Bridge Study of the Imputation Methods for the Public Libraries Survey (PLS): FY2006-08

Peter Schilling
Irene Brown
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Introduction

The Public Libraries Survey (PLS) provides a national census of public libraries and their public service outlets. PLS is designed as a universe survey, with approximately 9,200 libraries in the 50 states, the District of Columbia, and the outlying areas of Puerto Rico, Guam, the Northern Marianas, and the Virgin Islands. Unit response rates are high, generally around 98 percent, whereas response rates for some items on the questionnaire are below 85 percent at the state level. Based on research done in 1997, PLS began imputing for missing data for libraries in the 50 states and the District of Columbia, but not the outlying areas. Since the initial research, there has been the addition of new variables to the survey and definitional changes to other variables. Therefore it was decided to conduct another research study to see if the imputation methodology should be changed.

In 2009, a study to evaluate the way the imputation cells were defined and to evaluate current and new imputation methods was completed. This bridge study shows the magnitude of change in the national and state estimates due to the new imputation methods that were implemented in the FY 2008 PLS. This study will only show how different the estimates are, not which methods are better. Results of research comparing the different imputation methods can be found in Brown [2009].

Methodology

The initial stage of the study compares estimates of imputed totals using the former and the new methods for survey years 2006, 2007 and 2008. In order to make comparisons, missing data in the 2006-08 surveys were imputed with the new methods. Then missing data for the survey year 2008 were imputed using the former methods. Both old and new imputation methods along with related definitions of data variables can be found in Attachment A.

Estimates of the totals at the national and state levels were then computed for survey years 2006-08 using both old and new PLS imputation methods. "New" estimates were then compared to those published in 2006 and 2007 that used the former imputation methods. For survey year 2008, "old" estimates were compared against those published using the new imputation methods. The difference and percent difference between estimates were calculated. Frequency counts of how often an individual method was actually used for a PLS variable were also tabulated for the new imputation methods. To compare individual methods used for a PLS variable for both old and new imputations, fifteen flags were developed and the definitions can be found in Table 1; the situation was complicated by the introduction of corresponding four character flags introduced for FY2008 results. For the former methods, the old imputation flags were grouped into the fifteen general purpose flags used in this study.

Table 1: FY 2008 Imputation Flag Values and Definitions

Flag Number	FY08 Flag Value	Definition
0	R_08	Reported value no imputation done
1	IG0**n**	Method 1: Mean growth rate
2		Method 2: Hot-deck growth rate
3	IQ08	Method 3: Adjusted cell mean
4	IJ08	Method 4: Cell mean
5	IK0**n**	Method 5: Prior year ratio to another item
6	ID08	Method 6: Cell Median of ratio to another item
7	IP0**n**	Method 7: Direct Substitution of prior year data
8	IM08	Method 8: Cell median or adjusted cell median
9		Method 9: Sequential hot-deck
10	U_08	New Item or outlying areas: no imputation done
11	IT08	Obtained value by relationship of total to detail items
12	IB08	Raking of detail items to match total
13	IS08	Special impute for an item
14	IY08	Changed because of consistency check

Note: The 'n' for certain FY 2008 Flag Values represents one of four prior years the data could be pulled from [4,5,6,7].

Results

Frequency tables that show how often a method was used with the new imputations can be found in three PDF files (COUNTSFY2006_NEW.PDF, COUNTSFY2007_NEW.PDF, COUNTSFY2008_NEW.PDF), which are available on request. These files show a breakdown of imputation methods for each PLS variable, using the flags from Table 1.

The estimates of totals and differences between old and new imputation methods can be found in three Excel files (FY2006COMPARE.xls, FY2007COMPARE.xls, FY2008COMPARE.xls) that are also available on request. These Excel files are broken down by PLS variables, with one tab showing a table of all state estimates and the national estimate for one variable. These tables show total estimates obtained using both old and new imputation methods, the difference between total estimates, the percent difference and a flag that denotes if estimates of the total include imputed values (0: no imputed values, 1: imputed values).

Data from the comparison spreadsheets were compiled to create a table of combined national estimates for PLS variables. A compilation of national estimates can be found in Table 2.

Table 2: Comparing National Totals between Old/New PLS Imputation Methods: FY2006-08

