

STATISTICAL RESEARCH DIVISION REPORT SERIES
Statistical Research Report Number:
CENSUS/SRD/RR-85/23

Content Evaluation of the 1982 Economic Censuses
Petroleum Distributors

by

Carol Corby
Statistical Research Division
U.S. Bureau of the Census
Rm. 3554, F.O.B. #3
Washington, D.C. 20233

(301) - 763 - 7088

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Report Recommended by: Nash Monsour
Report Completed: April 8, 1986
Report Issued: April 9, 1986

1. Executive Summary

The primary objective of the content evaluation was to measure the accuracy of census data tabulated from respondents to the 1982 Census of Wholesale Trade. Wholesale petroleum distributors, Standard Industrial Classifications (SIC's) 5171 and 5172, were chosen for the study because they were known to have serious reporting problems.

Intensive personal visit interviews were conducted for a sample of establishments, during which highly accurate data (reinterview) were collected for three census items: total dollar volume of business (sales), operating expenses, and sales by commodity line. The reinterview data were compared to data tabulated in the census for sample establishments using ratio estimators. The table below shows the ratios for total sales and operating expenses.

Ratios of $\frac{\text{Reinterview Data}}{\text{Tabulated Data}}$
(standard errors)

Census item	SIC 5171	SIC 5172
Total Sales	1.129 (0.085)	0.868 (0.037)
Operating Expenses	1.111 (0.029)	0.793 (.056)

In the process of constructing the reinterview data, the kinds of errors made by respondents were examined. Errors made on individual components of sales and operating expenses did not contribute much to total response errors. However, respondents' use of estimates on census forms, or failure to report sales or operating expenses at all, accounted for the largest subtotals of reporting errors.

2. Results

The content evaluation studied the accuracy of several items on 1982 Economic Censuses forms completed by wholesale petroleum distributors, Standard Industrial Classifications (SIC) 5171 and 5172. These items were dollar volume of business (total sales) including commission business, operating expenses, and sales by commodity line. These items are reproduced from the census form CB-5109 in Appendix A. The main objective of the study was to estimate the accuracy of the respondent (to the census) portion of the published U.S. totals for each of these items. Highly accurate values, referred to as reinterview data, for these items were obtained from a sample of establishments during intensive personal visit interviews. Estimates of totals for an item were then computed from the sample using each of three versions of data: values as originally reported on census forms, called reported data, values as used in census tabulations after census processing, called tabulated data, and reinterview data. Ratios of pairs of these totals were then computed as shown in Table 1 for total sales and operating expenses.

Table 1. Ratios of Estimated Totals for
Dollar Volume of Business (Sales) and Operating Expenses

	$\hat{R}_1 = \frac{\text{reinterview}}{\text{tabulated}}$ (se of \hat{R}_1)	$\hat{R}_2 = \frac{\text{tabulated}}{\text{reported}}$ (se of \hat{R}_2)	$\hat{R}_3 = \frac{\text{reinterview}}{\text{reported}}$ (se of \hat{R}_3)
SIC 5171			
Sales	1.129 (.085)	0.987 (.026)	1.095 (.098)
Operating Expenses	1.111* (.029)	0.825* (.050)	0.948 (.044)
SIC 5172			
Sales	.868* (.037)	1.557# (.461)	0.908 (.132)
Operating Expenses	0.793* (.056)	0.425@* (.229)	0.889** (.061)

* Significantly different from 1.0 at $\alpha = 0.05$.

** Significantly different from 1.0 at $\alpha = 0.10$.

When one outlier is removed, this ratio becomes 1.107 (.029)*.

@ When one outlier is removed, this ratio becomes 1.160 (.090)**.

In looking at the first column of Table 1, the ratios of reinterview data to tabulated data, the respondent portion of the published totals for both sales and operating expenses are lower than they should be for SIC 5171 and higher than they should be for SIC 5172. Three of these ratios are significantly different from 1.0, indicating that the published data for respondents is significantly in error for these items.

The ratios \hat{R}_1 , \hat{R}_2 , and \hat{R}_3 can be looked at together as an indication of what may be happening in census processing. The combinations of ratios

indicate the following:

For SIC 5171 sales: tabulated data < reported data < reinterview data
 For SIC 5171 operating expenses: tabulated data < reinterview data < reported data,
 For SIC 5172 sales: reinterview data < reported data < tabulated data,
 For SIC 5172 operating expenses: reinterview data < reported data < tabulated data,

For SIC 5171 sales and SIC 5172 sales and operating expenses, census editing is changing the originally reported data in the wrong direction. For SIC 5171 operating expenses, census editing is changing the reported data in the right direction but too far.

The reinterview values for each census item were reconstructed from originally reported data by making corrections for errors made by respondents on individual components of the census item. For example, business insurance costs should be included in reported operating expenses according to the census definition of the item. When a respondent indicated that insurance costs had been left out, the amount of the error was obtained and later added in to reconstruct the reinterview operating expenses for the respondent's establishment. Tables 2 through 5 summarize the corrections made to originally reported values for sales and operating expenses.

The first two rows of Tables 2-5 represent the overall quality of the originally reported numbers. The first row represents cases which left the sales or operating expenses item on their census forms blank, but based on the content evaluation interviews, should have reported figures (item nonresponse on the census form). The second row represents cases which reported estimated figures on the census forms but gave the interviewer better total figures. The rest of the rows represent particular components of the census items. The census definitions for the sales and operating expenses items determined whether each component should have been included in or excluded from the reported total sales or operating expenses figures. The columns of the tables are defined as follows:

- Number Eligible - weighted count of establishments for which a component was applicable, and therefore the respondents had an opportunity to make a mistake.
- Number of Errors - weighted count of establishments which included or excluded a component erroneously when providing the total figure for sales or operating expenses.
- Number of Reliable Corrections - Weighted count of establishments which provided book figures or reliable estimates for corrections to the component. The difference between this column and Number of Errors represents the establishments which could not provide a correction or provided corrections that were judged by the interviewer to be unreliable.
- Total Correction - Weighted total, in thousands of dollars, of the book or reliable corrections for the component.

