of moves to a new address in Muncie in the interval between the publication of the Polk directory (November 1998) and the printing of the school list (January 1999), or the number of children who have a relative's phone number listed as the contact phone number. The 13 percent of students who matched on phone number and address may be members of step families. An additional 13 percent of students matched Polk directory entries on the name and address, but not the phone either due to the Polk directory entry not having a phone number (9%), the school directory not having a phone number (1%) or the phone numbers not matching (3%). The 10% of students not matching on phone number due to lack of a

phone number in either the Polk or school directory, most likely indicates an error in record keeping not the lack of a phone. The 3% whose last name, and address match, but whose phone number does not match, may indicate multiple phone lines within the household, or the presence of cell phones. About six percent of students only have a phone number that matched, and considering neither of the directories had missing address data, this seems to imply some error in the directories. The second largest overall group were the 27% of students who only matched on address. Some portion of these may be students in step families where the home has multiple phone numbers or a phone is lacking. Very few, five percent, had no address or phone information in the Polk directory. These five percent constitute the group that, no matter what method of interview, would be missed if the Polk Directory was used as sampling frame.

Despite the impressively high coverage rate of 95%, depending on the amount of information required to conduct the study, the coverage rate varies considerably. For example, if a researcher intended to conduct a face-to-face interview, but did not require the last name of the potential respondent (groups 1,3,4-6,8), the coverage rate would be 86% (if the last name was required it would drop to at most 73% due to dropping group 3). However, because some of these families most certainly are step families it is probably more correct to understand these as endpoints of a range. So that if a researcher intended to conduct a study using the address (either face to face or mail) the coverage rate would be between 73-86 percent. Similarly the coverage rate for a telephone study would range from 49%-56% (groups 1-3, 6,7). Of course, if the researcher intended to use multiple modes with each respondent (i.e. sending a letter and conducting a telephone survey), therefore requiring at least the phone number and the address, the coverage rate would be between 32%-46% (groups 1,3,6).

Table 2 about here

From a sampling perspective, the impact of an undercoverage would be less if the distribution of non covered were randomly distributed across the city and social groups. The Indiana Department of Education (2001) provided school profile data on each elementary and secondary school in the Muncie Public School System. Since the school directory list included both elementary and secondary students and elementary schools cover distinct neighborhoods in Muncie, we used school profile data on percent minority and percent free lunches for each elementary school as a rough indicator of race and level of

economic well-being within that geographic area. Using Mapquest® and a detailed map of elementary school boundaries, we located all 3,275 residences that did not have two or more information matches and identified the elementary school assigned to that address. The distribution of “no match” by geographic area is presented in Table 2. The expected percentage of “no match” represents the distribution of “no match” if they were equally distributed across all Muncie elementary schools and proportionate to the size of the school. The difference between the expected and the actual observed percentages indicates if the percentage is less than (-) or more than (+) would be expected. We excluded any cases that were either outside the school district or not found in the address search (1.3% and 1.6% of the “no match” total respectively).

In general we found that the higher the percent minority and percent free lunch in an elementary school the greater the amount of relative “no matches.” Although we only have twelve data points we ran Pearson correlations to document the strength and significance of the associations. The correlation between percent minority and the difference in percent of “no match” is significant and of moderate strength (Pearson $r = 0.58$, $p < .05$). Similarly, the correlation between receipt of a free lunch and “no match” was significant and even stronger (Pearson $r = 0.83$, $p < .001$). When taken together, if the data from the elementaries is representative of the area it serves, then areas with higher percent minority and a lower socioeconomic status have higher rates of being left off the Polk Directory¹. Of course a major issue is that some elementary schools do not represent all of the neighborhoods it serves. Thus, it may be that these relationships are stronger, or weaker (although the latter would probably be more difficult).

Screening

Table 3 about here

We now turn our attention from examining if the directory could be used as a general sampling frame to a sampling frame for a specific population. That is, can researchers use the information found in the directory to “prescreen” a population instead of using costly screening procedures in the field. First we examine how the directory compares on a macro level to the population. Table 3 provides

¹ It is important to note that percent minority and percent receiving free lunch was also strongly correlated (Pearson $r = 0.75$, $p < .01$).

demographic characteristics of Muncie based on two different sources: the Polk directory information and the 2000 Census (Table DP-1. Profile of General Demographic Characteristics: 2000, Geographic Area: Muncie city, Indiana). The 2000 Census reported 54,081 people age 18 or older living in Muncie whereas the Polk directory had listings for 48,050 people 18 an older, about 11 percent fewer than would be expected from the Census.

