ELSEVIER

Design and Simulation Study of the Immunization Data Quality Audit (DQA)

STACY WOODARD, LINDA ARCHER, ELIZABETH ZELL, OLIVIER RONVEAUX, AND MAUREEN BIRMINGHAM

The goal of the Data Quality Audit (DQA) is to assess whether the Global Alliance for Vaccines and Immunization–funded countries are adequately reporting the number of diphtheria-tetanus-pertussis immunizations given, on which the "shares" are awarded. Given that this sampling design is a modified two-stage cluster sample (modified because a stratified, rather than a simple, random sample of health facilities is obtained from the selected clusters); the formula for the calculation of the standard error for the estimate is unknown. An approximated standard error has been proposed, and the first goal of this simulation is to assess the accuracy of the standard error. Results from the simulations based on hypothetical populations were found not to be representative of the actual DQAs that were conducted. Additional simulations were then conducted on the actual DQA data to better access the precision of the DQ with both the original and the increased sample sizes.
Ann Epidemiol 2007;17:628–633.

KEY WORDS: Probability Sample, Sampling Error, Computer Simulation, Vaccination.

BACKGROUND

The Global Alliance for Vaccines and Immunization (GAVI) initiative was launched in the year 2000 to promote increases in childhood immunization levels in 75 developing countries and help support the introduction of new vaccines. After an initial investment in participating countries, GAVI provides a financial reward based on the increase (from one year to the other) in the reported number of diphtheria-tetanus-pertussis (DTP#) vaccinations (1, 2). The number of vaccinations administered in a country is typically tracked through a routine reporting system where the individual health facilities (HFs) report their numbers to a district (province and/or region) level, these numbers are compiled, and the totals reported to the national level. An auditing tool was needed to verify the countries' reported vaccinations. To this end, the Data Quality Audit (DQA) was developed (3–5). In this article, the DQA design will be described along with its verification measure, the results of a simulation study designed to evaluate the proposed standard error estimator will be presented, the impact of zero recounts on the standard error will be investigated, and the precision of the DQA verification measure will be assessed.

Design of the DQA

The sampling design for the DQA is a modified two-stage cluster sample in which 4 districts (clusters) are sampled with probability proportional to the number of reported DPT3 vaccinations. Within each of the selected districts, the HFs are stratified into 3 strata (large/medium/small) with respect to the number of reported DPT3 vaccinations. Within each of the three strata, two HFs are randomly selected—for a total sample size of 24 health facilities. At each of the selected HFs, the annual number of DPT3 vaccinations is recounted from the HF's record.

Calculation of the Verification Factor

The data collected in the DQA are used to calculate the national verification factor (VF). The VF is the ratio of the number of recounted, or verified vaccinations, to the number of vaccinations reported to the World Health Organization (WHO). The data collected include (1) the total numb of DPT3 vaccinations that the country **reported** to WHO for each of the four selected clusters (e.g., health districts), (2) the number of **reported** vaccinations for each of the selected HFs, and (3) the number of **recounted** vaccinations for each of the selected HFs. For the calculation of the VF we define the following notation:

Let

m = the number of clusters (health districts) selected = 4
i = cluster indicator (i = 1, 2, 3, 4)

From the Department of Biostatistic Quintiles Inc., Research Triangle Park, NC (S.W.); IntraHealth International, Nairobi, Kenya (L.A.); Centers for Disease Control and Prevention, Atlanta, GA (E.Z.); and Vaccine Assessment & Monitoring, Department of Immunization, Vaccines and Biologicals, World Health Organization, Geneva, Switzerland (O.R., M.B.).

Address correspondence to: Dr. Stacy Woodard, 5927 South Miami Blvd, Morrisville, NC 27560. Tel.: (919) 924-6399; fax: (919) 998-2389. E-mail: stacy.woodard@quintiles.com.

Received September 30, 2005; accepted January 16, 2007.

