

Capture-Recapture Method in the Epidemiology of Type 2 Diabetes

A contribution from the Verona Diabetes Study

GIUSEPPE VERLATO, MD
MICHELE MUGGEO, MD

OBJECTIVE — The present investigation used data from the Verona Diabetes Study to verify a main assumption of the capture-recapture method (source independence) and to characterize the subgroup of known diabetic patients missed by all sources whose number is estimated by the capture-recapture method.

RESEARCH DESIGN AND METHODS — The Verona Diabetes Study identified 7,148 type 2 diabetic patients on 31 December 1986 using 3 different sources: family physicians, a diabetes center, and a drug prescription database. Completeness of ascertainment was estimated with traditional methods based on the hypergeometric distribution and with a log-linear model.

RESULTS — Identification sources were not independent because the drug prescription database was positively related to family physicians and negatively related to the diabetes center ($P < 0.001$). Thus, completeness of ascertainment was overestimated (87.5% [95% CI 86.3–88.8]) when using only family physicians and the drug prescription database and underestimated (45.9% [43.9–48.1]) when using only the diabetes center and the drug prescription database. Because of characteristics contributing to variable “catchability” (probability of ascertainment), the estimated proportion of ascertainment increased with increasing time since diagnosis from 65.6% in the first tertile (<6 years) to 91.5% in the third tertile (>12 years); moreover, the ascertainment was estimated to be nearly complete (97.9%) for insulin-treated patients and scanty (28.9%) for diet-treated patients.

CONCLUSIONS — Because identification sources are likely to be dependent, the capture-recapture method should be used with caution in diabetes epidemiology and possibly when at least 3 sources are available. The subgroup of diabetic patients whose existence is inferred by this technique likely consists of newly diagnosed patients with mild disease severity.

Diabetes Care 23:759–764, 2000

The capture-recapture method allows researchers to correct the estimate of an unknown population size by combining the information from at least 2 sources of identification (1,2). When the degree of overlap between the different sources is low, the unknown population size is estimated to be much larger than the value actually observed (a low proportion of ascertainment); in contrast, when the degree of over-

lap is high, the proportion of ascertainment is also high. This method relies on 2 basic assumptions: 1) the sources of identification should be independent (2–4), and 2) the probability of identifying an individual from any particular source should remain constant in different subgroups within the studied population (constant catchability) (4–6).

The capture-recapture method has been extensively used in the epidemiology

of type 1 diabetes, and a high proportion of ascertainment was reported in most studies (usually >90%) (7–10). This method was seldom applied to the epidemiology of known diabetes as a whole (4,11,12), and lower proportions of ascertainment were found that ranged from 80 to 91%. So far, completeness of ascertainment has not been specifically addressed through the capture-recapture method when assessing the prevalence of type 2 diabetes, which is more likely to be underdiagnosed and undertreated by the health care system than type 1 diabetes (13).

Some authors maintain that the capture-recapture method can also be applied when the proportion of ascertainment is much lower and that this statistical method is more cost-effective than field surveys and patient registries in the assessment of disease prevalence (14). However, other authors (15–18) question the usefulness of the capture-recapture method for several reasons. First, the assumption of source independence is often violated in practice, and this can have deleterious consequences, particularly when only 2 sources are available. Second, the scientific validity of the subgroup of patients who are not ascertained by any source of identification and whose existence is inferred from the capture-recapture method is highly questionable. These patients probably have unusual characteristics because they are never seen by a physician, they never buy medication, and they do not attend specialty clinics (18). Finally, with most methods, the CI, which reflects random variability, is very small compared with potential systematic errors that could be hidden in a particular model (19). For example, in the survey performed in Casale Monferrato, Italy, in 1988, the number of prevalent diabetes cases was estimated to be 2,283 (95% CI 2,240.7–2,325.3) in one article using Chapman’s estimator (11) and 2,586 (2,341–2,830) in a subsequent article using the log-linear model, which allows researchers to consider source dependence and variable catchability (4). The CIs obtained with the 2 statistical approaches, although based on the same data, did not even overlap.

From the Chair of Medical Statistics (G.V.) and the Division of Endocrinology and Metabolic Diseases (M.M.), University of Verona Medical School, Verona, Italy.

Address correspondence and reprint requests to Giuseppe Verlato, MD, Sezione di Statistica Medica, Dipartimento di Medicina e Sanità Pubblica, Istituti Biologici II, Strada Le Grazie 8, 37134 Verona, Italy. E-mail: giuseppe@biometria.univr.it.

