volume 16 no 8 pp 1019–1041 august 2011

Systematic Review

Quality assessment of capture–recapture studies in resource-limited countries

Rob van Hest[1], Andrew Grant[2] and Ibrahim Abubakar[3]

1 *Tuberculosis Control Section, Rotterdam Public Health Service, Rotterdam, The Netherlands*
2 *Statistics, Modelling and Bioinformatics Department, Health Protection Agency, London, UK*
3 *Tuberculosis Section, Health Protection Agency, London, UK*

Summary

objectives Resource-limited countries often lack robust routine surveillance systems to accurately assess the burden of human attributes and diseases. In these settings capture-recapture analysis can be an alternative tool to obtain prevalence and incidence rates. Performance of capture-recapture analyses in resource-limited countries has not been systematically reviewed.

methods Systematic review of the performance of capture–recapture analyses in the categories of human attributes, non-infectious and infectious diseases in resource-limited countries, assessing individual study quality criteria and a minimum quality criterion per category, using PRISMA methodology.

results A total of 1671 potentially relevant PubMed citations were screened, resulting in 52 eligible publications: 36% in human attributes, i.e. hidden populations, injuries and mortality; 48% in non-infectious and 15% in infectious disease categories. Twenty-one per cent of selected studies were from low income countries, 40% from lower-middle-income countries and 38% from upper-middle-income countries. Thirteen per cent achieved good individual study quality criteria, 25% were intermediate and 19% were poor. Of the good studies, six were performed on human attributes and one on a non-infectious disease. The proportions of publications meeting the minimum quality criterion per category were 42%, 20% and 37%, respectively.

conclusions Few capture–recapture studies in resource-limited countries achieved good individual quality criteria and a minority met the minimum quality criterion per category. Capture–recapture techniques in these settings should be carefully considered and implemented rigorously and are not a panacea for strengthening of routine surveillance systems.

keywords resource-limited, systematic review, capture–recapture methods

Introduction

Meaningful quantification and description of the distribution of human attributes and disease within a community is an essential part of any control programme. Both in resource-rich and in resource-limited settings, surveillance should be a driver for identification of priorities and implementation and evaluation of interventions. Observing and monitoring health and behaviour trends requires a surveillance system that captures useful data on those persons correctly identified with the characteristic under study. For epidemiologists globally, prevalence and incidence rates are fundamental components of their discipline but these data often are inaccurate because of under-ascertainment of the true number of persons or events and false-positive cases. Standardised means to evaluate and adjust prevalence and incidence rates for the degree of under-ascertainment enable more accurate surveillance but are rarely used (Hook & Regal 1992; van Hest 2007).

The number of individuals with a certain condition or disease in a population can be ascertained directly, by counting every single person as attempted in a census, or it can be estimated exhaustively through active community-based finding or a sample survey (Glaziou *et al.* 2008). Notification or registration (passive case-finding) tends to yield an under-estimate of the burden. A statistical technique that estimates completeness of ascertainment of censuses, surveys and registers is capture–recapture analysis, first used in animal population biology by ecologists to adjust wildlife population estimates for undercount, and used later in epidemiology (Wittes & Sidel 1968; Hook & Regal 1995; International Working Group for Disease Monitoring and Forecasting 1995a,b). After record-linkage, i.e. comparing data across multiple registers containing information on persons with the condition or disease under study, and based on certain assumptions, capture–recapture methods use data from partially overlapping and possibly incomplete registers to estimate the number of cases unknown to all registers and thus the estimated total number of cases. The preferred capture–recapture method entails log-linear modelling of at least three linked registers, less compromised by almost unavoidable violation of the underlying assumptions than capture–recapture analysis based on two linked registers. A synopsis of the methodology and mathematical framework, global applications and publications can be found elsewhere (Hook & Regal 1995; International Working Group for Disease Monitoring and Forecasting 1995a,b; van Hest 2007).

Capture–recapture analysis was promoted 15–20 years ago as a cheap, quick and simple alternative tool to obtain reliable prevalence and incidence rates, also in resource-limited countries, although limitations were immediately described (Chiu *et al.*1993; Black *et al.* 1994; Desenclos & Hubert 1994; Gutteridge & Collin 1994; Laporte 1994; Waters 1994; Hook & Regal 1995, International Working Group for Disease Monitoring and Forecasting 1995a,b; Papoz *et al.* 1996). In resource-rich countries, capture–recapture analysis has been employed widely to estimate the size of hidden groups or events or the burden of diseases. However, a limited number of capture–recapture studies have been published from resource-limited countries, despite the burden of human attributes and disease, the lack of robust conventional monitoring and surveillance systems, prominent private (and some specific public) health care provision (with poor referral and notification to the national authorities) and methodology-attributed advantages (Chiu *et al.*1993; International Working Group for Disease Monitoring and Forecasting 1995b; Masjedi *et al.* 2007). Capture–recapture studies in resource-limited countries predominantly involve diseases but only a few address, in these settings often highly endemic, infectious and tropical diseases. In contrast, in resource-rich countries a multitude of capture–recapture studies concerning less endemic or imported infectious and tropical diseases have been performed, for example on HIV/AIDS, malaria or tuberculosis (Bernillon *et al.* 2000; van Hest *et al.* 2002, 2007; Iňigo *et al.* 2003; Pezzoti *et al.* 2003; Cathcart *et al.* 2009). Some aspects of capture–recapture analysis in resource-limited setttings have been discussed recently, in the context of WHO quantitative implementation and impact targets and United Nations' Millennium Development Goals (Bassili *et al.* 2010).

