PART B PHYSICIAN SAMPLE REDESIGN

By

James C. Beebe
Statistical Consultant
1345 Tydings Road
Annapolis MD 21401

July 18, 1995

Prepared for :
Ann Meadow, Sc.D., Project Officer
Office of Research and Demonstrations
Health Care Financing Administration

Table of Contents

List of Tables

Introduction

In 1992 HCFA designed a physician sample to replace the Part B Medicare Annual Data (BMAD) Provider File, which for several years supplied Medicare claims data to support numerous studies of physician payment and other issues. Based on the terminal digits of the Unique Physician Identification Number (UPIN), the new sample is self-weighting within each State and is intended to be representative of the physicians treating Medicare beneficiaries. The database comprises detailed line item information from all available claims of the sample physicians. An earlier evaluation of the initial sample design suggested that the variability in the physician data was overestimated, leading to sample sizes larger than needed to achieve precision requirements. Subsequently, new summary data on 100% of physician billings from the National Claims History yielded State variance estimates that are more reliable than previous estimates. The summary data also permitted better enumeration of physicians actively billing Medicare than did the original sampling frame (the National UPIN Registry).

This paper describes our use of this newly available data to customize the sample design of the Part B Physician Samples for each State. Variances and counts of the number of active physicians were available for 1991 and 1992 for nearly all States and territories. (1993 data became available later and were useful for verifying the consistency of the variances. More on this later.) Sample sizes are based on the variance of allowed charges per physician even though estimates of other quantities, such as caseload, will also be important. In earlier work, the variance of caseload was generally found to be smaller than the variance of allowed charges. Thus, sample sizes based on allowed charges should yield adequate precision for caseload estimates.

Terminal Digit Sampling and Universe Counts

We chose to use the terminal digits of the UPIN to select the sample because of the convenience of the method. The last digits of the UPIN are assigned at random under the direction of the Bureau of Program Operations. The Office of Research (OR), HCFA, has compared the distribution of specialities in the sample with their distribution in the

Registry and found no reason to doubt the randomness of the sample by specialty. OR also empirically verified that the last two terminal digits are uniformly distributed. Appendix A contains some information on problems in getting accurate State universe counts for design purposes.

To use terminal digit sampling, it is necessary to first determine the sample size needed and translate this to a percentage of the universe of active UPINS. For example, a State with a universe of 10,000 physicians and a calculated sample of 700 requires a 7% sample which could be achieved by randomly selecting 7 distinct pairs of terminal digits. It may be important for some users to note that, because independent samples are selected for each State, physicians who practice in more than one State can fall into more than one sample. Thus, when, for example, a mean per physician is calculated, it will not be an exact reflection of the mean for complete physician practices.

The sample is formed by selecting HCFA bill records submitted by physicians which have UPINs with the specified terminal digits. This means that only those physicians who treat Medicare patients in any given year will have an opportunity to fall into the sample. Under this sampling procedure, it is necessary to distinguish between the design sample size and the realized sample size. First note that the universe of interest is that group of physicians who are actively treating the Medicare patients in any given year. At the design stage we have either an actual count of the universe for a given year or an estimate of that count. The sample size to achieve a desired precision is then calculated and expressed as a percent of the universe in order to know how many terminal digits to select into the sample. If the universe count is incorrect (as it certainly will be to some extent in future years) the realized sample size will be different from the design sample size.

Thus, the key to getting the correct sample size is having an accurate count of the universe size of active UPINS. In designing the original sample some years ago, it was assumed that the count of UPINS in the Registry would be approximately equal to the number of active UPINS. An earlier analysis of the sample results for 18 States found that the actual samples fell an average of 36% short of the design specifications. We speculated that this was because the Registry of UPINS overstated the number of active UPINS by 36%, i.e., not all physicians on the Registry treated Medicare patients every year. Thus, it was decided at that time to increase all design sample sizes by 36%.

In the most recent National Claims History (NCH) summary data, we obtained what are believed to be accurate counts of valid and active UPINS for 1991 and 1992. We also had 1993 Registry counts for 18 states previously analyzed. When the 1992 NCH counts were inflated to allow for growth over time and the result compared to the 1993 Registry counts, we found the active UPINS to be from 34% to over 70% lower than Registry counts. This raised two related questions: (1) Should we use NCH universe counts or adjusted registry counts at the design stage; and (2) for computing estimates requiring universe counts (e.g., weighted national estimates), what is the best source of universe counts? Using the accurate NCH universe counts for 1991 and 1992 for design purposes seemed the best answer to question one, and this is what we did. We do not have an adequate explanation of why the large discrepancy exists between the active UPINs and the Registry counts. However, because it is so large, one would hope that our estimate of the universe size is understated rather than overstated. To the extent that it is understated, the sample sizes will be larger than targeted and, thus, the precision of estimates greater.

