The use of capture-recapture methods to monitor diabetes in Dominica, West Indies

Eugene S. Tull,[1] *Cleve Butler,*[2] *Leyale Gumbs,*[1] *and Sherida Williams*[1]

ABSTRACT *The aim of this project was to evaluate the utility of capture-recapture (CR) estimation methodology, together with easily accessible sources of information, for monitoring the frequency of diabetes mellitus in a developing country. On the Caribbean island of Dominica, from July 1995 through July 1996, the names were collected of all individuals who were listed in at least one of three sources as having been diagnosed with diabetes during 1995. The sources were the membership list of the Dominica Diabetic Association, a register of patients in seven regional health centers, and clinic records of the Princess Margaret Hospital, which is the main hospital on the island. Capture-recapture techniques were used to estimate the number of Dominicans with diagnosed diabetes who were missed by all three sources of information and to evaluate dependencies among the sources. An ascertainment-corrected prevalence estimate of diagnosed diabetes was then calculated. The three sources identified a total of 1 945 different individuals. When this number was compared with the ascertainment-corrected estimate obtained with CR methodology (2 688), it appeared to represent only 72.3% of the true prevalence of diagnosed diabetes. Most of the under-ascertainment by the data sources was due to cases that were missed in the urban area of the island. This approach to disease monitoring may be useful for other Caribbean countries. However, standardized procedures for collecting and reporting prevalence and morbidity data are needed to maximize its accuracy and cost-effectiveness.*

The global trend toward increasing rates of diabetes mellitus has raised concerns about the consequences and cost of associated morbidity and mortality in developing countries (*1*, *2*). Information about diabetes frequency is crucial for the effective allocation of limited resources to deliver needed services to diabetic individuals. In many countries throughout the world, however, data about diabetes prevalence are lacking at the regional and national levels. The countries of the Caribbean are noteworthy in this context. Sustained economic development combined with high genetic susceptibility for Type 2 diabetes suggests that current rates of the disease will continue to increase (*3*). Yet in most countries of the region, little is known about the population frequency of diabetes mellitus, and reliable systems for monitoring the disease have not been developed.

Capture-recapture (CR) methods, traditionally used in demography to adjust population estimates to reflect the level of census undercount, may be useful for establishing disease monitoring systems in the Caribbean (*4*). The basic assumption of CR is that the sources of ascertainment are independent. This assumption is difficult to achieve when working with human diseases. However, when three or more sources are employed to identify cases, log linear modeling techniques can be used to estimate the dependency among sources and the number of cases that are not counted. By adjusting for the undercount and source dependencies, a more accurate estimate of the total number of cases in the population can be determined with corresponding 95% confidence intervals (95%CI).

[1] University of Pittsburgh, Graduate School of Public Health, Pittsburgh, Pennsylvania, USA. All correspondence and reprint requests should be sent to Eugene Tull at the following address: Virgin Islands Center for Chronic Disease Research, P.O. Box 3414, Christiansted, St. Croix Virgin Islands, 00820. Telephone: 809-778-8400; fax: 809-778-8407.

[2] Dominica Diabetic Association, Roseau Health Center, Roseau, Dominica, West Indies.

Some epidemiologists have promoted the use of CR methods to monitor chronic illnesses such as diabetes (*5, 6*). Such monitoring would utilize existing national and local sources of information to provide raw data on diabetes frequency, and then apply CR to these data to adjust for the number of cases missed by the sources of ascertainment. The resulting ascertainment-corrected estimates would provide ministries of health with reliable bases for strategic planning (*7*). Moreover, additional information about the treatment and complications status of diabetic subjects could be collected to monitor the quality of care delivered to patients. To evaluate the potential for utilizing CR methods in determining the frequency of diagnosed diabetes in a developing Caribbean country, we implemented a system on the island of Dominica.

MATERIALS AND METHODS

The first step in CR estimation of the overall prevalence of diagnosed diabetes on Dominica was to collect data on all individuals who were listed as having diabetes during the period January 1, 1995 to December 31, 1995. Three sources were used to identify diabetic individuals: the membership list of the Dominica Diabetic Association (DDA); the combined case lists of the seven government-run regional health clinics; and the patient list from the diabetes clinic of the Princess Margaret Hospital, the main hospital on the island. With very active affiliate DDA chapters located throughout the island, the DDA maintains a membership that is geographically representative of the entire island population. The regional clinics are distributed geographically so as to provide the entire population with equal access to health care services. The Princess Margaret Hospital clinic serves individuals from throughout the island who have diabetes complications or comorbid conditions—such as hypertension and dyslipidemia—that put them at higher risk for mortality.