In the Polk data file, more than half are males (52%). In contrast, Census data indicated that about 54 percent of Muncie adults are female. The difference in the proportion male and female between Census and Polk data may be due to the high percentage of individual records in the Polk file with missing sex data (26%). It appears from census data that a much higher proportion of the missing sex data in the Polk records were actually female listings.

Another common demographic characteristic is marital status. The Polk directory reported that about 11 percent of individuals, age 18 and over, were married. In contrast, Census data reported that over 41% of people age 15 and older were married. A very small part of the percentage difference may be due to the different specification of age (18 versus 15). We submit that the main explanation for the substantial percentage difference is that, unlike sex information, cases where marital status was not known in the Polk directory were coded as “not married” instead of “missing”. Thus, a large proportion of individuals listed in the Polk directory as “not married” were probably married.

The final comparison by households rather than by individuals produced large differences. If one assumes that each address represents a unique household by census definitions (all the persons who occupy a housing unit), there was a substantial difference in the number of households between the two data sources (Polk = 40,102, Census = 27,322). Initial checks showed that the difference was not due to the inclusion of Ball State students who lived in dorms. Married student apartments were included in the Polk list, but students living in group quarters were not. The difference in the number of households may be mostly the result of a large number of individual Polk listings with “current resident” as the name for the listing, the absence of apartment numbers on addresses, and the large number of second phone lines (fax/internet connections, teen lines) with “current resident” as the name. There were numerous Polk listings that had the same address and “current resident” as the occupant. For example, there was one

address with five “current resident” listings for that address and no information about any of the occupants. In another case, there was 50 households at one address with 44 having different names and phone listings and 16 listed as “current resident.” These “current resident” listings could represent a different household or a vacant apartment.

There are some important differences between the Census and the Polk directory. First as mentioned above the Polk Directory was considered current in November 1998, while the census would have been current nearly 17 months later in April 2000. As well, the information from the Census was based on the people living within the city limits of Muncie (i.e. not a Metropolitan area) (Table DP-1. Profile of General Demographic Characteristics: 2000, Geographic Area: Muncie city, Indiana), while the data from the directory was based on people having Muncie listed as their city. Considering information from people in other towns near Muncie, were included in the directory listing, we propose that the City indicator is based on the mailing address. This would suggest that in addition to the people inside the city limits of Muncie, other people living near Muncie but not located in nearby towns, would have their data on the Muncie city directory. Despite the exact overlap, there is some evidence that the information that may be used in the Polk directory may not be completely accurate.

Table 4 about here

Although the macro comparisons to the census are indicative of the accuracy the Polk directory, by using the school directory and Polk directory comparison again we can better examine the accuracy. In a previous analyses we took the school list as the criterion to see how many students addresses could be located in the Polk directory. In Table 4, we shift the focus to Polk’s database of households with kids at home and try to find these households in the school list. There were 7,527 Muncie households with children in the Polk targeted subsample (kids=1). We were only able to find 28 percent of the entries in Polk’s kid subsample using any combination of name, address and phone number. However, about two-thirds of the found entries had the same last name, address and phone number. Most of the other found entries had only the same address and phone number or last name and address.

Other than accuracy issues in the directory data, there are two main reasons why the Polk kid matches might be lower than expected: children’s age and family size. First, the school list does not

include children under age six (first grade), while the Polk directory contains families of children younger than school age in addition to school age children. Based on the Census there are almost 3,943 children under the age of 5, but some of these young children may also have other siblings under 5, reducing the number considerably. Second, a single household in the Polk file may have two or more children in the public school system, resulting in just one match for multiple children. Thus the low matching rate may be partially explained by key differences in the lists.

Despite the differences in the lists, however, it is important to note that over one-third of the 8,438 public school students were matched to a household that Polk listed as having no children in the home. Or in other words, if a researcher had “prescreened” this sample to only include homes with children, a large proportion of the target sample would have been eliminated despite their correct contact information being in the directory. Thus on one hand, the large number of students matched in the “no kids” group indicates considerable error in this “screening” indicator. On the other hand, when the indicator correctly identifies a “kid” household there is a substantially higher percentage of exact matches on name, address, and phone number.