The aim of this study was to conducted a systematic review of the performance of capture–recapture analyses in the categories of human attributes, i.e. hidden populations, injuries and mortality, and non-infectious and infectious diseases, in resource-limited countries, assessing individual study quality criteria and a minimum quality criterion per category, using PRISMA methodology (Liberati *et al.* 2009).

Methods

Eligibility criteria

We only included studies in low-income and lower- and upper-middle-income countries as defined by the World Bank in 2008 (World Bank 2009). Although no general language restrictions were applied, studies with the full text of the paper in Chinese were excluded. Studies not published in PubMed/MEDLINE indexed journals, poster abstracts, academic theses or reports were also excluded (Woodward 1990). One recent capture–recapture study was not considered because of the involvement of the first author (Cojocaru *et al.* 2009).

Search strategy and selection process

To identify relevant studies, we conducted a literature search in the bibliographic database PubMed/MEDLINE from January 1990 to January 2010 using the keyword 'recapture'. The search was limited to capture–recapture studies conducted on human attributes or diseases involving humans. Potentially relevant studies were selected through examination of initially the title and subsequently the abstracts by two independent reviewers (R.v.H. and I.A.). When identified as possibly eligible, the full text of the article was obtained and reviewed by one of the researchers (R.v.H). References of the full-text articles were checked for additional studies.

Data extraction

The eligible studies were divided into three categories: (i) human attributes, i.e. hidden populations, injuries and mortality, (ii) non-infectious diseases and (iii) infectious diseases, although public use and need for rigour can be different. Characteristics of interest were the capture–recapture model used and seven recommended criteria for capture–recapture studies, i.e. availability of sufficient existing registers, discussion of the quality of record-linkage, sufficient capture of events or cases in each register

(≥15% of the total number of cases identified after record-linkage considered sufficient), sufficient overlap of the linked registers (≥15% considered sufficient), discussion of the underlying capture–recapture assumptions in the context of the study (yes: ≥4 assumptions, partly: 1–3 assumptions, no: assumptions not discussed), comparison to other studies estimating the number of cases or events in the same population under study and involvement of local researchers, i.e. from the low-income country, and also publication delay since study period. Data were extracted by one of the researchers (R.v.H.).

Data analysis

As a quality score per individual study, the capture–recapture studies were assessed for completeness of addressing assumptions and feasibility: good (defined as addressing ≥4 capture–recapture assumptions (Table 1) and ≥3 of the first four recommended criteria mentioned in Table 3), partly (defined as addressing 1–3 capture–recapture assumptions and ≥2 of the four recommended criteria) or not (defined as addressing none of the capture–recapture assumptions and <2 of the four recommended criteria).

The proportion of capture–recapture studies in resource-limited settings using the two-source model and fulfilling each of the seven recommended criteria for capture–recapture analysis was calculated, divided over the three categories. The proportion of capture–recapture studies in resource-limited settings meeting from 0 to 7 of the recommended criteria for capture–recapture studies was calculated, for all studies, studies in low-income and lower-middle-income countries, and studies in upper-middle-income countries. The minimum quality criterion per category was defined as meeting ≥4 recommended criteria.

Table 1 Capture–recapture assumptions and recommendations

Perfect record-linkage (i.e. no erroneous misclassification of records)
Closed population (i.e. no immigration or emigration in the time period studied)
Homogeneous population (i.e. no subgroups with markedly different probabilities of being observed and re-observed)
Independent registers (i.e. the probability of being observed in one register is not affected by being – positive dependence- (or not being – negative dependence-) observed in another)
Registers should not include false-positive records, i.e. the specificity and positive predictive value of the registers should ideally be 100%
Sources selected should have sufficient overlap and not be complementary or mutually exclusive
Individuals under study should be captured within the time and space defined by the investigation

Results

The search strategy resulted in 1671 citations, reduced to 87 potentially relevant studies after title inspection. Eight manuscripts were excluded because of full-text Chinese language. Of the remaining 79 articles after abstract inspection, or when in doubt after full-text inspection, three were found to be animal related, nine were not performed in resource-limited countries (French Guyana and La Reunion are considered French regions) (Aptel *et al.*1999; Guernier *et al.* 2006) and 18 appeared not to be capture–recapture studies. One article claimed to perform capture–recapture analysis but actually only performed record-linkage (Thompson-Cerna & Medina 2007). In addition to the 49 studies identified through the search, three relevant manuscripts were found in the references of these articles, bringing the total number of included studies to 52. The article selection process is summarised in Figure 1, and an overview of the identified capture–recapture studies in resource-limited countries is shown in Table 2A–C, with country, income category, year of publication, subject, first author, objectives, methodology, data sources and study outcomes.

In the hidden populations, injuries and mortality category there were 19 (36%) studies; in non-infectious 25 (48%) and in infectious disease categories, 8 (15%). The results of the data extraction per study in each of the three categories are shown in Table 3A–C, respectively. Eleven studies were performed in low-income countries (21%), 21 in lower-middle-income countries (40%) and 20 in upper-middle-income countries (38%).

