It is crucial to estimate the prevalence of type 2 diabetes mellitus in a given area accurately. First, it is essential for health care planners to establish public health actions. Second, by comparing the prevalence across populations and time, and analysing various factors affecting the prevalence, one might determine possible risk factors for the disease, allowing the planning of further epidemiological studies and preventive policies.1 A population-based survey using an oral glucose tolerance test (OGTT) as a primary examination is typically considered to offer the best estimate of the prevalence in a given area. However, only very limited areas are covered by population-based surveys worldwide.2 In Japan, there were only two areas, Funagata and Hisayama, where the prevalence was determined with this approach.3,4 The prevalence of diabetes is known for less than 0.01% of the Japanese adult population and it is less than 0.1% in the US. Major reasons for this extremely small coverage are the very expensive cost of conducting a population-based survey5 and the difficulty in obtaining a sufficiently high participation rate in the survey.

Recently, LaPorte et al. proposed applying capture-recapture methods for monitoring non-insulin dependent diabetes mellitus (NIDDM).6 These methods have been widely employed

---

**Evaluating the reported prevalence of type 2 diabetes mellitus by the Oguni diabetes registry using a two-sample method of capture-recapture**

A Sekikawa,a H Eguchi,b M Tominaga,c H Manaka,b H Sasaki,d YF Chang a and T Kato b

**Background** Capture-recapture methods have been widely employed in the study of wildlife populations and have recently been applied to count various human diseases and conditions. We have estimated the prevalence of type 2 diabetes mellitus by adjusting for the degree of undercount using a two-sample model of capture-recapture among men and women aged 50–69 in Oguni town, Japan.

**Methods** Oguni town diabetes registry data were utilized as the first source. In the registry, only those who had experienced fasting plasma glucose of \( \geq 7.8 \text{ mmol/l (140 mg/dl)} \) or 2 h plasma glucose after a 75 g oral glucose tolerance test (OGTT) of \( \geq 11.1 \text{ mmol/l (200 mg/dl)} \) were counted as having diabetes. A second source was a sample study selecting 200 men and 200 women aged 50–69 randomly, which was conducted in August 1991. A 75 g OGTT was done in the morning. The 1985 World Health Organization criteria were used to classify the diabetes status of the participants. A two-sample model of capture-recapture methods was employed to estimate the total number of cases of diabetes and determine the ascertainment rates of the registry.

**Results** The prevalence estimated by the diabetes registry was 7.1%. The prevalence from the sample study was 8.8% with a participation rate of 74%. Estimated prevalence employing the capture-recapture method was 13.1%. The ascertainment rate of the registry was 53.8%.

**Conclusions** Little is known about the prevalence of type 2 diabetes in local areas in Japan, the US and the world. Capture-recapture methods are likely to provide a means to accurately assess the prevalence of diabetes.

**Keywords** Capture-recapture methods, ascertainment rate, prevalence, disease registries

**Accepted** 6 September 1998

---

a Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, 5th floor 3460 Fifth Avenue, Pittsburgh, PA 15206, USA.

b The Third Department of Internal Medicine, Yamagata University School of Medicine, Yamagata, Japan.

c Department of Clinical Laboratory, Yamagata University School of Medicine, Yamagata, Japan.

d Diabetes Center, Nagaoka Red Cross Hospital, Nagaoka, Japan.
in the study of wildlife populations and have recently been applied to count various human diseases and conditions. These diseases and conditions included insulin-dependent diabetes mellitus, cancer, infectious diseases, birth defects, injuries, drug use, and others. Utilizing capture-recapture methods, the prevalence of type 2 diabetes can possibly be estimated much less expensively and much more accurately than the population-based OGTT survey. Moreover, continued monitoring can be done over time to assess incidence and time trends. In addition, areas where prevalence is estimated could be broadened dramatically.

The simplest model of the capture-recapture methods is a two-sample one. Using this model, one can estimate the total number of cases or conditions in a population under several assumptions. The primary assumption of the two-sample model is independence of sources. There is a readily available list of diabetic patients in many populations. If a random sample study for diabetes was conducted in such populations, one could regard the sample study as a source totally independent of the available list. In such a case, the two-sample model could be applied and one could obtain an ascertainment rate for the list and precise estimate of the cases in the population.

In this study, we estimated the total number of cases with diabetes among men and women aged 50–69 in Oguni town, Yamagata Prefecture, Japan, applying the two-sample model of the capture-recapture methods. We used a town diabetes registry as a first source. We conducted a random sample study for diabetes and used the results as a second source. We also determined the prevalence rates adjusted for the degree of undercount.

