Quality Control of Birth Defect Registry Data: A Case Study

JANE SCHULMAN, PhD
JUDITH A. HAHN, MA

At the time of the study, Dr. Schulman was Senior Biostatistician and Ms. Hahn was a Biostatistician, California Birth Defects Monitoring Program (CBDMP). Dr. Schulman is now a Senior Research Scientist with the Battelle Memorial Institute, Arlington, VA.

The authors are grateful to Deborah Gant of CBDMP for coordinating the field work and to the data collection staff at the CBDMP for performing the field work for this study.

Tearsheet requests to Dr. Schulman, Battelle Memorial Institute, 2101 Wilson Blvd., Suite 800, Arlington, VA 22201; tel. 703-875-2943.

Synopsis

The California Birth Defects Monitoring Program maintains a population-based registry of children born with congenital malformations. Trained data collectors routinely visit hospitals and genetics centers to identify cases and abstract information. These data are provided to the public health, medical, and lay communities and are used for conducting prevalence and case-control studies.

A stratified sample of each data collector's work for one data year was reevaluated to assess the quality of case ascertainment and record abstraction. The sample included data from 109 facilities (37 percent) and 729 abstracts (5 percent).

There are three steps in data collection: case-finding, the process of identifying potential cases; culling, the process of reviewing the charts of potential cases to determine which are reportable; and abstracting, the process of recording information from the charts of reportable cases. The probability that a potential case is missed during casefinding is 7 percent for small facilities, 4 percent for medium facilities, and 1 percent for large facilities. The probability that a reportable case is mistakenly classified as not reportable during culling is 3 percent for small and medium facilities and 1 percent for large facilities. The probabilities of incorrectly abstracting selected diagnoses and demographic items are slightly higher (8 percent for small and medium facilities and 6 percent for large facilities) because these are more complex processes than are casefinding and culling. Finally, the overall probability of missing a case from the registry is 3 percent.

Therefore, these data indicate that the information collected by this registry are both reliable and complete.

THE CALIFORNIA BIRTH DEFECTS Monitoring Program (CBDMP) maintains a population based registry of children born with major structural malformations. This program was begun in five San Francisco Bay Area counties in 1983, and now more than 300,000 births per year are monitored. Trained data collectors routinely visit all nonmilitary hospitals and genetics centers to identify children with major malformations diagnosed prior to their first birthday. Detailed demographic and diagnostic information for those children who meet the program's reportability criteria is abstracted. In addition, data for reportable children are abstracted from every facility in which the child is admitted as an inpatient (*1,2*).

The CBDMP provides information to the public health, medical, and lay communities, and responds to community concerns by investigating reported clusters. Registry data are also used to conduct prevalence and case-control studies. Therefore, the information collected by the registry needs to be complete and accurate to ensure that any conclusions drawn using these data are valid.

Many disease registries have been concerned with issues of completeness and accuracy. For example, some programs estimate the completeness of case ascertainment by re-casefinding a certain proportion of cases (*3–6*), while others estimate the accuracy or consistency of abstracting procedures by re-abstracting already known cases (*7–12*). A few programs have used a combined approach of re-casefinding and re-abstracting (*13–18*). Prevalences or rates observed in one registry have been compared with those found in other registries as a measure of completeness (*8–12,19,20*). The primary goals of this study were to estimate the chances that children who meet the CBDMP's reportability criteria are included in the registry and that the

STATISTICAL NOTE
Calculation of Probability Estimates in Table 1.

The probability of missing a case during culling is shown for illustration. The other probability estimates given in table 1 are calculated analagously. The subscript i refers to data collector ($i=1,2,\ldots,24$), the subscript j refers to facility size ($j=1,2,3$), and X_{ij} refers to the number of cases culled incorrectly by data collector i at facility size j during the quality control sample time period. Similarly, n_{ij} is the number of cases that should have been identified by data collector i at facility size j during the quality control sample time period. Finally, N_{ij} is the total number of reportable cases identified by data collector i at facility size j during 1988 and N is the total number of reportable cases identified by all data collectors at all facilities during 1988, that is,

$$N = \Sigma\Sigma N_{ij}.$$

The estimated probability of missing a case during culling for data collector i at facility size j is