Of the 7 (13%) best scoring studies (good) (Razzak & Luby 1998; Kruse *et al.* 2003; Tercero & Anderson 2004; Van *et al.* 2006; Chiang *et al.* 2007; Zavareh *et al.* 2008; Azevedo-Silva *et al.* 2009), six were performed on hidden populations, injuries and mortality (30%), all in low-income or lower-middle-income countries, and only one recent study assessed childhood acute leukaemia in Brazil (4%). The 13 (25%) intermediate (partly) scoring studies (Xia *et al.* 1995; Diallo *et al.* 1996; Ramachandran *et al.* 1996; Tull *et al.* 1998; Yang *et al.* 1998; Simondon & Khodja 1999; Chen *et al.* 2004; Khan *et al.* 2004; Platt *et al.* 2004; Luan *et al.* 2005; Mingoti *et al.* 2006; Cristiano *et al.* 2009; Odega *et al.* 2010) were proportionally equally divided over the three categories (around 25%). The proportions of best or intermediate scoring studies in the three categories were 58%, 28% and 25%, respectively. The proportions of the 10 (19%) poorest scoring studies (Sanghavi *et al.* 1998; Spichler *et al.* 2001; Martinucci *et al.* 2002; Sánchez *et al.* 2002, 2004; Tuchinda *et al.* 2002; Vázquez *et al.* 2008; Ronsmans *et al.* 2009; Sivamani *et al.* 2009; Negrato *et al.* 2010)

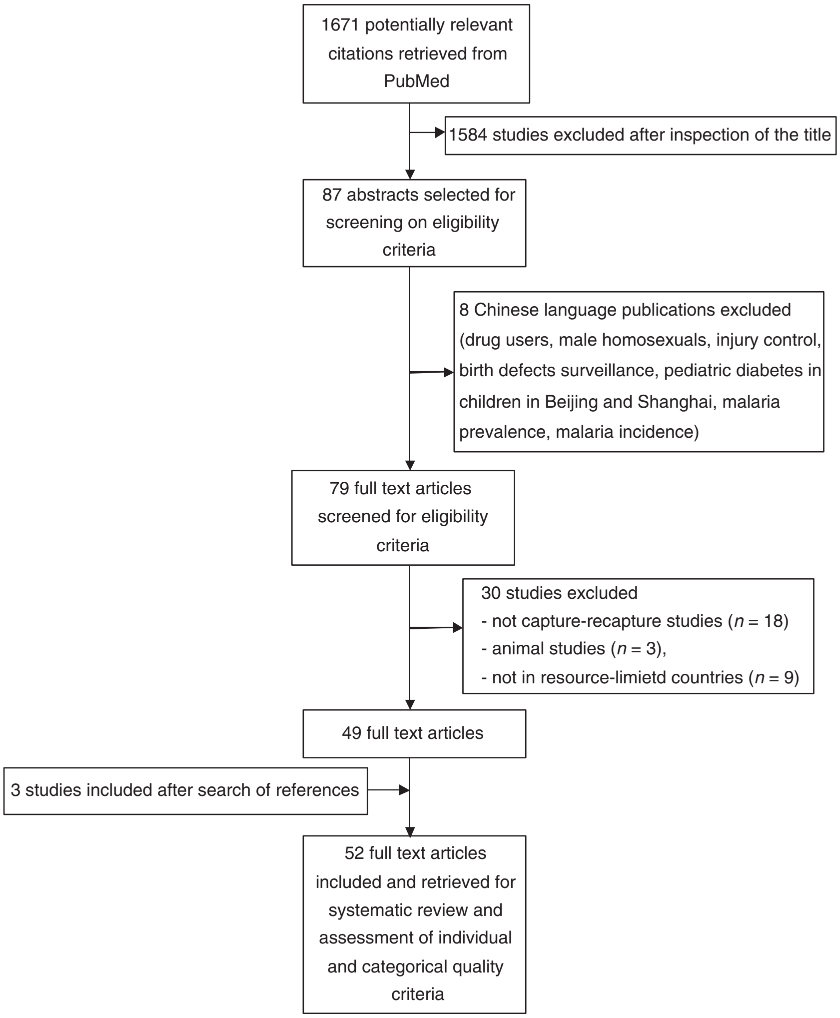

Figure 1 Flow diagram of selection process of publications included in the systematic review.

in the three categories were 11%, 28% and 13%, respectively. The proportions of indeterminate studies, i.e. not fitting the scoring system, were 32%, 44% and 63%, respectively.

In Tables 4, 5 and 6, the proportions of the 52 capture–recapture studies in the review meeting the characteristic under study are shown per category. Only 10 studies (19%) used a three-source capture–recapture model and with the exemption of one study applying a single-source model and one study using a non-defined model, others (n = 40) relied on a two-source model. Thirteen of these studies did not discuss the limitations of capture–recapture analysis. Three-source models were applied in one quarter of the disease-related studies. Just over half of the disease-related studies used existing record sources but almost all studies estimating hidden populations needed alternatives to create databases. In the non-disease-related studies, record-linkage received sufficient attention. Especially in the non-infectious disease category, good record-linkage was rarely discussed or constraints mentioned. Approximately two-thirds of all studies in all categories had sufficient capture by each of the registers used. A minority of the studies in all categories, especially infectious disease studies, had sufficient overlap between the various registers used. Only in the non-disease-related category, over half of the studies comment on more than four assumptions. Of the disease-related studies, a vast majority do not or only partly address possible limitations. Approximately a third of the studies in all categories compared capture–recapture outcomes to previous relevant results or consisted of only local researchers.