As to the second question concerning the calculation of estimates, the accurate NCH universe counts will not necessarily be available to all users in future years. For these users, the proportion (call this p/100 to be consistent with notation explained later) of the universe, N, being sampled and the realized sample size, n, can be used as an estimate of the universe size. Since $n = N(p/100)$ by definition, if N is unknown, it can be calculated as $N = n/(p/100)$. The p's are shown in column 7 of Table B1, and Appendix C gives more details on how to do this.

Finally there was the question of how much the universe would increase over time. This is important because we are sampling a flat proportion of the universe. As the universe grows, sample size is likely to get larger than is needed to achieve the desired precision. There was a 5% increase in the number of active UPINs between 1991 and 1992. The 1993 universe size could have been estimated by increasing the 1992 counts by 5%. However, since 1991 and 1992 are to be resampled using the new sampling rates, it was decided to simply use the 1992 universe counts for design purposes. Since understating the universe tends to result in oversampling beyond targeted levels, this is a conservative strategy for future years. If sample sizes begin to get much larger than is necessary, periodic adjustment could be made.

Determining Sample Size

It was decided early in the design stage that reasonably precise State estimates would be needed. Determining the level of precision needed for the estimates is an important decision that had to be made by the primary users of the samples. Precision is often expressed in terms of a confidence interval or in terms of the relative precision of the estimate. Both concepts depend on the standard error of the estimate and the estimate itself and, given these quantities, both can readily be calculated. Thus, if one estimated the mean allowed charge to be $100,000 with a standard error of $9,500, the 95% confidence interval would be $100,000 +/- 1.96($9,500) or from $81,380 to $118,620. The relative standard error (or, equivalently, the relative precision) of the estimated mean would be $9,500/$100,000 = 9.5%. If we wanted to narrow the confidence interval or decrease the relative standard error, the sample size would have to be increased. In the case of the UPIN samples, it was decided that a targeted relative precision of 7.5% for individual State estimates would strike a reasonable balance between precision and sample size.

There are several steps involved in determining sample size. These were performed for each State. The steps are:

- Calculate the mean allowed charge per physician and the standard deviation of allowed charges across physicians.

- Divide the standard deviation by the mean to get the coefficient of variation (CV) for allowed charges.

- Calculate the basic sample size, n', by dividing the square of the desired relative standard error ($.075^2 = .005625$) by the square of the CV.

- Let N = the universe size (active UPINS in a State) and adjust n' for the finite population correction (fpc): n" = n'/(1+n'/N).

- Calculate p' = 100(n"/N). Round p' upward to the next highest whole number to get p. If p < 2 then set p = 2. After this final adjustment, p is the percentage of the universe to be selected into the sample. It is also the number of pairs of terminal digits to be selected, as shown in Table B2.

- The final design sample size is n = N(p/100).

Details of these calculations for each State are shown in Appendix B, Table B1. As can be seen, p (column 7 of Table B1) is both the percent of a State's universe to be sampled and the number of pairs of terminal digits to be sampled. The minimum allowed value of p = 2 is to assure that a 2% national sample of physicians with identical terminal digits will be available. Because p' is rounded upward to get p, this procedure will, in most cases, result in a larger sample than would be obtained with no rounding. The pairs of terminal digits to be selected into the sample are shown in Appendix B, Table B2. To the extent possible, the terminal digits used to select an earlier version of the 1991 and 1992 samples were retained to provide longitudinal data from 1991 forward for those analysts who require it. The earlier samples were used by the Health Care Financing Administration in a recent analysis of physician services. (See "Access to Care Before and After Fee Schedule Implementation: A Physician-Based Analysis," Appendix VIII in Report to Congress: Monitoring the Impact of Medicare Physician Payment Reform on Utilization and Access, 1994, HCFA Pub. No. 03358, September 1994. The 1992 data were also available as a public use file.)

The procedures outlined above resulted in design State sample sizes ranging from 240 physicians in Wyoming to 1,094 physicians in California. The percent of the State universes sampled ranges from 2% to 44%. (These numbers exclude the Virgin Islands and Guam which have universes so small it was decided to sample 100% of their physicians.) The total national sample is 22,537 physicians out of the total universe of 470,373 active physicians with valid UPINs in 1992.

As can be seen in Table B2, the terminal digits are completely nested in the sense that the terminal digits for any State are a subset of the terminal digits for the State with the next larger number of terminal digits. Note also that all States use terminal digits 04 and 81. This makes it possible to select conveniently a 2% national sample from which unweighted estimates can be calculated. Weighting issues are discussed in a later section. Of course, for those States with more than a 2% sample, any pair of the available terminal digits could be used to create a valid 2% national sample. One would not have to use 04 and 81 in every State.

Consistency of Coefficients of Variation

We are calculating sample sizes with the expectation that the CVs of past years used to calculate sample sizes will be fairly representative of the CVs of future years. Preferably, future CVs will be no larger than past CVs, even though this might mean that we are selecting more cases than necessary in some States. This section investigates what information we have about the consistency of variances over time.

As stated earlier, sample sizes were determined before 1993 data became available. A comparison of 1991 and 1992 CVs led to the decision to use 1992 CVs only, as opposed to using, say, an average of 1991 and 1992. When the 1993 CVs became available, they provided further support for ignoring 1991 CVs. Table 1 shows various statistics to support this decision.