$$\hat{p}_{ij} = \frac{x_{ij}}{n_{ij}}$$

and the overall probability of missing a case during culling is

$$\hat{p} = \Sigma\Sigma \frac{N_{ij}}{N} \hat{p}_{ij}$$

Since X_{ij} is assumed to be a binomial random variable with parameters p_{ij} and n_{ij}, the estimated variance of this overall probability is

$$\hat{Var}(\hat{p}) = \Sigma\Sigma \frac{N^2_{ij}}{N} \left(\frac{p_{ij}(1-p_{ij})}{n_{ij}-1}\right)\left(\frac{N_{ij}-n_{ij}}{N_{ij}}\right)$$

with an associated 95 percent confidence interval of
$(p-1.96\sqrt{\hat{var}(\hat{p})}, \hat{p}+1.96\sqrt{\hat{var}(\hat{p})}$.

The Probability of Missing a Case from the California Birth Defects Monitoring Program Registry

Presented is an outline for estimating the probability of missing a case from the registry. The probability that a case will be incorrectly excluded from the registry can be represented as the sum of three probabilities

1. the probability that the data collector did not casefind the case at one facility and the case was not identified at another reporting source,
2. the probability that the data collector identified the case during casefinding, incorrectly culled the chart and the case was not identified at another reporting source, and
3. the probability that the data collector identified the case during casefinding, culled the chart correctly but the abstract was misplaced during data processing and was not identified elsewhere.

The chart represents these three possibilities using a flow chart and shows that the first probability was $(4 \div 1119) \times (2 \div 4) = .0018$. The second probability was $(1115 \div 1119) \times (32 \div 1115) \times (25 \div 32) = .0223$ and the third was $(1115 \div 1119) \times (1083 \div 1115) \times (9 \div 1083) = .0080$. Therefore, the sum $.0018 + .0223 + .0080 = .032$ would be the estimated probability of missing a case from the registry. The value of this probability presented in the text (.03; 95 percent confidence interval = .02,.04) was calculated using this general outline but incorporates the appropriate population weights at each step as described in the first part of this box.

information abstracted for each reportable case is accurate.

Methods

There are three steps in the CBDMP's data collection procedure—casefinding, culling, and abstracting. Casefinding refers to the process of reviewing the diagnostic index and all the logs (for example, pathology, autopsy, surgery, labor, and delivery) in a hospital or genetics center to identify potential cases. Potential cases are identified as those children (*a*) with a very broad spectrum of malformations as recorded on the diagnostic index based on ICD-9-CM codes, or (*b*) who have one or more of many conditions (for example, congenital syphilis, club foot) or notes written by their names in the logs such as "death, twin whose mate

is stillborn," that may reflect accompanying malformation(s).

The medical charts for all children identified as potential cases are then reviewed to determine who are reportable to the system. Approximately 30 percent of all those identified as potential cases meet the program's reportability criteria. This chart review process is referred to as culling. Finally, the diagnostic and demographic information for each reportable case is abstracted.

There are a variety of errors that might occur during casefinding, culling, and abstracting. A child may be mistakenly included or excluded by the data collector during casefinding or culling. Mistakenly including nonreportable children at the data collection stage is not of concern because this error will be detected during the coding stage of data processing. Mistakenly excluding reportable cases is of major concern, however. During abstracting, a reportable diagnosis may be missed, a reportable diagnosis may be mistakenly recorded, or a nonreportable diagnosis may be recorded. The first type of error is the most serious because it will result in underascertainment. Nonreportable diagnoses, on the other hand, are likely to be detected either during the coding stage of data processing or during data analysis. Diagnoses that were mistakenly recorded may be discovered during data analysis because of the presence of incompatible diagnoses for the same child (for example, spina bifida with and without hydrocephalus). Finally, an item of demographic information, such as sex, race, address, and so forth, might be incorrectly recorded during abstracting.

To estimate the chances of the errors considered to be serious, approximately 30 abstracts were re-abstracted for each data collector. These abstracts were obtained by first selecting a stratified random sample of all facilities for which a particular data collector was responsible. The strata were defined by the number of abstracts obtained from a particular facility in the previous year; small (less than 50 abstracts), medium (50–150), and large (more than 150). The number of facilities chosen from each stratum was proportional to the total number of facilities in that stratum. Abstracts were then sampled within each facility in proportion to the total number of abstracts obtained from that facility in the previous year.

