

## Completeness of notification of tuberculosis in The Netherlands: how reliable is record-linkage and capture–recapture analysis?

N. A. H. VAN HEST<sup>1,2\*</sup>, F. SMIT<sup>3,4</sup>, H. W. M. BAARS<sup>1</sup>, G. DE VRIES<sup>1,2</sup>,  
P. E. W. DE HAAS<sup>5</sup>, P. J. WESTENEND<sup>6</sup>, N. J. D. NAGELKERKE<sup>7</sup>  
AND J. H. RICHARDUS<sup>1,2</sup>

<sup>1</sup> Department of Infectious Disease Control, Rotterdam Public Health Service, Rotterdam, The Netherlands

<sup>2</sup> Department of Public Health, Erasmus University Medical Centre Rotterdam, Rotterdam, The Netherlands

<sup>3</sup> Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

<sup>4</sup> Department of Clinical Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

<sup>5</sup> Diagnostic Laboratory of Infectious Diseases and Prenatal Screening, Mycobacteria Reference Unit, National Institute of Public Health and the Environment, Bilthoven, The Netherlands

<sup>6</sup> Laboratory for Pathology Dordrecht, Dordrecht, The Netherlands

<sup>7</sup> United Arab Emirates University, Faculty of Medicine and Health Sciences, Department of Community Medicine, Al Ain, United Arab Emirates

(Accepted 4 October 2006; first published online 7 December 2006)

### SUMMARY

The aim of this study was to describe a systematic process of record-linkage, cross-validation, case-ascertainment and capture–recapture analysis to assess the quality of tuberculosis registers and to estimate the completeness of notification of incident tuberculosis cases in The Netherlands in 1998. After record-linkage and cross-validation 1499 tuberculosis patients were identified, of whom 1298 were notified, resulting in an observed under-notification of 13·4%. After adjustment for possible imperfect record-linkage and remaining false-positive hospital cases observed under-notification was 7·3%. Log-linear capture–recapture analysis initially estimated a total number of 2053 (95% CI 1871–2443) tuberculosis cases, resulting in an estimated under-notification of 36·8%. After adjustment for possible imperfect record-linkage and remaining false-positive hospital cases various capture–recapture models estimated under-notification at 13·6%. One of the reasons for the higher than expected estimated under-notification in a country with a well-organized system of tuberculosis control might be that some tuberculosis cases, e.g. extrapulmonary tuberculosis, are managed by clinicians less familiar with notification of infectious diseases. This study demonstrates the possible impact of violation of assumptions underlying capture–recapture analysis, especially the perfect record-linkage, perfect positive predictive value and absent three-way interaction assumptions.

### INTRODUCTION

Surveillance of infectious diseases, including tuberculosis, is vital for public health. Mandatory

notification is one of the mechanisms to carry out such surveillance but can be contaminated by false-positive cases while true-positive cases may be missed [1, 2]. For correct interpretation of tuberculosis figures and the longitudinal trends therein the quality of tuberculosis registers and the completeness of notification should be assessed [3].

\* Author for correspondence: N. A. H. van Hest, M.D., M.Sc., Tuberculosis Control Section, Department of Infectious Disease Control, Rotterdam Public Health Service, PO Box 70032, 3000 LP Rotterdam, The Netherlands.  
(Email: vanhestr@ggd.rotterdam.nl)

Of importance in this assessment is record-linkage, i.e. comparing patient data across registers. Record-linkage not only improves completeness of registration but cross-validation with other registers also improves the quality of the data [3, 4]. In The Netherlands multiple tuberculosis registers are available. Completeness of notification and other registers can then be assessed relative to the case ascertainment, i.e. the total number of patients observed in at least one register, or relative to an estimated number of patients through capture–recapture analysis. Based on certain assumptions capture–recapture methods use information on the overlap between registers to estimate the number of cases unknown to all registers and thus the estimated total number of cases [5]. The preferred capture–recapture method entails log-linear modelling of at least three linked registers, less compromised by possible violation of the underlying assumptions compared to capture–recapture analysis based on two linked registers [6–9]. Capture–recapture analysis has been used to assess the completeness of notification and other registers of various infectious diseases [10], including tuberculosis [11–15].

The primary objective of this study is to describe a systematic process of record-linkage of different tuberculosis registers, cross-validation, case ascertainment and capture–recapture estimation of incident tuberculosis cases in The Netherlands in 1998. The secondary objective is to assess the completeness of tuberculosis notification. Under-notification was expected to be low in a country with a well-organized system of tuberculosis control and with a previous estimate of 8% between 1995 and 1998 [16].