	FY2008	FY2007	FY2006		FY2008	FY2007	FY2006
popu_lsa*	0.00%	0.00%	0.00%	**fcap_rev**	-0.05%	0.37%	-27.56%
centlib*	0.00%	0.00%	0.00%	**ocap_rev**	-0.15%	-0.42%	-0.33%
branlib*	0.00%	0.00%	0.00%	**cap_rev**	-0.11%	-0.03%	-0.48%
bkmob*	0.00%	0.00%	0.00%	**capital**	-0.92%	0.08%	0.18%
master	-0.02%	0.02%	0.00%	**bkvol**	-0.02%	-0.01%	-0.06%
libraria	-0.06%	-0.05%	-0.06%	**ebook**	0.10%	0.01%	1.66%
othpaid	-0.02%	-1.21%	-0.38%	**audio**	0.00%	0.00%	0.01%
totstaff	-0.03%	-0.83%	-0.27%	**video**	-0.09%	-0.24%	-0.36%
locgvt	0.08%	0.08%	0.00%	**database**	0.17%	0.41%	0.93%
stgvt	0.04%	0.03%	0.03%	**subscrip**	0.06%	0.15%	0.08%
fedgvt	-0.03%	-1.11%	-0.68%	**esubscrp**	-2.13%	0.81%	-0.09%
othincm	-0.32%	-0.05%	-0.05%	**hrs_open**	-0.07%	-0.09%	-0.09%
totincm	0.04%	0.06%	0.00%	**visits**	0.19%	-0.24%	0.22%
salaries	-0.04%	0.03%	0.01%	**referenc**	0.69%	0.89%	-0.09%
benefit	0.00%	-0.05%	-0.20%	**totcir**	0.00%	0.02%	0.04%
staffexp	-0.03%	0.01%	-0.04%	**kidcircl**	0.55%	0.33%	0.23%
prmatexp	0.08%	0.72%	0.11%	**loanto**	-0.08%	-0.09%	0.24%
elmatexp	0.12%	0.50%	0.05%	**loanfm**	0.09%	0.11%	0.36%
othmatex	-0.18%	-2.49%	-0.25%	**totpro**	-0.09%	-0.31%	-0.22%
totexpco	0.03%	0.07%	0.04%	**kidpro**	0.07%	-0.06%	0.07%
othopexp	-0.22%	-0.22%	-0.17%	**totatten**	-0.19%	-0.14%	-0.04%
totopexp	-0.07%	-0.03%	-0.05%	**kidatten**	0.22%	0.22%	0.17%
lcap_rev	-0.09%	0.07%	0.03%	**gpterms**	0.14%	0.19%	0.15%
scap_rev	-0.14%	-0.16%	-0.43%	**pitusr**	1.44%	0.05%	-2.33%

SOURCE: Institute of Museum and Library Services, Survey of Public Libraries in the United States, Fiscal Years 2006-2008

Note: See Attachment B for full variable names

The national totals in Table 2 show roughly equal numbers of variables with positive/negative percentage change values when comparing old versus new PLS imputation methods. The percentage differences in Table 2 don't suggest that the change in imputation methods leads to a bias in either direction at the national level. However, the situation is different when considering the change at the individual variable level, as some PLS variables have consistently higher or lower values over the three years using the new PLS imputation methods.

Overall, the national level PLS variable results in Table 2 show relatively small differences in magnitude when comparing old and new imputation methods. Most PLS variables had absolute percentage differences below 1% and all variables had differences below 2.5% over the three fiscal years with the exception of FCAP_REV. During FY2006, a 27.6% difference was observed between old and new imputation methods for FCAP_REV. Closer examination revealed that this discrepancy could be traced back to a single imputed value of this variable for a library in Louisiana. Since the imputed value of FCAP_REV using the old methods was highly inflated, it eventually caused a large discrepancy at the national level. However, these sort of discrepancies are anomalies that should be relatively rare.

The three Excel comparison spreadsheets for FY2006-08 provide estimates of PLS variables at national and state levels across all three fiscal years. Though these estimates allow analysts to conduct rigorous analysis comparing the old and new PLS imputation methods, using the comparison spreadsheets can be a cumbersome process. Data for each PLS variable are separated by tabs in the Excel spreadsheets and aggregating data across multiple variables can require a series of copy-and-paste operations or import into a separate analysis package, such as SAS. Often, a more visual overview of the data can be useful for analysts, particularly to understand overall trends and to make high level comparisons within a complicated study like this one.

The first task in developing a visual display for the PLS comparison data was establishing a set of categories to measure the degree of consistency between old and new PLS imputation methods. To accomplish this, these categories were defined by absolute percentage difference in state estimates of PLS variables at boundaries that could be recognized as "important" by analysts. Deciding on these category boundaries was somewhat subjective;

aggregated data from the PLS comparison spreadsheets were used to guide these decisions. After a period of evaluation and data analysis, the category boundaries were set at differences of 2%, 5%, 10% and 20%. A final set of consistency categories was defined corresponding to letters A-F:

"A" No Difference

"B" 0% < Difference < 2%

"C" 2% <= Difference < 5%

"D" 5% <= Difference < 10%

"E" 10% <= Difference < 20%

"F" Difference >= 20%

The next task in developing the visual display was aggregating the categorical rankings into a simple representation so an analyst can quickly assess how consistent the state estimates are for a particular variable during a fiscal year. We decided to create a color coded display that would show a clear contrast for different levels of consistency between old and new imputations of PLS variables. It took several iterations and incorporating feedback from IMLS to determine how to show consistency levels between old and new imputation methods for different estimates of PLS variables. We finally decided to use two basic displays: one display (Table 4) shows color coded consistency levels for PLS variables over FY2006-08 with a breakdown of consistency categories among the 50 states and Washington DC; the second display (Table 5) shows color coded consistency levels for the 50 states and Washington DC with a breakdown of consistency categories among the 48 PLS variables.

The third task in the design of the visual display was the creation of consistency levels using categorical rankings and then assigning a color code for each consistency level. Establishment of consistency levels for a fiscal year can be somewhat subjective and open to interpretation. We decided to use absolute differences of 10% ('E' consistency category) as an important threshold, knowing from Table 2 that one discrepancy between old and new imputation methods in state estimates of a PLS variable can be sufficient to cause a large (>25%) discrepancy at the national level. Thus, if a PLS variable or state had no corresponding 'E' or 'F' consistency rankings, it would be assigned to a "High" consistency level, but one 'E' or 'F' ranking would result in an assignment to the "Medium" or even "Low" consistency level.