In each facility, a reviewer re-casefound all cases diagnosed during a particular time period—two months, for example—culled all the charts to determine which were reportable, and then abstracted data for a certain number of reportable cases. The number of cases re-abstracted was proportional to the size of the facility. The length of the casefinding period was determined by the number of abstracts to be re-abstracted from the facility. The start date of the casefinding period was chosen randomly.

'The CBDMP provides information to the public health, medical, and lay communities, and responds to community concerns by investigating reported clusters. Registry data are also used to conduct prevalence and case-control studies.'

Because of practical considerations, each data collector was trained to serve as the reviewer for one other data collector for this study. The number 30 was also chosen primarily because of practical considerations. However, 30 abstracts per data collector allowed us to estimate precisely the probabilities of interest. As an example, from a sample size point of view, the mistake that is the most difficult to estimate is the probability of a culling error because it has the smallest denominator of all the error probabilities considered; that is, the denominator is based on the number of abstracts versus the number of potential cases or the number of diagnoses. Our sampling scheme yielded approximately 225 abstracts in the facility size stratum with the smallest number of abstracts, the facility size 1 stratum. Even in this "worst case scenario," we have 95 percent confidence that our estimate differs from the true probability of a culling error by no more than .02 if the true probability was .05 (*21*).

Separate probabilities for missing a potential case during casefinding, missing a reportable case during culling, missing diagnoses during abstracting, and incorrectly abstracting demographic items were estimated. Each of these probabilities is a weighted average of all the data collectors' estimated probabilities, with weights proportional to the total number of abstracts contributed to the data base by each data collector (see box).

The overall probability of missing a case from the registry was also estimated (see box and figure for details). The CBDMP's data collection procedures specify that data for any reportable case be abstracted from every facility in which the child is admitted as an inpatient. Therefore, this probability is basically the sum of three probabilities

Table 1. Estimated error probabilities and 95 percent confidence intervals (CI) by facility size, California Birth Defects Monitoring Program study, 1988

Error	Small		Medium		Large		Total	
	Estimated probability	CI	Estimated probability	CI	Estimated probability	CI	Estimated probability	CI
Missing a potential case during casefinding..	.069	.046,.091	.039	.029,.050	.011	.005,.016	.037	.029,.044
Missing a case during culling.....	.028	.007,.050	.034	.017,.051	.008	.001,.015	.022	.014,.030
Missing selected reportable diagnoses[1]	.058	.002,.115	.102	.062,.141	.047	.021,.072	.070	.048,.091
Incorrectly abstracting a demographic item ..	.061	.055,.066	.074	.070,.078	.051	.047,.053	.061	.059,.064

[1] Probability of missing a diagnosis from 1 of 46 categories (see text).

1. the probability that the data collector missed the case during casefinding and the case was not identified at another reporting source,
2. the probability that the data collector identified the case during casefinding but incorrectly culled the chart and the case was not identified at another reporting source, and
3. the probability that the data collector identified the case during casefinding, culled the chart correctly but the abstract was never key-entered and was not identified elsewhere.

To estimate error probabilities for specific defects, individual diagnoses were coded using a modified version of the British Pediatric Association system and classified into 46 malformation categories using a scheme developed by the Metropolitan Atlanta Congenital Defects Program and the Centers for Disease Control. Also, to account for the fact that abstracts are often updated with new or additional diagnostic information, we excluded from analysis any diagnosis identified by a reviewer that was made after the original data collector abstracted the chart.

Results

Included in this quality control sample were 109 facilities, or approximately 37 percent of all facilities that the CBDMP monitored during 1988. The quality control reviewers re-casefound 3,654 potential cases, of which 1,119 were declared reportable during culling. A total of 729 abstracts, approximately 5 percent of all abstracts obtained during 1988, were re-abstracted and contained 3,089 individual diagnoses.

The probability of missing a potential case during casefinding decreases as facility size increases (table 1)—.07 for small facilities (95 percent confidence interval [CI] .046,.091), .04 for medium facilities (95 percent .029,.050), and .01 for large facilities (95 percent .005,.016). No such pattern is evident for the probabilities of missing a case during culling, missing a reportable diagnosis, or incorrectly abstracting a demographic item.

During casefinding, the original data collectors failed to identify 137 of the 3,654 potential cases identified by the quality control reviewers. These 137 mistakes fell into approximately 60 categories. The most common mistake was failure to identify 34 children who had the notation "transfer to tertiary care" associated with their names in the logs. The other two most frequently missed notations or conditions were 8 "amniotic bands" and 6 "urinary tract infection." We use the notation "urinary tract infection" as a casefinding trigger because children who have these infections during the first year of life often have underlying renal anomalies. The frequency in all the other categories was less than 5.