## METHODS

Permission for this study was obtained from the Medical Ethics Committee of the Erasmus Medical Centre in Rotterdam and the data protection committees of the tuberculosis registrations.

### Data sources and patient identifiers

Three registers of tuberculosis cases in The Netherlands in 1998 were examined:

- (1) Patients notified by tuberculosis physicians to the Register of Notifiable Infectious Diseases of the Health Care Inspectorate (Notification).
- (2) Patients with a positive culture for *Mycobacterium tuberculosis* complex known to the Mycobacteria Reference Unit at the National

Institute for Public Health and the Environment (Laboratory).

- (3) Hospitalized patients recorded by the National Morbidity Registration with an International Code for Diseases (ICD-9) for active tuberculosis (ICD-9 codes 010–018) (Hospital).

Duplicate entries in each register and laboratory contamination records were deleted. Three other tuberculosis-related registers used for cross-validation (exclusion of false-positive tuberculosis cases or verification of assumed true-positive tuberculosis patients among non-culture-confirmed tuberculosis cases) or acquisition of additional patient variables, will be discussed later. For each patient date of birth, postal code, sex, and date of notification, first culture sample or hospital admission were collected as personal identifiers to be used in all record-linkage procedures.

### Study year

The reference year chosen was 1998 as from 1 April 1999 only the year of birth is recorded among the mandatory notification data, effectively ruling out reliable record-linkage between the Notification and other registers [17]. Patients with a date of notification, hospital admission or culture sampling (in order of primacy) between 1 January 1998 and 1 January 1999 were included. To correct for misclassification due to late notification or positive bacteriological results, all three registers were examined between 1 July 1997 and 1 July 1999.

### Case-definition

Tuberculosis cases are defined as all observed (by notification, culture confirmation or hospital admission) and unobserved cases of active tuberculosis (excluding *Mycobacterium bovis* BCG infection). Culture-confirmed patients are assumed true-positive tuberculosis patients.

### Record-linkage

Record-linkage was performed manually using the patient identifiers and proximity of date of notification, first culture sample or hospital admission. First the Notification and Laboratory registers were linked. For perfect linkage all patient identifiers should be identical and date of notification and first culture sample should differ by <1 month. To avoid misclassification of near links with a minor

discrepancy in one of the identifiers, e.g. due to clerical errors such as typing mistakes, near links and cases with a date difference of >1 month were checked using the surname of the patient. Since the researchers did not know the patients' names due to privacy regulations, a 'trusted third party' ascertained match or mismatch. Finally, the Hospital register was linked to the two other registers, using human judgement and consensus in case of near links.

### Cross-validation of cases and collection of additional variables

To improve the positive predictive value of the linked tuberculosis registers, non-culture-confirmed cases were examined through record-linkage with three tuberculosis-related datasets in The Netherlands. Cross-validation was conducted in four steps. First, cases with disease actually caused by non-tuberculous mycobacteria (NTM) were identified and excluded through record-linkage with the national register for NTM cultures at the Mycobacteria Reference Unit, after a representative check in a large regional laboratory demonstrated that 80% (143/179) of the local NTM isolates could be found in the national NTM register. Second, patients later diagnosed with disease other than tuberculosis or NTM were identified and excluded through record-linkage with a dataset of such patients secondary to The Netherlands Tuberculosis Register (NTR), an extensive system of voluntary reporting by tuberculosis physicians [18]. Third, non-culture-confirmed patients possibly diagnosed by histopathology examination were verified through the Pathological Anatomy Laboratory Computerized Archive (PALGA), the nationwide network and registry of histopathology and cytopathology results in The Netherlands. Excerpts of the histopathology reports of linked patients were reviewed by a pathologist and cases with inconsistent results discarded. Finally, the total set of linked tuberculosis registers was linked to the NTR for verification of the remaining non-culture-confirmed tuberculosis patients and collection of additional variables for cases in any of the linked registers: nationality (Dutch, non-Dutch), location of tuberculosis (pulmonary, extrapulmonary) and infectiousness (sputum smear-positive, sputum smear-negative). Although more complete in data the NTR was expected to have a complete overlap with the Notification register (both registers are maintained by the same tuberculosis physicians) and was deliberately used for the purpose

of validation of the conventional notification, laboratory and hospital tuberculosis registers [3].

### Case ascertainment, capture–recapture analysis and observed and estimated register-specific coverage rates

The total and stratified observed register-specific coverage rates are defined as the number of tuberculosis patients in each register divided by the total or stratified case ascertainment, expressed as percentage.