Materials and Methods

Oguni town

Oguni town, Yamagata Prefecture is located in the northeast part of Honshu Island, which is the main island of Japan (Figure 1). Yamagata Prefecture is predominantly rural and one of the major rice-producers. Oguni town lies in the southeastern part of Yamagata Prefecture. More than 95% of the area of the town is covered by mountains. The total population of Oguni town was 11,386 in June 1990. The proportion of inhabitants aged ≥65 was 20%, which was higher than the national average of 13%. There are one hospital and five clinics in the town. Most of the patients in the town are treated at these medical facilities.

Oguni town diabetes registry

A diabetes patient list at the town municipal building was utilized as the first source. Data on name, age, sex, and address are available from the list. The men and women aged 50–69 on the list in August 1991 were counted in this study.

Public health nurses visited every household annually since 1982 to ask whether they had diabetes mellitus. If the answer was yes, they were asked where they had been diagnosed and their current treatment. Based on the information obtained, medical records were confirmed by two of the authors (AS, HE). Only those who experienced a fasting plasma glucose level of ≥7.8 mmol/l (140 mg/dl) or a 2-h plasma glucose level after 75 g OGTT of ≥11.1 mmol/l (200 mg/dl) were classified as known cases of diabetes mellitus and listed in the registry. These criteria comply with the 1985 WHO criteria for diabetes. Informed consent was obtained from all respondents.

Figure 1 Oguni Town, Yamagata Prefecture, Japan

Sample study—second source

The random sample study was conducted in 1991. Based on the residence registration of Oguni town, 200 men and 200 women aged 50–69 were selected by using random digits in June. Equal sample size for men and women was chosen as the number of residents in this age group is not much different between men and women (1543 for men and 1775 for women). In July, the method and significance of the study were explained to those selected. They were instructed to abstain from food or drink beginning at 8 p.m. on the day before the OGTT. The OGTT was conducted at Oguni town sports centre starting at 7 a.m. in August. Informed consent was obtained from all the participants. Before the OGTT, blood samples were taken for testing the fasting plasma glucose. Within 10 min of the initial blood collection, the OGTT was conducted with 75 g of glucose (Trelan-G Shimizu Pharmaceutical, Shimizu, Japan). Blood samples were also taken to measure plasma glucose levels 2 h after glucose loading. Glucose levels were measured by the glucose oxidase method (Auto Stat GA-1122, Kyoto Dai-ichi Kagaku, Kyoto). The 1985 WHO criteria were employed to classify the status of OGTT.

We determined the sample size of 400 based on the Oguni registry data as well as the Funagata diabetes data. From the Oguni diabetes registry data in the past, we assumed that the prevalence of diagnosed diabetes in this age group was around 7% and there were no differences in prevalence between men and women. From our experience of the Funagata study which we had conducted in the same prefecture, the ratio of undiagnosed diabetes to diagnosed diabetes was expected to be 1 to 1. Therefore we expected the prevalence to be around 14%. We would like to have the prevalence with a precision of 4%, alpha level of 0.05 and 95% CI. From these numbers the sample size was determined to be 290. We also assumed that the participation rate would be about 75%, yielding the sample size of 387.
Table 1  Prevalence of type 2 diabetes mellitus in Oguni in men and women aged 50–69 by the registry, the sample study, and the two-sample method of capture-recapture

<table>
<thead>
<tr>
<th>ODRa</th>
<th>SSb</th>
<th>No. overlapped cases</th>
<th>Estimated no. total cases</th>
<th>Estimated point prevalence (%)</th>
<th>Ascertainment rate of the ODR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No. cases</td>
<td>Prevalence (%)</td>
<td>No. cases</td>
<td>Prevalence (%)</td>
</tr>
<tr>
<td>Men</td>
<td></td>
<td>110</td>
<td>7.1</td>
<td>10</td>
<td>8.2</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td>124</td>
<td>7.0</td>
<td>16</td>
<td>9.2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>234</td>
<td>7.1</td>
<td>26</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Table 1 summarizes the capture-recapture analysis. The estimated total number of cases was 435, which was almost two times the number of cases listed on the registry. The number was 220 for men and 220 for women, both of which almost doubled the respective cases on the registry. The estimated total prevalence was 13.1%, which was substantially higher than the prevalence in the registry as well as by the sample study. No significant difference was observed between the estimated prevalence of men and women. The overall ascertainment rate of the registry was 53.8%.

Statistical methods

The total number of cases was estimated using a two-sample model of capture-recapture. The formula is described below: 16

\[ N = M \times \frac{n}{m} \]

N is the estimated number of cases, M is the number of cases identified by the registry, n is the number of cases identified by the sample study, and m is the number of cases identified in both sources. The 95% CI of point estimates was computed by a goodness-of-fit method. 17 Goodness-of-fit CI is based upon the global behaviour of likelihood function and changes in the goodness-of-fit statistics associated with changes in trial value of the population total. This CI has been proved to be more appropriate than the symmetric standard-error based CI in the capture-recapture application. The CI takes accounts of the variability of the data. When two rates were compared, the \( \chi^2 \) was used. A difference was considered significant when the \( P \)-value was < 0.05.