Received for publication 22 July 1999 and accepted in revised form 8 February 2000.

A table elsewhere in this issue shows conventional and Système International (SI) units and conversion factors for many substances.

The current research joins this debate by applying the capture-recapture method to the Verona Diabetes Study, which is a population-based survey on the prevalence of type 2 diabetes in Verona, Italy (20–22). In particular, the present study aims to verify the assumption of source independence and evaluate possible related bias, verify the assumption of constant catchability to estimate the characteristics of those patients not identified by any source whose existence is inferred from the capture-recapture method, and compare the range of CIs to the variability of point estimates resulting from possible systematic errors.

RESEARCH DESIGN AND

METHODS — The design of the Verona Diabetes Study has been described elsewhere (20–22). Briefly, this survey used 3 different sources of ascertainment to assess the prevalence of type 2 diabetes in the Social Health Unit of Verona on 31 December 1986: family physicians, the local diabetes center, and the drug prescription database for 1986. The family physicians who agreed to participate in the survey (78%) were asked to provide a list of all of the diabetic patients they had in charge and to report all of the diabetic patients they visited in the subsequent 4 months. The diabetes center provided a list of all the patients registered in the clinical records. The first referral to the diabetes center was determined by family physicians according to their personal opinion, and subsequent visits were scheduled by the physicians at the diabetes center. In the drug prescription database, only those patients with at least 2 prescriptions within 6 months were identified as having diabetes (20). The present investigation focused on the 7,148 type 2 diabetic patients identified by the 3 sources altogether. These patients were means ± SD 66.6 ± 11.2 years of age, had a slightly larger proportion of women ($n = 3,782$, 52.9%) than men, mostly lived in Verona ($n = 6,456$, 90.4%), and had a mean time since diagnosis of 10.0 ± 7.2 years. Most patients were treated with oral agents ($n = 5,821$, 81.4%), 860 patients (12.0%) were treated with diet, and 437 patients (6.1%) were treated with insulin alone or with insulin in combination with oral agents.

The capture-recapture procedure was applied first with simple traditional methods such as Chapman's estimators (1,2) and then with a more sophisticated technique, the log-linear model (3,4).

Traditional methods provided an estimate of the unknown size of the diabetic population by 3 different procedures. The first procedure involved only 2 sources at a time and used Chapman's equations (1):

$$n' = (n_1 + 1)(n_2 + 1)/(m + 1) - 1$$

$$\text{variance } (n') = (n_1 + 1)(n_2 + 1)(n_1 - m)(n_2 - m) / [(m + 1)^2(m + 2)]$$

where n' is the estimated number of cases in the target population; n_1 and n_2 are the cases identified by the first and second data sources, respectively; and m is the number of cases ascertained by both sources. The second procedure involved contrasting 1 source with the combination of the other 2 sources and applying Chapman's estimator. The third procedure involved all 3 sources separately (2):

$$(n' - n_1)(n' - n_2)(n' - n_3) = n'^2(n' - n)$$

$$\text{variance } n' = n' / [n' / (n' - n) + 2 - n' / (n' - n_1) - n' / (n' - n_2) - n' / (n' - n_3)]$$

where n' is the real number of cases (identified and not identified) in the target population estimated by the 3-source model; n_1 , n_2 , and n_3 are the cases identified by the first, second, and third data sources, respectively; and n is the number of distinct cases ascertained by all sources. The equation to compute n' is a second-degree equation, and hence it has 2 solutions, 1 of which is rejected as being lower than n .

The proportion of ascertainment was the actual number of cases observed divided by the estimate of the total population size according to the capture-recapture method.

Independence between 2 sources was tested with a χ^2 test using the subjects identified by the third source (2). The assumption of an equal probability of listing for a given source across the study population was checked for sex, age (<62.9, 62.9–72.8, and >72.8 years), site of residence (downtown vs. rural areas), time since diagnosis (<6, 6–12, and >12 years), and treatment (diet, oral agents, or insulin). The hypothesis of statistical independence between each source of identification and each stratifying variable was tested with a χ^2 test.

The extent and the consequences of source dependence and different degrees of catchability were also evaluated by a log-linear model that investigated not only first-

order but also second-order interactions among sources and stratifying variables. Moreover, this approach allowed an accounting of these potential sources of bias when estimating the unknown population size.