1047-2797/07/$–see front matter
doi:10.1016/j.annepidem.2007.01.038

Selected Abbreviations and Acronyms

DQA = data quality audit
DTP3 = diphtheria-tetanus-pertussis
GAVI = Global Alliance for Vaccines and Immunization
HF = health facility
SE = standard error
VF = verification factor
WHO = World Health Organization

N = the total number of **reported** vaccinations in the country
h = stratum (large/medium/small) indicator ($h = 1, 2, 3$)
N_{hi} = the number of **reported** vaccinations in the hth stratum of the ith cluster
j = HF indicator ($j = 1, 2, \ldots, q$)
q = number of HFs selected per cluster
x_{hij} = the number of **reported** vaccinations in the jth HF in the hth stratum of the ith cluster
y_{hij} = the number of **recounted** vaccinations in the jth HF in the hth stratum of the ith cluster.

To obtain an estimate of the actual number of vaccinations administered in the ith cluster, we use a *separate ratio estimator* (6). This estimator incorporates the fact that, although the reported number of vaccinations in the individual HFs will vary from sample to sample, the reported number of vaccinations for the stratum and the cluster are fixed quantities, regardless of which HFs are selected.

The estimated number of actual vaccinations administered in the ith cluster is given by the equation (6):

$$\hat{A}_i = \sum_{h=1}^{L} N_{hi} \left(\frac{\sum_{j=1}^{q} y_{hij}}{\sum_{j=1}^{q} x_{hij}} \right) \tag{1}$$

Once the estimated number of actual vaccinations are obtained for each of the four clusters, the estimates are combined to obtain an overall estimate for the country. The overall ratio (R) of the estimated number of actual vaccinations to the number of reported vaccinations is given by the equation:

$$\hat{R} = \frac{\sum_{i=1}^{m} \hat{A}_i}{\sum_{i=1}^{m} N_i} = \frac{\text{Sum of the Estimated Number of Vaccinations in the 4 clusters}}{\text{Sum of the Reported Vaccinations in the 4 clusters}} = \text{VF} \tag{2}$$

This ratio will be referred to as the "verification factor," denoted by VF.

Standard Error Estimator

In general, the standard error for an estimate from a cluster sample is given by:

$$S\hat{E}(\hat{R}) \approx \sqrt{\frac{\sum_{i=1}^{m} (\hat{R}_i - \hat{R})^2}{m(m-1)}} = SE_1 \tag{3}$$

$$where\ \hat{R}_i = \frac{\hat{A}_i}{N_i} = \frac{\text{Estimated of Actual Vaccinations for cluster } i}{\text{\# of Reported Vaccinations in cluster } i}. \tag{4}$$

This standard error estimator assumes that the clusters were sampled with probability proportional to size and that the HFs within each of the chosen clusters were **randomly** selected. However, in the DQA, the HFs in the clusters are not randomly selected, but selected using a stratified sample. This standard error estimator only incorporates the between-cluster variability and does not incorporate the within-cluster (or within-strata) variability. Also note that this standard error estimator is only an approximation to the true standard error of a ratio estimator; the actual standard error of a ratio estimator is based on a Taylor series expansion and would incorporate the correlation between the recounted and reported numbers of immunizations (6). An approximation to the Taylor Series estimate of the standard error of the VF from the DQA would be:

$$S\hat{E}(\hat{R}_{VF}) \approx \sqrt{\frac{1}{\bar{x}^2} \frac{\sum_{i=1}^{m} (y_i - \hat{R}x_i)^2}{m(m-1)}} = SE_2 \tag{5}$$

where m = the number of selected clusters,
x_i = the number of reported vaccinations in cluster i,
y_i = the estimated number of actual vaccinations in cluster i,
$\bar{x}$ = the average of the reported values from the selected clusters, and
$\hat{R}$ = the estimated ratio of actual to reported vaccinations.