Of the 1,119 cases that the reviewers declared reportable during culling, 32 were incorrectly culled by the original data collectors. These 32 cases had a total of 72 diagnoses. No consistent pattern accounted for these missed cases.

The probability that a reportable diagnosis was missed during abstracting by the original data collector is shown in table 2 for each of the 46 categories. As opposed to the probabilities shown in table 1, these estimates do not take the stratified sampling scheme into account. In addition, we have assumed that an error was not committed if a data collector abstracted an individual diagnosis differently than the reviewer, provided that both diagnoses were classified into the same category. The number of diagnoses re-abstracted was very small for many of the 46 categories, making the corresponding 95 percent wide and interpretation difficult. For example, the estimated probability of missing a diagnosis of choanal atresia is .50 but this number is based on missing two of four, giving a 95 percent CI of .07 to .93. Therefore, probabili-

ties and 95 percent CIs are also given for groups of categories, such as nervous system diagnoses, chromosomal anomalies, and so on.

Table 2 indicates that 5 of the 8 most common defects monitored by the CBDMP are not likely to be missed—Down Syndrome (.02; 95 percent CI = .00,.11), hip dislocation (.05; 95 percent CI = .01,.18), obstruction of the kidney or ureter (.04; 95 percent CI = .00,.23), pyloric stenosis (.02; 95 percent CI = .00,.11), and cleft lip with or without cleft palate (.02; 95 percent CI = .00,.11). Higher error probabilities were observed for hydrocephalus (.17; 95 percent CI = .06,.39), microcephalus (.14; 95 percent CI = .05,.32), and cleft palate (.12; 95 percent CI = .03,.32).

The CBDMP assigns a measure of precision to each diagnosis that is abstracted. This measure refers to the degree of certainty associated with the diagnosis as it appears in the chart. For example, diagnoses with low precision have adjectives such as "possible" or "consider" associated with them, and diagnoses with high precision have associated adjectives such as "most likely" or "probable." The CBDMP generally excludes diagnoses with low precision from analyses and prevalence reports because experience has shown that most of these represent false-positive diagnoses.

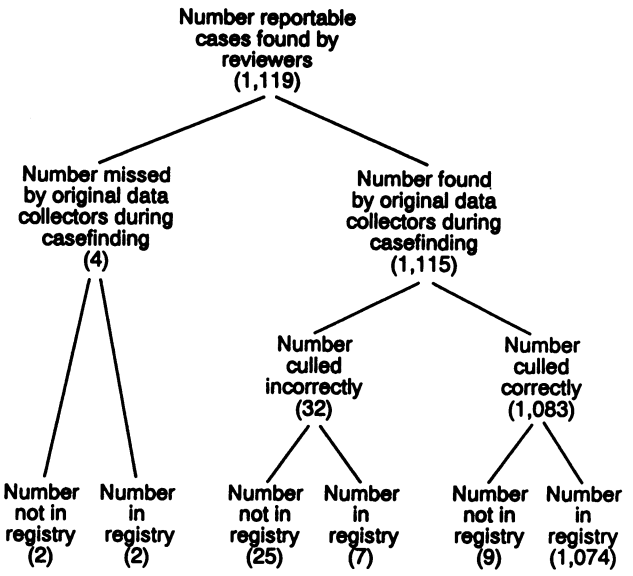
Table 3 shows the frequency of diagnostic errors separately for low and high precision diagnoses. Fifteen (30 percent) of the 50 missed diagnoses shown in table 2 had low precision codes implying that, although these diagnoses were technically missed, the consequences of these errors are minimal. One of the three missed diagnoses of cleft palate and one of the four missed diagnoses of hydrocephalus were low precision diagnoses.

In addition, the errors in table 3 are displayed separately by the type of error—abstracting, coding, and key-data entry. For example, of the 35 errors among high precision diagnoses, 31 could be attributed to incomplete abstracting by the original data collector, 3 resulted from mistakes in the coding of the original diagnosis, and 1 was an error in key-data entry. The error in key-data entry resulted in an apparently missed diagnosis of hydrocephalus.