The total number of unobserved tuberculosis cases was estimated on the basis of the cross-validated distribution of the observed cases over the Notification, Laboratory and Hospital registers. The independence of registers and other assumptions underlying capture–recapture analysis have been described previously [10]. Interdependencies between the three tuberculosis registers are probable, causing possible bias in two-source capture–recapture estimates. Three-source log-linear capture–recapture analysis was employed to take possible interdependencies into account [12, 15]. Estimated register-specific coverage rates are defined as the number of tuberculosis patients in each register divided by the estimated total number of tuberculosis patients by capture–recapture analysis.

## RESULTS

Table 1 shows the initial number of cases, the number of cases excluded from the study before and after record-linkage and the final number of cases in the three tuberculosis registers in The Netherlands in 1998. The hospital admission of 12 cases in 1997 and eight cases in 1999, all notified in 1998, was included in the data.

Among the 295 near links between the Notification and Laboratory registers, the 'trusted third party' confirmed 267 candidate pairs as true links. Among the confirmed links, 133 candidate pairs had administrative discrepancies, predominantly (63.8%) in the postal code.

Record-linkage of all 537 non-culture-confirmed cases to the NTM register and the subset of the NTR revealed that despite NTM infection or any other diagnosis than tuberculosis 26 out of 426 non-culture-confirmed cases on the Notification register (6.1%) were not de-notified and 25 out of 217 non-culture-confirmed cases on the Hospital register (11.5%) were still recorded with an ICD-9 tuberculosis code. Figure 1 shows the distribution of the final

Table 1. The initial number of cases, the number of cases excluded from the study before and after record-linkage and the final number of cases in the three tuberculosis registers in The Netherlands in 1998

	Tuberculosis registers		
	Notification	Laboratory	Hospital
Patients initially found in the different tuberculosis registers in The Netherlands	1334	1074	658
Patients excluded from the analysis			
Patients lost during matching process	3	0	0
Duplicate entry laboratory register	0	1	0
Duplicate entry notification register	1	0	0
Laboratory contamination	6	19	2
Culture of <i>Mycobacterium bovis</i> BCG	0	14	1
Subtotal before record linkage	1324	1040	655
Patients with a laboratory sample date in 1998 but notified in 1997	0	3	2
Patients with a laboratory sample date in 1998 but notified in 1999	0	29	6
Patients only known to the hospital in 1998 but notified in 1997	0	0	10
Patients only known to the hospital in 1998 but notified in 1999	0	0	2
Patients not notified with a laboratory sample date in 1998 but admitted to the hospital in 1999	0	2	0
Patients with initial tuberculosis notification in 1998 but diagnosis later withdrawn because of non-tuberculous mycobacteria ( $n = 35$ ; 7 patients appear in both registers)	19	0	23
Patients with initial tuberculosis notification in 1998 but diagnosis later withdrawn because of other reasons than non-tuberculous mycobacteria	7	0	2
Patients included in the capture–recapture analysis	1298	1006	610