The derivation of this calculation is provided in the Appendix. This standard error formulation is a function of the squared differences of the estimated number of actual vaccinations to the number of reported vaccinations, after the reported number has been "adjusted" by the estimated VF. This standard error equation, however, also assumes that the health facilities within a cluster were randomly selected. It still ignores the fact that in this sampling design the health facilities within a cluster are selected by a stratified random sample. A formulation for a standard error estimator that incorporates the stratification of the sample was

AEP Vol. 17, No. 8
August 2007: 628–633

also investigated, but was found to yield an invalid estimate of the variance. This was because the estimated variance from the actual number of vaccinations (which incorporates the stratification) was much smaller than the variance for the number of reported vaccinations, producing a negative value.

SIMULATION

Given that both of the proposed estimated standard errors of the VF (equations 3 and 5) are approximations, a simulation study was conducted to assess their accuracy. The first step in conducting the simulation was to create populations from which repeated samples could be selected. A total of nine different populations were created.

Populations

Data for the creation of the populations were obtained from different sources. The first source was data on the districts and HFs in the country of Burkina Faso. The second source was the data collected from countries where the DQA had been pilot tested (Pakistan, Uganda, Sri Lanka, Kenya, Mali, Tanzania, Côte d'Ivoire, Liberia). The number of recounted DPT3 vaccinations was assigned by simulating VFs from both normal and beta distributions, and then multiplying the number of reported vaccinations for each health facility by the simulated VF. Differing values of the VF and associated standard deviation were used to generate a total of eight populations.

In the pilot DQA samples, approximately 12% of the HFs and zero recounted values (i.e., no vaccinations could be verified at the HF usually due to missing records). Because of this, the number of recounted vaccinations for the eighth population was simulated such that 12% of the HFs had a recount value of zero. To investigate the effect of zero recounted values, a ninth population was created such that all of the HFs (even those with an observed zero recount) had non-zero simulated recount values.

Methods

From each of the nine populations, 1000 samples were selected using the sampling design for the DQA. The VF was calculated from each sample and the sampling distribution of the VF was estimated. The "true" standard error of the estimated VF was then estimated by the standard deviation of the sampling distribution of the VF. In addition to the VF, the approximated standard errors, bias, and absolute bias were calculated from each of the samples. Descriptive statistics (mean, standard deviation, minimum, maximum values from the 10,000 samples) were calculated for each of the parameters from each of the simulation populations.

Given that the width of the confidence intervals from the actual DQA samples were much larger than expected, it was of interest to determine how much the precision could be increased with a larger sample size. To this end, additional simulations were conducted, increasing the number of clusters selected and/or the number of HFs selected per cluster and the precision calculated. Precision is a measure of the width of the 9th confidence interval about the VF (i.e., Precision $= t_{(\alpha/2),df} * S\hat{E}(\hat{R})$).

The simulations were programmed in C++ (7). The statistical analyses were conducted using Stata Version 7.0 (8). Ancillary Table 1 provides a description of each of the stimulated populations.

RESULTS

The results of the simulations to estimate the standard error of the VF found that both of the approximated standard errors (SE_1, SE_2) are very close to the true standard error of the estimated VF. Even though the difference in magnitude between the estimated and "true" standard errors is small, the percentage difference ranged from 4% to 21%. Note that the standard error approximation based on the ratio estimator is also a little more variable than the cluster estimator. The average bias is very close to zero, as would be expected, confirming that the estimated ratio of actual to reported vaccinations is an unbiased estimate of the true population value of the VF.

The zero recounts were found not to have an impact on the estimated standard error of the estimate or the bias. A reason that the zero recounts do not have much of an impact on the standard error of the VF is that the zero recounts are not the only reason for large discrepancies between the reported and recounted number of vaccinations.