The probability of incorrectly abstracting particular demographic items depends on the item being abstracted (table 4). For example, the probability of incorrectly abstracting the child's date of birth is .01 (95 percent CI .00,.02), whereas the probability of incorrectly abstracting the child's first name is .10 (95 percent CI .08,.12).

Finally, we estimate using the method outlined in the box that the probability of excluding a reportable case from the registry is .03 (95 percent CI .02,.04); for small facilities .03 (95 percent CI .01,.05), for medium .05 (95 percent CI .03,.07), and .01 (95 percent CI .00,.02) for large.

Schematic representation for estimating probability of missing a case from the California Birth Defects Monitoring Program registry



Discussion

The chances of missing a case during casefinding and culling are smaller than the chances of incorrectly abstracting demographic and diagnostic information (that is, 2–4 percent versus 6–8 percent). In general, casefinding and culling are simpler and more straightforward processes than abstracting. For example, in order for a child to meet the program's reportability criteria, the data collector must identify only one reportable diagnosis in the chart, assuming that the child meets the program's age and geographic inclusion criteria. When the data collector abstracts a chart, however, he or she must identify every reportable diagnosis that the child has. This can be a difficult task, because there are often multiple volumes to be reviewed for a child who has numerous malformations.

Thirty-four of the 137 casefinding errors arose because a data collector failed to identify a child as a potential case if the notation "transfer to tertiary care" was associated with his or her name in a log. However, the consequences of this error are not serious because these children are usually identified at the tertiary care centers to which they are

Table 2. Probability of missing selected[1] reportable diagnoses by category with 95 percent confidence intervals, California Birth Defects Monitoring Program study, 1988

Condition	Number missed	Number reabstracted	Estimated probability	95 percent confidence interval
Nervous system	9	73	.12	.06,.23
Anencephalus	0	6	.00	.00,.46
Encephalocele	...	0	...	...
Hydrocephalus	4	23	.17	.06,.39
Microcephalus	4	29	.14	.05,.32
Spina bifida	1	15	.07	.01,.32
Eye	8	25	.32	.16,.53
Anophthalmia	...	0	...	...
Cataract	1	6	.17	.00,.64
Glaucoma	3	5	.60	.15,.95
Microphthalmia	4	14	.29	.08,.58
Circulatory system	4	38	.10	.03,.26
Aortic stenosis	0	3	.00	.00,.71
Hypoplastic left heart	2	8	.25	.03,.65
Single ventricle	...	0	...	...
Tetralogy of fallot	0	13	.00	.00,.25
Total anomalous pulmonary venous return	0	1	.00	.00,.97
Transposition of the great vessels	1	10	.10	.00,.44
Truncus arteriosus	1	3	.33	.01,.91
Respiratory system	4	20	.20	.07,.44
Agenesis of lung	2	16	.12	.01,.38
Choanal atresia	2	4	.50	.07,.93
Alimentary system	6	164	.04	.01,.08
Biliary atresia	0	3	.00	.00,.71
Cleft lip with no cleft palate	1	46	.02	.00,.11
Cleft palate	3	24	.12	.03,.32
Hirschsprung disease	1	8	.12	.00,.53
Malrotation of intestine	0	8	.00	.00,.37
Pyloric stenosis	1	48	.02	.00,.11
Stenosis/atresia duodenum	0	12	.00	.00,.26
Stenosis/atresia rectum	0	7	.00	.00,.41
Alimentary system, continued				
Stenosis/atresia small intestine	0	4	.00	.00,.60
Tracheo-esophageal fistula/esophageal atresia	0	4	.00	.00,.60
Genitourinary system	4	46	.09	.03,.22
Bladder/urethra obstruction	1	6	.17	.00,.64
Hypospadius 2nd or 3rd degree	1	6	.17	.00,.64
Obstruction kidney/ureter	1	22	.04	.00,.23
Renal agenesis	1	12	.08	.00,.38
Musculoskeletal system	8	82	.10	.05,.19
Arm/hand limb reduction	1	10	.10	.00,.46
Arthrogryposis multiplex congenita	4	13	.31	.09,.61
Diaphragmatic hernia	0	8	.00	.00,.37
Hip dislocation-dysplasia	2	37	.05	.01,.18
Leg/foot limb reduction	0	3	.00	.00,.71
Scoliosis/lordosis	1	11	.09	.00,.41
Chromosomal anomalies	2	62	.03	.06,.12
Trisomy 13	0	6	.00	.00,.46
Trisomy 18	1	10	.10	.00,.44
Trisomy 21 (Down syndrome)	1	46	.02	.00,.11
Others	5	44	.11	.04,.25
Amniotic bands	2	15	.13	.02,.40
Fetal alcohol syndrome	2	8	.25	.03,.65
Gastroschisis	0	5	.00	.00,.52
Hearing loss with ear anomaly	1	14	.07	.00,.36
Omphalocele	0	2	.00	.00,.84