The proportions of all publications meeting more than half of the seven recommendations for capture–recapture studies in the three categories were 42%, 20% and 37%, respectively (Table 7). Stratified for low-income/lower-

Table 2 Published capture–recapture studies on (A) hidden populations, injuries and mortality, (B) non-infectious diseases, (C) infectious diseases in resource-limited countries

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
A							
Thailand (LMI)	1994	HIV-positive injection drug users	Mastro *et al.*	To estimate the number of injection drug users infected with HIV in Bangkok, Thailand, in 1991	Two-source CRC model	Methadone treatment records Police urine test records	Estimated number of HIV-infected injection drug users was 12 000
Burkina Fasso (LI)	1996	Child mortality	Diallo *et al.*	To estimate completeness of child mortality registration in a rural area of Burkina Fasso, 1994	Two-source CRC model	Census Community informants	Estimated sensitivities of census and community informants were 97% and 76%, respectively
Pakistan (LMI)	1998	Road traffic accidents	Razzak & Luby	To estimate death and injury rates because of road traffic accidents in Karachi, Pakistan, in 1994	Two-source CRC model	Police records Ambulance records	Estimated injury and death rates were 185 and 11.2/100 000, respectively. Official sources counted only 56% of deaths and 4% of serious injuries
Brazil (UMI)	2003	Injecting drug users	Caiaffa *et al.*	To estimate the number of street injecting drug users and those infected with HIV and hepatitis C virus, in a syringe exchange programme in Porto Alegre, 1998	Two-source CRC model	Outreach interview 1 Outreach interview 2	Estimated number of injecting drug users was 317 (95% CI 235–467), of which 151 and 168 were HIV and hepatitis C virus infected, respectively
Madagascar (LI)	2003	Sex workers	Kruse *et al.*	To estimate the size of a mobile sex-worker population in Diego-Suarez, Madagascar, in 2000	Two-source CRC model	Cross-sectional survey 1 Cross-sectional survey 2	Estimated number of sex workers was 2684 (95% CI 2588–2780)
Thailand (LMI)	2004	Drug users	Böhning *et al.*	To estimate the number of drug users in Bangkok in 2001	Single-list truncated CRC model	Drug treatment surveillance records of Narcotics Control Board	Estimated heroin and methamphetamine users was 11 296 (95% CI 8964–13 628) and 32 105 (95% CI 24 647–39 563) and register comple-teness 38.4% and 9.4%, respectively
Brazil (UMI)	2004	Street children	Gurgel *et al.*	To estimate the number of street children in Aracaju, north-east Brazil	Three-source log-linear CRC model	NGO records Crosssectional survey 1 Crosssectional survey 2	Observed and estimated street children were 526 and 1456, respectively
Bangladesh (LI)	2004	Mobile male sex workers	Khan *et al.*	To estimate the number of mobile male sex workers in a city of Bangladesh, 2000	Two-source CRC model	Field survey 1 Field survey 2	Observed and estimated sex workers was 230 and 248, respectively

Table 2 (Continued)

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
Vietnam (LI)	2004	Sex workers	Minh *et al.*	To assess the number of sex workers in Na Trang City, Vietnam, in 2000	Two-source CRC model	NGO records Survey 1 Survey 2	Estimated number of sex workers was 930 (95% CI 891–969)
Russian Federation (UMI)	2004	Injecting drug users (IDU's)	Platt *et al.*	To estimate prevalence of IDU's in Togliatti City, Russian Federation, 2001	Three-source covatiate log-linear CRC model	Narcology records HIV positive test records Police arrest records	Estimated number of IDU's was 20 226 (95% CI 16 971–24 749), translating in a prevalence of (2.7%) 5.4% in the (registered) population
Nicaragua (LMI)	2004	Transport injuries	Tercero & Andersson	To estimate transport-related injury incidence in Leon municipality, Nicaragua, in 1993	Two-source CRC model	Hospital records Traffic police records	Estimated coverage of police and hospital records was in non-fatal cases 2.6 and 19.0%, respectively, and in fatal cases 56.1 and 22.8%, respectively
China (LMI)	2005	Men having sex with men (MSM)	Luan *et al.*	To estimate the population size of MSM, Chendu city, China, unknown year	Two-source CRC model	Survey 1 with tag 1 Survey 2 with tag 2 Survey 3	Estimated MSM number was 1408 (95% CI 1116–1908), 1207 (95% CI 932–1712) and 949 (95% CI 757–1272)
Brasil (UMI)	2006	Injecting drug users (IDU's)	Mingoti *et al.*	To estimate the number of IDU's attending a syringe exchange programme in three cities in Brasil, 2000–2001	Two-source CRC model	Outreach interview 1 Outreach interview 2	Observed and estimated number of IDU's were 624 and 1020, respectively
Vietnam (LI)	2006	Non-fatal road traffic accidents	Van *et al.*	To estimate non-fatal road traffic injuries in Thai Nguyen City in 2000–2004	Two-source CRC model	Hospital records Police records	Estimated number of non-fatal road traffic incidents and register completeness were 11 140 (95% CI 10 626–11 654) and between 21.9–60.1%
Taiwan (LMI)	2007	Heroin and methampheta-mine male users (15–54 years)	Chiang *et al.*	To estimate the numbers of male heroin and methamphetamine users in Taoyuan County, northern Taiwan, 1999–2002	Two-source CRC model	Medical records Judiciary records	Observed numbers in the years under study were 2809, 2486, 1661 and 1440 drug users, respectively. Estimated numbers were 16 192, 14 532, 16 844 and 11 783 drug users, respectively
Kenya (LI)	2007	Sex workers	Geibel *et al.*	To estimate the number of men who have sex with men (MSM) in and around Mombassa, Kenya, in 200	Two-source CRC model	Survey of MSM locations 1 Survey of MSM locations 2	Estimated number of was 739