Table 1: Summary of CVs for 1991-1993 Across 50 Carriers

Year	Mean CV	Standard Deviation of the CVs
1991	1.70	0.371
1992	1.54	0.131
1993	1.50	0.125
1991*	1.63	0.146

*1991 statistics recalculated with two largest carrier values deleted.

The mean CV for 1991 was considerably higher and the standard deviation much higher than for 1992 or 1993. This suggests a lack of stability in the first year which may be due to problems associated with the establishment of a new data base rather than intrinsically high variation in the data. To be fair, however, 1991 had two carriers with unusually large CVs. When these are deleted (see 1991*), the mean and standard deviation for 1991 are closer to the other two years but still considerably higher. Table 2 shows the mean difference, mean absolute difference (MAD), and correlation of the CVs between the three years (with the two 1991 outliers omitted).

Table 2: Mean Differences, MADs, and Correlations of the CVs

Years	Mean Difference*	MAD	Correlation
91-92	+0.15	0.16	0.67
91-93	+0.20	0.21	0.58
92-93	+0.05	0.05	0.92

*early year minus later year

The mean difference column shows that the 1991 CVs were considerably higher than the other two years, with 1992 and 1993 being much closer together. MAD is very close to the mean difference because nearly all of the mean differences were positive. The correlations also suggest an improvement in stability in the latter two years.

Finally, a regression of the later year CVs on the previous year would show a 0 intercept and a slope of 1 if there were no difference between the two years. Table 3 shows the results of such regressions.

Table 3: CV Regressions: 1992 vs 1991 and 1993 vs 1992

Years	Intercept	Slope
92 vs 91	0.614	0.563
p-value*	(.000)	(.000)
93 vs 92	0.144	0.877
p-value*	(.091)	(.014)

*tests the hypotheses that the intercept equals 0 and the slope equals 1

Again, we see that 1992 and 1993 are much more alike than 1991 and 1992. The intercept and slope for 93 vs. 92 are much closer to 0 and 1 respectively than for 92 vs 91.

In summary, the evidence suggests that the CVs are becoming more stable, both over time and between areas. This gives us some confidence that sample sizes will be adequate to meet the targeted precision level in future years.

Weighted and Unweighted Estimates When Combining States' Data

Two or more independent simple random samples (SRSs) from two or more universes can be combined and estimates calculated as though the combined sample were one SRS from the combined universes if the sampling fractions of the original samples are the same. In our case the State samples are treated as independent SRSs with sampling fractions equal to the number of terminal digits being selected divided by 100. Thus, FL, OH, PA, TX, NY, and CA all have sampling fractions of .02. These States could be combined and estimates calculated without weighting. However, if CA and AK, with sampling fractions of .02 and .44 respectively were combined for some reason, weighting would be necessary to get an unbiased combined estimate for the two States. This can be seen intuitively by considering the numbers. CA has a universe of 54,717 and sample size of 1,094. AK has a universe of 797 and a sample of 351. If estimates intended to represent the combined universes of the two States were made on the basis of the combined, unweighted samples, AK physicians would get a relative weight of 351/(1,094 + 351) = .24 and CA physicians a relative weight of 1,094/(1,094 + 351) = .76. However, their proper relative weights, based on a similar calculation for universe sizes, would be .01 for AK and .99 for CA. Thus, if States are to be combined without discarding any data, it will in most cases be necessary to weight the estimates by treating the States as strata and using estimation methods appropriate to this kind of sampling. Appendix C contains more details on methods for calculating both individual State estimates and estimates for combined States.

On the other hand, combining States will result in larger sample sizes than are needed to achieve the targeted relative standard error of 7.5%. For example, if four States of about equal size are combined, the formulas of Appendix C show that the expected relative standard error of a mean would be cut by half to 3.75%. Thus, it may be convenient for some analysts to discard data in order to make the file more manageable and , at the same time, avoid the complexity of weighting. The nested terminal digit pattern is designed to facilitate this.

The nested nature of the terminal digit pattern can be seen in Appendix B, Table B2. The State with the highest proportion of cases in the sample, AK with 44%, is shown at the top. Its terminal digits were selected at random. The next State, VT with 30%, has terminal digits selected at random from AK's 44%. Thus, if these two States were combined and a subsample of AK physicians selected based on VT's 30 terminal digits, we would have a SRS from the two States with a sampling fraction of .30 and no weighting would be required to calculate combined estimates. A similar procedure can be used to form SRSs for any number and combinations of States. The sample will be comprised of those terminal digits in the State with the smallest number terminal digits. A national 2% SRS of physicians would result from subsampling the two terminal digits 04 and 81.

If it were desired to get a 5% national sample, this would not be possible without supplementing the sample because the larger States already have a smaller proportion than 5%. However, one could approximate a 5% sample by a combination of subsampling and weighting. Referring to Table B2, if all of the States above NC were subsampled on the basis of the terminal digits of the States with p = 5 (indicating a 5% sample), a single estimate for these States could be calculated without weighting. This single estimate for these States would then get a weight equal to the sum of their universe counts divided by the total universe when combined with the remaining States. The remaining States would get individual weights as described in Appendix C.