Initially, we planned to use "High", "Medium" and "Low" consistency levels, but after closer examination of the data, two additional levels were added to distinguish among units in the "High" and "Low" levels. Five consistency levels were created, as shown in Table 3 below:

Table 3: Consistency Levels for Single Year Comparison between Old and New PLS Imputations Methods

Consistency Level	**Color**	**Criteria**
Highest	Dark Green	All units in A and B categories
High	Light Green	All units in A, B, C and D categories
Medium	Yellow	All units in A, B, C and D categories, except one or two units in E category
Low	Red	Three or four units in E category AND/OR one to four units in F category
Lowest	Black	Five or more units in E/F cateogry

Note: 'units' could refer to PLS variables or states, depending on the display.

The first display of comparison data between old and new PLS imputation methods is given in Table 4, which examines PLS variables and provides a breakdown of consistency category rankings over FY2006-08.

Table 4: Consistency between PLS Variables for Old/New PLS Imputations Methods: FY2006-08

	FY2008						FY2007						FY2006					
	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
popu_lsa*	51						51						51					
centlib*	51						51						51					
branlib*	51						51						51					
bkmob*	51						51						51					
master	38	13					38	13					36	15				
libraria	31	19	1				31	18	2				31	18	2			
othpaid	35	16					32	17	1		1		31	19	1			
totstaff	35	16					28	22			1		30	19	2			
locgvt	33	17	1				34	15	1		1		32	18	1			
stgvt	37	11	3				33	15	2	1			38	9	4			
fedgvt	45	4	1			1	47	2				2	40	7		2	1	1
othincm	34	16	1				33	17	1				31	18	1	1		
totincm	34	16	1				35	14	1	1			30	20	1			
salaries	31	19	1				29	21	1				28	23				
benefit	29	21	1				27	22	1	1			25	23	3			
staffexp	33	17	1				30	21					30	21				
prmatexp	29	22					30	17	2	1	1		26	25				
elmatexp	34	15	1	1			33	14	3	1			31	19	1			
othmatex	31	16	4				27	19	2		1	2	30	17	3	1		
totexpco	34	17					31	19	1				29	22				
othopexp	33	16	1		1		33	15	1	1	1		27	20	3		1	
totopexp	34	15	2				29	21	1				25	25	1			
lcap_rev	40	11					38	10	1	1	1		44	6				1
scap_rev	39	8		3		1	39	3	1	2	1	5	41	4	1	1	1	3
fcap_rev	43	3	1			4	39	7				5	42	2	1	1		5
ocap_rev	39	8	1	3			39	7	3	1	1		42	5	3			1
cap_rev	38	13					39	10	2				40	9				2
capital	32	15	2	1	1		31	17	2	1			29	19	3			
bkvol	33	18					33	18					37	13	1			
ebook	38	9	2		1	1	40	10	1				43	7				1
audio	34	17					34	15	1	1			36	15				
video	29	18	4				28	17	5			1	26	18	6			1
database	30	15	5		3		29	13	3	3		3	29	10	8	2	1	1
subscrip	27	18	6				27	19	4	1			26	21	3	1		
esubscrp	32	7	3	1	1	7	34	14		1		2	32	12	2	3		2
hrs_open	32	18	1				34	16	1				32	18	1			
visits	16	31	3	1			22	25	2	2			18	30	2		1	
referenc	15	21	8	5	2		14	20	12	4		1	15	18	13	2	3	
totcir	35	16					32	19					34	17				
kidcircl	20	30		1			18	29	3	1			12	33	4	1	1	
loanto	21	23	3	1	3		21	24	2	1	2	1	17	28	1	2	3	
loanfm	23	20	4	2	2		22	21	3	3	1	1	19	25	1	4	2	
totpro	29	19	1	2			26	20	3	1	1		21	27	3			
kidpro	30	19	2				23	23	4	1			26	20	4	1		
totatten	24	25	1		1		25	23	2		1		21	27	2	1		
kidatten	28	20	2	1			26	20	4	1			23	23	4	1		
gpterms	36	12	3				33	14	4				35	14	2			
pitusr	14	25	8	3	1		13	17	11	4	4	2	12	15	10	6	5	3

*Not Imputed

Categories:

A: No Difference
B: 0% < Difference < 2%
C: 2% <= Difference < 5%
D: 5% <= Difference < 10%
E: 10% <= Difference < 20%
F: Difference >= 20%

Consistency Levels:

Highest: All states in A and B categories
High: All states in the A, B, C, and D categories
Medium: All states in the A, B, C and D categories except one or two states in E category
Low: Three or four states in E category AND/OR One to four states in F category
Lowest: Five or more states in E/F category

SOURCE: Institute of Museum and Library Services, Survey of Public Libraries in the United States, Fiscal Years 2006-2008

The breakdown in Table 4 shows how many state estimates fell into each consistency category ranking for a PLS variable over the three year period. A summary of the results from Table 4 is given below.