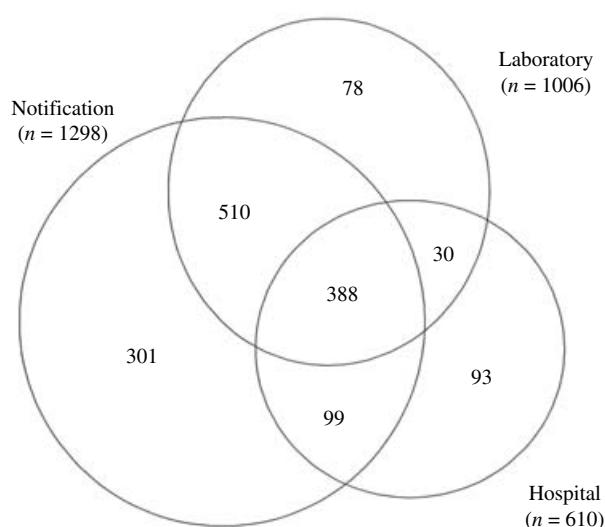


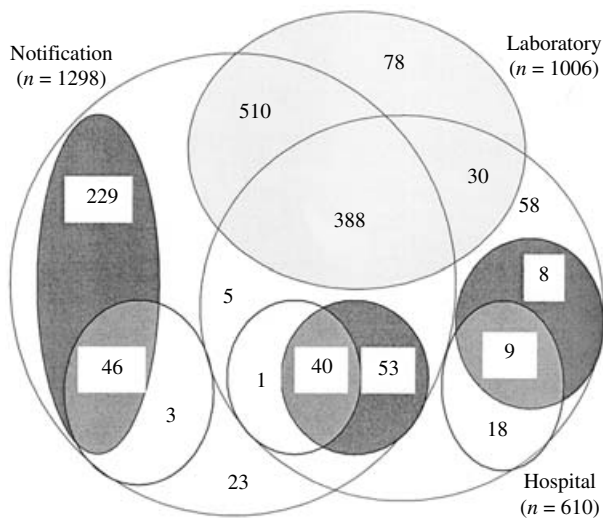
Fig. 1. Schematic view of the distribution of observed number of tuberculosis patients in The Netherlands in 1998, after record-linkage of three tuberculosis registers (total number of observed cases is 1499).

number of 1499 cases over the different tuberculosis registers. Of the 1006 culture-confirmed tuberculosis patients 108 patients (10.7%) could not be found in the Notification register.

Verification through PALGA of the remaining 493 non-culture-confirmed cases in the linked registers identified 117 patients (23.7%) with a histopathology report consistent with active tuberculosis. Verification through the NTR identified 385 patients (78.1%). Both exercises combined verified 407 patients (82.6%). Figure 2 shows the distribution of the PALGA- and NTR-verified non-culture-confirmed cases over the three linked tuberculosis registers. In total 94.3% (1413/1499) of all patients were culture confirmed or verified but only 37.6% (35/93) of the unlinked hospital patients.

Record-linkage of patients observed in any of the three linked tuberculosis registers with the NTR resulted in a coverage of 91.1%, 84.7% and 78.9% of the Notification, Laboratory and Hospital registers respectively. Of the 108 culture-confirmed tuberculosis patients not found in the Notification register 38 (35%) were voluntarily reported to the NTR.

The total and stratified observed number of tuberculosis patients and register-specific coverage rates of the three tuberculosis registers are shown in Table 2. Observed completeness of notification, culture confirmation and hospitalization is 86.6%,



**Fig. 2.** Schematic view of the distribution of observed number of tuberculosis patients in The Netherlands in 1998, after record-linkage of three tuberculosis registers (light grey = culture-positive), and the number of validated tuberculosis patients among the culture-negative cases (dark grey = Netherlands Tuberculosis Register; white = Pathological Anatomy Laboratory Computerized Archive).

67.1% and 40.7% respectively. The completeness of the Notification register is consistent over the strata, with non-culture-confirmed patients least likely to be notified. The Laboratory and Hospital registers have higher proportions of sputum smear-positive patients and both registers show a trend of culture confirmation and hospitalization increasing with age. If only culture-confirmed or otherwise verified cases were included the verified observed completeness of the Notification register would be 89.9%. The observed and verified observed under-notification is 13.4% and 10.1% respectively. When all 58 non-verified unlinked hospital cases are considered false-positive and the 38 culture-confirmed patients reported to the NTR considered notified, the adjusted observed under-notification is 7.3% (105/1441).

Based on the Akaike Information Criterion (AIC) the log-linear capture–recapture procedure initially selected the saturated model (see Discussion) as the best-fitting model which estimated 554 unobserved tuberculosis cases, resulting in an estimated total number of 2053 [95% confidence interval (CI) 1871–2443] tuberculosis cases. This translates into an estimated completeness of case ascertainment of 73.0% (1499/2053) and estimated register-specific coverage rates of 63.2%, 49.0% and 29.7% for the Notification, Laboratory and Hospital registers

respectively. The estimated under-notification is 36.8% (95% CI 30.6–46.9).

After adjustment for the 58 possibly false-positive unlinked hospital cases and the 38 possibly misclassified laboratory patients (Fig. 3) the selected, most parsimonious, log-linear capture–recapture model was the model with two two-way interactions between Notification and Laboratory and between Notification and Hospital. The small likelihood ratio,  $G^2$ , compared with the number of degrees of freedom (D.F.), shows that this model fits the data well ( $G^2=0.053$ , D.F. = 2,  $P=0.974$ ,  $AIC=-3.95$ ) and estimates 1547 (95% CI 1513–1600) tuberculosis patients. The completeness of case ascertainment after adjustment is 93.1% (1441/1547) and the estimated register-specific coverage rates are 86.4%, 65.0% and 35.7% for the Notification, Laboratory and Hospital registers respectively. Adjusted estimated under-notification is 13.6% (95% CI 11.7–16.5).