Because of empty cells, simultaneously analyzing the effect of the 3 sources of identification and the stratifying variables was not possible, so 6 different initial models were analyzed. They included the 3 sources of identification alone or in combination with each of the stratifying variables (sex, age, site of residence, time since diagnosis, and treatment). The best-fitting log-linear model was selected through a backward selection strategy (4) using a modified version of the program published by Cormack (3). The estimates of the number of diabetic subjects in the empty cells were computed as the antilog of a linear combination of the parameter estimates, whose variance was the algebraic sum of the elements of the variance-covariance matrix.

RESULTS — The distribution of type 2 diabetic patients according to identification sources is presented in Fig. 1. The ellipse representing the drug prescription database is largely superimposed over the ellipse representing family physicians, but it is rather apart from the ellipse representing the diabetes center. This pattern was confirmed when dependence between 2 sources was checked for the patients identified by the third source. In particular, a strong positive dependence was observed between family physicians and the drug prescription database ($P < 0.001$) with the number of cases identified by both sources exceeding by 96% the number expected under the null hypothesis. In addition, a remarkable negative dependence occurred between the diabetes center and the drug prescription database ($P < 0.001$) because the number of patients identified by both sources was lower than expected by 25%. Furthermore, a mild departure from the independence assumption was also found for the relationship between family physicians and the diabetes center because the number of cases ascertained by both sources exceeded the number expected by 5.8% ($P = 0.004$). These source dependencies were also significant in the log-linear model.

The consequences of source dependence on the estimation of population size are depicted in Fig. 2, which reports the observed value ($n = 7,148$) together with the estimates of the number of diabetic

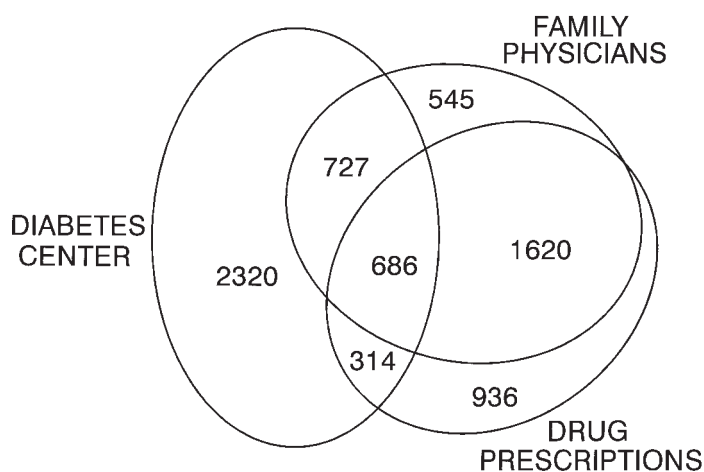


Figure 1—Prevalent cases of known type 2 diabetes in the Verona Diabetes Study by source of identification.

patients when applying traditional capture-recapture methods to 2 sources of data, to 1 source versus the other 2 combined, and to 3 sources or when applying the log-linear model to 3 sources. A great uncertainty in the results was evident when using just 2 sources of identification because the estimates of the prevalence of type 2 diabetes varied from 5,517 (5,438–5,597) when considering family physicians and the drug prescription database to 14,383 (13,728–15,038) when considering the diabetes center and the drug prescription database (Fig. 2). Likewise, completeness of ascertainment was overestimated (87.5% [86.3–88.8]) when using only family physicians and the drug prescription database and underestimated (45.9% [43.9–48.1]) when using only the diabetes center and the drug prescription database. The 95% CIs for the estimates were very narrow regarding variations in the point estimates, which suggests that random variability may be much lower than systematic biases. The uncertainty in the point estimates decreased with increasing amounts of information implemented in the traditional capture-recapture method, and a final estimate of 8,875 cases (8,743–9,007) was obtained when using all possible information. The log-linear model yielded a slightly larger estimate ($n = 9,345$; 8,954–9,820) that suggests that, in the Verona Diabetes Study, positive dependencies among sources slightly exceeded negative dependencies.

The correction methods mentioned above remarkably affected the estimates of type 2 diabetes prevalence. The crude prevalence of type 2 diabetes in the whole

population of Verona when assuming full ascertainment was 2.37% (2.32–2.42). (A slightly higher value, 2.49% [2.43–2.54] had been reported previously [20] because the analysis had been restricted to that part of the Verona population [81%] cared for by family physicians who agreed to participate in the survey.) The traditional capture-recapture method applied to the 3 sources yielded an estimate of 2.94% (2.90–2.99),

which increased to 3.10% (2.97–3.26) when using the log-linear model.