Figure 1 about here

The lack of timely updates to list mailing files is a common explanation for errors and omissions in lists. If this is the case, listings that remain on the data file for longer periods of time should have more chances to have information about the household verified and any omissions or errors in the household information corrected. As a result, there should be a higher proportion of matches for households that have been on the list file for a longer time. In a similar vein, those households that do not change for several years should have greater exposure to being recorded in the database. To examine this issue we checked the percent found (with at least two pieces of information) by the Polk city directory’s variable for number of years in the household (see Figure 1). There was a modest increase in the percent found in the targeted kids data file as the number of years in the household increased. The highest found percentage (33%) was for the longest durations in the household (15+ years), and the lowest were in years 1 and 2 (19% and 18% respectively).

DISCUSSION AND CONCLUSIONS

In recent years there has been a tremendous increase the number and scope of information sources on the American public. Although there are seemingly limitless possibilities for accessible and varied samples covering millions of people, researchers have not adequately assessed the sampling implications associated with use of these lists. The lack of reliability checks has been largely due to the absence of criterion indicators and the lack of access to the actual databases. Innovations, such as Polk's *InfoTyme* database, provide researchers the opportunity to assess data quality against available criterion indicators.

The check of a single point of information (kids present) against school enrollment information raises concerns about the accuracy of personal information attached to households. The comparison indicates that the list service had a listing that matched 61 percent of the currently enrolled students in a mid-western community. A plot of the geographic dispersion of the non-matches demonstrates that the non-matches are not randomly distributed. Non-matches are significantly more likely in older sections of town where minority and lower socioeconomic residents predominate. Thus, the listing misses a large proportion of the population of children in public schools and introduces a socioeconomic bias.

The target key, kids present, provides little advantage in identifying listings with kids in public schools. About half of the public school children were found in the "no kid" subfile. It was surprising to find that the length of time on the file did not markedly improve the match rate. We expected listings to improve in information availability and accuracy the longer the listing went without major change in residence. Additional time should permit the posting of new information about a person or household, particularly phone numbers. This was not the case.

In sum, the results highlight the benefits and shortcomings associated with the reliance on recent information sources to construct probability samples with special characteristics. The coverage of households provided by list service providers has improved greatly over the last decade. It is clear from this one comparison of household information that there is considerable room for improvement in the demographic data attached to these lists, particularly for targeted samples.

Finally, the limitations and assumptions of this study are important to note. First, we took only one Polk Directory (Muncie, IN) from one time period (November of 1998) to examine. This one

directory may not be a typical directory in that it may be more or less accurate than other directories. As well we examine only one screening variable (children present), which may have different amounts of error from other screening variables. Clearly much more work is needed both in other locations, on other screening variables and other database lists. Thus, we acknowledge that there is some chance that our findings are outliers. Nevertheless we suggest that these lists may become more available and useful, but that their accuracy, flexibility and scope have not been studied in much depth within the survey methodological literature, and thus this is an initial step toward better understanding professional listing databases.

Table 1. Percentage of Muncie School Children Matched to Polk Directory Household Listing (n = 8,438)

1) Perfect match (last name, address, phone number)	2,704	32%
2) Same last name and phone number	250	3%
3) Same address and phone number	1,102	13%
4) Same last name and address - no phone number in Polk Directory	745	9%
5) Same last name and address - phone number doesn't match	276	3%
6) Same last name and address - no phone number in school directory	86	1%
7) Telephone only	543	6%
8) Address only	2,312	27%
9) No Match	335	4%
10) Youth Opportunity Center	85	1%

Table 2. Distribution of Little or No Information Students by Muncie Elementary School Boundaries (School Characteristics: percent minority, free lunch program)

Elementary School					School Characteristics	
	School size	Percent of No Matches			Percent Minority (2000)	Percent Receiving Free Lunch (1999/2000)
		expected %	observed %	difference (o-e)		
South View	526	12.7	15.1	2.4	8%	53%
Morrison-Mock	166	4.0	2.6	-1.4	8	26
West View	329	7.9	6.3	-1.6	9	31
Grissom	502	12.1	11.6	-0.5	20	50
North View	307	7.4	6.6	-0.8	21	27
Storer	353	8.5	5.3	-3.2	22	29
Claypool	213	5.1	4.8	-0.3	22	43
Mitchell	225	5.4	3.7	-1.7	23	17
Sutton	494	11.9	10.5	-1.4	26	56
Washington-Carver	412	9.9	13.6	3.7	38	71
Garfield	384	9.3	11.9	2.6	54	73
Longfellow	241	5.8	8.1	2.3	59	77