Appendix A: Adjustments for Data Deficiencies

There were some irregularities in the available data. These were handled as explained here.

We did not obtain any universe counts or CVs for Washington State. The universe count was estimated to be 8,000 which is about 60% of the 1993 registry count of 13,737. A CV of 1.6 was used to determine sample size. This is toward the upper end of other States' CVs.

Our State universe counts are actually counts of physicians in areas covered by HCFA carriers. In most cases, carrier areas are confined to single State boundaries. However, carrier 00740 covers parts of both Kansas and Missouri. We could have allocated its 3,544 physicians between the two States, but had no reasonable basis for making the allocation. So 00740 was ignored. Kansas is represented by carrier 00650 and Missouri by 11260. This is a conservative approach from the point of view of precision because underestimating the universe size tends to result in a larger sample size.

There are three pairs of States or territories (NH/VT, ND/SD, and PR/VI), each pair of which is covered by one carrier. We allocated universe sizes to the individual State and territories based on their relative frequencies in the 1993 registry. The CV from each combined unit was rounded upward and assigned to the individual areas.

Appendix B: Sample Size Calculations and Terminal Digits To Be Selected

Table B1: Calculation of Sample Size and
Proportion of Universe To Be Selected

Carrier	1992 UPINS	CV	n'	n''	n''/N %	p rnd up	n
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Total	470,373		23,146	19,922			22,537
00000_CA	54,717	1.68	500	495	0.9	2	1,094
00000_MN	8,556	1.53	415	396	4.63	5	428
00000_NY	39,941	1.73	531	524	1.31	2	799
00000_VA	8,737	1.34	321	309	3.54	4	349
00510_AL	8,703	1.44	367	348	5.19	6	402
00520_AR	3,843	1.47	383	348	9.06	10	384
00528_LA	7,032	1.59	452	425	6.04	7	492
00550_CO	6,386	1.47	383	361	5.65	6	383
00570_DE	1,181	1.42	360	276	23.37	24	283
00580_DC	7,492	1.57	441	416	5.55	6	450
00590_FL	27,101	1.59	447	439	1.62	2	542
00621_IL	21,156	1.87	620	602	2.85	3	635
00630_IN	9,767	1.65	487	464	4.75	5	488
00640_IA	5,026	1.61	463	424	8.44	9	452
00650_KS	3,104	1.63	471	409	13.18	14	435
00655_NE	2,702	1.72	525	439	16.25	17	459
00660_KY	6,341	1.44	367	347	5.47	6	380
00690_MD	7,770	1.55	427	405	5.21	6	466
00700_MA	17,149	1.67	496	482	2.81	3	514
00710_MI	14,788	1.56	434	422	2.85	3	444
00751_MT	1,543	1.55	430	336	21.78	22	339
00780_NH	1,512	1.44	370	297	19.64	20	302
00820_ND	1,206	1.38	337	263	21.81	22	265
00825_WY	800	1.39	341	239	29.88	30	240
00860_NJ	15,563	1.44	370	361	2.32	3	467
00865_PA	28,243	1.44	369	364	1.29	2	565
00870_RI	2,187	1.42	361	310	14.17	15	328
00880_SC	5,235	1.37	336	316	6.04	7	366
00900_TX	27,224	1.65	484	476	1.75	2	544
00910_UT	2,752	1.65	486	413	15.01	16	440
00951_WI	9,010	1.41	354	341	3.78	4	360
00973_PR	4,580	1.71	518	466	10.17	11	504
01020_AK	797	1.86	613	346	43.41	44	351
01030_AZ	6,652	1.85	610	559	8.4	9	599
01040_GA	10,377	1.49	397	383	3.69	4	415
01120_HI	2,250	1.66	490	402	17.87	18	405
01290_NV	2,056	1.57	437	360	17.51	18	370
01360_NM	2,578	1.64	478	403	15.63	16	412

Table B1 (cont'd): Calculation of Sample Size and Proportion of Universe To Be Selected

01370_OK	5,027	1.54	423	390	7.76	8	402
01380_OR	5,601	1.5	401	374	6.68	7	392
05130_ID	1,583	1.46	380	306	19.33	20	317
05440_TN	8,918	1.47	383	367	4.12	5	446
05535_NC	11,181	1.6	457	439	3.93	4	447
10230_CT	7,473	1.35	324	310	4.15	5	374
10250_MS	3,381	1.47	384	345	10.2	11	372
11260_MO	6,922	1.45	373	354	5.11	6	415
16360_OH	20,100	1.43	362	355	1.77	2	402
16510_WV	3,382	1.5	402	360	10.64	11	372
21200_ME	2,558	1.41	355	312	12.2	13	333
WA	8,000	1.6	455	431	5.39	6	480
VT	966	1.5	400	283	29.3	30	290
SD	1,158	1.4	348	268	23.14	24	278
VI	22	1.7	514	21	95.45	100	22
GU	44	1.7	514	41	93.18	100	44