Table 1 presents the results of the simulations to show the precision of the VF for the original sampling design as well as for increases in sample size. These simulations were based on the population using the data from the eight pilot DQAs (population 8). The number of clusters was increased to 10, 15, 20, and 30, with the selection of one, two, and six HFs per stratum in each cluster. Little difference in precision was found when the number of HFs was increased from one to two to six per stratum. Given that it is more time- and cost-effective to increase the number of HFs per stratum than to increase the number of clusters (districts), this result is disappointing. The selection of two HFs per stratum in each cluster appears to be the optimal number, regardless of the number of clusters selected. The original sample size (of four clusters and two HFs per stratum in each cluster) yields an average precision of ± 45 percentage points. Doubling the sample size to 8 clusters (with two HFs per stratum) would increase the average precision by almost half (i.e., to

TABLE 1. Estimated precision of the verification factor simulated from 10,000 samples selected from Population 8 with increases in the number of districts and/or the number of health facilities selected per cluster

			No. of districts (clusters)					
			4	8	10	15	20	30
Population	True VF (SD)	No. of HFs per stratum	Estimated precision					
8	0.59 (0.70)	1	0.472	0.249	0.212	0.164	0.138	0.110
		2	**0.446**	0.238	0.202	0.156	0.131	0.104
		6	0.469	0.245	0.209	0.159	0.134	0.106

HF = Health facility; SD = standard deviation; VF = verification factor.
Note: Precision is based on the estimated standard error multiplied by the critical value from a 95% confidence interval based on a *t* distribution with *m*-1 degrees of freedom (where *m* is the number of districts sampled).
Bolded value is the precision using the original sampling design.

±24). However, almost doubling the sample size again (i.e., to 15 clusters) only increases the average precision by about one third (to ±16). Thirty clusters are needed before an average precision of ±10 points are observed. Ancillary Table 2 provides the results from all 8 of the stimulated populations.

CONCLUSION

The original goal of this simulation study was to determine how well the proposed standard error estimator for the VF approximates the true standard error. The results indicate that even though the proposed standard error estimator does not incorporate the fact that a stratified random sample (rather than a simple random sample) was taken within each of the selected clusters, the proposed formula provides a reasonable approximation to the true standard error.

The objective of the DQA is to verify the accuracy of the number of DTP3 vaccinations reported, so that the GAVI rewards can be fairly and objectively awarded using consistent rules. In the eight DQAs that were conducted in 2001, the major differences in the reported and recounted number of vaccinations were due to missing records. However, the zero recount values were not found to have a significant impact on the estimated standard error of the VF. To improve the precision of the VF, a sample size of 45 HFs would be needed. To obtain an estimated VF with a precision of ±10 points, 30 clusters would be needed. Given that the DQA is to be completed in a 2-week period by two evaluation teams, the large increase in sample size is not feasible (because of logical and financial limitations).

Although the original concept of the DQA and associated VF was to "adjust" the GAVI shares, the large variability of the estimated VF made this impracticable. However, the VF has been used to categorize the reliability of the countries reporting system as follows: 85% ⩽ VF < 115% indicates consistent reporting, VF ⩾ 115% indicates underreporting, and VF < 85% indicates overreporting (5). In this manner, the DQA has been used to help countries identify problems in their reporting system. The populations with consistent reporting have smaller standard errors, so that once the countries increase their reliability, the VF may be precise enough to adjust shares, even without increasing the sample size.

The DQA is an innovative tool; its use allows the international donors to be reassured about the reliability of country-reported figures. It can be used for a variety of monitoring systems, in a short time-frame, with relatively small sample sizes. Although the initial goal of the DQA was to enable an external entity to monitor reporting systems, countries can use this tool themselves to monitor their own systems.

