Epidemiol. Infect. (2002), **129**, 371–377. © 2002 Cambridge University Press
DOI: 10.1017/S0950268802007306 Printed in the United Kingdom

Underreporting of malaria incidence in the Netherlands: results from a capture–recapture study

N. A. H. VAN HEST[1]*, F. SMIT[2] AND J. P. VERHAVE[3]

[1] *Department of Tuberculosis Control, Municipal Health Service Rotterdam, P.O. Box 70032, 3000 LP Rotterdam, The Netherlands*
[2] *Department of Monitoring and Epidemiology, Trimbos Institute* (*Netherlands Institute of Mental Health and Addiction*), *P.O. Box 725, 3500 AS Utrecht, The Netherlands*
[3] *Department of Medical Microbiology, University Hospital Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands*

(*Accepted 8 April 2002*)

SUMMARY

The aim of this study was to estimate the completeness of notification of malaria by physicians and laboratories in the Netherlands in 1996. We used a capture–recapture (CRC) analysis of three incomplete, partially overlapping registers of malaria cases: a laboratory survey, the Notification Office and the hospital admission registration. The response of the laboratories was 83·2%. In 1996 the laboratories microscopically identified 535 cases of malaria, 330 patients with malaria were admitted to hospital and physicians notified 311 malaria cases. 667 malaria cases were recorded in at least one register. CRC analysis estimated the total number of malaria cases at 774 (95% CI of 740–821). This implies a completeness of notification of 40·2% for physicians and 69·1% for the laboratories. It can be concluded that laboratory-based notification can considerably increase the number of officially reported malaria cases as compared to notification by physicians. However, possibly one-third of the cases may still go unreported.

INTRODUCTION

Malaria is one of the most frequently imported diseases in the Netherlands. The number of notified malaria cases increased over 25 years from 19 patients in 1972 to 311 in 1996. This increase was mainly the result of a rise in the number of reported cases of *Plasmodium falciparum* malaria, a potentially fatal disease. A similar trend has recently been described in 23 European countries, Australia, Canada, New Zealand and the United States [1].

* Author for correspondence.
Part of this article was previously published in Dutch in the Nederlands Tijdschrift voor Geneeskunde (van Hest NAH, Smit F, Verhave JP. Considerable underreporting of malaria in the Netherlands; a capture–recapture analysis. Ned Tijdschr Geneeskd 2001; **145**: 175–9).

Under previous legislation regarding infectious diseases in the Netherlands, malaria was placed in category B. This group of infectious diseases had to be notified nominally (that is with the name and other particulars of the patient) within 24 h to the Municipal Health Service by the diagnosing physician. The Municipal Health Service forwarded this information to the Register of Notifiable Infectious Diseases (RNID) at the Office of the Chief Medical Officer where national data were aggregated for analysis, monitoring, public health intervention or policy making. Meaningful surveillance of imported malaria, such as trends in number of patients or type of plasmodium, identification of groups at risk (e.g. immigrants from malaria endemic countries or last-minute tourists with tropical destinations), evaluation

of chemoprophylaxis advice, and implementation of adequate interventions, should preferably be based on data without bias due to incompleteness or underreporting. However, substantial underreporting of malaria was suspected [2]. After comparing hospital admission and notification data this was estimated at 59% over the years 1988–92 [3]. To reduce underreporting, laboratory-based notification was recommended because of the laboratory's crucial role in the diagnosis of malaria. On 1 April 1999 a new Contagious Diseases Act came into force in the Netherlands. Under this law malaria and nine other infectious diseases (brucellosis, yellow fever, leptospirosis, anthrax, ornithosis/psittacosis, Q fever, rubella, *E. coli*-infection and trichinosis) are placed into category C, which introduces mandatory notification by the head of the diagnosing laboratory instead of the physician.