Results

Response rates to the house visits were more than 95% every year. The population aged 50–69 was 3320 in August 1991. There were 1545 men and 1775 women. Among these, 234 people were on the list of the Oguni Diabetes Registry (110 men, 124 women). The overall prevalence in this age group was 7.1%. The sex-specific prevalence was 7.0% for men and 7.1% for women. No statistically significant difference of prevalence was observed between men and women.

Among the 400 people who were selected for the sample study, 14 people (5 men and 9 women) were listed on the registry. They were counted as overlapped cases and excluded from the subsequent OGTT. Among the remaining 386 participants, 282 people (117 men and 165 women) underwent the OGTT. The overall participation rate for the sample study was thus 74% (296/400). Among these 282, 12 people (5 men and 7 women) were diagnosed as having diabetes. The prevalence of diabetes in the sample study was 8.8% (26/296).

Table 1 summarizes the capture-recapture analysis. The estimated total number of cases was 435, which was almost two times the number of cases listed on the registry. The number was 220 for men and 220 for women, both of which almost doubled the respective cases on the registry. The estimated total prevalence was 13.1%, which was substantially higher than the prevalence in the registry as well as by the sample study. No significant difference was observed between the estimated

Discussion

It might appear that our finding of the overall estimated prevalence as being 13.1% was too high because the prevalence of diabetes among adults in Japan has been reported to be 1.7–3.9% in the early 1980s. However, recent population-based studies have shown that the prevalence exceeded 9% among adults age ≥40 when an OGTT was employed as a primary examination.4,5 Our population-based diabetes study in Funagata, which is located in the same prefecture and more rural than Oguni, showed that prevalence in ages 50–69 was 10.0%. Another population-based diabetes study from Hisayama, which is more urban than Oguni, reported that the prevalence was 13.5% (this figure is calculated using Table 4 of the article).4 Our finding that the overall ascertainment was 53.8% is not surprising as NIDDM manifests no symptoms in its early stage and, in many cases, patients in this early phase can not be diagnosed unless they undertake an OGTT. This portion is often called undiagnosed diabetes in contrast to diagnosed or known diabetes. The proportion of diagnosed diabetes among all diabetes in a population varies and has been reported to be 33–53%.

In the Funagata Study, among people aged 50–69, known cases of diabetes constituted 44% of all diabetic cases (unpublished data).

It was difficult to determine the sample size necessary in the two-sample model of capture-recapture because the CI of an estimate of the method depends more on the character of the sources rather than the sample size.

Age- and gender-specific prevalence rates provide valuable information. Capture-recapture can also be used in a specific age group. However, in our study, due to the small number of new diabetic cases from the second source, we decided not to perform age-specific capture-recapture analysis.

The crucial assumption of the two-sample capture-recapture method is that of independence between both sources. In the health field, it is an extremely difficult assumption to achieve when existing data sources are utilized, because being on one list is often associated with being on another. For example, the patients on the lists of health check-up centres are very likely to be associated with the list of the patients at hospitals, as the patients at health check-up centres are often referred to the hospitals. However, if data from a random sample study were employed as one source, it is not difficult to meet the requirement of independence.
The other assumptions of the two-sample model of capture-recapture method are the following: (1) there is no change to the population during the investigation, (2) individuals can be matched from capture to recapture, (3) for each sample, each individual has the same chance of being included in the sample.29 Our study obviously met the first two assumptions. The random sample study satisfied the last assumption. As for the registry data, a person with high risks for developing diabetes such as a person with obesity or family history of diabetes was more likely to take a diagnostic test—an OGTT. This is because health workers advise people with such risk factors to take examinations for diabetes more frequently than people without these factors.30 It is possible that we have overestimated the prevalence to some extent.

Incidence and prevalence rates are central to epidemiology, health policies, and their evaluations. However, it appears that for traditional forms of disease monitoring, ascertainment (which is one of the most important determinants of these rates) has largely been ignored.31 We evaluated the ascertainment rate of existing registry data by conducting a random sample study and applying the two-sample model of capture-recapture. To our knowledge, there was no capture-recapture application using random sample survey as the secondary source. This approach satisfies the crucial assumption of the method: independence between two sources.27 Though further investigation must be necessary on the methodology including sample size determination, it could potentially be utilized to evaluate the ascertainment rate of wide varieties of registries.

Acknowledgements
We thank the officials of Oguni Municipal Office, who co-operated with the survey of diabetes mellitus, and the residents of the town who participated in the survey.

References
26 Harris MI, Hadden WC, Knowler WC, Bennett PH. Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in US population aged 20–74 years. Diabetes 1987;36:523–34.