## DISCUSSION

### Main findings

This study shows that, even in a country with a well-organized system of tuberculosis control, record-linkage and cross-validation improve the data quality of tuberculosis registration and case ascertainment. These findings underscore the need for scrutiny of all tuberculosis registers, especially with regard to hospital-based data. Total and verified observed under-notification of tuberculosis in The Netherlands in 1998 was 13.4% and 10.1% respectively. The latter was slightly higher than a previously reported under-notification of 8%. After correction for possibly misclassified laboratory patients and remaining false-positive hospital cases the adjusted observed under-notification of 7.3% is similar to this previous estimate. The 36.8% under-notification estimated by a log-linear capture–recapture model before adjustments were made is highly inconsistent with the prior report. Adjustment for possible misclassification of laboratory patients and remaining false-positive hospital cases had a considerable impact on the log-linear capture–recapture estimate.

### Possible causes of poor data quality

The quality of the tuberculosis registers is mainly determined by the proportion of administrative discrepancies causing possible record-linkage misclassification (8.6% between Notification and Laboratory)

Table 2. Total and stratified number of tuberculosis cases identified by three tuberculosis registers and observed register-specific fractions

	Observed cases <i>n</i> (%)	Notification		Laboratory		Hospital	
		Freq.	Fraction	Freq.	Fraction	Freq.	Fraction
Total	1499	1298	86.6%	1006	67.1%	610	40.7%
Male*	849 (57.2)	747	88.0%	580	68.3%	357	42.0%
Female*	635 (42.8)	541	85.2%	411	64.7%	251	39.5%
Dutch†	389 (32.0)	372	95.6%	250	64.3%	157	40.4%
Non-Dutch†	826 (68.0)	790	95.6%	588	71.2%	316	38.3%
Pulmonary tuberculosis‡	770 (62.2)	734	95.3%	545	70.8%	296	38.4%
Extra-pulmonary tuberculosis‡	467 (37.8)	448	95.9%	307	65.8%	185	39.6%
Sputum smear-positive§	276 (42.3)	265	96.0%	243	88.0%	149	54.0%
Sputum smear-negative§	376 (57.7)	358	95.2%	237	63.0%	105	28.0%
<15 yr	101 (6.7)	89	88.1%	40	39.6%	37	36.6%
≥15 yr to <65 yr	1150 (76.7)	1000	90.0%	790	68.7%	450	39.1%
≥65 yr	248 (16.5)	209	84.3%	176	84.3%	123	49.6%
Culture-confirmed cases	1006 (67.1)	896	89.0%	1006	100%	418	41.6%
Non-culture-confirmed cases	493 (32.9)	402	81.1%	0	0%	192	38.9%
Metropolitan	477 (31.8)	418	87.6%	333	69.8%	182	38.2%
Non-Metropolitan	1022 (68.2)	880	86.1%	673	65.9%	428	41.9%

Freq., Frequency.

\* For 15 cases no information was available.

† For 284 cases no information was available.

‡ For 262 cases no information was available.

§ For 847 cases no information was available or they were non-pulmonary tuberculosis.

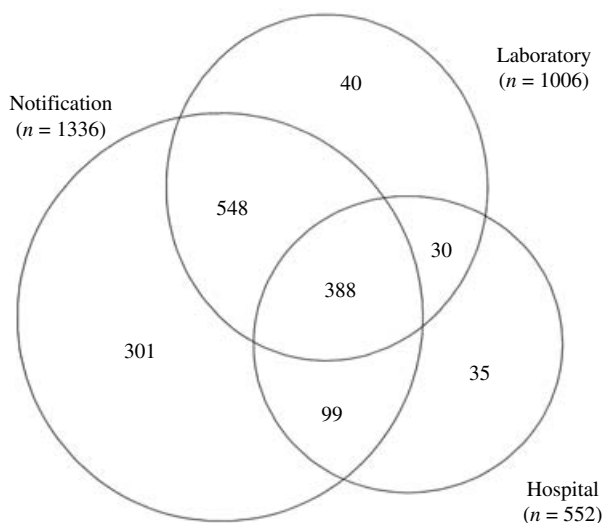


Fig. 3. Schematic view of the distribution of observed number of tuberculosis patients in The Netherlands in 1998, after record-linkage of three tuberculosis registers, and correction for possible misclassification of culture-positive patients and remaining false-positive unlinked hospital cases (total number of observed cases is 1441).

and the proportion of false-positive cases (8.2% among non-culture-confirmed cases in this study after previous elimination of laboratory contamination records and exclusion of *M. bovis* BCG isolates).