At the regional health clinics and the Princess Margaret Hospital, public health nurses responsible for the care of diabetic patients confirmed that their cases met WHO criteria (*8*) for the diagnosis of diabetes. To assess the diagnostic validity of the cases on the DDA list, we randomly selected a 10% sample and checked their 1995 medication status and blood sugar values. The use of insulin or oral hypoglycemic medication or, alternatively, two or more fasting glucose values >7.8 mmol/L (140 mg/dL) were regarded as confirmation of the diagnosis of diabetes. By these criteria, the entire sample had diabetes.

The lists from all sources were examined for mortality status. At the time the lists were initially obtained from the clinics and diabetic association, they were reviewed by a study representative and clinic nurses to identify individuals who were deceased. At the end of the study period, a second review was conducted to determine if additional individuals had died. The mortality status of individuals identified as deceased was confirmed by examination of medical records at the Queen Elizabeth Hospital. Individuals known to be deceased were removed from all lists. All other subjects were assumed to be alive on December 31, 1995, the date to which the prevalence estimates apply.

The information collected on each diabetic individual included his or her name, gender, age, and address. For regional information, the case listings were checked to ensure that all addresses corresponded to appropriate regions. Pair-wise comparisons of the sources were performed by health region, matching cases on name, age, and address.

Using the technique described by LaPorte et al. (*9*), log linear modeling with the General Linear Interactive Models (GLIM) software (*10*) was used to estimate the total number of individuals with diagnosed diabetes and to examine the dependence among the data sources. First, each diabetic individual was assigned to a cell in an eight-cell[3] contingency table corresponding to the source or sources in which he or she appeared—the DDA list (S1), the regional clinic records (S2), the hospital registry (S3), or a combination of these (S1+S2, S1+S3, S2+S3, or S1+S2+S3). The remaining open cell in the contingency table corresponds to the number of cases that are present in the population but do not appear in any of the three lists. Log linear models were then fitted that predict the open cell based on the data in the seven filled cells. Interaction terms were included in the models to assess whether the appearance of a name on any one list increased or decreased the likelihood of its appearing on any other (dependency).

It was anticipated that the proportions of cases ascertained by the sources would be different in Roseau, which is urban, than it was in the other regions, which are rural. Therefore, in addition to the analysis of the overall data, log linear models were fitted to the data from the Roseau area alone, and to data from the combination of the other regions excluding Roseau. The significance of parameters in the models was assessed with the chi-square test statistic. Population denominators for the calculation of prevalence rates were obtained from the 1991 Dominica census (*11*).

RESULTS

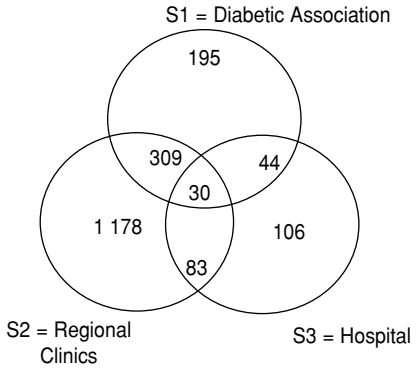
The numbers of individuals with diagnosed diabetes tallied in the three sources in each of Dominica's health regions are presented in Table 1. After matching, the number of different individuals was found to total 1 945 (Figure 1). The best fitting log linear model for the overall data (Table 2) indicated that approximately 743 cases (27.6% of the island's diabetics) were missing from all three sources. Therefore, the ascertainment-corrected estimate of the total number of diagnosed

[3] 2^K, with K being the number of sources.

TABLE 1. Total number of diabetic subjects identified, by source and region

Region	Diabetic Association	Regional clinics	Hospital
Castle Bruce	16	59	3
Grand Bay	35	171	15
La Plaine	52	112	12
Marigot	86	215	3
Portsmouth	54	175	3
Roseau	270	820	214
St. Joseph	65	48	13
Total	578	1 600	263

FIGURE 1. Number of diabetic subjects identified by source and the number common to multiple sources



diabetic subjects on Dominica in 1995 was 2 688 (95%CI: 2 548 to 2 828). The corresponding crude and ascertainment-corrected estimates of the prevalence of diagnosed diabetes were 2.71% and 3.75%, respectively.

In every source, the majority of diabetic individuals were identified in the Roseau region. The best fitting log linear model for the data from Roseau indicated that 546 cases (95%CI: 396 to 696) in that region were missed by all three sources of ascertainment. When compared to the observed number of cases, 1 044, this figure represents an under-ascertainment of approximately 34.3%. This suggests that most of the missed cases on the island were in the Roseau region. When the Roseau region was excluded from analysis, the best fitting log linear model indicted that the three sources of ascertainment failed to ascertain only 50 (95%CI: 12 to 87) diabetic individuals in the other six health regions. Thus, with 901 observed cases of diabetes in the rural areas, underascertainment was only 5.3%.