All sources presented variable catchability regarding some stratifying variables (Table 1) according to the χ^2 test. Type 2 diabetic women were more likely to be identified by family physicians and the drug prescription database than men. The diabetes center preferentially ascertained patients in the low and middle tertiles of age, whereas the drug prescription database preferentially ascertained patients in the upper tertile of age. Patients living outside of the main town had a greater likelihood of being identified by family physicians and the drug prescription database, and patients with a longer duration of diagnosed disease had a greater likelihood of being detected by the diabetes center and the drug prescription database. The diabetes center and family physicians preferentially ascertained insulin-treated patients; 16.2% of patients treated with diet were identified by the drug prescription database because of occasional drug use. Analysis with a log-linear model not only reached statistical significance for all first-order interactions between each source and each stratifying variable but also highlighted significant second-order interactions involving 2 sources (diabetes center and

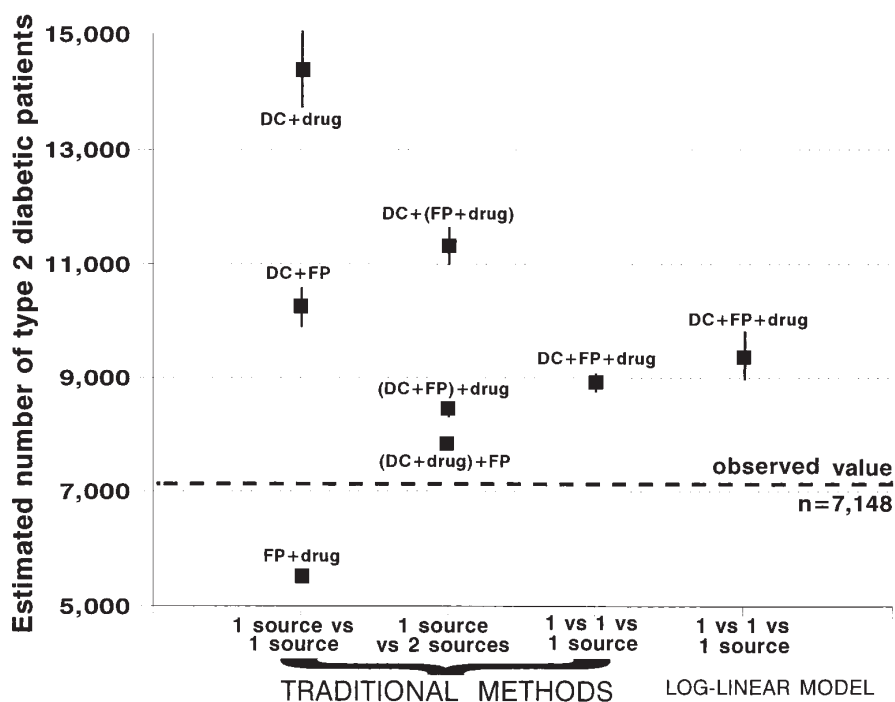


Figure 2—Estimates of the number of type 2 diabetic patients in the Verona Diabetes Study with the traditional capture-recapture method based on 2 or 3 samples and with the log-linear model. The vertical bars are CIs of prevalence estimates. DC, diabetes center; drug, drug prescription database; FP, family physicians.

Table 1—Type 2 diabetic patients ascertained by each source and overall proportion of ascertainment in the different levels of sex, age, site of residence, time since diagnosis, and treatment