The concept of underreporting (i.e. incomplete coverage) is often mentioned in the literature but seems to be based upon different definitions and correspondingly involves different calculations. Instead of quantifying underreporting in one register relative to other registers a more accurate picture is portrayed by assessing the completeness of the different registrations relative to an estimated total number of cases (i.e. the number of registered cases plus an approximated number of unobserved cases). The unobserved cases can be estimated with a statistical technique known as 'capture–recapture (CRC) analysis'. CRC analysis has been used to assess the completeness of registration of various infectious diseases [4–14], including malaria [15, 16]. We have performed a CRC analysis using three malaria registrations and estimated the completeness of notification by physicians and laboratories, followed by separate analyses for each type of plasmodium, because of a special interest in the underreporting of the most severe form, *P. falciparum* malaria.

METHODS

Nearly all Dutch laboratories involved in parasitology participate in the national quality assessment for parasitological diagnosis. In January 1996 these laboratories ($n = 107$) were asked to report to us all microscopically confirmed cases of malaria found in that year through standardized questionnaires, with specific identifiers for patient (date of birth, sex and postal code) and parasite (*Plasmodium* species). Checks were carried out to exclude the possibility that a number of malaria cases would be diagnosed outside the laboratories in the survey, but notified to the RNID. Information from the Centres for Asylum-Seekers, the Central Military Hospital, the Medical Service for Merchant Sailors and the Occupational Health Service of Amsterdam Airport and KLM Royal Dutch Airlines assured us that all these institutions perform malaria diagnosis through the regular laboratories.

Using the individual identifiers, the laboratory survey data were matched to two other national registers for malaria in the Netherlands: the RNID and the hospital admission data (ICD9 code for malaria) from the National Morbidity Registration, after elimination of duplicate reports within each of the registrations. Two authors matched the data files by hand and in case of doubt consensus was sought.

The total number of individuals present in one or more registrations does not necessarily reflect a reliable approximation of the true number of cases. The purpose of CRC analysis is to assess, on the basis of the available information, the number of cases that are not registered. In an article published in 1972, Stephen Fienberg demonstrated how this number of unobserved cases could be estimated, using log-linear analysis [17]. For CRC analysis according to Fienberg the availability of data from at least three different, possibly incomplete, partially overlapping and preferably, but not necessarily, independent sources is needed [6, 17–20]. The data can be put in a $2 \times 2 \times 2$ contingency table, indicating the absence or presence of a case in each of the registers. This table has one empty cell, corresponding to the number of cases never registered. CRC aims at obtaining an estimate of the unregistered number of patients in the empty cell from the available data in the other cells. This estimate can be found under the best fitting and most parsimonious log-linear model. Finally, the total number of individuals is the number of registered cases plus the estimated number of non-registered patients.

Starting from a saturated model non-significant interaction terms were eliminated one after the other until the best fitting, most parsimonious, log-linear model was obtained by stepwise analysis as implemented in SPSS® (version 8.0 for Windows), with the procedure for hierarchical log-linear analysis. The coefficients and thus the final estimates were calculated with the SPSS® procedure for generalized log-linear analysis. The 95% confidence interval around the

estimated number of malaria cases was calculated assuming a Poisson distribution. Four assumptions must be met for the 3-sample CRC model to be valid. We will return to these assumptions in the discussion. After estimating the total number of malaria cases we performed a stratified CRC analysis by type of plasmodium to find out if the four malaria types have different CRC probabilities. This was done to assess whether underreporting occurred to a lesser or greater extent in relationship with the dangerous *P. falciparum* malaria or the more benign types.

RESULTS

The response rate of the laboratories in the survey in 1996 was 83·2%. Some of the participating laboratories reported not performing microscopic diagnosis of malaria (4·7%) or did not identify any malaria case (5·6%), resulting in 72·9% of the laboratories reporting at least one case of malaria. In the laboratory survey *P. falciparum* accounted for 57·0% of the malaria cases. The distribution of the different malaria parasites in the laboratory survey is shown in Table 1. In the RNID 60% *P. falciparum* could be found against 69% among the hospital admissions.