[1] Probability of missing a diagnosis from one of 46 categories (see text).

referred. In fact, the CBDMP ceased using "transfer to tertiary care" as a casefinding trigger in subsequent data years as part of a program-wide effort to make the data collection procedures more efficient.

The probability of missing a potential case during casefinding is inversely related to facility size. The number of logs available to the data collector is usually larger at large facilities compared with small facilities, which often maintain only a diagnostic index. Therefore, the data collector has more opportunities to identify a potential case at the large facilities compared with small facilities.

The three diagnoses that had relatively high error probabilities for which sufficient data were available for interpretation were hydrocephalus (.17), microcephalus (.14), and cleft palate (.12). The diagnosis of microcephalus is subject to considerable diagnostic variability, making interpretation of prevalence estimates difficult. Therefore, the high error rate for this diagnosis is not of particular concern. Serious consequences could result from missing diagnoses of cleft palate and hydrocephalus. Of the seven missed diagnoses of hydrocephalus and cleft palate combined, however, two were low precision diagnoses and would not have affected prevalence estimates.

The 46 diagnostic categories used for this analysis include approximately half the individual diagnoses in the registry. These categories were chosen

because they (*a*) represent a variety of anatomic abnormalities, (*b*) include mostly major malformations, and (*c*) represent the malformations most frequently addressed in the scientific literature.

The CBDMP routinely matches registry data to California vital statistics birth certificate files to verify inclusion criteria and incorporate additional demographic information into the registry. The demographic items chosen for this analysis (child's name, date of birth, sex, mother's first name, and zip code) are the variables used in matching the cases in the registry to the vital statistics files and are, therefore, of special interest in a quality control investigation.

Three assumptions underlie this analysis. First, we assume that a random sample of abstracts was obtained from each facility. It was not possible, however, to sample abstracts randomly within facilities, because facilities do not maintain master lists of children's names from which to sample. Therefore, the reviewer re-casefound all cases diagnosed during a particular period (for example, two months) and then culled the charts to determine which were reportable. The length of the period was determined by the number of abstracts to be re-abstracted from any particular facility. The start date of the time period was chosen randomly to eliminate any unknown effects of seasonality.

Second, the reviewer was considered to be the gold standard, that is, the information obtained by the data collector who performed the review was used as the baseline for comparison. For practical reasons, each data collector was specially trained to serve as the reviewer for one other data collector. Although it is likely that this gold standard assumption is not totally satisfied, this scheme ensures a lack of systematic bias in the results.

Third, it is assumed that all reportable cases in a facility can potentially be identified by reviewing the diagnostic index and all the logs in a given facility. It is well documented that the diagnostic index is not a suitable source, by itself, for casefinding (*22*). The CBDMP uses *all* the logs in a facility in conjuction with the diagnostic index for casefinding purposes. Nonetheless, because it is infeasible for our data collectors to review every chart in every facility to determine reportability, it is possible that some cases may have been missed because they were not registered in either the diagnostic index or on any log. Based on our field experience, however, we do not believe that this event is very likely.

In summary, the CBDMP has implemented an ongoing quality control program to ensure that complete and accurate data are collected. The results from this study indicate that the chances of missing a case during casefinding and culling are approximately 2–4 percent and the chances of incorrectly abstracting demographic and diagnostic information are approximately 6–8 percent. Overall, the data base is estimated to be 97 percent complete for the 1988 birth year. Therefore, it appears that the CBDMP is correctly identifying children to include in the registry and that reliable data are being collected.

Table 3. Frequency of diagnostic errors, California Birth Defects Monitoring Program study, 1988

Error	High precision diagnoses[1]	Low precision diagnoses[1]
Abstracting	31	14
Coding	3	1
Key data entry	1	0
Totals	35	15

[1] High precision diagnoses are included in prevalence estmates; low precision diagnoses are not (see text).