The patterns of interaction in the national data and the Roseau data were similar. In both best fit models, there was positive interaction between the lists from the DDA and the hospital, and negative interaction between the lists from the regional clinics and the hospital. However, for the data from the rural areas excluding Roseau, there was negative interaction both between the DDA and the regional clinics and between the regional clinics and the hospital.

DISCUSSION

In many countries, monitoring systems that rely on counting cumulative cases are too inaccurate to serve as a basis for the effective allocation of resources to address the problem of diabetes (*12*). An alternative monitoring strategy, blood glucose screening, is usually too costly for these countries to use to obtain national estimates of the prevalence of diabetes. Capture-recapture with log linear modeling appears to be a useful means of obtaining cost-efficient, reliable estimates of diabetes frequency (*9*). However, the feasibility of using CR to monitor diabetes in developing countries depends to a great extent on the ease of access to sources of information and the amount of manpower required to obtain the necessary information.

On the island of Dominica, we were able to find three sources of information that could be used to identify diabetic subjects across the island. Similar sources may be available in many of the other Caribbean countries in the region. The major manpower costs associated with this study involved the time required to travel to various clinics to collect the lists, and follow-up of diabetic subjects to confirm diagnosis of diabetes or mortality status. We anticipate that once the available sources have been integrated into an ongoing monitoring system, the costs associated with data collection and follow-up will be minimized.

TABLE 2. Best-fit log linear models for the data from the three sources (S1, S2, S3)[a] of ascertainment with parameter coefficients and estimated number of missed cases with 95% CI, by region

Region	Model	Cases Missed
All regions	$(-1.338)S1 + (0.460)S2 + (-1.931)S3 + (0.400)S1.S3^b + (-0.7436)S2.S3^b$	743(95%CI: 603 to 883)
Urban (Roseau)	$(-1.648)S1 + (0.1082)S2 + (-1.777)S3 + (0.4072)S1.S3^b + (-0.3335)S2.S3^b$	546(95%CI: 395 to 696)
Rural regions (combined)	$(0.5978)S1 + (2.441)S2 + (-1.504)S3 + (-1.676)S1.S2^b + (-2.241)S2.S3^b$	50(95%CI: 12 to 87)

[a] S1 = Diabetic Association; S2 = regional health clinics; S3 = hospital clinic.
[b] Interaction parameters.

Similar to other reports (*13, 14*), the results of our CR study indicated that the pooled data from the sources of ascertainment significantly underrepresented the number of diabetic individuals on the island. They accounted for only 72.4% of the total diabetic population as estimated with CR. It is interesting to note that the portion of estimated cases that was missed by the three sources of ascertainment was greatest (34.2%) in the Roseau region, the urban center of the island. This may reflect the fact that, compared to the rural areas, there are larger numbers of private physicians in the Roseau region who are more accessible as alternative sources of care for diabetic subjects. Additional sources of ascertainment (for example, lists from pharmacies that dispense diabetic medications) may be needed in that region to provide a more stable national estimate of the frequency of diagnosed diabetes on the island. In the six other regions, the three sources of ascertainment appear to be adequate, as they identified approximately 94.7% of the estimated total number of individuals with diagnosed diabetes.

Given that the majority of diabetic individuals were identified in the Roseau region, it is not surprising that the pattern of interactions among the three sources of ascertainment was parallel in Roseau and nationwide. The finding in Roseau of a positive interaction between the DDA and clinic lists—such that an individual's presence on one increased the likelihood that he or she would also be represented on the other—may reflect an increased awareness of the need for specialized care among the DDA members, and Roseau residents' relative ease of access to the hospital clinic, compared to their more remote compatriots. The negative interactions observed among the sources of information in the rural areas suggest that the rural clinics may have access to relatively unique populations of diabetic individuals. This may indicate that the DDA is not reaching the rural areas as well as it is the urban population, and that rural residents are dissuaded from using the Princess Margaret Hospital clinic by the distance they would have to travel.

At the start, it was our intention to use CR to monitor diabetes treatment and complications issues as well as disease prevalence. However, we found the regional clinics did not all record diabetic complications in the same way, which put this objective beyond reach. Nevertheless, in principle, CR methods can be used for monitoring important disease features in addition to prevalence.

In summary, to our knowledge, this is the first use of CR methods to monitor the frequency of diabetes in the Caribbean. To optimize accuracy and cost-effectiveness, these methods should be used with a surveillance system that incorporates standardized guidelines for data collection. Such a system would assist ministries of health with strategic planning for the allocation of resources to meet the needs of diabetic individuals in the Caribbean.