In the participating laboratories 535 cases of malaria were microscopically identified in 1996, while physicians officially notified 311 malaria cases and 330 malaria patients were admitted to a hospital. To increase the validity of the CRC analysis, the matched data file was corrected for 12 cases notified to the RNID in 1997 but found to be diagnosed in the laboratories in 1996 and 15 cases notified in 1996 but actually diagnosed in the laboratories in 1995. After this correction for late notification a total of 667 malaria patients were known in at least one of the registers (Table 2). For two cases in the laboratory survey insufficient identifiers for perfect matching were available and these patients were excluded from the CRC analysis. Figure 1 shows the distribution of the 665 malaria patients over the different malaria registrations and the overlap between these lists, as used in the CRC analysis. A substantial number of malaria patients are only known to the RNID or the hospital admission register and do not appear in the laboratory survey.

A log-linear model with two 2-way interactions, N*H, L*H (L = laboratory survey, N = Notification Office and H = hospital admissions) was obtained. These interactions represent pair-wise dependencies between the different registers [N and H] and [L and H]. The small likelihood ratio, G^2, compared to the number of degrees of freedom (D.F.), shows that this model fitted the data well ($G^2 = 0{\cdot}741$; D.F. = 1; $P = 0{\cdot}785$) and gave an estimate of 774 (95% CI of 740–821) malaria cases. The completeness of notification in 1996 can now be estimated at 40·2% for physicians (311/774 cases) and 69·1% for the laboratories (535/774 cases).

The case ascertainment (the number of malaria patients known in at least one of the three malaria registers) for 1996 can be estimated as 86·2% (95% CI 81·2–90·1%).

Table 1. *Distribution of diagnosed malaria parasites* (plasmodium *species*) *and their percentage of the total number of malaria cases identified in the Netherlands in 1996*

Plasmodium species	Malaria patients (%)
P. falciparum	305 (57·0%)
P. vivax	165 (30·8%)
P. ovale	43 (8·0%)
P. malariae	7 (1·3%)
Parasite unknown	15 (2·8%)

Table 2. *The number of malaria patients identified in each of the three malaria registrations and the number of malaria patients registered in at least one of malaria registers in 1996*

Notified to the RNID*	311
Hospital admissions*	330
Diagnosed in the laboratories*	535
Registered in at least one of the registers*†	667

* After correction for duplicate reports.
† After correction for late notification.

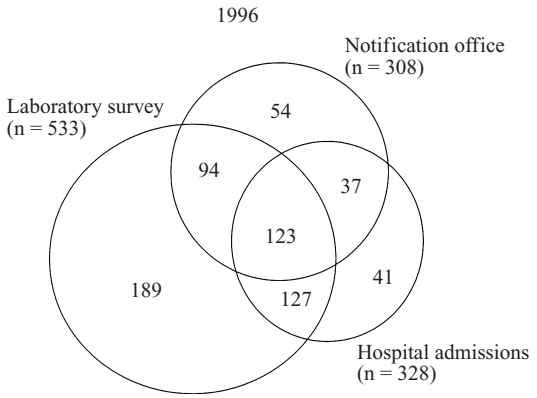


Fig. 1. Distribution of malaria cases in the Netherlands between three registers in 1996.

Table 3. *Stratified capture–recapture analysis of malaria cases in the Netherlands in 1996, according to the* plasmodium *species. The detection rate is calculated as obs(N)/est(N)*

Plasmodium species	Cases observed*	Cases estimated	Detection rate
P. falciparum	383	438	0·87
P. vivax	195	222	0·88
P. ovale	50	56	0·89
P. malariae	8	8	1·0
Parasite unknown	29	64	0·45
Total malaria cases	665	788	0·84

* After exclusion of two laboratory cases with insufficient identifiers for perfect matching.