- PLS variables generally held to similar consistency levels over the three fiscal years.
- As expected, the four non-imputed variables showed no differences between old and new PLS imputation methods among any of the PLS variable state estimates (popu_lsa, centlib, branlib, bkmob).
- Low consistency was observed between old and new PLS imputation methods for the PLS capital revenue variables (lcap_rev, scap_rev, fcap_rev, ocap_rev, cap_rev), which are grouped together during imputations. Table 4 shows all five variables had at least one 'F' consistency ranking in one or more of the three years. This level of inconsistency is not unexpected for capital variables because of the variability in this type of data.
- Table 4 shows high consistency over the three years for stgvt variable, but low consistency (at least one 'F' consistency ranking per year) for fedgvt.
- State estimates of PLS income and benefits related variables (othincm, totincm, salaries, benefit, staffexp) were highly consistent over the three years.
- Table 4 shows high consistency was observed for state estimates of print serial subscriptions (subscrip), but low consistency for electronic serial subscriptions (esubscrp), which agrees with the expected results.
- Consistency problems for state estimates of PLS inter-library loans variables (loanto, loanfm).
- Very low consistency in 2006 and 2007 between old and new imputation methods for the number of users of public internet computers per year (pit_usr).

The second display of PLS imputation method comparison data is provided in Table 5, which examines states and gives a breakdown of consistency category rankings. The breakdown determines how many PLS variable estimates for a particular state fell into each consistency category over the three-year period.

Table 5: Consistency between States for Old/New PLS Imputations Methods: FY2006-08

	FY2008						FY2007						FY2006					
	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
AL	34	11	1	1		1	36	9	2	1			35	8	2	3		
AK	47	1					47	1					47	1				
AZ	11	34	2	1			9	32	1	2	2	2	23	21	1	2	1	
AR	13	32	2	1			14	31	2	1			5	34	4	3	1	1
CA	24	22	1	1			22	25	1				32	16				
CO	39	9					35	12		1			40	6	1		1	
CT	6	27	9	2	2	2	6	27	12	2	1		8	30	9		1	
DC	47		1				42	2	1	1		2	41	2	1		3	1
DE	44	1			1	2	46			1		1	37	3	1			7
FL	18	29	1				10	38					10	35	2	1		
GA	44	1	1	1	1		43	4		1			47	1				
HI	48						48						48					
ID	18	30					18	29	1				18	25	3			2
IL	11	37					6	31	3	4	2	2	10	31	2	3	2	
IN	44	4					41	6	1				38	8	1			1
IA	8	37	2			1	12	35				1	20	28				
KS	24	23	1				37	6	2			3	14	30	2		1	1
KY	48						45	3					41	7				
LA	47	1					48						35	9	2			2
ME	11	34	2	1			7	22	10	6	1	2	11	28	6	3		
MD	45	2				1	47	1					46	2				
MA	13	32	3				12	34	2				11	35	1	1		
MI	9	31	6	1	1		17	30			1		7	35	4	4		1
MN	42	5		1			32	15				1	38	9		1		
MS	48						48						47	1				
MO	42	5	1				33	11	2		1	1	27	15	3	2	1	
MT	47	1					48						47	1				
NE	7	27	9	3	2		7	25	8	4	4		11	25	9	1	2	
NV	47	1					47	1					48					
NH	12	33	2		1		7	32	4	4		1	6	33	4	3		2
NJ	4	40	1	2		1	6	39	2		1		6	39	2		1	
NM	45	3					18	28	2				23	24	1			
NY	45	3					45	3					45	3				
NC	42	5	1				44	4					43	5				
ND	33	10	1	3		1	23	22	1	2			23	21	2	1	1	
OH	36	11	1				34	13	1				34	13	1			
OK	48						47	1					48					
OR	31	17					43	3	2				25	21	1	1		
PA	33	15					33	14	1				35	11	2			
RI	15	22	11				25	18	5				16	24	8			
SC	47	1					47	1					47		1			
SD	11	29	2	3	1	2	11	21	10	2	1	3	11	21	8	4	2	2
TN	46	2					36	12					27	21				
TX	44	3		1			36	11			1		31	17				
UT	17	29	2				41	7					41	7				
VT	8	25	9	1	3	2	8	23	8	3	2	4	20	23	3		2	
VA	37	9		1	1		40	8					35	11	2			
WA	30	9	5	1	3		11	28	6		1	2	24	16	7			1
WV	41	7					43	4	1				42	5	1			
WI	32	14		1		1	36	11	1				38	9	1			
WY	48						46	2					44	4				

Note: Each fiscal year compared old and new imputation methods for 48 variables

Categories:
A: No Difference
B: 0% < Difference < 2%
C: 2% <= Difference < 5%
D: 5% <= Difference < 10%
E: 10% <= Difference < 20%
F: Difference > 20%

Consistency Levels:
Highest: All variables in A and B categories
High: All variables in the A, B, C, and D categories
Medium: All variables in the A, B, C and D categories except one or two variables in E category
Low: Three or four variables in E category AND/OR One to four variables in F category.
Lowest: Five or more variables in E/F categories

SOURCE: Institute of Museum and Library Services, Survey of Public Libraries in the United States, Fiscal Years 2006-2008

A summary of the results from Table 5 is given below.

- More variation in consistency levels was observed for states compared to PLS variables in Table 4, which indicates that states have a greater tendency to switch consistency levels from one year to the next.
- As expected, certain states had chronic consistency issues between old and new PLS imputation methods, landing in the "Low" or "Lowest" consistency levels over the three year period. Some of these states confirmed anecdotal findings from the comparison spreadsheets, including: DC, DE, SD, VT, and NH.
- Some other states showed a high degree of consistency between old and new PLS imputation methods, landing in the "High" or "Highest" consistency levels in each of the three years. These states included: AK, CA, FL, HI, KY, MA, MS, MT, NV, NM, NC, OH, OK, OR, PA, RI, SC, TN, UT, WV, WY.
- Table 5 has some limitations since it can't show the geographic boundaries among the US states. Analysts must try to visualize the layout of the different states and how high or low consistency states may cluster together since no US map is provided.