REFERENCES

1. Wittet S. Introducing GAVI Global Fund for Children's Vaccines. Vaccine. 2000;91:385–386.
2. Brugha R, Starling M, Walt G. GAVI, the first steps: lessons for the Global Fund. Lancet. 2002;359:435–438.
3. World Health Organization. The immunization data quality audit (DQA) procedure [WHO document WHO/V&B/03.19]. Geneva: World Health Organization; 2003.
4. The LATH Consortium. Immunization data quality audit evaluation report: final report from Deloitte Touche Tohmatsu Emerging Markets, Euro Health Group, and Liverpool Associates in Tropical Health. Geneva: World Health Organization; 2001.
5. Ronveaux O, Rickert D, Hadler S, Groom H, Lloyd J, Bchir A, Birmingham M. The Immunization Data Quality Audit: verifying the quality and consistency of immunization coverage monitoring systems. Bull World Health Organ. 2005;83:503–510.
6. Levy P, Lemeshow S. Sampling of populations: methods and applications. 3rd ed. New York: Wiley; 1999.
7. Visual C++. Microsoft Corporation; 1994.
8. Intercooled Stata 7.0 for Windows 98/85/NT. College Station (TX): Stata Corporation; 2000.

APPENDIX

The standard error presented in equation 3 is not only an approximated standard error for this sampling design, it is also an approximation of the standard error of a ratio estimator. The variance of the ratio of two random variables

AEP Vol. 17, No. 8
August 2007: 628–633

($R = Y/X$), as obtained by a Taylor Series Expansion, is given by the equation:

$$Var(R) = \frac{1}{\mu_x^2}\left(\sigma_x^2\frac{\mu_y^2}{\mu_x^2} + \sigma_y^2 - 2\rho\sigma_x\sigma_y\frac{\mu_y}{\mu_x}\right) \quad (1)$$

From this equation the estimated standard error of a general ratio estimator can be calculated as

$$S\hat{E}(\hat{R}) = \sqrt{\frac{1}{\bar{x}^2}\left(s_x^2\frac{\bar{y}^2}{\bar{x}^2} + s_y^2 - 2\hat{\rho}s_x s_y\frac{\bar{y}^2}{\bar{x}^2}\right)} \quad (2)$$

$$= \sqrt{\frac{1}{\bar{x}^2}\,\frac{\sum_{i=1}^{n}(y_i - \hat{R}x_i)^2}{n(n-1)}} \text{ where } \hat{R} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \quad (3)$$

For our modified two-stage cluster sample, an alternative standard error formula for the verification factor can therefore be estimated as

$$S\hat{E}(\hat{R}_{VF}) \approx \sqrt{\frac{1}{\bar{x}^2}\,\frac{\sum_{i=1}^{m}(y_i - \hat{R}x_i)^2}{m(m-1)}} \quad (4)$$

where m = the number of selected clusters,
x_i = the number of reported vaccinations in cluster i,
y_i = the estimated number of actual vaccinations in cluster i,
$\bar{x}$ = the average of the reported values from the selected clusters, and
$\hat{R}$ = the estimated ratio of actual to reported vaccinations.

ANCILLARY TABLE 1. Description of the simulation populations

Population	Country(s)	True VF (SD)	No. of districts (clusters)	No. of HFs	Strata*	No. of vaccinations reported at HF†	No. of vaccinations recounted at HF‡
1	Burkina Faso	0.97 (0.05)	46	1046	10% Large	Large—*N* (1500, 300†)	*N* (0.97, 0.5)
2		0.75 (0.10)			50% Medium	Medium—*N* (600, 150†)	*N* (0.75, 0.1)
					40% Small	Small—*N* (150, 45†)	*N* (0.50, 0.15)
3	Sri Lanka, Kenya	0.50 (0.15)	8		Known	Known	*N* (0.97, 0.5)
4		0.97 (0.05)					*N* (0.75, 0.1)
5		0.75 (0.10)		107			*N* (0.50, 0.15)
6	Burkina Faso	0.50 (0.15)	46	1046	10% Large	Large—*N* (1500, 300†)	*N* (0.70, 0.5)
7		0.70 (0.50)			50% Medium	Medium—*N* (600, 150†)	
					40% Small	Small—*N* (150, 45†)	
8	8 DQAs (with zero recounts)	0.59 (0.70)	32	1043	Known	Known	190 HFs known (23 with zero recount), 853 HF Beta ($\alpha = 1$, $\beta = 1$), 12% with zero recount
9	8 DQAs (without zero recounts)	0.56 (0.70)	32	1043	Known	Known	167 HFs known (no zero recount), 876 HF Beta ($\alpha = 1$, $\beta = 1$)