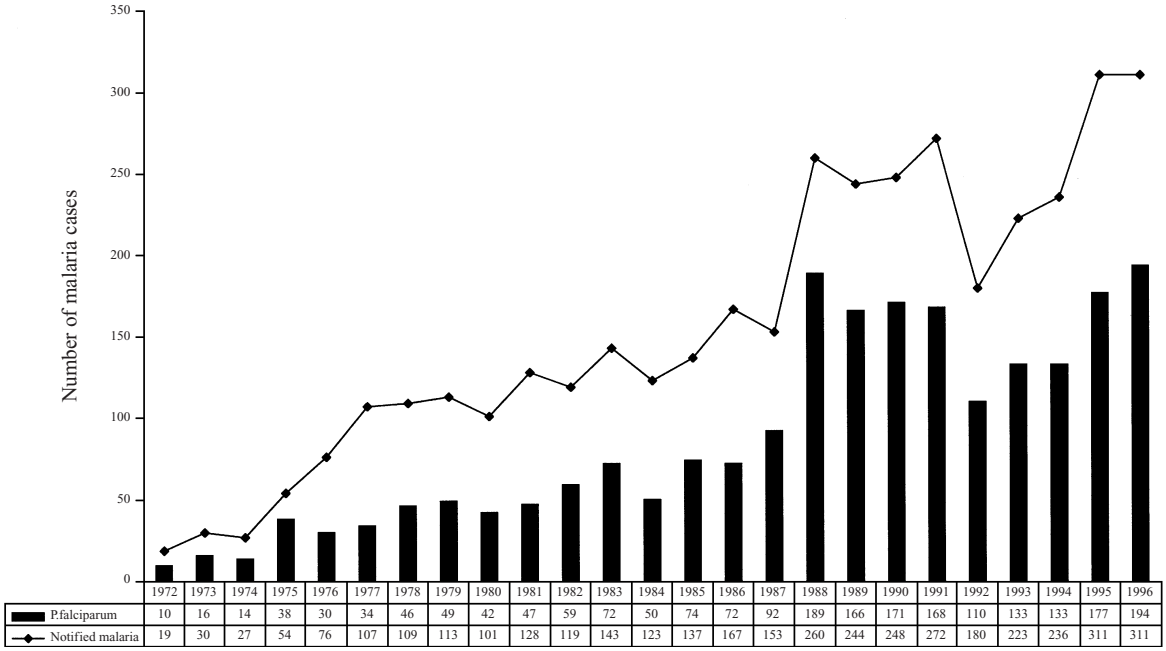
The stratified CRC analysis by type of plasmodium (Table 3) resulted in a slightly higher total number of estimated malaria cases of 788 patients (within 95% CI of original estimate). The detection rates of patients with different plasmodia do not show very much variation.

DISCUSSION

This study confirms that more malaria cases occur in the Netherlands than are reflected by the numbers officially notified by physicians in the past. Furthermore it demonstrates that the three different malaria registers are all substantially incomplete. Of particular interest is the observation that a considerable number of patients could only be found in the records of the RNID and/or the hospital admission register. They were unknown to the laboratories, although malaria diagnosis by thick or thin smear is often considered as the gold standard, especially at the time of this study when antigen tests were only used for research purposes. These cases could partly be explained by non and incomplete reporting of laboratories and cases could also have occurred in the (few) laboratories not participating in the national quality assessment of parasitological diagnosis. Other patients might have been diagnosed abroad and started with anti-malarials before their arrival in the Netherlands, clearing the parasites from the peripheral blood and subsequently reported to us as negative by the Dutch laboratories. A number of patients may have been notified or admitted solely on clinical grounds, without laboratory verification. Although unlikely, some physicians (for example former tropical doctors) could still prepare and microscopically examine the blood films themselves.