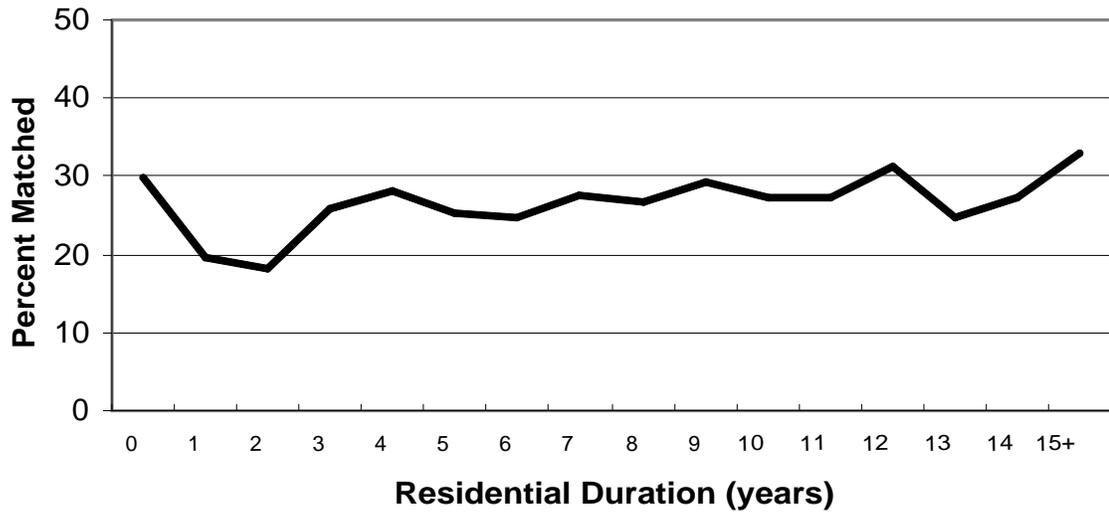
Table 3. Census and Polk Demographic Comparisons for Muncie City

Individuals - sex	Polk		Census	
males	18,429	52%	25,083	46%
females	16,928	48%	28,998	54%
no sex information	12,693	-	-	
Individuals - marital status				
married	5,245	11%	23,203	41%
not married	42,805	89%	33,042	59%
Households				
kids present	7,265	18%	7,241	27%
no kids	32,837	82%	20,081	74%

Table 4. Percentage of “Kid” Households from Polk Directory that Had Multiple Pieces of Information on the Muncie School District List

Zero or One Piece of Information	5,450	72%
Multiple Pieces of information	2,077	28%
perfect match (last name, address, phone number)	1,302	63%
same name and phone number	68	3%
same address and phone number	303	15%
same name and address - no phone number in Polk Directory	306	15%
same name and address - phone number doesn't match	75	3%
same name and address - no phone number in school directory	23	1%

Figure 1. Percent of Polk Households with Children found with multiple pieces of information in Muncie School Lists by Polk Recorded Residential Duration



REFERENCES

- Caplow T, H.M. Bahr, B.A. Chadwick, R. Hill, & M.H. Williamson. (1982). Middletown Families: Fifty Years of Change and Continuity. Minneapolis: University of Minnesota Press.
- Dillman, D. A. (2000). Mail and Internet Surveys: The Tailored Design Method. New York: John Wiley & Sons.
- Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology* New York: Wiley.
- Indiana Department of Education. 2001.
<http://mustang.doe.state.in.us/SEARCH/s3.cfm?city=Muncie>. Last accessed June 15, 2002.
- Lynd, R.S. & H. M. Lynd (1929). Middletown: A Study in Contemporary American Culture. New York: Harcourt, Brace and Company.
- Polk City Directories. 2003. *Sales Leads & Mailing Lists* (February), Livonia, Michigan.
- R. L. Polk & Co. 1998. *User=s Guide for InfoTYME*. WexTec Systems, Inc.
- St. John, Craig and Frieda Clark. 1984. "Race and Social Class Differences in the Characteristics Desired in Residential Neighborhoods" *Social Science Quarterly*, 65(3):803-814.
- U.S. Bureau of the Census. 2001. *Profiles of General Demographic Characteristics: 2000 Census of Population and Housing*. Geographic area: Muncie city, Indiana. Washington, D.C.: U.S. Government Printing Office.
- Welch, Michael R., Charles R. Tittle, and Harold G. Gramsick. 2006. "Christian Religiosity, Self-Control and Social Conformity" *Social Forces* 94(3):1605-1623