Notes:
Col. (4) = sample size based on the CV and 7.5% relative precision
Col. (5) = (4)with finite population correction
Col. (6) = (5)divided by universe size (N), col. (2)
Col. (7) = (6) rounded upward, minimum of 2
Col. (8) = Col. (7)/100 multiplied by Col. (2) = design sample size

Table B2: Universe Size, Sampling Percentage, Design Sample Size, and Terminal Digits

	Count of Tarminal Digits			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
Carrier	N	p	n	Tarminal Digits																																											
Total	470,373		22,537																																												
01020_AK	797	44	351	4	6	7	8	14	16	18	21	23	25	29	30	31	35	36	37	39	40	42	47	49	51	52	53	54	55	57	62	66	68	69	70	72	75	77	79	81	84	88	91	92	95	96	98
00825_WY	800	30	240	4	6	7	8			18		23	25	29		31	35	36		39	40		47			52	53		55		62	66	68	69		72		77	79	81	84	88	91	92	95		
VT	966	30	290	4	6	7	8			18		23	25	29		31	35	36		39	40		47			52	53		55		62	66	68	69		72		77	79	81	84	88	91	92	95		
00570_DE	1,181	24	283	4	6	7	8			18		23	25	29		31		36		39	40						53		55		62	66	68	69		72		77	79	81		88	91				
SD	1,156	24	278	4	6	7	8			18		23	25	29		31		36		39	40						53		55		62	66	68	69		72		77	79	81		88	91				
00820_ND	1,206	22	265	4	6	7	8			18		23	25	29				36		39	40						53		55		62	66		69		72		77	79	81		88	91				
00751_MT	1,543	22	339	4	6	7	8			18		23	25	29				36		39	40						53		55		62	66		69		72		77	79	81		88	91				
05130_ID	1,583	20	317	4	6	7	8			18			25	29				36		39	40						53		55		62	66		69		72			79	81		88	91				
00780_NH	1,512	20	302	4	6	7	8			18			25	29				36		39	40						53		55		62	66		69		72			79	81		88	91				
01120_HI	2,250	18	405	4	6	7	8			18			25	29				36			40						53		55		62	66		69					79	81		88	91				
01290_NV	2,056	18	370	4	6	7	8			18			25	29				36			40						53		55		62	66		69					79	81		88	91				
00655_NE	2,702	17	459	4	6	7				18			25	29				36			40						53		55		62	66		69					79	81		88	91				
00910_UT	2,752	16	440	4	6	7				18			25	29				36			40						53		55			66		69					79	81		88	91				
01360_NM	2,578	16	412	4	6	7				18			25	29				36			40						53		55			66		69					79	81		88	91				
00870_RI	2,187	15	328	4	6	7				18			25	29				36			40						53					66		69					79	81		88	91				
00650_KS	3,104	14	435	4	6	7				18			25	29				36			40						53					66		69					79	81		88					
21200_ME	2,558	13	333	4	6	7							25	29				36			40						53					66		69					79	81		88					
16510_WV	3,382	11	372	4	6									29				36			40						53					66		69					79	81		88					
00973_PR	4,580	11	504	4	6									29				36			40						53					66		69					79	81		88					
10250_MS	3,361	11	372	4	6									29				36			40						53					66		69					79	81		88					
00520_AR	3,843	10	384	4										29				36			40						53					66		69					79	81		88					
00640_IA	5,026	9	452	4										29				36			40						53					66		69						81		88					
01030_AZ	6,652	9	599	4										29				36			40						53					66		69						81		88					
01370_OK	5,027	8	402	4										29				36			40											66		69						81		88					
00880_SC	5,235	7	366	4										29				36			40											66		69						81							
00528_LA	7,032	7	492	4										29				36			40											66		69						81							
01380_OR	5,601	7	392	4										29				36			40											66		69						81							
00660_KY	6,341	6	380	4										29							40											66		69						81							
00690_MD	7,770	6	466	4										29							40											66		69						81							
WA	8,000	6	480	4										29							40											66		69						81							
00550_CO	6,386	6	383	4										29							40											66		69						81							
11260_MO	6,922	6	415	4										29							40											66		69						81							
00580_DC	7,492	6	450	4										29							40											66		69						81							
00510_AL	6,703	6	402	4										29							40											66		69						81							
00630_IN	9,767	5	488	4										29																		66		69						81							
00000_MN	8,556	5	428	4										29																		66		69						81							
05440_TN	8,918	5	446	4										29																		66		69						81							
10230_CT	7,473	5	374	4										29																		66		69						81							
05535_NC	11,181	4	447	4																												66		69						81							
01040_GA	10,377	4	415	4																												66		69						81							
00951_WI	9,010	4	360	4																												66		69						81							
00000_VA	8,737	4	349	4																												66		69						81							
00710_MI	14,788	3	444	4																												66								81							
00621_IL	21,156	3	635	4																												66								81							
00700_MA	17,149	3	514	4																												66								81							
00860_NJ	15,563	3	467	4																												66								81							
00590_FL	27,101	2	542	4																																				81							
00000_NY	39,941	2	799	4																																				81							
00000_CA	54,717	2	1,094	4																																				81							
00865_PA	28,243	2	565	4																																				81							
16360_OH	20,100	2	402	4																																				81							
00900_TX	27,224	2	544	4																																				81							
VI	22	100	22	Take All																																									
GU	44	100	44	Take All																																									