The number of malaria notifications in the Netherlands showed an increasing trend until 1996 (Fig. 2). According to Wetsteyn and De Geus [21] the incidence of imported malaria is determined by the level of endemicity in the malarious areas visited, the exposure to infected Anopheles mosquitoes (in turn, related to duration of stay, way of travelling and practising anti-mosquito measures) and the success of chemoprophylaxis (determined by compliance and prophylactic drug resistance). The increase of imported malaria in the Netherlands in the second half of the 1970s could be explained by growing tourism to tropical Africa and a further rise during the 1980s is expected to be the result of the spread of resistance against chloroquine and other commonly used anti-malarial drugs. Apart from that, malaria transmission itself seems to have increased in certain areas, such as West Africa [22]. Participation in peace-keeping operations or elections [23, 24], the number and nationalities of immigrants and asylum-seekers [24–26] or the extent of certain marginalized groups [25, 27] can also alter the incidence of imported malaria cases over a certain period of time, as well as influence the proportion of the different malaria parasites. In the Netherlands an increase could be observed in the proportion of malaria caused by *P. falciparum*. Around 1980 falciparum malaria was responsible for approximately 40% of all notified malaria cases but 10 years later this had increased to almost 70% [21]. In the 1990s the proportion of falciparum malaria stabilized around 60%.

Estimates of underreporting are frequently derived from different settings. They can be based upon surveys performed at the national level [28] or among small groups [25, 29]. The background of the data that are compared with the official notification register



	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
P.falciparum	10	16	14	38	30	34	46	49	42	47	59	72	50	74	72	92	189	166	171	168	110	133	133	177	194
Notified malaria	19	30	27	54	76	107	109	113	101	128	119	143	123	137	167	153	260	244	248	272	180	223	236	311	311

Fig. 2. The total number of notified malaria cases and the number of malaria cases caused by *P. falciparum* in the Netherlands between 1972 and 1996.

may be different, and can vary between hospital admission data [3], laboratory-based information [25, 28], physicians consulting a Reference Laboratory [29] or travellers [30]. The different registers were sometimes matched at the individual level [25], at times in a stratified manner [3] or in another way [28]. In this study we used a well-described and replicable method and estimated the completeness of notification of three different malaria registers through CRC analysis.

For the 3-sample CRC technique to be valid, four assumptions must hold [17–20]. First, overlap between registers must be established without erroneously misclassifying people as observed in only one, two or all three registers. This can be achieved when cases can be uniquely identified. We used individual identifiers for each of the patients and only two patients could not be identified beyond doubt due to (partially) missing identifiers. It is important that only true cases are counted. Ideally both the positive predictive value (PPV) and the negative predictive value (NPV) of the registrations should be 100%. None of our registrations will meet this condition, although in the case of malaria specifically, we assume that the PPV will be high. The large overlap of the hospital records and the notification data with one or two of the other registrations also supports this view. When the PPV of registrations is low CRC analysis will result in overestimating the number of cases.

Second, the registers should preferably, but not necessarily, be 'independent', meaning that the probability of being recorded in one register is not affected by being (or not being) registered in another. Such dependence can result from co-operation between the agencies that keep the different registrations, exchange of information or a more or less predictable flow of patients along various institutions due to referral. In 2-sample CRC methods this assumption is crucial and dependencies can cause under- or overestimation. In the 3-sample CRC approach pair-wise dependencies between registers can be handled analytically. In the log-linear model they can be identified as interactions in the model. Since we could not rule out pair-wise dependencies, we decided not to rely on the 2-sample CRC analysis but instead to use Fienberg's method.

Third, the population should be 'homogeneous' meaning that the population under consideration should not be composed of segments that have markedly different CRC-probabilities. One way of handling the homogeneity assumption is to stratify the population into more homogeneous strata and then to carry out CRC analyses for each of the distinct subgroups. We performed a stratified CRC analysis by type of plasmodium. This resulted in a slightly, but not significantly, higher total number of estimated malaria cases of 788 patients. The detection rates of patients with different plasmodia do not show considerable variation. This may indicate the absence of

a violation of the homogeneity assumption. However, we cannot exclude the possible presence of other (but unmeasured) sources of heterogeneity.