Table 2. Components for Total Sales - SIC - 5171
(549 unweighted, 8,886 weighted establishments included in table)

Component	Number of Eligible Establishments	Number of Errors	Number of Reliable Corrections	Total Correction (\$1,000's)
1. Item Nonresponse on Census Form	324	324	324	\$ 13,297,824
2. Correction for Estimation on Census Form	8524	1279	1279	9,024,642
3. Receipts for services (should be included)	2899	225	174	8,036
4. Receipts for goods delivered in 1982 but not paid in 1982 (should be included)	6495	76	22	2,129
5. Goods delivered before 1982, paid for in 1982 (should be excluded)	5623	459	333	-158,207
6. Carrying or other credit charges (should be excluded)	4142	1438	1267	-46,960
7. Nonoperating income (should be excluded)	3563	938	926	-47,913
8. Sales abroad of goods that never entered U.S. territory (should be excluded)	316	1	0	0
9. Export sales (should be included)	111	0	0	0
10. Sales or excise taxes collected from customers by establishment (should be excluded)	6326	2591	2012	-997,562
11. Excise taxes paid before or at time that establishment purchased goods (should be included)	2474	144	86	81,282
12. Refunds, discounts, allowances (should be deducted)	4648	1137	632	-104,724
13. Value of trade-ins accepted as partial payment for goods (should be included)	369	67	33	270
14. Gross selling value of commission business including commissions, brokerage, or agent fees (should be included)	340	195	195	18,981
TOTAL				\$ 21,077,258

Table 3. Components for Total Sales - SIC - 5172
(144 unweighted 1,129 weighted establishments included in table)

Component	Number of Eligible Establishments	Number of Errors	Number of Reliable Corrections	Total Correction (\$1,000's)
1. Item Nonresponse on Census Form	22	22	22	\$ 2,160,171
2. Correction for Estimation on Census Form	1094	94	94	-1,445,840
3. Receipts for services (should be included)	268	13	11	844
4. Receipts for goods delivered in 1982 but not paid in 1982 (should be included)	800	26	26	3,018
5. Goods delivered before 1982, paid for in 1982 (should be excluded)	677	76	41	-5,575
6. Carrying or other credit charges (should be excluded)	172	56	46	-3,802
7. Nonoperating income (should be excluded)	363	58	58	-10,301
8. Sales abroad of goods that never entered U.S. territory (should be excluded)	52	12	4	-1,290,774
9. Export sales (should be included)	63	1	0	0
10. Sales or excise taxes collected from customers by establishment (should be excluded)	502	139	137	-49,378
11. Excise taxes paid before or at time that establishment purchased goods (should be included)	210	23	0	0
12. Refunds, discounts, allowances (should be deducted)	411	75	68	-11,386
13. Value of trade-ins accepted as partial payment for goods (should be included)	26	0	0	0
14. Gross selling value of commission business including commissions, brokerage, or agent fees (should be included)	150	128	128	<u>3,080</u>
TOTAL				\$ -649,943

Table 4. Components of Operating Expenses - SIC - 5171
 (537 unweighted, 8,550 weighted establishments included in table)

Component	Number of Eligible Establishments	Number of Errors	Number of Reliable Corrections	Total Correction (\$1,000's)
1. Item nonresponse on census form	791	791	791	\$ 218,272
2. Correction for estimation on census form	7759	1009	1009	-2,070,548
3. Withdrawals by or payments to proprietors or partners (should be excluded)	582	63	63	-1,631
4. Gross payroll (should be included)	8005	266	233	41,833
5. Employer's contributions to legally required governmental programs - FICA etc. (should be included)	8148	192	80	2,219
6. Employer's contributions to voluntary programs or negotiated benefits (should be included)	5363	245	211	4,401
7. Business insurance costs (should be included)	7415	179	95	941
8. Sales or excise taxes collected from customers (should be excluded)	5954	594	408	-16,568
9. Fines, license fees, taxes (should be included)	7738	325	257	3,246
10. Cost of goods bought for resale (should be excluded)	8376	154	86	-252,098
11. Costs of office supplies (should be included)	8468	43	23	487
12. Costs for storing or shipping inventory (should be included)	5149	427	353	60,085
13. Costs of utilities (should be included)	8470	87	46	413
14. Losses from theft, damage, bad debts (should be included)	5094	631	578	15,668
15. Costs for fuel purchased and consumed by establishment (other than highway vehicles) (should be included)	4255	1	0	0

Table 4. Components of Operating Expenses - SIC - 5171
(Continued)

Component	Number of Eligible Establishments	Number of Errors	Number of Reliable Corrections	Total Correction (\$1,000's)
16. Costs for purchased advertising (should be included)	6214	216	169	566
17. Commissions paid to employees not included in payroll (should be included)	642	32	32	64
18. Commissions or subcontract fees paid to persons or firms outside the company (should be included)	1675	211	140	15,783
19. Repair service payments to persons or firms outside the company (should be included)	7183	104	103	1,462
20. Capital leasing payments (should be excluded)	982	585	470	-12,416
21. Depreciation charges against capital leases (should be included)	1676	241	182	5,141
22. Depreciation or amortization charges (should be included)	5528	465	386	9,413
23. Value of or payments for depreciable assets (should be excluded)	5331	253	201	-15,685
24. Lease or rental payments by multi-unit establishments to parent company (should be excluded)	596	219	208	-7,235
25. Rent for buildings or equipment owned by another company (should be included)	2565	130	109	2,419
26. Payment for leasing of land (should be excluded)	1334	1094	1023	-9,252
27. Costs of repairs covered by lease payments (should be excluded)	444	76	76	-203
28. Costs for utilities covered by lease payments (should be excluded)	213	63	42	0
TOTAL				\$ -2,013,505

Table 5. Components of Operating Expenses - SIC - 5172
(146 unweighted, 1,163 weighted establishments included in table)