Table 2 (Continued)

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
Iran (LMI)	2008	Fatal road traffic injuries	Zavareh *et al.*	To estimate the number of fatal road traffic injuries in the est Azarbaijan Province, Iran, 2004–2005	Two-source CRC model	Death register Forensic medicine register	The observed number the number of fatal road traffic injuries was 897 and the estimated number was 1018, or a rate of 34/100 000 population
India (LMI)	2009	Childhood injuries	Sivamani *et al.*	To validate passive and active childhood injuries surveillance systems in Tamilnadu India, 2001	Two-source CRC model	Passive surveillance system Cross-sectional survey	Estimated passive and active childhood injury rates were 13,6 (CI 11.7–15.3) and 341.9 (CI 254.5–429.3)/1000 child-years, respectively
Indonesia (LMI)	2009	Maternal death	Ronsmans *et al.*	To estimate the number and rate of maternal deaths in two districts in West Java, Indonesia, 2004–2005	Two-source CRC model	Two informant networks	Estimated overall maternal mortality ratio was 435 (95% CI 376–498)/100 000 live births
B							
Sudan (LMI)	1992	Childhood type 1 diabetes	Elamin *et al.*	To determine incidence of type 1 diabetes in children in Khartoum, Sudan, 1987–1990	Two-source CRC model	Hospital-based register Second source	Estimated number of cases was 327 and completeness of sources and case-ascertainment 73%, 82% and 95%, respectively
China (LMI)	1994	Type 1 diabetes (0–14 years)	Fu *et al.*	To determine incidence of type 1 diabetes in children in Shanghai, China, 1980–1991	Two-source CRC model	Pediatric hospital departments Primary and middle schools	Crude annual incidence rate was 0.61/100 000 (95% CI 0.48–0.77). Ascertainment corected annual incidence rate was 0.72/100 000 (95% CI 0.57–0.91)
China (LMI)	1995	Leukopenia after benzene exposure	Xia *et al.*	To assess risk of benzene-exposure in small-scale industires in Shanghai, China, unknown year	Two-source CRC model	Cross-sectional investigation 1 Cross-sectional investigation 2	Estimated relative risk of leuk openia after benzene-exposure was 2.9
India (LMI)	1996	Type 1 diabetes (0–14 years)	Ramachandran *et al.*	To estimate incidence of type 1 diabetes in children 0–14 years of age in Madras, India, 1991–1994	Two-source CRC model	Hospital records Diabetes camp and private diabetologist records	Estimated overall incidence of type 1 diabetes was 10.5 cases per 100 000 children years
China (LMI)	1996	Type 1 diabetes (0–14 years)	Shen *et al.*	To estimates number of children with type 1 diabetes in urban Shanghai, China, 1989–1993	Two-source CRC model	Hospital records School records	Observed and estimated cases of type 1 diabetes were 58 and 67, respectively

Table 2 (Continued)

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
India (LMI)	1998	Childhood epilepsy	Pal *et al.*	To ascertain sensitivity, efficacy and costs of key informants and survey methods of ascertainment of childhood epilepsy in west Bengal, India, 1995	Two-source CRC model	House-to-house survey Key informants	The survey had an absolute sensitivity of 59% but was four times as sensitive as the key informants
Dominica, West Indies (UMI)	1998	Diabetes mellitus	Tull *et al.*	To evaluate the utility of capture–recapture analysis for monitoring the frequency of diabetes mellitus in Dominica, West Indies, in 1995	Three-source log-linear CRC	Dominica diabetic association records Regional health clinic records Hospital diabetes clinic records	Case-ascertainment after record-linkage represented 72.3% of the estimated prevalence of diagnosed diabetes mellitus
China (LMI)	1998	Type 1 diabetes (0–14 years)	Yang *et al.*	To estimate incidence of type 1 diabetes in children 0–14 years of age in China, 1985–1994	Two-source CRC model	Medical records School health, family planning, insurance, education, child care or anti-epidemic records	Overall incidence of type 1 diabetes was estimated at 0.51 cases per 100 000 children year
Brazil (UMI)	1999	Primary cerebral neoplasms in children 1–14 years	Argollo & Lessa	To estimate prevalence and fatality rate of primary cerebral neoplasms in children 1–14 years in Bahia State, Brazil, 1995.	Two-source CRC model	Neuro-imaging and pathology records Pediatric oncology and radiotherapy records	Estimated prevallence was 1.85/100 000 (95% CI 1.4–2.2) in Bahia State and 10.1/100 000 (95% CI 7.6–12.8) in the capital
Benin (LI)	2000	Epilepsy patients	Debrock *et al.*	To estimate prevalence of epilepsy in two villages in Benin in April and May 1997	Three-source log-linear CRC model	Cross-sectional study Key informants information Health centre records	Epilepsy prevalence rate increased from 15.9/1000 to 21.1/1000 and to 35.1/1000 after record-linkage and CRC, respectively
Colombia (UMI)	2000	Multiple sclerosis	Sánchez *et al.*	To determine prevalence of multiple sclerosis in five provinces of Columbia, 1995–2000	Two-source CRC model	Different combinations of sources in different provinces	Estimated multiple sclerosis prevalence varied between 1.48 and 4.98/100 000 population
South Africa (UMI)	2001	Non-communi cable disease (NCD)	Gill *et al.*	To enumerate NCD patients in a rural district of Kwazulu Natal, South Africa (unkown period)	Two-source CRC model	Hospital records Peripheral clinic records	Estimates of asthma and epilepsy patients failed because of a lack of register overlap and for hypertension and diabetes poor overlap gave high estimates with wide 95% CI's