The majority of administrative discrepancies were found in the postal code. Apart from clerical errors, this could be due for example to frequent transfers of asylum seekers, notification of home address of prisoners vs. laboratory postcode of prison region or assigning a random local postal code to records with missing data in some registers. Patients with a culture of *M. bovis* BCG were excluded because of an expected low positive predictive value for systemic disease as all were either infants (probably with a post-BCG vaccination abscess) or older males (with probable urological *M. bovis* BCG instillation).

Despite maximum efforts to eliminate administrative discrepancies and false-positive records, our results still indicate imperfect record-linkage as, assuming a negligible number of lost reports, only 91.1% of all tuberculosis cases in the Notification register could be linked to the NTR. Since tuberculosis physicians report to both registers the expected overlap is 100%. A proportion of the tuberculosis cases in the final dataset not present in the Notification register could be explained by imperfect record-linkage because, remarkably, 38 culture-confirmed but not notified patients were voluntarily reported to the NTR, suggesting notification as well. After adjustment the number of patients in the

Notification register (1336) is almost similar as the number reported by the NTR in 1998 (1341). However, 70 culture-confirmed patients may not have been notified, reflecting the most serious public health aspect of under-notification, i.e. preventing possibly indicated contact investigations around potentially infectious patients.

In almost one-quarter of the non-culture-confirmed patients histopathology examination contributed to the diagnosis of tuberculosis. The majority of these patients were found in the Hospital register which is plausible because histopathology examination is more likely to be performed as part of a diagnostic work-up in patients with extrapulmonary tuberculosis requiring hospital admission. In The Netherlands, the contribution of PALGA to case verification in addition to the NTR was limited.

Despite the availability of additional tuberculosis-related registers, the majority (62.4%) of unlinked hospital cases could not be verified, compared to 7.6% of the unlinked notified cases. Although often used as a third data source in capture–recapture studies on human disease incidence, in the case of tuberculosis the data quality of hospital registers should be judged critically. A local capture–recapture study in the United Kingdom found 27% of all tuberculosis cases in the hospital register to be false-positive and in a regional capture–recapture study in Italy this was as high as 80% among unlinked hospital tuberculosis cases [12, 15].

### Limitations

The findings have to be placed in the context of the limitations of this study. The estimated coverage of the tuberculosis registers was based on three-source log-linear capture–recapture models. These models are only valid in the absence of violation of their underlying assumptions: perfect record-linkage (i.e. no misclassification of records), a closed population (i.e. no immigration or emigration in the time period studied) and a homogeneous population (i.e. no subgroups with markedly different probabilities to be observed and re-observed). In two-source capture–recapture methods one must also assume independence between registers [i.e. the probability of being observed in one register is not affected by being (or not being) observed in another]. In the three-source capture–recapture approach dependencies between two registers can be identified and incorporated in the log-linear model [5]. The three-way interaction

however, i.e. dependency between all three registers, cannot be incorporated in the model and its absence must be assumed. Nevertheless, violation of this assumption may occur, rendering capture–recapture analysis outcomes less valid. This and other limitations of capture–recapture analysis are described elsewhere in more detail [8, 19–25].

In this study, the possible remaining false-positive cases and violation of the perfect record-linkage assumption have already been discussed. Violation of the closed population assumption is presumably limited as with tuberculosis the opportunities for notification, culture confirmation or hospitalization are largely determined within a short period of time but could result in overestimation of the number of patients. More likely is violation of the absent three-way interaction assumption. Tuberculosis services in The Netherlands are organized around close collaboration between clinicians, microbiologists and public health professionals such as tuberculosis physicians and tuberculosis nurses. Examples of this collaboration are laboratory pre-notification, clinical isolation, contact investigations and referrals, explaining the two two-way interactions identified in the final log-linear capture–recapture model. The initial log-linear capture–recapture model with the best goodness-of-fit was the saturated model, i.e. including all two-way interactions. Violation of the absent three-way interaction assumption, which biased our estimates of the true population size, cannot be ruled out [8, 21, 23, 26]. Also more likely is violation of the homogeneity assumption: age, location of disease and infectiousness, among others, can account for different probabilities of being seen in a tuberculosis register. Although at least as vulnerable as log-linear models to the violation of underlying assumptions, to investigate possible bias as a result of violation of the homogeneity assumption, we have examined the data again with alternative estimators, as described in the capture–recapture analysis literature [8, 27]. These estimators reportedly perform well when compared to log-linear capture–recapture estimates [28], are arguably more robust to violation of the homogeneity assumption [29] and have been used in social sciences to estimate the size of hidden populations such as illicit drug users and homeless persons [29–32]. We applied Chao's heterogeneity and bias-corrected homogeneity models on the adjusted observed distribution of tuberculosis patients [33–35]. Both models estimate a total of 1545 tuberculosis patients (95% CI 1519–1580), very similar to the log-linear model, with

an estimated case ascertainment of 93.3% (1441/1545) and an estimated under-notification of 13.5% (95% CI 12.0–15.4). The CI of the adjusted log-linear and alternative estimates does not contain the expected value of 8%.