Component	Number of Eligible Establishments	Number of Errors	Number of Reliable Corrections	Total Correction (\$1,000's)
1. Item nonresponse on census form	54	54	54	\$ 20,635
2. Correction for estimation on census form	1109	157	157	-2,729
3. Withdrawals by or payments to proprietors or partners (should be excluded)	37	5	5	-1,857
4. Gross payroll (should be included)	1037	0	0	0
5. Employer's contributions to legally required governmental programs - FICA etc. (should be excluded)	978	6	6	13
6. Employer's contributions to voluntary programs or negotiated benefits (should be included)	590	11	7	167
7. Business insurance costs (should be included)	890	8	5	535
8. Sales or excise taxes collected from customers (should be excluded)	501	37	27	-1,569
9. Fines, license fees, taxes (should be included)	955	5	1	30
10. Cost of goods bought for resale (should be excluded)	1140	18	16	-33,660
11. Costs of office supplies (should be included)	1163	0	0	0
12. Costs for storing or shipping inventory (should be included)	432	34	24	13,120
13. Costs of utilities (should be included)	1163	0	0	0
14. Losses from theft, damage, bad debts (should be included)	640	60	33	9,046
15. Costs for fuel purchased and consumed by establishment (other than highway vehicles) (should be included)	410	1	0	0

Table 5. Components of Operating Expenses - SIC - 5172
(Continued)

Component	Number of Eligible Establishments	Number of Errors	Number of Reliable Corrections	Total Correction (\$1,000's)
16. Costs for purchased advertising (should be included)	639	23	1	14
17. Commissions paid to employees not included in payroll (should be included)	79	0	0	0
18. Commissions or subcontract fees paid to persons or firms outside the company (should be included)	464	21	15	4,845
19. Repair service payments to persons or firms outside the company (should be included)	832	2	1	320
20. Capital leasing payments (should be excluded)	106	46	46	-9,457
21. Depreciation charges against capital leases (should be included)	319	1	0	0
22. Depreciation or amortization charges (should be included)	583	41	41	14,331
23. Value of or payments for depreciable assets (should be excluded)	542	14	11	-184
24. Lease or rental payments by multi-unit establishments to parent company (should be excluded)	75	38	25	-1,414
25. Rent for buildings or equipment owned by another company (should be included)	484	1	1	325
26. Payment for leasing of land (should be excluded)	62	60	25	-870
27. Costs of repairs covered by lease payments (should be excluded)	51	0	0	0
28. Costs for utilities covered by lease payments (should be excluded)	182	12	12	-268
TOTAL				\$ 11,373

In order to interpret Tables 2 through 5, it should be noted that the total correction for each component is understated due to Component 2, correction for estimation on census form. When respondents provided a new total figure for a reported census item the difference between the new and old total figures was not broken down into smaller components. Therefore Component 2 represents a composite of corrections for other components.

The reason for compiling Tables 2 through 5 was to determine whether any individual components were responsible for a large part of reporting error on each census item. In looking at Tables 2 and 3, however, components of total sales for SIC's 5171 and 5172, most of the individual components are of little importance and cancel out with each other. Component 10, sales or excise taxes, needed the largest correction for SIC 5171 sales, and Component 8, sales abroad of goods that never entered U.S. territory, needed the largest correction for SIC 5172 sales. Component 2 for these tables is interesting in that, for SIC 5171, the correction for estimation on the census form added nine billion dollars to total sales, almost half of the grand total addition to sales, while for SIC 5172, this correction subtracted one billion dollars from sales, part of which was cancelled out by additions to sales.

In looking at Tables 4 and 5, only one definitional component stands out, which is Component 10, cost of goods bought for resale, in Table 4, SIC 5171 operating expenses. Also in Table 4, the correction for estimation on the census form subtracted two billion dollars from total operating expenses, which is about the same amount as the grand total correction to reported operating expenses.

After looking at the ratios and components of sales and operating expenses, it appears that the best way to improve the accuracy of the reported published sales and operating expenses is to first encourage respondents to

report book figures rather than estimates. After that, census processing must be designed to avoid compounding respondents' errors.

The sales by commodity line (Item 13, see Appendix A) item on the census form requests a breakdown of each establishments total sales into separate figures for each commodity commonly handled by petroleum wholesalers, such as aviation gasoline, motor gasoline, jet fuel, and related products like marketing equipment or tires and tubes. Tables 6 and 7 show ratios of weighted reported, tabulated and reinterview sales in dollars and sales in gallons where appropriate, for each commodity line. To arrive at the published census total for a commodity line, say aviation gasoline, data were tabulated for dealers in aviation gasoline which reported figures that appeared to be reliable. This tabulated figure was then inflated to reflect all aviation gasoline dealers, including nonrespondents to the census. The ratios in Tables 6 and 7 are computed from establishments which were used in census tabulations for a commodity line and provided reinterview data to the content evaluation survey. For some commodity lines, this set of establishments was very small and for the commodity lines where less than 10 establishments fell into this category, the ratios were dropped from the tables.

The most interesting thing to be seen in Tables 6 and 7 is the wide variability in the ratios from one commodity line to another. For some lines, such as aviation gasoline for SIC 5171, reported, tabulated and reinterview data were very close together, while for others, such as lubricating oil and grease for SIC 5171 the three versions of the data were much farther apart. Also, the accuracy of sales in gallons varied differently from sales in dollars. And last, the gallons and dollars ratios were neither consistently above nor below 1.0 but part of these is explained by the fact that different

establishments were in each ratio. So Tables 6 and 7 provided a warning that ensuring the accuracy of published totals for individual commodity lines would be a very difficult task.

Table 6. Ratios of Reinterview, Tabulated and Reported Data
for Sales by Commodity Line
SIC 5171