Table 2 (Continued)

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
Brazil (UMI)	2001	Lower extremity amputation	Spichler *et al.*	To estimate rates of lower extremity amputations in Rio de Janeiro, Brazil, 1992–1994	Three-source log-linear CRC model	Hospital records Limb fitting records Rehabilitation records	Estimated number of amputations was 3954, translating into an estimated rate of 13.9/100 000 compared to the routine surveillance rate of 5.4/100 000
Belarus (UMI)	2002	Type 1 diabetes (0–14 years)	Martinucci *et al.*	To study the incidence of childhood type 1 diabetes in Gomel region, Belarus, 1976–1999	Two-source CRC model	Regional hospital records Pharmaceutical dispensary records	The average crude incidence was 4.6/100 000 but in the years after the Chernobyl disaster 5.7/ 100,000, with most increase among 10–14 years old
Colombia (UMI)	2002	Myastenia gravis	Sánchez *et al.*	To determine prevalence of myastheia gravis among the inhabitants of Antioquia, Columbia, 1995–2000	Two-source CRC model	Neurology hospital records University hospital records	Estimated prevalenve was 27.7 cases per million inhabitants (95% CI 23.2–32.2)
Thailand (LMI)	2002	Type 1 diabetes (0–14 years)	Tuchinda *et al.*	To estimate the incidence of type 1 diabetes in children 0–14 years of age in Bangkok-Noi, 1991–1995	Unspecified CRC model	Hospital records Pediatrician and endocrinologist records School records	Estimated overall incidence of type 1 diabetes was 1.65 cases per 100 000 children per year
Kenya (LI)	2003	Demographic surveillance system for <5 years children	Eisele *et al.*	To evaluate completeness of registration of vital events in children (<5 years) in western Kenya, in 2000 and 2001	Two-source CRC model	Demographic surveillance system Household survey	Estimated completeness was 62% and 49% for neonatal deaths, 72% and 78% for post-neonatal child deaths and 88% and 78% for newborns
Taiwan (LMI)	2004	Infantile spasms	Chen *et al.*	To estimate the prevalence, incidence and case-fatality of infantile spasms in Taiwan, 1985–1997	Two-source CRC model	National university hospital records National health insurance records	Estimated infantile spasms prevalence and icidence was 4.6 and 6/100 000. Case-fatality rate was 11,6%
Brazil (UMI)	2004	Diabetes mellitus	Coeli *et al.*	To estimate completeness of a surveillance system for diabetes in the elderly, 1994–1995	Three-source log-linear CRC model	Hospital morbidity register Out-patient morbidity register Mortality register	Observed and estimated number of patients were 740 and (implausible) 22 925 (95% CI 11 354–57 269) patients.
Colombia (UMI)	2004	Parkinson's disease	Sánchez *et al.*	To estimate prevalence of Parkinson's disease and parkinsonism in Antioquia, Colombia, 1996–2000	Two-source CRC model	University hospital records Neurology hospital records	Estimated prevalence of Parkinson's disease and parkinsonism was 30 (95% CI 29.2–32.2) and 42.1 (95% CI 40.3–43.8), respectively

Table 2 (Continued)

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
Nigeria (LMI)	2006	Diabetic foot ulceration	Ogbera *et al.*	To estimate the prevalence of diabetic foot ulceration in a university hospital in Lagos, Nigeria, 1998–2000	Two-source CRC model	Ward records Medical records Operating theatre records	The prevalence of diabetic foot ulceration among all diabetes mellitus patients was estimated at 9.5%
Uruguay (LMI)	2008	Amyotrophic lateral sclerosis	Vázquez *et al.*	To determine incidence and prevalence of amyotrophic lateral sclerosis in Uruguay, 2002–2003	Three-source log-linear CRC model	Hospital and medical records Neurophysiology records Death certificates	Observed and estimated incidence for amyotrophic lateral sclerosis was 1.37/100 000 and 1.42/100 000 population respectively. Crude prevalence was 1.9/100 000
Argentina (UMI)	2009	Multiple sclerosis	Cristiano *et al.*	To determine multiple sclerosis prevalence in Greater Buenos Aires Metropolitan area, Argentina, in 1996	Three-source CRC model	Multiple sclerosis association records Hospital records 1 Hospital records 2 Hospital records 3	Estimated number of multiple sclerosis cases was between 1833 and 2359 and estimated prevalence rate 14–19.8/100 000
Brasil (UMI)	2009	Type 1 diabetes (0–14 years)	Negrato *et al.*	To assess the long-term trends in the incidence of type 1 diabetes in Bauru, Brazil, 1986–2006	Two-source CRC model	Clinical records Diabetes association, school, hospital and health unit records	Observed and estimated incidence increased from 2.8/100 000 to 18.5/100 000 and 2.8/100 000 to 27.2/100 000 in 2002
Brazil (UMI)	2009	Childhood acute leukemia	Azevedo-Silva *et al.*	To analyse the number of cases and incidence of childhood acute lymphoblastic leukemia in three Brazilian cities: Salvador, Recife and Belo Horizonte, in 2001	Two-source CRC model	Population-based registry Reference laboratory database	Estimated incidence of childhood acute lymphoblastic leukemia was 5.8, 6.3 and 5.5/100 000 inhabitants for Salvador, Recife and Belo Horizonte, respectively. Estimated completeness of population-based registies was 11.5, 35.4 and 29.2%, respectively
C							
Peru (UMI)	1998	Tuberculosis (TB)	Sanghavi *et al.*	To estimate pulmonary TB incidence in Las Pampas de San Juan de Miraflores, Peru, 1989–1993	Two-source CRC model	Survey records Laboratory smear records	Estimated incidence of tuberculosis was 364 (95% CI 293–528), translating in 63.2% under-reporting
Senegal (LI)	1999	Pertussis vaccine efficacy	Simondon & Khoda	To estimate relative vaccine efficacy in the Senegal pertussis trial, 1990–1994	Three-source log-linear CRC model	Bacteriology records Serology records Epilink records	Relative vaccine efficacy was estimated at 1.50 (95% CI 1.24–1.82)
Nigeria (LMI)	2001	Disabled leprosy patients	Van den Broek *et al.*	To estimate unknown disabled leprosy patients in four States in northern Nigeria, 1997–1998	Two-source CRC model	Hospital admission records Disabled patients sample survey	Estimated disabled leprosy patients were 1262 (95% CI 991–1553) and completeness of the survey at 31%