Acknowledgments. This research was supported by a Minority International Research Training grant, NIH 5 T37 TW00038-01, from the Fogarty International Center, National Institutes of Health, USA. We would like to thank Ron LaPorte and Yue-Fang Chang from the University of Pittsburgh for the technical assistance they provided to this project. We also appreciate the efforts of staff members at the regional health clinics and the Princess Margaret Hospital on Dominica who assisted us in obtaining the lists of diabetic subjects.

REFERENCES

1. Zimmet P. Type 2 (noninsulin-dependent) diabetes—an epidemiological overview. *Diabetologia* 1982;22:399–411.
2. Alberti KGM. World aspects of diabetes. *Diabetic Med* 1986;3:103–105.
3. Tull ES. Diabetes in the West Indies: current aspects and future prospects. *IDF Bulletin* 1993;38:21–23.
4. Hook EB, Regal RR. The value of capture-recapture methods even for apparent exhaustive surveys: the need for adjustment for source of ascertainment intersection in attempted complete prevalence studies. *Am J Epidemiol* 1992;135:1060–1067.
5. LaPorte RE, McCarty DJ, Tull ES, Tajima N. Counting birds, bees, and NCDs (letter). *Lancet* 1992;339:495.
6. Bruno G, LaPorte RE, Merletti F, Biggeri A, McCarty DM, Pagano G. National diabetes programs: application of capture-recapture to count diabetes? *Diabetes Care* 1994;17:548–556.
7. LaPorte RE, McCarty D, Bruno G, Tajima N, Baba S. Counting diabetes in the next millennium: application of capture-recapture technology. *Diabetes Care* 1993; 16:528–534.
8. World Health Organization. *WHO Expert Committee on Diabetes Mellitus*. Geneva: WHO; 1985.
9. LaPorte RE, Dearwarter SR, Chang Y, Songer TJ, Aaron DJ, Anderson RL, Olsen T. Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods to injury monitoring. *Am J Epidemiol* 1995;142:1069–1077.
10. Barker RJ, Nelder JA. *GLIM Manual* (release 3.77). Oxford, England: Royal Statistical Society; 1987.
11. Central Statistical Office. *Commonwealth of Dominica, Population and Housing Census, 1991*. Roseau, Dominica: CSO; 1991.
12. International Working Group for Disease Monitoring. Capture-recapture and multiple-record systems estimation II: applications in human disease. *Am J Epidemiol* 1995;142:1059–1068.
13. McCarty DJ, Tull ES, Moy CS, Kwoh CK, LaPorte RE. Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol* 1993;22:559–565.
14. LaPorte RE, Tull ES, McCarty D. Monitoring the incidence of myocardial infarctions: applications of capture-mark-recapture technology. *Int J Epidemiol* 1992; 21:258–262.

Manuscript received on 22 September 1997. Revised version accepted for publication on 12 December 1997.

RESUMEN

Uso de métodos de captura y recaptura para monitorear la diabetes en Dominica, Indias Occidentales

El propósito de este trabajo fue evaluar la utilidad de la metodología de estimación por captura y recaptura (CR), aplicada junto con fuentes de información fácilmente accesibles, para monitorear la frecuencia de diabetes mellitus en un país en desarrollo. En la isla caribeña de Dominica, se recogieron de julio de 1995 a julio de 1996 los nombres de todas las personas que figuraban como mínimo en una de tres fuentes diferentes por haber recibido un diagnóstico de diabetes en el transcurso de 1995. Las fuentes fueron la lista de miembros de la Asociación de Diabetes de Dominica, un registro de los pacientes atendidos en siete centros de salud regionales y las fichas clínicas del Hospital Princess Margaret, que es el más importante de la isla. Se aplicaron técnicas de CR para estimar el número de dominicanos con diabetes diagnosticada que no fueron detectados por ninguna de las tres fuentes, y para evaluar las discrepancias entre las distintas fuentes. Se procedió a calcular la prevalencia estimada de diabetes diagnosticada, corrigiéndola según los datos obtenidos. Las tres fuentes identificaron a un total de 1 945 individuos diferentes. Cuando se comparó con la estimación corregida que se obtuvo por el método de CR (2 688), se observó que esta cifra representaba solamente 72,3% de la verdadera prevalencia de diabetes diagnosticada. La mayor parte de la subdetección en las fuentes de datos correspondió a casos que no fueron captados en la zona urbana de la isla. Esta forma de monitorear la enfermedad podría resultar útil en otros países caribeños. No obstante, hacen falta procedimientos estandarizados de recolección y notificación de datos sobre prevalencia y morbilidad para lograr que este enfoque rinda al máximo en exactitud y costo-efectividad.