	<u>Sales in Dollars</u>			<u>Sales in Gallons</u>		
	$\hat{R}_1 = \frac{\text{rein}}{\text{tab}}$	$\hat{R}_2 = \frac{\text{tab}}{\text{rep}}$	$\hat{R}_3 = \frac{\text{rein}}{\text{rep}}$	$\hat{R}_1 = \frac{\text{rein}}{\text{tab}}$	$\hat{R}_2 = \frac{\text{tab}}{\text{rep}}$	$\hat{R}_3 = \frac{\text{rein}}{\text{rep}}$
Aviation Gasoline						
Ratio	1.0005	1.0018	.9989	.9155	1.0564	.9999
No. of Cases (weighted)	389	810	298	426	417	351
No. of Cases (unweighted)	35	47	26	36	31	29
Motor Gasoline						
Ratio	1.0032	1.0251	.9994	1.0034	.9926	1.0000
No. of Cases (weighted)	5166	5167	4892	5266	4535	4588
No. of Cases (unweighted)	291	263	246	296	237	236
Special Naptha						
Ratio	.8534	.9958	.6947	.7964	.8905	1.0000
No. of Cases (weighted)	655	671	421	708	498	485
No. of Cases (unweighted)	38	37	25	39	27	26
Jet Fuel						
Ratio	1.1466	1.2076	1.0000	1.1697	.8724	1.0000
No. of Cases (weighted)	378	712	287	350	355	320
No. of Cases (unweighted)	41	52	34	41	36	35
No. 2 Distillate fuel						
Ratio	1.0640	1.1061	1.0183	1.0220	.9885	.9980
No. of Cases (weighted)	4717	4779	4370	4869	4241	4272
No. of Cases (unweighted)	273	247	225	277	224	220
All other Distillate fuel						
Ratio	1.0657	.9624	.9826	1.0037	.9951	.9913
No. of Cases (weighted)	3200	2991	2698	3253	2661	2698
No. of Cases (unweighted)	179	155	140	183	144	144
Residual fuel oil						
Ratio	.9756	1.0103	.9971	1.0104	.9822	.9969
No. of Cases (weighted)	274	691	287	268	262	244
No. of Cases (unweighted)	36	48	33	34	30	31
Lubricating oil and grease						
Ratio	1.0320	.9863	.9837	.9758	.9035	.9692
No. of Cases (weighted)	4506	4249	3963	3855	3531	3375
No. of Cases (unweighted)	244	192	179	207	161	151
Liquefied petroleum gases						
Ratio	.9620	1.1551	.9992	.9577	1.2109	.9992
No. of Cases (weighted)	460	715.	459	476	375	370
No. of Cases (unweighted)	34	39	26	36	22	23
Crude oil						
Ratio	1.0294	.9992	---	.9932	---	---
No. of Cases (weighted)	71	357	---	64	---	---
No. of Cases (unweighted)	13	17	---	11	---	---

Table 6. Ratios of Reinterview, Tabulated and Reported Data
for Sales by Commodity Line
SIC 5171 Continued

	<u>Sales in Dollars</u>		
	$\hat{R}_1 = \frac{\text{rein}}{\text{tab}}$	$\hat{R}_2 = \frac{\text{tab}}{\text{rep}}$	$\hat{R}_3 = \frac{\text{rein}}{\text{rep}}$
Other petroleum products			
Ratio	1.5373	.7668	.9915
Number of Cases (weighted)	116	522	198
Number of Cases (unweighted)	16	31	19
Automotive parts and supplies			
Ratio	1.1128	.9970	.9752
Number of Cases (weighted)	1491	1222	832
Number of Cases (unweighted)	88	62	47
Petroleum products marketing equipment			
Ratio	1.6000	.9327	---
Number of Cases (weighted)	332	483	---
Number of Cases (unweighted)	21	25	---
Tires and Tubes			
Ratio	1.0266	1.0001	.9358
Number of Cases (weighted)	1439	1532	1238
Number of Cases (unweighted)	82	71	55
Chemicals and Allied products			
Ratio	2.2456	1.0170	---
Number of Cases (weighted)	228	516	---
Number of Cases (unweighted)	13	26	---
Farm Supplies			
Ratio	.9521	1.2438	.9932
Number of Cases (weighted)	313	527	215
Number of Cases (unweighted)	15	26	12
Rental receipts			
Ratio	1.6272	.9877	1.0090
Number of Cases (weighted)	1627	730	542
Number of Cases (unweighted)	108	44	36
Service receipts and labor charges			
Ratio	1.0412	.9829	.9738
Number of Cases (weighted)	1150	1060	739
Number of Cases (unweighted)	78	54	42

Table 7. Ratios of Reinterview, Tabulated and Reported Data
for Sales by Commodity Line
SIC 5172

	<u>Sales in Dollars</u>			<u>Sales in Gallons</u>		
	$\hat{R}_1 = \frac{\text{rein}}{\text{tab}}$	$\hat{R}_2 = \frac{\text{tab}}{\text{rep}}$	$\hat{R}_3 = \frac{\text{rein}}{\text{rep}}$	$\hat{R}_1 = \frac{\text{rein}}{\text{tab}}$	$\hat{R}_2 = \frac{\text{tab}}{\text{rep}}$	$\hat{R}_3 = \frac{\text{rein}}{\text{rep}}$
Motor Gasoline						
Ratio	1.0073	.9815	1.0001	.9390	1.1181	.9766
No. of Cases (weighted)	491	339	187	475	243	168
No. of Cases (unweighted)	47	54	37	45	39	32
No. 2 Distillate fuel						
Ratio	1.0390	.9586	1.0060	1.0118	.9904	.9988
No. of Cases (weighted)	463	282	172	434	213	147
No. of Cases (unweighted)	47	53	38	44	39	33
All other distillate fuel						
Ratio	1.0149	.9730	1.0010	1.0032	.9999	1.0026
No. of Cases (weighted)	308	162	79	263	109	65
No. of Cases (unweighted)	30	32	19	26	20	16
Residual fuel oil						
Ratio	.9660	1.6388	1.0007	1.0105	1.0000	1.0004
No. of Cases (weighted)	105	121	75	92	52	52
No. of Cases (unweighted)	17	21	15	15	12	12
Lubricating oil and grease						
Ratio	.9883	1.0084	.9882	1.0339	.9988	.9995
No. of Cases (weighted)	423	306	161	253	202	105
No. of Cases (unweighted)	45	54	35	32	34	25
Liquefied petroleum gases						
Ratio	.9764	1.0006	.9815	1.0412	.9619	.9965
No. of Cases (weighted)	65	63	56	65	53	56
No. of Cases (unweighted)	11	16	12	11	12	12
Crude oil						
Ratio	.9704	1.3505	1.0000	---	.6553	---
No. of Cases (weighted)	80	121	80	---	44	---
No. of Cases (unweighted)	17	26	17	---	12	---

Table 7. Ratios of Reinterview, Tabulated and Reported Data
for Sales by Commodity Line
SIC 5172 Continued

Sales in Dollars

$$\hat{R}_1 = \frac{\text{rein}}{\text{tab}} \quad \hat{R}_2 = \frac{\text{tab}}{\text{rep}} \quad \hat{R}_3 = \frac{\text{rein}}{\text{rep}}$$