Stratifying variables	Cases	Diabetes center	Family physicians	Drug prescription database	Ascertainment rate
Sex					
Men	3,366	1,919 (57.0)	1,607 (47.7)	1,599 (47.5)	74.0 (69.9–77.8)
Women	3,782	2,128 (56.3)	1,971 (52.1)	1,957 (51.7)	78.8 (75.2–82.1)
P		NS	<0.001	<0.001	
Age at baseline (years)					
<62.9	2,383	1,367 (57.4)	1,234 (51.8)	1,083 (45.4)	74.7 (70.4–78.5)
62.9–72.8	2,382	1,431 (60.1)	1,155 (48.5)	1,133 (47.6)	77.6 (71.0–83.1)
≥72.9	2,383	1,249 (52.4)	1,189 (49.9)	1,340 (56.2)	77.5 (71.4–82.5)
P		<0.001	0.074	<0.001	
Site of residence					
Main town	6,456	3,665 (56.8)	3,169 (49.1)	3,178 (49.2)	78.1 (71.6–83.4)
Rural areas	685	381 (55.6)	409 (59.7)	372 (54.3)	76.8 (73.1–80.2)
P		NS	<0.001	<0.001	
Time since diagnosis (years)					
<6	2,057	1,247 (60.6)	1,174 (57.1)	763 (37.1)	65.6 (61.4–69.5)
6–12	2,024	1,317 (65.1)	1,163 (57.5)	883 (43.6)	79.5 (76.0–82.6)
>12	1,965	1,472 (74.9)	1,086 (55.3)	849 (43.2)	91.5 (89.2–93.3)
P		<0.001	NS	<0.001	
Therapy					
Diet	860	484 (56.3)	452 (52.6)	139 (16.2)	28.9 (22.7–36.0)
Oral agents	5,821	3,236 (55.6)	2,847 (48.9)	3,176 (54.6)	84.2 (81.1–86.9)
Insulin	437	325 (74.4)	251 (57.4)	229 (52.3)	97.9 (94.3–99.2)
P		<0.001	<0.001	<0.001	

Data are n, n (%), or % (95% CIs). Significance of differences was evaluated by a χ^2 test. Information on site of residence, time since diagnosis, and treatment was lacking in 7, 1,102, and 30 type 2 diabetic patients, respectively.

drug prescription database) and either site of residence, time since diagnosis, or treatment. However, considering stratifying variables did not substantially affect the final estimates for the unknown population size, which were close to estimates obtained when considering only the 3 sources of identification in the log-linear model ($n = 9,345$): final estimates were 9,343 (8,934–9,846), 9,338 (8,791–10,079), 9,286 (8,879–9,793), 8,932 (8,615–9,319), and 9,293 (8,905–9,774), respectively, when stratifying for sex, age, site of residence, time since diagnosis, and treatment.

As shown in Table 1, the probability of ascertainment estimated by the capture-recapture method was markedly affected by time since diagnosis and treatment. The completeness of ascertainment was estimated to be 65.6, 79.5, and 91.5%, respectively, in the first, second, and third tertiles of time since diagnosis. Moreover, according to the capture-recapture method, the survey identified nearly all of the patients treated with insulin (97.9%) and most of the patients treated with oral agents (84.2%) but only a minor fraction of

patients treated with diet (28.9%). Probability of identification was slightly higher for women and patients >63 years of age, whereas it was comparable for people living in the main town or in rural areas.

CONCLUSIONS — The main results of the present study are as follows. First, assumption of source independence was not met in the Verona Diabetes Study. In particular, the type 2 diabetic patients in the drug prescription database were positively related to the patients on the lists of family physicians who prescribe drugs. Instead, a negative dependence occurred between the diabetes center and the drug prescription database, probably because patients attending the diabetes center could obtain drugs directly from the center itself without prescriptions from a family physician. Thus, this drug consumption was missed by the drug prescription database, which recorded only prescriptions made by family physicians and used by patients to obtain drugs freely at the pharmacy.

Second, source dependence had a great effect on point estimates of population size

(i.e., on estimates based on a single value). The variability of point estimates elicited by this bias largely overwhelmed the range of the CIs, especially when the capture-recapture method was based on only 2 sources of identification (Fig. 2). In most articles dealing with diabetes epidemiology, this method had been applied using only 2 sources of identification (7,10), which does not allow evaluation of completeness of ascertainment.

Third, variable catchability took place across the Verona diabetic population because the probability of identification changed as a function of all of the stratification variables considered (sex, age, site of residence, time since diagnosis, and treatment). When these findings were extrapolated to type 2 diabetic patients who were not identified by any source and whose existence was inferred from the capture-recapture method, the latter were described as patients with newly diagnosed diabetes of mild severity. Interestingly, the degree of ascertainment by treatment method was proportional to the number of visits per year among patients attending the diabetes center: from 1986 to 1987, this number

was 2.06 ± 1.27 (range 0–9) in patients treated with diet, 3.13 ± 1.38 (0–12) in patients treated with oral agents, and 5.40 ± 2.52 (0–20) in patients treated with insulin.