Table 2 (Continued)

Country (income)*	Year†	Subject	Researchers	Objective	Method	Data-source	Outcome
Argentina (UMI)	2001	Leishmaniasis	Yadon *et al.*	To estimate completeness of cutaneous leishmaniasis surveillance in Santiago del Estero province, Argentine, 1990–1993	Two-source CRC model	National leishmaniasis surveillance system Hospital records and case-control study records	Estimated number of leishmaniasis patients was 210 (95% CI 202–218) and completeness of reporting 44.8% (95% CI 43.2–46.4%)
Brazil (UMI)	2005	Human leptospirosis	Brum & Kupek	To estimate cases of human leptospirosis in the district of Santa Maria in southern Brazil in 2001–2002	Three-source log-linear CRC model	Laboratory records Hospital records Surveillance records	Record-linkage identified >20 times leptospirosis cases than officially reported. CRC showed only few cases were missed by all sources
Brazil (UMI)	2007	Visceral leishmaniasis cases	Maia-Elkhoury *et al.*	To estimate coverage of national information systems for visceral leishmaniasis in Brazil, 2002–2003	Two-source CRC model	Notification records Mortality records Hospital records	Estimated completeness of the notification system 57.8% and 55%, respectively
Cameroon (LI)	2009	Buruli ulcer cases	Porten *et al.*	To estimate Buruli Ulcer prevalence in Akonoloinga District, Cameroon, 2007	Two-source CRC model	Central location screening Active case finding	Estimated sensitivity of active case-finding was at 84%
Nigeria (LI)	2010	Measles	Odega *et al.*	To determine completeness of measles reporting in a State in Southern Nigera, 2007–2008	Two-source CRC model	Health facilities register State surveilance records	Estimated overall completeness of reporting was 11.5% (95% CI 8.1–19.6%)

CRC, capture–recapture; CI, confidence interval; HIV, human immunodeficiency virus.
*LI, low-income; LMI, lower-middle income; UMI, upper-middle income.
†Year, year of publication.

Table 3 Consideration of key assumptions and recommendations in published capture–recapture studies on (A) hidden populations, injuries and mortality, (B) non-infectious diseases, (C) infectious diseases in resource-limited countries

Author, subject, country	Year*	Model	Sufficient registers	Good record-linkage†	Sufficient capture	Sufficient overlap	Limitations discussed‡	Comparison to other studies§	Local researchers	Publication delay (years)
A										
Mastro *et al.*, HIV-positive injection drug users, Thailand	1994	Two-source CRC model	No	Yes	No	No	Yes	Yes	No	3
Diallo *et al.*, child mortality Burkina Fasso	1996	Two-source CRC model	No	Not discussed	Yes	Yes	Partly	Yes	No	2
Razzak *et al.*, road traffic accidents, Pakistan	1998	Two-source CRC model	Yes	No (L,D)	Yes	Yes (deaths) No (injuries)	Yes	No	No	4
Caiaffa *et al.*, injecting drug users, Brazil	2003	Two-source CRC model	No	Yes	Yes	No	Yes	Yes	Yes	5
Kruse *et a.l*, sex workers, Madagascar	2003	Two-source CRC model	No	Yes (tagging)	Yes	Yes	Yes	No	No	3
Böhning *et al.*, drug users, Thailand	2004	Single list CRC model	Not applicable	Yes	Not applicable	Not applicable	Yes	No	No	3
Gurgel *et al.*, street children, Brazil	2004	Three-source CRC model	No	Not discussed	Yes	No	Partly	No	No	Unknown
Khan *et al.*, mobile male sex workers, Bangladesh	2004	Two-source CRC model	No	Yes (tagging)	Yes	Yes	Partly	Yes	Yes	4
Minh *et al.*, sex workers, Vietnam	2004	Two-source CRC model	No	Yes	Not disclosed	Yes	Yes	Yes	No	4
Platt *et al.*, injecting drug users, Russian Federation	2004	Three-source covariate model	Yes	Not discussed	Yes	Partly	Partly	Yes	No	3
Tercero *et al.*, transport injuries, Nicaragua	2004	Two-source CRC model	Yes	Yes	Yes (Fatal) Partly (Non-fatal)	Yes (Fatal) No (Non-fatal)	Yes	Yes	No	11
Luan *et al.*, men having sex with men, China	2005	Two-source CRC model	No	Yes (tagging)	Yes	No	Partly	No	Yes	Unknown
Mingoti *et al.*, injecting drug users, Brasil	2006	Two-source CRC model	No	Yes	Yes	Yes	Partly	Yes	Yes	5–6
Van *et al.*, non-fatal road traffic accidents, Vietnam	2006	Two-source CRC model	Yes	Exact questionnable (L,D) Relaxed good	Yes	No	Yes	No	No	2–6
Chiang *et al.*, heroin and methamphetamine male users (15–54 years), Taiwan	2007	Two-source CRC model	Yes	Not discussed	Partly	No	Partly	No	Yes	5–8
Geibel *et al.*, sex workers, Kenya	2007	Two-source CRC model	No	Yes (tagging)	Yes	Yes	Yes	No	No	1
Zavareh *et al.*, fatal road traffic injuries, Iran	2008	Two-source CRC model	Yes	Yes	Yes	Yes	Yes	No	No	3–4