Appendix C: Formulas and Examples

The Health Care Financing Administration (as well as the Social Security Administration) has, for many years, relied on terminal digit sampling of the Social Security Numbers of their beneficiaries. These have generally been treated as simple random samples ignoring the fact that the sample size, n, is a random variable. This practice will be adopted for the purpose of presenting formulas for the Part B Physician Sample in this appendix. The formulas given below can be found in, or derived from, formulas given in one or more of the three references.

Notation

Unless otherwise specified, this notation will apply to the full sample and universe of all physicians in any given State or combination of States as well as any subgroup of physicians in the State or States. Thus, n and N can refer to the sample size and universe size of all physicians in one or more States or of any subgroup of physicians, such as family practitioners, in one or more States.

n = realized sample size. Note that this is the actual number of physicians falling into the sample, not the design number shown in Table B1.

M = the number of states for which an estimate is to be calculated.

f = the proportion of the universe selected into the sample. f is also equal to p/100 where p is defined in the main text and shown on Table B2.

N = the universe size. In general, we will assume that N = n/f because the actual count of active physicians will not automatically be part of the sample files. However, if 100% counts of active physicians are available from other sources, as was the case for this redesign, then the 100% counts can be used in place of n/f. N can also represent the size of any subdomain of interest--general practitioners, for example. If the subdomain count of general practitioners is not known, it can be calculated as $N = n/f$ where n is now the realized sample size of general practitioners. Note that f remains the same whether the estimates of interest are for the universe of all physicians or some subdomain of that universe.

x = any measurable characteristic of the sample. Similarly for y.

i = subscript used to index physicians.

h = subscript used to index states.

Σ = indicates summation from $i = 1$ to n or $h = 1$ to M unless otherwise indicated. Whether we are summing over physicians (i) or states (h) will be clear from the context.

SE = Standard error.

The quantity $1-f$ is called the finite population correction factor or fpc. In finite population sampling, the fpc acts to reduce the variance of estimates of the types shown in formulas 1 through 16. See Cochran, page 24, for more information on this subject. However, when testing for statistically significant differences, the fpc is not used in the standard error formulas (see Cochran, page 39 and Deming, Chapter 7). Thus, it should be understood that, beginning with formula (17) all standard errors are to be calculated without the fpc, even though the notation has not been changed.

The next section presents formulas for simple random sample (SRS) estimates. These formulas can be used for individual State estimates or for estimates of combined States which have been sampled or subsampled in a manner such that the number of terminal digits is equal in each state being combined.

Formulas for Simple Random Sample Estimates

(1). *MEAN*: $\bar{x} = \dfrac{\Sigma x_i}{n}$

(2). *STANDARD DEVIATION*: $s = \sqrt{\dfrac{\Sigma x_i{}^2}{(n-1)} - \dfrac{(\Sigma x_i)^2}{n(n-1)}}$

(3). *SE OF MEAN*: $s_{\bar{x}} = s\sqrt{\dfrac{(1-f)}{n}}$

(4). *ESTIMATED TOTAL*: $\hat{X} = N\bar{x}$

(5). *SE OF ESTIMATED TOTAL*: $s_{\hat{X}} = Ns_{\bar{x}}$

(6). *Estimated Ratio*: $r = \dfrac{\bar{x}}{\bar{y}} = \dfrac{\hat{X}}{\hat{Y}} = \dfrac{\Sigma x_i}{\Sigma y_i}$

(7). *Estimated Covariance*: $s_{xy} = \dfrac{\Sigma x_i y_i - \dfrac{\Sigma x_i \Sigma y_i}{n}}{n-1}$

(8). *SE of Estimated Ratio*: $s_r = \sqrt{\dfrac{(1-f)}{n\bar{y}^2}(s_x^2 + r^2 s_y^2 - 2 r s_{xy})}$

The above statistics may be used in one-sample t tests with degrees of freedom of n-1.

Formulas for Stratified Samples

The basic principle in forming stratified estimates with the Part B Physician Sample is to form weighted combinations of the SRS sample estimates. To show summation over States, we introduce the subscript h to represent States. The weights will depend on the N_h physicians in the State universes. For 1992, N_h can be obtained from Column 2 of Table B1. Then the h^{th} State's weight is computed as the State universe count, N_h, divided by 470,373 as shown in formulas (9) and (10). For other years N_h is defined as n_h/f_h if the 100% count of active physicians is unknown. Note that N_h and n_h can also represent universe and sample counts of a subdomain such as general practitioners.