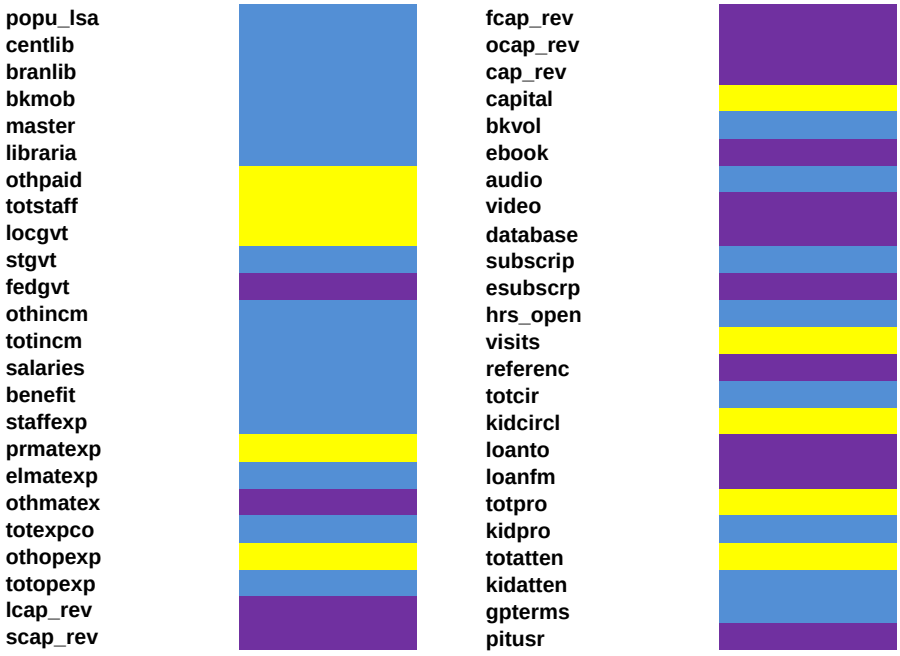
Tables 4 & 5 provide a useful display for summarizing the comparison between old and new PLS imputation methods, and explain what consistency rankings led to assignment of a consistency level to a particular PLS variable or state. By aggregating the consistency rankings over the three fiscal years, a more concise overview of the comparison data can be produced. The main issue to be resolved in creating a consolidated display is determining how to define a set of consistency levels similar to those used in Tables 4 & 5, but extending over the three-year period. After some deliberation, three consistency levels ("High", "Medium" and "Low") were developed for the three-year comparison display.

Table 6: Consistency Levels for Three Year Comparison between Old and New PLS Imputations Methods

Consistency Level	Color	Criteria
High	Blue	Differences between state estimates of PLS variables were below 10% over all three years.
Medium	Yellow	If one or two differences between state estimates of PLS variables were 10% or higher but below 20% in at least one of the three years.
Low	Dark Red	Three or more differences between state estimates of PLS variables were 10% or higher but below 20% in at least one of the three years AND/OR at least one difference was 20% or higher.

Consistency levels in Table 6 were first applied to PLS variable comparisons over FY2006-08, as given in Table 4. The resulting consolidated display is shown in Table 7.

Table 7: Overall Consistency between PLS Variables for Old/New PLS Imputation Methods: FY2006-08



Consistency Levels:

High: All differences below 10% during all three years.

Medium: In at least one year, one or two state differences were 10% or higher, but below 20%

Low: In at least one of the three years: three or more state differences were 10% or higher but below 20% AND/OR at least one state difference was 20% or higher