Finally, the population should be 'closed' such that the true population size is neither affected by people entering the population (e.g. through in-migration of cases and disease onset) nor by people leaving the population (e.g. through out-migration, recovery or mortality). The closed population assumption should be given critical attention because the aim of this study was to obtain an estimate of the incidence of imported malaria cases and violation might have resulted in overestimation (because incident cases may be late entries who have, therefore, a smaller probability of being captured more than once). When an open population is assumed, this could be handled in two different ways. One method is to perform the analysis of the different registrations within a short period of time and therefore the population could be considered as 'closed' during this interval. For imported malaria, a relatively rare disease with a short course, this approach does not seem feasible. An alternative is to use more complicated models, allowing for migration, birth and death to take place, such as the Jolly–Seber model [31]. The design of capture–recapture studies, the data requirements, the validity of the outcome of the different analyses and the selection of the most appropriate model to estimate the incidence of imported malaria and other infectious diseases should be given thought in further studies. In the context of these considerations our results suggest that laboratory-based notification can considerably increase the number of reported malaria cases as compared to notification by physicians. Since we actively approached the laboratories their level of underreporting found in this study cannot necessarily be extrapolated to the level of underreporting for laboratory notification. However, malaria was notified 571 times in 2001. Assuming a similar number of cases of imported malaria, this figure lies well within the range of our laboratory results for 1996. But possibly one-third of the malaria cases may still go unreported.

ACKNOWLEDGEMENTS

This study was performed in co-operation with the Office of the Chief Medical Officer in the Netherlands, the Netherlands Society of Parasitology (NVP), the Foundation for Parasitology Laboratory Diagnosis (SPLD), the Foundation for Quality Assessment in Medical Microbiology (SKMM) and the Centre for Communicable Disease Epidemiology of the National Institute of Public Health and the Environment (RIVM). Permission was obtained from the Inspector-General for Infectious Diseases at the Office of the Chief Medical Officer and the Privacy Commission of the National Morbidity Registration (LMR). We thank the heads of the participating laboratories for their co-operation.

REFERENCES

1. Muentener P, Schlagenhauf P, Steffen R. Imported malaria (1985–95): trends and perspectives. Bull W H O 1999; **77**: 560–6.
2. Sprenger MJW, Schrijnemakers PM. More information on infectious diseases provided by a national information system. Ned Tijdschr Geneeskd 1998; **142**: 1923–6.
3. Reep-van den Bergh CMM, Docters van Leeuwen WM, Kessel RPM van, Lelijveld JLM. Malaria: under-notification and risk assessment for travellers to the tropics. Ned Tijdschr Geneeskd 1996; **140**: 878–82.
4. Hubert B, Desenclos JC. Evaluation of the exhaustiveness and representativeness of a surveillance system using the capture-recapture method. Application to the surveillance of meningococcal infections in France in 1989 and 1990. Rev Epidémiol Santé Publique 1993; **41**: 241–9.
5. Abeni DD, Brancato G, Perucci CA. Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. Epidemiol 1994; **5**: 410–4.
6. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record estimation I: History and theoretical development. II: Applications in human diseases. Am J Epidemiol 1995; **142**: 1047–68.
7. Dromer F, Mathoulin S, Dupont B, Laporte A. Epidemiology of cryptococcosis in France: a 9-year survey (1985–1993). French Cryptococcosis Study Group. Clin Infect Dis 1996; **23**: 82–90.
8. Infuso A, Hubert B, Etienne J. Underreporting of legionnaires' disease in France: the case for more active surveillance. Eurosurveill 1998; **3**: 48–50.
9. Devine MJ, Bellis MA, Tocque K, Syed Q. Whooping cough surveillance in the north west of England. Commun Dis Publ Hlth 1998; **1**: 121–5.
10. Reintjes R, Termorshuizen F, van de Laar MJ. Assessing the sensitivity of STD surveillance in the Netherlands: an application of the capture-recapture method. Epidemiol Infect 1999; **122**: 97–102.
11. Dechant EJ, Rigau-Perez JG. Hospitalizations for suspected dengue in Puerto Rico, 1991–1995: estimation by capture-recapture methods. The Puerto Rico Association of Epidemiologists. Am J Trop Med Hyg 1999; **61**: 574–8.