The estimates for stratified samples use the SRS estimates calculated in the previous section. The subscript i is no longer needed because all summations over physicians have already been accomplished. We now introduce the h to indicate summation over states.

(9). *Total Universe of States Being Combined*: $N = \Sigma N_h$

(10). *State Weight*: $W_h = \dfrac{N_h}{N}$

(11). *Stratified Mean*: $\overline{x}_{st} = \Sigma W_h \overline{x}_h$

where $\overline{x}_h$ is equation (1) for the h^{th} State.

(12). *SE of Stratified Mean*: $s_{\overline{x}_{st}} = \sqrt{\Sigma W_h^2 s_{\overline{x}_h}^2}$

where $s_{\overline{x}_h}^2$ is equation (3) squared for the h^{th} State.

(13). *Stratified Total:* $\hat{x}_{st} = \Sigma\hat{x}_h = N\bar{x}_{st}$

(14). *SE of Stratified Total:* $s_{\hat{x}_{st}} = \sqrt{\Sigma s^2_{\hat{x}_h}}$

where $s^2_{\hat{x}_h}$ *is equation (5) squared for the* h^{th} *state.*

(15) *Stratified Ratio:* $r_{st} = \frac{\hat{x}_{st}}{\hat{y}_{st}}$

(16). *SE of Stratified Ratio:* $s_{r_{st}} = \sqrt{\Sigma W^2_h s^2_{rh}}$

where s^2_{rh} *is equation (8) squared for the* h^{th} *State.*

Note that s^2_{rh} is calculated for the h^{th} State using the square of formula (8) with the important exception that r_{st} is substituted for r. In other words, r_{st} is constant across all strata as opposed to being equal to the strata ratios. Thus, the variance of the stratified ratio estimate is not simply a weighted sum of the strata variances.

When performing one-sample t tests for stratified estimates, the degrees of freedom is $\Sigma(n_h-1)$ or, equivalently, the total sample size across all strata minus the number of strata.

Testing for Differences

This section will discuss some basic procedures to test for significant differences between estimated means (or, equivalently, estimated totals) and for temporal change in a variable. There are two situations in which testing means will commonly arise: the difference between means of independent groups or the difference between means of dependent groups. Examples of the first situation are testing for the difference between the means of two states or the means of two different specialities within a state. The dependent situation arises because the Part B Physician Sample is longitudinal, permitting the testing of measures of change between years. Sample physicians who are active in both years contribute to the mean in both years, although the overlap will not, in general, be 100%.

We remind the user that the standard errors of the mean used in equations (17), (19), and (21) have the fpc omitted when used for statistical testing. Thus, for these equations:

$$s_{\bar{x}_1} = \frac{s_1}{\sqrt{n}} \textit{ is correct, } s_1\sqrt{\frac{(1-f)}{n}} \textit{ is incorrect}$$

Similar changes are needed for $s^2_{\bar{x}_1}$, $s_{\bar{x}_2}$, $s^2_{\bar{x}_2}$

Differences Between Independent Groups

This test is from Snedecor and Cochran, page 96ff. It uses the t distribution with degrees of freedom given by (19) and assumes that the variances of the two means are not known to be equal.

(17) *SE Difference (Indep. Gps.)*: $s_{d_I} = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}}$

(18) *Test Statistic*: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_{d_I}}$

(19) *Degrees of Freedom*: $df = \dfrac{(s_{\bar{x}_1} + s_{\bar{x}_2})^2}{\dfrac{s^2_{\bar{x}_1}}{n_1 - 1} + \dfrac{s^2_{\bar{x}_2}}{n_2 - 1}}$

For stratified samples, the stratified means and standard errors can be substituted in the formulas.

Differences Between Overlapping Groups

For overlapping groups the standard error of the difference involves a covariance term (Kish, page 457ff). Here we use the notation n_c to indicate the size of the overlap (number of physicians in common).

$$(20)\quad \textit{SE Overlap Diff:}\quad s_{d_c} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2 - \frac{2n_c s_{x_1x_2}}{n_{x_1}n_{x_2}}}$$

$$(21)\quad \textit{Test Statistic (Overlapping Gps.):}\quad t = \frac{\bar{x}_1 - \bar{x}_2}{s_{d_c}}$$

The author is not aware of any theoretical work on the appropriate df in the case of overlapping samples. However, if sample sizes are reasonably large and the percentage overlap high, then using (19) as the degrees of freedom formula should be satisfactory. There are two reasons for this: (1) For sample sizes over 30 or so, the t distribution changes very little with changing degrees of freedom. (2) With a large overlap, the sample sizes will be nearly equal, and one would not expect the variances to change much from one year to another. With equal variances and n's, (19) reduces to 2(n-1) which, in most cases, will be greater than 30.