### Improving tuberculosis surveillance systems

Some ways of improving the performance of tuberculosis (and other infectious disease) surveillance systems could be:

- As an alternative to log-linear three-source capture–recapture analysis to estimate tuberculosis incidence, record-linkage (preferably web-based), between the two most relevant sources for tuberculosis surveillance, namely Notification and Laboratory registers, with both registers having a high positive predictive value, will improve timeliness of reporting, completeness of demographic, microbiological and epidemiological variables of the patients, and completeness of the number of patients and hence observed tuberculosis incidence.
- Treatment of all tuberculosis patients, including extrapulmonary cases, by a limited group of experienced specialist physicians, such as tuberculosis physicians, chest physicians or infectiologists, familiar with notification procedures, will improve completeness of notification.
- The introduction of pre-notification of positive laboratory test results for tuberculosis to the public health physicians responsible for processing the notifications from the local clinicians to the Health Care Inspectorate at the national level, with subsequent follow-up of unreported cases, as implemented in some regions of The Netherlands, will improve completeness of notification.

### CONCLUSION

Tuberculosis under-notification in The Netherlands in 1998 is probably around 8% and possibly around 13.6%. This study demonstrates the need for assessment of tuberculosis registers for quality of the data and completeness, and the importance of record-linkage [22]. It underscores that ‘as for the results of all epidemiological investigations, the credibility of any capture–recapture estimate will be enhanced to the extent that the investigator may be able to confirm the accuracy of all information used, such as diagnosis, location of the case within the space–time interval

analysed, and appropriate case matching, as with capture–recapture methods, errors are highly likely to have a more than additive effect on estimates’ [8, 36].

### ACKNOWLEDGEMENTS

We thank Nico Kalisvaart of the KNCV Tuberculosis Foundation, Dr Bert Mulder and Karel Nolsen of the Regional Laboratory for Microbiology Twente, Matty Meijer of the Register of Notifiable Infectious Diseases of the Health Care Inspectorate, Willem Hoogen Stoevenbelt of the National Morbidity Registration, Dr Mariel Casparie of PALGA, and all Departments of Tuberculosis Control of the Public Health Services in the Netherlands for technical assistance and cooperation.

### DECLARATION OF INTEREST

None.

### REFERENCES

1. **Dye C, et al.** Global burden of tuberculosis. *Journal of the American Medical Association* 1999; **282**: 677–686.
2. **WHO.** *Global Tuberculosis Control: surveillance, planning, financing.* WHO Report 2004. Geneva: World Health Organization, 2004.
3. **Migliori GB, et al.** Validation of the surveillance system for new cases of tuberculosis in a province of northern Italy. Varese Tuberculosis Study Group. *European Respiratory Journal* 1995; **8**: 1252–1258.
4. **Mukerjee AK.** Ascertainment of non-respiratory tuberculosis in five boroughs by comparison of multiple data sources. *Communicable Diseases and Public Health* 1999; **2**: 143–144.
5. **International Working Group for Disease Monitoring and Forecasting.** Capture-recapture and multiple-record estimation. I: history and theoretical development. *American Journal of Epidemiology* 1995; **142**: 1047–1058.
6. **Fienberg SE.** The multiple-recapture census for closed populations and the  $2^k$  incomplete contingency table. *Biometrika* 1972; **59**: 591–603.
7. **Bishop YMM, Fienberg SE, Holland PW.** *Discrete Multivariate Analysis.* Cambridge, MA: MIT Press, 1975.
8. **Hook EB, Regal RR.** Capture-recapture methods in epidemiology: methods and limitations. *Epidemiologic Reviews* 1995; **17**: 243–263.
9. **International Working Group for Disease Monitoring and Forecasting.** Capture-recapture and multiple-record estimation. II: applications in human diseases. *American Journal of Epidemiology* 1995; **142**: 1059–1068.