	$\hat{R}_1 = \frac{\text{rein}}{\text{tab}}$	$\hat{R}_2 = \frac{\text{tab}}{\text{rep}}$	$\hat{R}_3 = \frac{\text{rein}}{\text{rep}}$
Other petroleum products			
Ratio	.9377	1.5427	---
Number of Cases (weighted)	50	54	---
Number of Cases (unweighted)	12	15	---
Automotive parts and supplies			
Ratio	1.5453	.9910	1.3831
Number of Cases (weighted)	80	50	43
Number of Cases (unweighted)	17	17	14
Tires and Tubes			
Ratio	1.0182	.9943	.9981
Number of Cases (weighted)	45	71	37
Number of Cases (unweighted)	12	15	10
Rental Receipts			
Ratio	1.0969	---	---
Number of Cases (weighted)	77	---	---
Number of Cases (unweighted)	12	---	---
Service Receipts and labor charges			
Ratio	1.0657	.9854	1.0000
Number of Cases (weighted)	47	120	56
Number of Cases (unweighted)	13	22	15

During editing, the SIC and single unit/multiunit status of each sample case was checked. It is interesting to note how often these classifications were wrong, since the processing of the census and the accuracy of the final tabulations depend on them. The two tables below show how the sample cases were classified for mailout of CB-5109 forms, in final census tabulations, and by the reinterview. Table 8 shows the SIC breakdown for 885 establishments for which a reinterview classification could be made (refusals, for example, are left out). Table 9 shows the single unit/multiunit status for the 759 cases for which interviews were completed.

**Table 8. Mailout, Tabulations, and Reinterview Classifications
(number of establishments)**

REINTERVIEW SIC 5171

	Tabulation SIC 5171	Tabulation SIC 5172	Tabulation SIC Out-of-Scope	Tabulation SIC Unknown
Mailout SIC 5171	<u>364</u>	44	10	3
Mailout SIC 5172	130	42	6	7

REINTERVIEW SIC 5172

	Tabulation SIC 5171	Tabulation SIC 5172	Tabulation SIC Out-of-Scope	Tabulation SIC Unknown
Mailout SIC 5171	10	26	1	1
Mailout SIC 5172	10	<u>98</u>	5	2

REINTERVIEW SIC OUT-OF-SCOPE

	Tabulation SIC 5171	Tabulation SIC 5172	Tabulation SIC Out-of-Scope	Tabulation SIC Unknown
Mailout SIC 5171	25	6	20	6
Mailout SIC 5172	13	24	26	6

Notice that only 462 of the 885 establishments (364 + 98, underlined) were mailed out and tabulated in the correct SIC, as determined by reinterview. This means that 423 cases, almost half, either were reclassified during processing or should have been reclassified.

**Table 9. Single Unit/Multiunit Classifications
(number of sample cases)**

REINTERVIEW SINGLE UNITS		
	Tabulation Single Units	Tabulation Multiunits
Mailout Single Units	<u>274</u>	11
Mailout Multiunits	0	8
REINTERVIEW MULTIUNITS		
	Tabulation Single Units	Tabulation Multiunits
Mailout Single Units	71	30
Mailout Multiunits	0	<u>365</u>

In Table 9, 120 of the sample cases were or should have been reclassified during census processing. Reclassification of a single unit to a multiunit is often difficult to do correctly during census processing, because data for a whole company must be split into subtotals for each establishment. Both Tables 8 and 9 demonstrate the extra problems associated with producing accurate tabulations for individual SIC's and counties beyond respondent errors as shown in Tables 2 through 5.

3. The Sample Design

The universe for this study was the set of establishments which were classified in SIC's 5171 or 5172 before the census, and after responding to the census, were still in SIC's 5171 and 5172. The first level of

stratification of establishments was by single-unit and multi-unit status. Establishments in the single-unit stratum were arranged in 128 primary sampling units (PSU's) which were formed from groups of contiguous counties. A small number of single units were separated from the PSU's and included in the sample with certainty due to their large size. The PSU's were stratified, based on employment and first quarter payroll, into 14 strata. One stratum, containing two PSU's, was designated the certainty stratum and both PSU's were later subject to within-PSU sampling. One PSU was sampled from each of the 13 non-certainty strata using probability proportional to first quarter payroll. Within the sample PSU's and the PSU's from the certainty stratum, establishments were further stratified into two to five substrata for each of SIC 5171 and 5172 based on annual payroll. The substrata of establishments with the largest payroll were designated certainty substrata within the PSU's and all establishments were included in the sample. Establishments were sampled systematically within the noncertainty substrata.

For multiunits, large establishments were first identified by payroll cutoffs and assigned to certainty strata. The rest of the multiunits, designated the noncertainty strata for SIC's 5171 and 5172, were further stratified by payroll size. Systematic sampling was used within these substrata.

In order to conduct interviews within a reasonable time frame, it was necessary to draw the samples from the file of names and addresses used for mailing out census forms. Because of this, the sample included establishments which never responded to the census or which were transferred to other SIC's during census processing, both types being out-of-scope of the content evaluation. In order to ensure that a sufficient number of in-scope establishments would be sampled, the sample sizes needed for estimation were

inflated to cover the anticipated percentage of out-of-scope cases. Table 10 below summarizes the sample sizes and the weights associated with noncertainty establishments.

Table 10. Sample Sizes and Weights

<u>Stratum</u>		Number of Establishments in Inflated <u>Sample</u>	Number of In-scope Establishments <u>in Sample</u>	Range of Weights Over Noncertainty <u>Substrata</u>
SIC 5171	Single units	434	267	2.25 - 41.93
	Multiunits	294	227	2.50 - 67.08
	Total	728	494	2.25 - 67.08
SIC 5172	Single units	285	166	2.80 - 18.5
	Multiunits	293	166	2.50 - 22.03
	Total	578	332	2.50 - 22.03

Details of the universe and sample design can be obtained from a sampling and estimation memorandum from Glenn Wolfgang to Carol Corby, July 16, 1985.

4. Methodology

Early in 1982, planning for the content evaluation began with the choice of wholesale petroleum distributors, a kind of business (KB) known to have serious reporting problems, as the target of the study. Questionnaires were developed and interviewers were trained. Interviews were conducted between July, 1983 and July, 1984, followed by processing and analysis of the data. The following subsections describe these activities.