In the Verona Diabetes Study, the probability of ascertainment was enhanced by a high probability of seeking services for diabetes and by a high probability of reporting a diagnostic code for diabetes for those services. Indeed, in Italy during the 1980s, diabetic patients could obtain virtually all health services and drugs free of charge through the national health care system. Moreover, the use of the diagnostic code for diabetes was necessary to receive services and drugs free of charge.

The analysis of source dependence highlighted a poor integration among different levels of diabetes care in Verona. Indeed, most of the patients registered by the diabetes center were not reported by family physicians either because they were treated by family physicians not collaborating in the Verona Diabetes Study or because they did not visit their family physicians for 4 consecutive months. This observation prompted us to implement a project of integrated care involving both family physicians and the diabetes center that is now in progress.

Generalizing the present findings to different populations and to different health care systems is difficult because many features of the Verona Diabetes Study were specific not only to a particular area but also to a particular time period. The pattern of source dependence observed in the Verona Diabetes Study was similar to the pattern observed in other surveys on known diabetes prevalence performed in central northern Italy at the end of the 1980s (4,23). However, different patterns of source dependence were observed in other countries (8,9). Moreover, the negative dependence between diabetes centers and the drug prescription database observed in 1986 has probably faded away because diabetes centers are no longer allowed to deliver drugs freely to their patients because of financial restrictions. Regardless, a strong recommendation resulting from the present findings is to apply the capture-recapture method with extreme caution and only after acquiring deep knowledge of the local health care system and social habits.

In the Verona Diabetes Study, completeness of ascertainment for type 1 diabetic patients defined by age at diagnosis and treatment (20) was much higher than

for type 2 patients (97.2 vs. 75%). According to the capture-recapture method, the number of type 1 diabetic patients was 211 (206–216), a value close to the number of patients identified by the 3 sources ($n = 205$). Hence, applying the capture-recapture method to the type 1 and type 2 diabetic populations separately seems wise versus applying it to the overall population of known diabetic patients. Moreover, to estimate the proportion of ascertainment for type 2 diabetic patients, mixing at least 3 sources of identification seems preferable, as does using a log-linear model. This model allows researchers to consider dependence between sources and variable catchability (4) and yields a broader range of CIs that are more adequate to reflect the uncertainty inherent to the capture-recapture method. Interestingly, when using this procedure, the prevalence estimates obtained in the Verona Diabetes Study were consistent with other estimates obtained in central northern Italy at the end of the 1980s (4,23,24).

In conclusion, the assumptions of the capture-recapture method were not met in the Verona Diabetes Study because the sources of identification were not independent, and the probability of identification changed as a function of sex, age, site of residence, time since diagnosis, and treatment. Source dependence elicited a great instability in prevalence estimates, particularly when using just 2 sources of identification, and instability in point estimates largely exceeded the range of CIs. People missed by all sources seemed to be patients with newly diagnosed diabetes of mild severity.

The present results suggest that application of the capture-recapture method to diabetes epidemiology requires preliminary knowledge of the social context and health care system. The assumptions of source independence and constant catchability should be verified, which has been seldom done in the current literature (4,8,9,23). Because no mathematical way exists to verify the independence of 2 sources of identification without a third source (4,8), the combination of at least 3 sources is recommended. However, 2 sources may work as well when the investigators are absolutely certain of their independence according to past experience, deep knowledge of the local organization of the health care system, or current literature. Until an external validation of this method is available, the estimates obtained should be viewed with caution.

Acknowledgments — This study was supported by grants from the Italian National Research Council, the Italian Ministry of the University and Scientific and Technological Research and the Veneto Region (Progetto Sanitario Finalizzato).