10. **Van Hest NA, Smit F, Verhave JP.** Underreporting of malaria incidence in The Netherlands: results from a capture-recapture study. *Epidemiology and Infection* 2002; **129**: 371–377.
11. **Sanghavi DM, et al.** Hyperendemic pulmonary tuberculosis in a Peruvian shantytown. *American Journal of Epidemiology* 1998; **148**: 384–389.
12. **Tocque K, et al.** Capture recapture as a method of determining the completeness of tuberculosis notifications. *Communicable Diseases and Public Health* 2001; **4**: 141–143.
13. **Mayoral Cortes JM, et al.** Incidence of pulmonary tuberculosis and HIV coinfection in the province of Seville, Spain, 1998. *European Journal of Epidemiology* 2001; **17**: 737–742.
14. **Cailhol J, et al.** Incidence of tuberculous meningitis in France, 2000: a capture-recapture analysis. *International Journal of Tuberculosis and Lung Disease* 2005; **9**: 803–808.
15. **Baussano I, et al.** Undetected burden of tuberculosis in a low-prevalence area. *International Journal of Tuberculosis and Lung Disease* 2006; **10**: 415–421.
16. **Van Loenhout-Rooijackers JH, et al.** Pyrazinamide use as a method to estimate under-reporting of tuberculosis. *International Journal of Tuberculosis and Lung Disease* 2001; **5**: 1156–1160.
17. **Klein S, Bosman A.** Completeness of malaria notification in the Netherlands 1995–2003 assessed by capture-recapture method. *Eurosurveillance* 2005; **10**: 244–246.
18. **KNCV Tuberculosis Foundation.** *Index Tuberculosis 2001–2002 – Netherlands*. The Hague: KNCV Tuberculosis Foundation, 2005.
19. **Desenclos JC, Hubert B.** Limitations to the universal use of capture-recapture methods. *International Journal of Epidemiology* 1994; **23**: 1322–1323.
20. **Brenner H.** Use and limitations of the capture-recapture method in disease monitoring with two dependent sources. *Epidemiology* 1995; **6**: 42–48.
21. **Cormack RM.** Problems with using capture-recapture in epidemiology: an example of a measles epidemic. *Journal of Clinical Epidemiology* 1999; **52**: 909–914.
22. **Papoz L, Balkau B, Lellouch J.** Case counting in epidemiology: limitations of methods based on multiple data sources. *International Journal of Epidemiology* 1999; **25**: 474–478.
23. **Hook EB, Regal RR.** Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources. *American Journal of Epidemiology* 2000; **152**: 771–779.
24. **Jarvis SN, et al.** Children are not goldfish-mark-recapture techniques and their application to injury data. *Injury Prevention* 2000; **6**: 46–50.
25. **Tilling K.** Capture-recapture methods-useful or misleading? *International Journal of Epidemiology* 2001; **30**: 12–14.
26. **Regal RR, Hook EB.** Marginal versus conditional versus structural source models: a rationale for an alternative to log-linear methods for capture-recapture estimates. *Statistics in Medicine* 1998; **17**: 69–74.
27. **Wilton RM, Collins MF.** Capture-recapture estimation with samples of size one using frequency data. *Biometrika* 1992; **79**: 543–553.
28. **Hook EB, Regal RR.** Validity of Bernoulli census, log-linear and truncated binomial models for correcting for underestimates in prevalence studies. *American Journal of Epidemiology* 1982; **116**: 168–176.
29. **Smit F, Reinking D, Reijerse M.** Estimating the number of people eligible for health service use. *Evaluation and Program Planning* 2002; **25**: 101–105.
30. **Smit F, Toet J, van der Heijden P.** Estimating the number of opiate users in Rotterdam using statistical models for incomplete count data. In: Hay G, McKeganey N, Birks E, eds. *Final Report EMCDDA Project Methodological Pilot Study of Local Level Prevalence Estimates*. Lisbon: EMCDDA, 1997, pp. 47–66.
31. **Bohning D, et al.** Estimating the number of drug users in Bangkok 2001: a capture-recapture approach using repeated entries in one list. *European Journal of Epidemiology* 2004; **19**: 1075–1083.
32. **Hay G, Smit F.** Estimating the number of hard drug users from needle-exchange data. *Addiction Research and Theory* 2003; **11**: 235–243.
33. **Chao A.** Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 1987; **43**: 783–791.
34. **Chao A.** Estimating animal abundance with capture frequency data. *Journal of Wildlife Management* 1988; **52**: 295–300.
35. **Chao A.** Estimating population size for sparse data in capture-recapture experiments. *Biometrics* 1989; **45**: 427–438.
36. **Seber GA, Huakau JT, Simmons D.** Capture-recapture, epidemiology, and list mismatches: two lists. *Biometrics* 2000; **56**: 1227–1232.