4.1 The Questionnaires

The census items studied in the content evaluation were from the CB-5109 census form: Item 6, Dollar Volume of Business (sales), Item 9, Operating Expenses, and Item 13, Sales by Commodity Lines (see Appendix A). The general style of the questionnaire was as follows: first, the originally reported figure for an item was verified and determined to be a book figure or an estimate. When an estimate had been reported, the interviewer requested a book figure or a better estimate. Then it was determined whether the figure

included each component that it should according to the census definition of the item, and similarly, that it excluded components that should have been left out of the census item. Each time an error was found, the amount of the error was requested. For example, if costs for purchased advertising were not included in Item 9, Operating Expenses, a value for advertising costs was requested. The amounts for each error were later used to correct the reported figure. It was typical to find that any one establishment made errors on at most one or two components of an item.

Each time an amount was requested and obtained, a subjective judgment was made of the quality of the amount, using probing questions and visual cues, such as whether the respondent actually took the figure from accounting records or guessed at the amount. Codes of B, R, or U for book figure, reliable estimate, or unreliable estimate were assigned to the amount based on the subjective judgment. In the event that the figure was unreliable or unavailable, or the interviewer was unsure of the quality of the figure, a request was made for a range of values that would indicate the lowest and highest values the true figure could take. During editing by SRD staff figures were compared to their ranges, and final judgements of the quality of each were made. If, after editing, the reinterview figure was still unreliable, it was not used in the analysis.

4.2 The Interviews

Voluntary personal visit interviews were conducted with the establishment or company employees who actually filled out the census forms or who were most familiar with the numbers reported on the forms. Personal visits were required mainly because of the need for judging the quality of numbers provided by respondents. These judgments could not be made using telephone interviews or a mailout-mailback questionnaire. After a two and one-half day

training session, Field Division interviewers conducted interviews at most single-unit establishments. Members of the professional staff from Census Bureau headquarters, primarily from Statistical Research Division and Business Division, conducted interviews for multi-unit establishments and large single units. The response rates for the content evaluation interviews are summarized below.

TABLE 11. Response Rates

<u>Stratum</u>		<u>Number of In-Scope Establishments#</u>	<u>Number of Completed Interviews (% of in-scope)</u>	<u>Number of Refusals (% of in-scope)</u>	<u>Number of Other Noninterviews* (% of in-scope)</u>
SIC 5171					
single units	unweighted	267	240 (89.9%)	11 (4.1%)	16 (6.0%)
	weighted	4712	4232 (89.8%)	160 (3.4%)	320 (6.8%)
multiunits	unweighted	227	219 (96.5%)	5 (2.2%)	3 (1.3%)
	weighted	4989	4773 (95.7%)	87 (1.7%)	129 (2.6%)
total	unweighted	494	459 (92.9%)	16 (3.2%)	19 (3.8%)
	weighted	9701	9005 (92.8%)	247 (2.5%)	449 (4.6%)
SIC 5172					
single units	unweighted	166	146 (88.0%)	9 (5.4%)	11 (6.6%)
	weighted	1110	1004 (90.5%)	48 (4.3%)	58 (5.2%)
multiunits	unweighted	166	154 (92.8%)	7 (4.2%)	5 (3.0%)
	weighted	1007	926 (92.0%)	20 (2.0%)	61 (6.1%)
total	unweighted	332	300 (90.4%)	16 (4.8%)	16 (4.8%)
	weighted	2117	1930 (91.2%)	68 (3.2%)	119 (5.6%)

from Table 10

* out-of-business, other cases that couldn't be reached

Given that the content evaluation interviews were voluntary, and that interviews lasted from 20 minutes to 90 minutes, the response rates shown in Table 9 were quite good. The effect of item nonresponse can be seen in Tables 2-5 in Section 2.

Several stages of processing were completed to get from raw questionnaire data to the final analysis. The first activity was a screening of the questionnaires for data from several establishments combined on one census form, cases which were actually in a different SIC, and major problems with the quality of the reinterview data. The more important problems that were uncovered by screening are documented in a memorandum from Glenn Wolfgang to Carol Corby, March 28, 1985.

The keying stage involved further editing of the questionnaires. Data were keyed on a microcomputer using a program that edited the data as they were keyed, prompting the keyer to type in answers to each appropriate question (the system resembles Computer Assisted Telephone Interviewing without the telephone). Invalid verbal responses were refused and skip patterns were checked. Then the data were keyed a second time, independently of the first keying. The two versions for each record were matched and any differences, usually due to keying errors, were reconciled.

The next stage completed two more reviews of the reinterview data: inspection of the high and low ranges on unreliable figures and verification of values for cases showing large differences between the reinterview figures and reported and tabulated figures. These edits ensured that only reliable reinterview data would be compared to tabulated data in the analysis. The last stage of data processing consisted of computer programming to prepare the file of establishment records for analysis (sorting, adding on weights, etc.) and to complete the ratio estimates.

4.3 The Estimators

This section presents a brief summary of the estimators used for the content evaluation excerpted from a full description included in a sampling and estimation memorandum from Glenn Wolfgang to Carol Corby, July 16, 1985. As was described in Section 2, the main results were a series of ratio estimates obtained from estimates of totals for reported data, reinterview data, and tabulated census data, for a given census item and SIC code. Let X and Y represent the numerator and denominator values appropriate for one of these ratios. Then

$$\hat{R} = \frac{\hat{X}}{\hat{Y}} \quad (4.1)$$

The general form for \hat{X} (and similarly for \hat{Y}) is

$$\hat{X} = \sum_{K=1}^2 \sum_{h=1}^18 \frac{1}{P_{Kh}} \sum_{j=1}^J W_{Kjh} \sum_{i=1}^{n_{Khj}} a_{Khji} X_{Khji} \quad (4.2)$$

where X_{Khji} is the census item value for establishment i in substratum j within stratum h of frame K . K marks the mail-out list (SIC 5171 or SIC 5172) from which the establishment was originally sampled. This allows use of data from establishments that belong in the kind of business being analyzed even when originally classified in the other frame.

a_{Khji} is an indicator variable (1 if case is valid for SIC being analyzed; 0 otherwise)