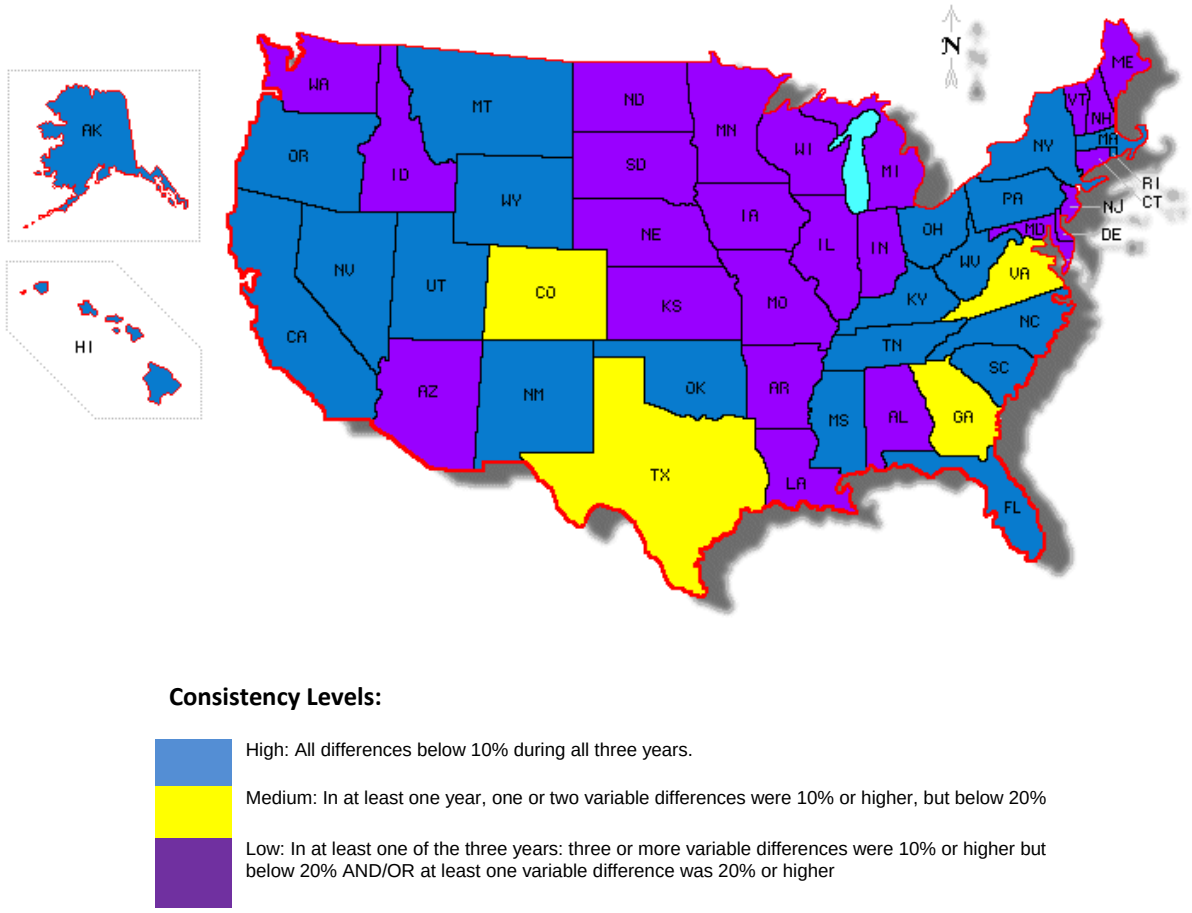
SOURCE: Institute of Museum and Library Services, Survey of Public Libraries in the United States, Fiscal Years 2006-2008

Table 7 complements the detailed display in Table 4, providing a concise overview and sharper contrast of the trends in comparing old and new imputation methods for PLS variables during FY2006-08. Some of the findings are similar to those observed for Table 4, but the simplified display allows trends to be spotted more easily.

- Though there was high consistency between old and new PLS imputation methods over the three fiscal years for the state government revenue (stgvt) variable, the results also showed medium consistency for local government revenue (locgvt) and low consistency for federal government revenue (fedgvt). This finding is consistent with prior observation of these revenues, since allocation of state and local government revenues tend to be more even than federal revenues, which can vary with economic conditions and/or election results.
- Using the old PLS imputation methods, some variables were grouped together so that all detail variables were imputed to a zero value if a total variable had a reported value of zero. However, this practice was *not* continued in the new imputation methods. These variables included four capital revenue variables (lcap_rev, scap_rev, fcap_rev, ocap_rev) and two materials expenditures variables (prmatexp, elmatexp). Table 7 shows a low consistency level for four of the six variables and a medium consistency level for one of the six variables over the three fiscal years. The above findings are consistent with prior experience for capital expenditures, since these tend to be among the most volatile variables in economic surveys.
- Confirming earlier observations, the PLS variable for print subscriptions (subscrp) shows a high consistency between old & new imputation methods over the three fiscal years, but the PLS variable for electronic subscriptions (esubscrp) shows low consistency. These findinds could be explained since electronic publications are still a relatively new medium while print publications are a mature medium. It should be noted that the PLS electronic subscriptions variable will be phased out by FY2010.
- As expected, highly reported PLS variables for total librarians (libraria) and librarians with master's degrees (master) were assigned to the high consistency level over the three fiscal years, but similar variables for all other paid staff (othpaid) and total paid employees (totstaff) were assigned to the medium consistency level. These results can be explained by better records available for the number of librarians and masters degrees versus other paid staff and total employees.
- Although the PLS variable for number of general public internet computers (gpterms) was assigned to the high consistency level, the similar variable for number of public internet computer users per year (pitusr) was assigned to the low consistency level. These results are not surprising since the number of public internet computer users per year has been difficult to count in the past.

Consistency levels in Table 6 were then applied to the aggregated PLS variable comparisons from Table 5 over FY2006-08 to create an image map of the US. The resulting image is shown in Figure 1.

Figure 1: Overall Consistency between States for Old/New PLS Imputation Methods: FY2006-08



SOURCE: Institute of Museum and Library Services, Survey of Public Libraries in the United States, Fiscal Years 2006-2008

The geographic display in Figure 1 complements Table 5 and is helpful in locating where clustering of the consistency levels is occurring among the states. Some of the findings from Figure 1 include:

- Generally, a higher degree of consistency was observed between old and new PLS imputation methods for eastern states, with the exception of New England and mid Atlantic states.
- Higher consistency was also observed between old and new PLS imputation methods for western states (along with Alaska and Hawaii).
- Lower consistency was observed between old and new PLS imputation methods for Midwestern states.