References

1. Chapman DG: Inverse, multiple and sequential censuses. *Biometrics* 8:286–306, 1952
2. Wittes JT, Colton T, Sidel VW: Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *J Chronic Dis* 27:25–36, 1974
3. Cormack RM: Log-linear models for capture-recapture. *Biometrics* 45:395–413, 1989
4. Bruno G, LaPorte RE, Merletti F, Biggeri A, McCarty D, Pagano G: National diabetes programs: application of capture-recapture to count diabetes? *Diabetes Care* 17:548–556, 1994
5. Hook EB, Regal RR: Effect of variation in probability of ascertainment by sources (“variable catchability”) upon “capture-recapture” estimates of prevalence. *Am J Epidemiol* 137:1148–1166, 1993
6. McCarty DJ, Tull ES, Moy CS, Kwok CK, LaPorte RE: Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol* 22:559–565, 1993
7. Verge CF, Silink M, Howard NJ: The incidence of childhood IDDM in New South Wales, Australia. *Diabetes Care* 17:693–696, 1994
8. Wadsworth E, Shield J, Hunt L, Baum D: Insulin dependent diabetes in children under 5: incidence and ascertainment validation for 1992. *BMJ* 310:700–703, 1995
9. Rangasami JJ, Greenwood DC, McSpornan B, Small PJ, Patterson CC, Waugh NR: Rising incidence of type 1 diabetes in Scottish children, 1984–93: the Scottish Study Group for the Care of Young Diabetics. *Arch Dis Child* 77:210–213, 1997
10. Vandewalle CL, Coeckelberghs MI, De Leeuw IH, Du Caju MV, Schuit FC, Pipeleers DG, Gorus FK: Epidemiology, clinical aspects, and biology of IDDM patients under age 40 years: comparison of data from Antwerp with complete ascertainment with data from Belgium with 40% ascertainment: the Belgian Diabetes Registry. *Diabetes Care* 20:1556–1561, 1997
11. Bruno G, Barger G, Vuolo A, Pisu E, Pagano G: A population-based prevalence survey of known diabetes mellitus in Northern Italy based upon multiple independent sources of ascertainment. *Diabetologia* 35:851–856, 1992
12. Berger B, Stenstrom G, Chang Y-F, Sundkvist G: The prevalence of diabetes in a Swedish population of 280,411 inhabitants: a report from the Skaraborg Diabetes

- Registry. *Diabetes Care* 21:546–548, 1998
13. Harris MI, Hadden WC, Knowler WC, Bennett PH: Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in the U.S. population aged 20–74 yr. *Diabetes* 36:523–534, 1987
 14. LaPorte RE, McCarty D, Bruno G, Tajima N, Baba S: Counting diabetes in the next millennium: application of capture-recapture technology. *Diabetes Care* 16:528–534, 1993
 15. Desenclos J-C, Hubert B: Limitations to the universal use of capture-recapture methods (Letter). *Int J Epidemiol* 23:1322–1323, 1994
 16. Schouten LJ, Straatman H, Kiemeneij LALM, Gimbrere CHF, Verbeek ALM: The capture-recapture method for estimation of cancer registry completeness: a useful tool? *Int J Epidemiol* 23:1111–1116, 1994
 17. Hay G, McKeganey N: Estimating the prevalence of drug misuse in Dundee, Scotland: an application of capture-recapture methods. *J Epidemiol Commun Health* 50:469–472, 1996
 18. Papoz L, Balkau B, Lellouch J: Case counting in epidemiology: limitations of methods based on multiple data sources. *Int J Epidemiol* 25:474–478, 1996
 19. Hook EB, Regal RR: Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *Am J Epidemiol* 145:1138–1144, 1997
 20. Muggeo M, Verlato G, Bonora E, Bressan F, Girotto S, Corbellini M, Gemma ML, Moghetti P, Zenere M, Cacciatori V, Zoppini G, de Marco R: The Verona Diabetes Study: a population-based survey on known diabetes mellitus prevalence and 5-year all-cause mortality. *Diabetologia* 38:318–325, 1995
 21. Verlato G, Muggeo M, Bonora E, Corbellini M, Bressan F, de Marco R: Attending the Diabetes Center is associated with increased 5-year survival probability of diabetic patients: the Verona Diabetes Study. *Diabetes Care* 19:211–213, 1996
 22. de Marco R, Locatelli F, Zoppini G, Verlato G, Bonora E, Muggeo M: Cause-specific mortality in type 2 diabetes: the Verona Diabetes Study. *Diabetes Care* 22:756–761, 1999
 23. Fabietti PG, Santeusano F, Iorio M, Messini B, Giovannini G, Nicolucci A, Carinci F, Massi Benedetti M, Brunetti P: Prevalence of known diabetes mellitus in a central region of Italy. *Diabetes Nutr Metab* 9:3–8, 1996
 24. Di Cianni G, Benzi L, Giannarelli R, Orsini P, Villani G, Ciccarone AM, Cecchetti P, Fedele O, Navalesi R: A prevalence study of known diabetes mellitus in Tuscany assessed from pharmaceutical prescriptions and other independent sources. *Acta Diabetol* 31:87–90, 1994