$X'_{Khji} = a_{Khji} X_{Khji}$, a simplification useful in later formulas

n_{Khj} number of establishments selected within substratum j of stratum h and frame K

N_{Khj} = $\sum a_{Khj}$; the total number of valid (for that SIC) establishments in substratum j of stratum h in frame K

W_{Khj} substratum weight (based on all p_{Khjt})

J_{Kh} number of substrata in stratum h of frame K

P_{Kh} inclusion probability for stratum h of frame K

It is useful to partition \hat{X} (and \hat{Y}) into five parts

$$\hat{X} = \hat{X}_S + \hat{X}_{SCP} + X_{SC} + \hat{X}_M + X_{MC} \quad (4.3)$$

where \hat{X}_S is the single-unit total derived from the first 13 strata, which exclude first-cut certainty cases and certainty PSU's

\hat{X}_{SCP} single-unit total from two certainty PSU's, strata 14 and 15

X_{SC} single-unit total for first-cut certainty cases, collected into stratum 16

\hat{X}_M multi-unit total derived from stratum 17 which excludes certainty cases

X_{MC} multi-unit certainty case total, collected in stratum 18

To derive a variance estimation formula in a form easy to compute, the overall estimator is first broken into components:

$$\hat{\text{Var}}(\hat{R}) = \hat{R}^2 \left[\frac{\hat{\text{Var}}(\hat{X})}{\hat{X}^2} + \frac{\hat{\text{Var}}(\hat{Y})}{\hat{Y}^2} - 2 \frac{\hat{\text{Cov}}(\hat{X}, \hat{Y})}{\hat{X}\hat{Y}} \right] \quad (4.4)$$

The numerators above may also be broken into components, i.e.

$$\hat{\text{Var}}(\hat{X}) = \hat{\text{Var}}(\hat{X}_S) + \hat{\text{Var}}(\hat{X}_{SCP}) + \hat{\text{Var}}(X_{SC}) + \hat{\text{Var}}(\hat{X}_M) + \hat{\text{Var}}(X_{MC}) \quad (4.5)$$

and similarly for $\widehat{\text{Var}}(\widehat{Y})$ and

$$\widehat{\text{Cov}}(\widehat{X}, \widehat{Y}) = \widehat{\text{Cov}}(\widehat{X}_S, \widehat{Y}_S) + \widehat{\text{Cov}}(\widehat{X}_{SCP}, \widehat{Y}_{SCP}) + \widehat{\text{Cov}}(\widehat{X}_{SC}, \widehat{Y}_{SC}) + \widehat{\text{Cov}}(\widehat{X}_M, \widehat{Y}_M) + \widehat{\text{Cov}}(\widehat{X}_{MC}, \widehat{Y}_{MC})$$

Since establishments providing \widehat{X}_{SC} , \widehat{X}_{MC} , \widehat{Y}_{SC} , and \widehat{Y}_{MC} are sampled with certainty, the variances of those sums have zero sampling error and their terms may be dropped from the formulae as in

$$\widehat{\text{Var}}(\widehat{X}) = \widehat{\text{Var}}(\widehat{X}_S) + \widehat{\text{Var}}(\widehat{X}_{SCP}) + \widehat{\text{Var}}(\widehat{X}_M) \quad (4.6)$$

When the variance and covariance expansions are substituted in (4.4)

$$\begin{aligned} \widehat{\text{Var}}(\widehat{R}) = \widehat{R}^2 & \quad \frac{\widehat{\text{Var}}(\widehat{X}_S)}{\widehat{X}^2} + \frac{\widehat{\text{Var}}(\widehat{X}_{SCP})}{\widehat{X}^2} + \frac{\widehat{\text{Var}}(\widehat{X}_M)}{\widehat{X}^2} \\ & + \frac{\widehat{\text{Var}}(\widehat{Y}_S)}{\widehat{Y}^2} + \frac{\widehat{\text{Var}}(\widehat{Y}_{SCP})}{\widehat{Y}^2} + \frac{\widehat{\text{Var}}(\widehat{Y}_M)}{\widehat{Y}^2} \\ & - 2 \frac{\widehat{\text{Cov}}(\widehat{X}_S, \widehat{Y}_S)}{\widehat{X} \widehat{Y}} - 2 \frac{\widehat{\text{Cov}}(\widehat{X}_{SCP}, \widehat{Y}_{SCP})}{\widehat{X} \widehat{Y}} - 2 \frac{\widehat{\text{Cov}}(\widehat{X}_M, \widehat{Y}_M)}{\widehat{X} \widehat{Y}} \end{aligned} \quad (4.7)$$

Because of differences in sampling techniques used for S, SCP, and M strata, variance and covariance formulae appropriate for these parts of the sample differ but, as shown, may be added to produce $\widehat{\text{Var}}(\widehat{R})$. The following shows how the numerators of (4.7) may be computed.

A collapsed stratum variance estimator was used for $\widehat{\text{Var}}(X_S)$ as follows:

$$\widehat{\text{Var}}(X_S) = \sum_{K=1}^2 \sum_{g=1}^6 \frac{L_{Kg}}{L_{Kg}-1} \sum_{l=1}^{L_{Kg}} \left(\widehat{X}_{Kgl} - \frac{\widehat{X}_{Kg}}{L_{Kg}} \right)^2 \quad (4.8)$$

is the model for $\widehat{\text{Var}}(\widehat{Y}_S)$ and also for

$$\widehat{\text{Cov}}(\widehat{X}_S, \widehat{Y}_S) = \sum_{K=1}^2 \sum_{g=1}^6 \frac{L_{Kg}}{L_{Kg}-1} \sum_{l=1}^{L_{Kg}} \left(\widehat{X}_{Kgl} - \frac{\widehat{X}_{Kg}}{L_{Kg}} \right) \left(\widehat{Y}_{Kgl} - \frac{\widehat{Y}_{Kg}}{L_{Kg}} \right)$$

where l is the index of the L_{Kg} (2 or 3) PSU's in group g
 g is the index of the 6 groups of the 13 single unit non-certainty PSU's.

$$\hat{X}_{Kgl} = \left(\frac{1}{P_{Kgl}} \right) \sum_{j=1}^{J_{Kgl}} W_{Khj} \sum_{i=1}^{n_{Kglj}} X'_{Kglji}$$

i.e. the stratum total, \hat{X}_{Kh} , for the l th PSU in group g .