DQA = Data quality audit; HF = health facility; SD = standard deviation; VF = verification factor.
*Where applicable, strata were assigned on the basis of simulated values from a uniform distribution.
†Where applicable, the number of reported vaccinations were simulated from normal distributions with the parameters depending on the stratum.
‡The number of recounted vaccinations were calculated as the number of reported vaccinations multiplied by the simulated VF, where the VF was simulated from the specified distribution.

ANCILLARY TABLE 2. Descriptive statistics for simulation parameters from 10,000 samples selected from each of the simulated populations

Population						
True VF (SE)	Statistic	VF	SE_1	SE_2	Bias	Abs Bias
1	Mean (SE)	0.97 (0.01)	0.010 (0.004)	0.010 (0.005)	0.000 (0.01)	0.01 (0.01)
0.97 (0.01)	Min, Max	0.93, 1.02	0.00, 0.03	0.00, 0.03	−0.04, 0.05	0.00, 0.05
2	Mean (SE)	0.76 (0.02)	0.022 (0.009)	0.020 (0.009)	0.003 (0.02)	0.02 (0.01)
0.75 (0.02)	Min, Max	0.67, 0.85	0.00, 0.06	0.00, 0.07	−0.08, 0.10	0.00, 0.10
3	Mean (SE)	0.52 (0.04)	0.034 (0.015)	0.031 (0.014)	0.006 (0.04)	0.03 (0.02)
0.50 (0.03)	Min, Max	0.36, 0.65	0.00, 0.11	0.00, 0.10	−0.15, 0.14	0.00, 0.15
4	Mean (SE)	0.97 (0.02)	0.014 (0.005)	0.014 (0.007)	−0.001 (0.02)	0.01 (0.01)
0.97 (0.01)	Min, Max	0.93, 1.03	0.00, 0.04	0.00, 0.04	−0.04, 0.06	0.00, 0.06
5	Mean (SE)	0.75 (0.03)	0.028 (0.010)	0.027 (0.013)	−0.001 (0.03)	0.03 (0.02)
0.75 (0.02)	Min, Max	0.65, 0.87	0.00, 0.07	0.00, 0.09	−0.10, 0.11	0.00, 0.12
6	Mean (SE)	0.50 (0.05)	0.041 (0.015)	0.041 (0.020)	−0.002 (0.05)	0.04 (0.03)
0.50 (0.03)	Min, Max	0.36, 0.68	0.00, 0.10	0.00, 0.13	−0.15, 0.17	0.00, 0.17
7	Mean (SE)	0.74 (0.11)	0.093 (0.039)	0.090 (0.043)	0.001 (0.11)	0.09 (0.06)
0.70 (0.10)	Min, Max	0.37, 1.18	0.00, 0.27	0.00, 0.32	−0.37, 0.44	0.00, 0.44
8	Mean (SE)	0.59 (0.18)	0.140 (0.046)	Not done	−0.004 (0.18)	0.15 (0.10)
0.59 (0.70)	Min, Max	0.16, 0.98	0.00, 0.28		−0.44, 0.39	0.00, 0.44
9	Mean (SE)	0.56 (0.18)	0.140 (0.048)	Not done	0.005 (0.18)	0.15 (0.10)
0.56 (0.70)	Min, Max	0.12, 0.98	0.00, 0.29		−0.44, 0.38	0.00, 0.44

VF = Estimated verification factor; SE_1 = estimated standard error based on the cluster sample formula (equation 3); SE_2 = estimated standard error based on the Taylor Series Expansion (equation 5); Bias = difference between the true and estimated VF; Abs Bias = absolute difference between the true and the estimated VF.
Note: The true standard error is calculated as the population standard deviation divided by the square root of the sample size (n = 24).