Differences Between Overlapping Groups with Stratification

The formulas needed for testing for differences with stratified sampling are similar to those developed for stratified ratio estimates. Within-strata estimates of differences are calculated, weighted, and combined to form the stratified estimates. However, because the universe sizes will, in general, not be equal in any two years, a new definition of the weights will be needed. We recommend that the weights be based on the average of the universe sizes in the two years for which the difference is calculated. Thus, with subscripts 1 and 2 representing years 1 and 2, we have:

(22) *Average Universe Size*: $\overline{N}_h = \dfrac{N_{1h} + N_{2h}}{2}$

(23) *Stratum Wts. for Est. Diff.*: $W_{dh} = \dfrac{\overline{N}_h}{\Sigma \overline{N}_h}$

(24) *Est. Diff. within Stratum*: $d_h = \overline{x}_{1h} - \overline{x}_{2h}$

(25) *Est. of Stratified Diff*: $d_{st} = \Sigma W_{dh} d_h$

(20) is used to calculate the standard error of the difference within strata. The standard error of the stratified difference is then:

(26) *SE Strat. Overlap Diff.*: $s_{d_c st} = \sqrt{\Sigma W_{dh}^2 s_{hd_c}^2}$

with degrees of freedom given by formula (19) in which the squared standard errors are now those appropriate for stratified samples (formula 12).

Testing the Difference in Change Over Time

One of the questions that comes up frequently is whether subgroups are changing differently over time. For example, it may be important to know whether some new Medicare policy changes the mean caseload per physician differently for family practitioners than for general surgeons between 1991 and 1992. Let the differences in mean visits between the two years (change scores) be denoted as d_{fp} for family practitioners and d_{gs} for general surgeons. These estimates have standard errors, s_{fp} and s_{gs} respectively, calculated from (20). Further, suppose that because of changing demographics, one would expect the change in mean visits between family practitioners and general surgeons to be of magnitude D even without the policy change. Then one can form the following test statistic:

$$(27)\quad \textit{Test for Change Diff:}\quad t = \frac{d_{fp} - d_{gs} - D}{\sqrt{s_{fp}^2 + s_{gs}^2}}$$

with degrees of freedom given by formula (19) in which the squared standard errors are now those appropriate for stratified samples (formula 12).

An alternative is to use only those sample physicians who are active in both years, ignoring data for physicians active in only one year. Then one could calculate individual physician differences, d_i's, and use the simpler methods of testing for paired differences as described in Snedecor and Cochran. This has the advantage of testing directly for changes in behavior but, unless the percentage overlap is very high, it may not be a very good estimate of overall differences in two years.

Examples

Following are a few brief examples to illustrate the use of the formulas. Assume there are three States for which we need to estimate the mean and total allowed charges both individually and for the three combined. We start with the following data. The means and standard deviations are actual 1992 allowed charge data for these States. The remaining data is taken from or derived from Table B2.

	CA	GA	WY
Mean allowed charge (x)	55,357	65,477	30,221
Standard deviation of allowed charge (s)	92,827	97,873	41,858
Universe size (N)	54,717	10,377	800
Sample size (n)	1,094	415	240
Percent of universe selected (p)	2	4	30
Proportion of universe selected (f)	.02	.04	.30
Stratum (h)	1	2	3

Note that if the Ns were not already known, then N = n/f within rounding error. The following calculations are numbered to indicate which formulas from above were used. Details of the first 5 calculations are shown only for WY.

(1) Mean: 30,221 (given)
(2) Standard deviation: 41,858 (given)
(3) SE of mean: 41,858*sqrt[(1-.30)/240] = 2,261
(4) Estimated total: 800*30,221 = 24,176,800
(5) SE of estimated total: 800*2,261 = 1,808,800

Similar calculations were made for the other States to yield the following (means and standard deviations are not repeated):

	CA	GA	WY
SE of estimated mean	2,778	4,707	2,261
Estimated total	3,028,968,969	679,454,829	24,176,800
SE of estimated total	152,003,826	48,844,539	1,808,800

This provides the information needed to make the following stratified calculations for the 3 States.

(9) Total universe: 54,717 + 10,377 + 800 = 65,894

(10) Weights: W_1 = 54,717/65,894 = .830
W_2 = 10,377/65,894 = .157
W_3 = 800/65,894 = .012

(11) Stratified mean: .830*55,357 + .158*65,477 + .012*30,221
= 56,654

(12) SE of stratified mean: $\text{sqrt}[(.830*2{,}778)^2 + (.158*4{,}707)^2 + (.012*2{,}261)^2] = 2{,}423$

(13) Stratified total: 3,028,968,969 + 679,454,829 + 24,176,800
= 3,732,600,598

(14) SE of stratified total: $\text{sqrt}[152{,}003{,}826^2 + 48{,}844{,}539^2 + 1{,}808{,}800^2] = 159{,}669{,}107$

References

Cochran, William G., Sampling Techniques. Third Edition. John Wiley & Sons, New York, 1977.

Deming, William Edwards, Some Theory of Sampling. Dover Publications, New York, 1950.

Kish, Leslie, Survey Sampling. John Wiley & Sons, New York, 1965.

Snedecor, S. W. and Cochran, W. G., Statistical Methods. Seventh Edition. The Iowa State University Press, Ames Iowa, 1980.