$$\hat{X}_{Kg} = \sum_{l=1}^{L_{Kg}} \hat{X}_{Kgl}, \text{ the group total.}$$

First difference estimators were used for $\text{Var}(X_{SCP})$ and $\text{Var}(X_M)$ as follows:

$$\widehat{\text{Var}}(\hat{X}_{SCP}) = \sum_{K=1}^2 \sum_{h=1}^5 \sum_{j=1}^{J_{Kh}} b_{Khj} N_{Khj}^2 \left(\frac{1}{n_{Khj}} - \frac{1}{N_{Khj}} \right) \sum_{i=2}^{n_{Khj}} \frac{(X'_{Khji} - X'_{Khj(i-1)})^2}{2(n_{Khj} - 1)} \quad (4.9)$$

where b_{Khj} is an indicator variable equal to 1 if some cases from that substrata were not selected and equal to 0 if all cases were selected as in certainty substrata.

$$\widehat{\text{Var}}(\hat{X}_M) = \sum_{K=1}^2 \sum_{j=1}^{J_{K17}} b_{K17j} N_{K17j}^2 \left(\frac{1}{n_{K17j}} - \frac{1}{N_{K17j}} \right) \sum_{i=2}^{n_{K17j}} \frac{(X_{K17ji} - X_{K17j(i-1)})^2}{2(n_{K17j} - 1)}$$

$\widehat{\text{Var}}(\hat{Y}_{SCP})$, $\text{Cov}(\hat{X}_{SCP}, \hat{Y}_{SCP})$, $\widehat{V}(\hat{Y}_M)$ and $\text{Cov}(\hat{X}_M, \hat{Y}_M)$ are similar.



U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

FORM

CB-5109

1982 CENSUS OF DISTRIBUTIVE TRADES

PETROLEUM AND PETROLEUM PRODUCTS

Item 6 - DOLLAR VOLUME OF BUSINESS IN 1982		Mil.	Thou.	Dol.
		010		
a. Sales of merchandise and other operating receipts				
		100		
		1	<input type="checkbox"/> YES-Go to c	
b. Did this establishment sell merchandise on a commission or brokerage basis?		2	<input type="checkbox"/> NO - SKIP to e	
		Mil.	Thou.	Dol.
c. Gross selling value of business conducted for the account of others (Include in item 6a.)		101		
d. Dollar volume of commissions or brokerage received on transactions reported in item 6c		102		
e. What percentage of the products sold by this establishment did your company (including firms under common ownership or control) manufacture or mine in the United States or its possessions?		Percent		
		103		
		104		
		1	<input type="checkbox"/> YES-Go to g	
f. Did this establishment have transfers (billings) to other establishments within your company?		2	<input type="checkbox"/> NO - SKIP to h	
		Mil.	Thou.	Dol.
g. Dollar volume of transfers (billings) to other establishments within your company (DO NOT include in item 6a.)		105		
		109		
		1	<input type="checkbox"/> YES	
h. Were 50% or more of the products sold by this establishment imported from a foreign country?		2	<input type="checkbox"/> NO	

Item 9 - OPERATING EXPENSES		Mil.	Thou.	Dol.
TOTAL 1982 operating expenses, including payroll but excluding cost of goods sold		040		

Item 13 - SALES BY COMMODITY LINES		HOW TO REPORT PERCENTS				If figure is 38.76% of total sales:			Mil.	Thou.	Dol.	Per-cent													
Report sales either in dollar figures (see example on page 1) or as percent (in whole percents) of the total (see example at right).									•Report whole percents → 39 Not acceptable → 38.76																
									Commodity lines (a)				Estimated sales during 1982 (b)		Census use			Number of gallons sold during 1982 (c)			Storage capacity (including underground) in gallons at end of 1982 (d)				
		Census use		Mil.		Thou.		Dol.		Per-cent		Census use		Mil.		Thou.		000		Mil.		Thou.		000	
Petroleum																									
1. Petroleum products																									
a. Aviation gasoline																									
5611																									
159																									
179																									
b. Motor gasoline																									
5612																									
161																									
181																									
c. Special naphtha																									
5613																									
162																									
182																									
d. Jet fuel (naphtha or kerosine type)																									
5614																									
163																									
183																									
e. No. 2 distillate fuel oil (No. 2 diesel and No. 2 heating oil)																									
5615																									
164																									
184																									
f. All other distillate fuel oil (No. 1 and No. 4 diesel and heating oil, kerosine, etc.)																									
5616																									
165																									
185																									
g. Residual fuel oil (No. 5, No. 6, heavy diesel, etc.)																									
5617																									
166																									
186																									
h. Lubricating oil and grease																									
5618																									
167																									
187																									
i. Total - Sum of lines 1a through 1h																									
5600																									
190																									
200																									
2. Liquefied petroleum gases (Exclude natural gas.)																									
5640																									
168																									
188																									
3. Crude oil																									
5650																									
169																									
189																									
4. Other petroleum products (Include asphalt, road oil, wax, etc.)																									
5660																									
Other commodity lines																									
5. Automotive parts and supplies, new and rebuilt (Exclude parts installed in repair work.)																									
0200																									
6. Petroleum products marketing equipment																									
0280																									
7. Tires & tubes																									
0300																									
8. Lumber; rough, dressed, and finished																									
0600																									
9. Ferrous metals (Include nails.)																									
1300																									
10. Nonferrous metals																									
1370																									
11. Coal and coke																									
1400																									
12. General-purpose industrial machinery, equipment, and parts																									
2520																									
13. Grain and beans																									
5100																									
14. Chemicals and allied products (Exclude agricultural, plastics, gases, and petroleum.)																									
5500																									
15. Farm supplies																									
5900																									
16. Miscellaneous commodities - Specify																									
a.																									
9811																									
b.																									
9812																									
c.																									
9813																									
17. Rental receipts - Specify																									
9940																									
18. Service receipts and labor charges (Include installed parts.)																									
9700																									
19. TOTAL sales and storage capacity in 1982 - Sum of lines 1 through 18 should equal the entry in item 6a if reporting in dollars																									
9990																									
100%																									
160																									
180																									