Conclusions

We've prepared a hierarchy of tools to help analysts sort through the large volume of information that compare the results of old and new PLS imputation methods in this bridge study. The comparison Excel spreadsheets form the foundation, while the year-by-year displays in Tables 4 and 5 are positioned at a second level. Finally, the consolidated displays in Table 7 and Figure 1, which aggregate comparison data over all three fiscal years, are at the top of the hierarchy. These comparison tools should help analysts quickly assess the differences between old and new imputation methods and help answer the most common questions.

Though a great deal of effort was made to allow for a smooth transition from old to new imputation methods, important differences remain and these can't be overlooked when comparing the bridge study results over FY2006-08. Some of these changes include:

- Differences in forming imputation cells: the old methods form cells using manual methods while the new methods use an automated approach (cumulative root method) to form imputation cells.

- Changes to the imputation methods used at the PLS variable level: The transition from old to new PLS imputation methods affects some variables more than others. If imputation methods for a particular variable underwent drastic changes and that variable had a large number of missing values, that could lead to large discrepancies in state estimates and eventually to low consistency levels in Tables 4 and 7. More information about differences between old and new PLS imputation methods by variable are provided in Attachment A.

- Differences in the implementation of detail checks: the old methods performed detail versus total checks for some variables at the outset of imputation while new methods perform similar checks for the same variables after an initial round of imputation. As mentioned earlier, the results from Tables 4 and 7 show that these variables (lcap_rev, scap_rev, fcap_rev, ocap_rev, prmatexp, elmatexp) can have relatively large discrepancies between old and new imputation methods.

These differences should be considered when evaluating the results of this bridge study.

References

Brown, I. (2009), "Evaluation of Alternative Imputation Methods for the Public Libraries Survey (PLS)". Internal report. U.S. Census Bureau.

Comparing Old and New Imputation Methods by PLS Variable

Imputation Methods:

Method 1: Prior year data with a cell mean growth rate: Prior year reported values are pulled forward and growth rates are applied to them. The growth rate is determined by calculating the mean of the growth rates of all respondents in an imputation cell. When reported values for the immediate prior year are not available, reported values from prior years two, three, or four are pulled forward and the growth rate for that year is applied to them.

Method 2: Prior year data with a hot-deck growth rate: The hot-deck procedure uses the growth rate of a respondent in the same imputation cell that is next in order to the non-respondent when ordered by population of legal service area. Prior year reported values are pulled forward and hot-deck growth rates are applied to them. When reported values for the immediate prior year are not available, reported values from prior years two, three, or four are pulled forward and the growth rate for that year is applied to them.

Method 3: An adjusted cell mean: The mean of all respondent values in an imputation cell for that library is adjusted for size by taking the ratio of the library's population to the cell mean of the population.

Method 4: The cell mean: The mean of all respondent values in an imputation cell for that library is used to calculate the imputed value.

Method 5: Prior year ratio to another item: The current year value of a highly correlated item is multiplied by the ratio of the prior year value of the item to be imputed to the prior year value of the highly correlated item. When reported values for the immediate prior year are not available, we compute the ratio using reported values from two, three, or four years prior.

Method 6: Cell median ratio with another item: The current year value of a highly correlated item is multiplied by the cell median ratio. The cell median ratio is calculated by finding the median of all ratios of the item to be imputed to the highly correlated item for all respondents in a cell.

Method 7: Prior year data with no growth rate (direct substitution): The prior year reported value is pulled forward for this item. When reported values for the immediate prior year are not available, pull forward reported values from two, three, or four years back.

Method 8: Cell median: The median of all respondent values in an imputation cell for that library is used to calculate the imputed value.

Method 9: Sequential hot-deck: The response from the next respondent, when ordered by population of legal service area, after the non-respondent from the same imputation cell is used.

Variable	Old Methods	New Methods
Central Library (CENTLIB)	1,3,4	1,3,4
Branch Library (BRANLIB)	1,3,4	1,3,4
Bookmobiles (BKMOB)	1,3,4	1,3,4

Variable	Old Methods	New Methods
Master (MASTER)	1,3,4	1,3,4 (sum)
Librarians (LIBRARIA)	1,3,4	1,3,4 (rake)
Other Paid Employees (OTHPAID)	1,3,4	1,3,4 (rake)
Total Staff (TOTSTAFF)	=LIBRARIA+OTHPAID	=LIBRARIA+OTHPAID [7,3]
Local Government (LOCGVT)	2,5,8	1,3,4
State Government (STGVT)	2,5,8	1,8
Federal Government (FEDGVT)	2,5,8	8
Other Revenue (OTHINCM)	2,5,8	1,8
Total Operating Revenue (TOTINCM)	=LOCGVT+STGVT+FEDGVT+OTHINCM	1,3,4
Salaries (SALARIES)	5, (special imputation method), 3,4	1,3,4 (sum)
Benefits (BENEFIT)	=STAFFEXP-SALARIES	5(STAFFEXP), 3,4 (rake)
Total Staff Expenditures (STAFFEXP)	1,3,4	1,3,4 (sum)
Print Materials Expenditures (PRMATEXP)	Set to 0 if total is 0,5,5(different item), 1,7,8	6(TOTEXPCO), 6(imputed TOTEXPCO), 3, 4 (sum)
Electronic Materials Expenditures (ELMATEXP)	Set to zero if total is 0,1,8,3,4	5(TOTEXPCO),3,4 (sum)
Other Materials Expenditures (OTHMATEX)	Special impute, 5,1,7,8 or =TOTEXPCO- (PRMATCXP+ELMATEXP)	1,3,4 (sum)
Total Collection Expenditures (TOTEXPCO)	1,3,4	1,3,4 (sum)
Other Operating Expenditures (OTHOPEXP)	=TOTOPEXP-(TOTSTAFF+TOTEXPCO)	1,3,4 (sum)
Total Operating Expenditures (TOTOPEXP)	1,3	1,3,4 (sum)
Local Capital Revenue (LCAP_REV)	If total is 0 set to 0,5,8	1,8, 11(rake)