12. Bernillon P, Lievre L, Pillonel J, Laporte A, Costagliola D. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990–1993. The Clinical Epidemiology Group from Centres d'Information et de Soins de l'Immunodeficience Humaine. Int J Epidemiol 2000; **29**: 168–74.
13. Gallay A, Vaillant V, Bouvet P, Grimont P, Desenclos JC. How many foodborne outbreaks of salmonella infection occurred in France in 1995? Application of the capture-recapture method to three surveillance systems. Am J Epidemiol 2000; **152**: 171–7.
14. Tocque K, Bellis MA, Beeching NJ, Davies PD. Capture recapture as a method of determining the completeness of tuberculosis notifications. Commun Dis Publ Hlth 2001; **4**: 141–3.
15. Deparis X, Pascal B, Baudon D. Evaluation of the completeness of the epidemiological surveillance systems for malaria by the capture-recapture system in the French armies in 1994. Trop Med Int Health 1997; **2**: 433–9.
16. Barat LM, Barnett BJ, Smolinski MS, et al. Evaluation of malaria surveillance using retrospective, laboratory-based active case detection in four southwestern states, 1995. Am J Trop Med Hyg 1999; **60**: 910–4.
17. Fienberg SE. The multiple-recapture census for closed populations and the 2^k incomplete contingency table. Biometrika 1972; **59**: 591–603.
18. Bishop YMM, Fienberg SE, Holland PW. Discrete multivariate analysis. Cambridge: MIT-Press, 1975.
19. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev 1995; **17**: 243–63.
20. LaPorte RE, Dearwater SR, Chang Y-F, et al. Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods to injury monitoring. Am J Epidemiol 1995; **142**: 1069–77.
21. Wetsteyn JCFM, De Geus A. Falciparum malaria, imported into the Netherlands, 1979–1988. I. Epidemiological aspects. Trop Geogr Med 1995; **47**: 53–60.
22. Philips-Howard PA, Porter J, Behrens RH, Bradley DJ. Epidemic alert: malaria infections in travellers from West Africa. Lancet 1990; **335**: 119–20.
23. Kachur SP, Reller ME, Barber AM, et al. Malaria surveillance-United States, 1994. In: CDC surveillance summaries. MMWR 1997; **46** (No. SS-5): 1.
24. Wetsteyn JCFM, Kager PA, Van Gool. The changing pattern of imported malaria in the Academic Medical Centre, Amsterdam. In: Wetsteyn JCFM, ed. Imported malaria in the Netherlands, an uninvited guest (dissertation). Amsterdam: University of Amsterdam, 1993: 117–33.
25. Lambeth, Southwark and Lewisham Health Authority. The surveillance of communicable disease and non-infectious environmental hazards in Lambeth, Southwark and Lewisham. London, United Kingdom: Directorate of Health Policy and Public Health, 1996.
26. Bradley DJ, Warhurst DC. Malaria imported into the United Kingdom during 1991. Commun Dis Rep 1993; **3** (Review No. 2): R25–R28.
27. Centers for Disease Control and Prevention. Summary of notifiable diseases United States, 1996. MMWR 1997; **45**: 1–87.
28. Legros F, Fromage M, Ancelle T, Burg E, Danis M. Enquête nationale de recensement des cas de paludisme d'importation en France métropolitaine pour l'année 1997. Paris, France: Centre National de Référence pour les Maladies d'Importation (CNMRI), 1998; Bulletin No. 14.
29. Davidson RN, Scott JA, Behrens RH, Warhurst D. Underreporting of malaria, a notifiable disease in Britain. J Infect 1993; **26**: 348–9.
30. Steffen R, Heuser R, Machler R, et al. Malaria chemoprophylaxis among European tourists in tropical Africa: use, adverse reactions, and efficacy. Bull W H O 1990; **68**: 313–22.
31. Cormack RM. Loglinear models for capture–recapture experiments in open populations. In: Hiorns RW, Cooke D, eds. The mathematical theory of the dynamics of biological populations II. London: Academic Press, 1981.