Table 4. Probability of incorrectly abstracting selected[1] demographic items, with 95 percent confidence intervals, California Birth Defects Monitoring Program study, 1988

Item	Estimated probability	95 percent CI
Child's first name	.100	.077,.123
Child's last name	.061	.043,.079
Date of birth	.008	.001,.015
Sex	.008	.001,.015
Zip code	.071	.052,.089
Mother's first name	.048	.031,.064

[1] These demographic variables are used to match cases to vital statistics files (see text).

References

1. Croen, L., Schulman, J., and Roeper, P.: Birth defects in California, January 1, 1983–December 31, 1986: CBDMP Report Series BDR #3(90), California Birth Defects Monitoring Program, Emeryville, CA 1990.
2. Croen, L., Shaw, G., Jensvold, N., and Harris, J.: Birth defects monitoring in California: a resource for epidemiologic research. Paediatr Perinat Epidemiol 5: 423–427 (1991).
3. Rice, M. J., and Kaplan, E. L.: Rheumatic fever in Minnesota, II. Evaluation of hospitalized patient and utilization of a state rheumatic fever registry. Am J Public Health 69: 767–771 (1979).
4. Gordis, L., Lilienfeld, A., and Rodriguez, R.: An evaluation of the Maryland Rheumatic Fever Registry. Public Health Rep 84: 333–339, April 1969.
5. Goldberg, J., Gelfand, H. M., Levy, P. S., and Mullner, R.: An evaluation of the Illinois Trauma Registry: the

completeness of case reporting. Med Care 18: 520–531 (1980).
6. Division of Epidemiologic Studies, Adverse Pregnancy Outcomes Reporting System, Illinois Department of Public Health: Casefinding study. Epidemiologic Report Series 89: 6, October 1988.
7. Pollsar, L., et al.: Accuracy of basic cancer patient data: results from an extensive recoding survey. J Natl Cancer Inst 72: 1007–1014 (1984).
8. Calzavara, L. M., et al.: Underreporting of AIDS cases in Canada: a record linkage study. Can Med Assoc J 142: 36–39 (1990).
9. Stone, D. H.: A method for the validation of data in a register. Public Health 100: 316–324 (1986).
10. Yeike, B. B., and Jussawalla, D. J.: An assessment of reliability and completeness of Bombay Cancer Registry data (1963–1985). Indian J Cancer 25: 177–190 (1988).
11. Schnatter, A. R., et al.: An analysis of death ascertainment and follow-up through Statistics Canada's Mortality Data Base system. Can J Public Health 81: 60–65 (1990).
12. Lowry, R. B., Thunem, N. Y., and Anderson-Redick, S.: Alberta Congenital Anomalies Surveillance System. Can Med Assoc J 141: 1155–1159 (1989).
13. Feigl, P., Polissar, L., Lane, W. W., and Guinee, V.: Reliability of basic cancer patient data. Stat Med 1: 191–204 (1982).
14. Hanszel, W., and Lourie, W. I.: Quality control of data in a large-scale cancer registration program. Methods Inf Med 5: 67–74 (1966).
15. Demlo, L. K., Campbell, P. M., and Brown, S. S.: Reliability of information abstracted from patients' medical records. Med Care 16: 995–1005 (1978).
16. West, R. R.: Accuracy of cancer registration. Brit J Prev Soc Med 30: 187–192 (1976).
17. Beard, C. M., Bergstralh, E. J., and Klee, G. G.: Interobserver variability in collecting data from medical records. Arch Path Lab Med 112: 594–596 (1988).
18. Herrmann, N., et al.: Interobserver and intraobserver reliability in the collection of emergency medical services data. Health Serv Res 1: 127–143 (1980).
19. Pole, D. J.: Myocardial infarction incidence: methods, validity and preliminary experience of a coronary register. Med J Aust 2: 1481–1484 (1972).
20. Chiazze, L.: Morbidity survey and case register estimates of cancer incidence. Natl Cancer Inst Monograph 19: 373–384 (1966).
21. Lemeshow, S., and Levy, P. S.: Sampling for health professionals. Lifetime Learning Publications, Belmont, CA, 1980.
22. Calle, E. E., and Khoury, M. J.: Completeness of the discharge diagnoses as a measure of birth defects recorded in the hospital birth record. Am J Epidemiol 134: 69–77 (1991).