Variable	Old Methods	New Methods
State Capital Revenue (SCAP_REV)	If total is 0 set to 0,5,8	8 (rake)
Federal Capital Revenue (FCAP_REV)	If total is 0 set to 0,5,8	8 (rake)
Other Capital Revenue (OCAP_REV)	If total is 0 set to 0 or =CAP_REV-(LCAPREV+SCAP_REV+FCAP_REV)	8 (rake)
Total Capital Revenue (CAP_REV)	1,5,8	1,8 (sum)
Total Capital Expenditures (CAPITAL)	1on TOTEXP then =TOTEXP-TOTOPEXP	7 on TOTEXP then =TOTEXP-TOTOPEXP or if no prior data then 3 for CAPITAL
Print Materials (BKVOL)	1,3,4	1,3,4
Electronic Books (EBOOK)	1,8	7,8
Audio (AUDIO)	1,3,4	1,3,4
Video (VIDEO)	3,4,8	1,3,4
State Databases (DB_ST)	New item	6(DATABASE), 3,4 (rake)
Local Databases (DB_LOC)	New item	8 (rake)
Other Databases (DB_OTH)	New item	1,8 (rake)
Databases (DATABASE)	1,8	3,4 (sum)
Print Subscriptions (SUBSCRIP)	1,3,4	3,4
Electronic Subscriptions (ESUBSCRP)	1,7,8	1,3,4
Hours Open (HRS_OPEN)	If outlets same 1, not same mean hrs by outlet type	5(outlets), mean hrs_open adjusted by outlets, 3,4
Library Visits (VISITS)	1, Population * total visits/total population, 4	1,3,4
Reference Transactions (REFERENC)	1,8	3,4

Variable	Old Methods	New Methods
Total Circulation (TOTCIR)	1,8	1,3,4
Children's Circulation (KIDCIRCL)	5,6	6(TOTCIR), 6(imputed TOTCIR), 3,4
Loans To (LOANTO)	5,8	5(LOANFM), 6(imputed LOANFM), 3,4
Loans From (LOANFM)	5,8	6(LOANTO), 3, 4
Total Library Programs (TOTPRO)	5,6,1,8	6(KIDPRO), 6(imputed KIDPRO) 3,4
Children's Programs (KIDPRO)	5,6,1,8	6(TOTPRO), 1,3,4
Total Attendance at Library Programs (TOTATTEN)	5,1,8	1,3,4
Children's Program Attendance (KIDATTEN)	5,6,1,8	6 (TOTATTEN), 6(imputed TOTATTEN) 3, 4
General Public Terminals (GPTERMS)	7,8	5(VISITS), 3,4
Electronic Users (PITUSR)	If no GPTERMS then zero, if yes 5,6	If GPTERMS=0 then 0, if yes 6(GPTERMS), 3,4
Registered Borrowers (REGBOR)	New item	1,3,4

Attachment B: PLS Variable Glossary

popu_lsa	Population of the Legal Service Area
centlib	Number of Central Libraries
branlib	Number of Branch Libraries
bkmob	Number of Bookmobiles
master	Librarians with master's degrees
libraria	Total Librarians
othpaid	All other Paid Staff
totstaff	Total Paid Employees
locgvt	Local Government Revenue
stgvt	State Government Revenue
fedgvt	Federal Government Revenue
othincm	Other Operating Revenue
totincm	Total Operating Revenue
salaries	Salaries & Wages Expenditures
benefit	Employee Benefit Expenditures
staffexp	Total Staff Expenditures
prmatexp	Print Materials Expenditures
elmatexp	Electronic Materials Expenditures
othmatexp	Other Materials Expenditures
totexpco	Total Collection Expenditures
othopexp	Other Operating Expenditures
totopexp	Total Operating Expenditures
lcap_rev	Local Government Capital Revenue
scap_rev	State Government Capital Revenue

fcap_rev	Federal Government Capital Revenue
ocap_rev	Other Capital Revenue
cap_rev	Total Capital Revenue
capital	Total Capital Expenditures
bkvol	Print Materials
ebook	Electronic Books (E-Books)
audio	Audio
video	Video
db_loc	Local licensed Databases
db_st	State licensed Databases
db_oth	Other licensed Databases (cooperative agreements within state/region)
database	Total licensed Databases
subscrip	Current Print Serial Subscriptions
esubscrp	Current Electronic Serial Subscriptions
hrs_open	Public Service Hours per Year
visits	Library Visits
referenc	Reference Transactions
regbor	Number of Registered Borrowers
totcir	Total Circulation
kidcircl	Circulation of Children's Materials
loanto	Inter-Library Loans Provided To
loanfm	Inter-Library Loans Received From
totpro	Total Number of Library Programs
kidpro	Number of Children's Programs
totatten	Total Attendance at Library Programs

kidatten	Children's Program Attendance
gpterms	Number of Internet Computers used by General Public
pitusr	Number of users of Public Internet Computers per Year