Table 3 (Continued)

Author, subject, country	Year*	Model	Sufficient registers	Good record-linkage†	Sufficient capture	Sufficient overlap	Limitations discussed‡	Comparison to other studies§	Local researchers	Publication delay (years)
Sivamani *et al.*, childhood injuries, India	2009	Two-source CRC model	No	Yes	No	No	No	No	Yes	8
Ronsmans *et al.*, maternal death, Indonesia	2009	Two-source CRC model	No	Not discussed	Not discussed	Not discussed	No	No	No	4–5
B										
Elamin *et al.*, childhood type 1 diabetes, Sudan	1992	Two-source CRC model	Yes	Not discussed	Yes	Yes	No	Yes	No	2–5
Fu *et al.l*, childhood type 1 diabetes, China	1994	Two-source CRC model	No	Not discussed	Yes	Yes	No	No	No	3–14
Xia *et al.*, leukopenia after benzene exposure, China	1995	Two-source CRC model	Yes	Not discussed	Yes	Yes	Partly	No	No	Unclear
Ramachandran *et al.*, child-hood type 1 diabetes, India	1996	Two-source CRC model	Yes	Not discussed	Yes	No	Partly	Yes	Yes	2–5
Shen *et al.*, childhood type 1 diabetes, China	1996	Two-source CRC model	No	Not discussed	Yes	Yes	No	Yes	No	3–7
Pal *et al.*, childhood epilepsy, India	1998	Two-source CRC model	No	Not discussed	Yes	Yes	No	No	No	3
Tull *et al.*, diabetes mellitus, Dominica, West Indies	1998	Three-source CRC model	Yes	Not discussed	No	Yes	Partly	No	No	3
Yang *et al.*, childhood type 1 diabetes, China	1998	Two-source CRC model	Yes	Not discussed	Yes	Yes	Partly	No	No	4–13
Argollo *et al.*, primary cere bral neoplasms, Brazil	1999	Two-source CRC model	Yes	Not discussed	Yes	Yes	No	No	Yes	4
Debrock *et al.*, epilepsy patients, Benin	2000	Three-source CRC model	No	No (L,D)	No	Yes	Yes	No	No	3
Sánchez *et al.*, multiple sclerosis, Colombia	2000	Two-source CRC model	Yes (in four provinces)	Not discussed	Unclear	Unclear	Yes	No	Yes	0–5
Gill *et al.*, non-communicable disease, South Africa	2001	Two-source CRC model	No	No (L, D)	Partly	No	Partly	No	No	Unclear
Spichler *et al.*, lower extremity amputation, Brazil	2001	Three-source CRC model	Yes	Not discussed	No	No	No	No	No	7–9
Martinucci *et al.*, childhood type 1 diabetes, Belarus	2002	Two-source CRC model	Yes	Not discussed	Unclear	Unclear	No	No	No	3–27
Sánchez *et al.*, mysthenia gravis, Colombia	2002	Two-source CRC model	Yes	Not discussed	Unclear	Unclear	No	No	Yes	5–7
Tuchinda *et al.*, childhood type 1 diabetes, Thailand	2002	Unclear	Unclear	Not discussed	Unclear	Unclear	No	No	Yes	7–11
Eisele *et al.*, demographic surveillance system for <5 years children, Kenya	2003	Two-source CRC model	No	No (L, D)	Yes	Partly	Yes	No	No	2–3
Chen *et al.*, infantile spasms, Taiwan	2004	Two-source CRC model	Yes	Not discussed	Yes	Yes	Partly	No	Yes	7–19

Table 3 (Continued)

Author, subject, country	Year*	Model	Sufficient registers	Good record-linkage†	Sufficient capture	Sufficient overlap	Limitations discussed‡	Comparison to other studies§	Local researchers	Publication delay (years)
Coeli *et al.*, diabetes mellitus, Brazil	2004	Three-source CRC model	No	Yes	Yes	No	Yes	Yes	Yes	9–10
Sánchez *et al.*, Parkinson's disease, Colombia	2004	Two-source CRC model	No	Not discussed	Unclear	Unclear	No	No	No	4–8
Ogbera *et al.*, diabetic foot ulceration, Nigeria	2006	Two-source CRC model	No	Not discussed	Yes	Partly	No	Yes	Yes	6–8
Vázquez *et al.*, amyotrophic lateral sclerosis, Uruguay	2008	Three-source CRC model	No	Not discussed	Unclear	Unclear	No	No	No	5–6
Cristiano *et al.*, multiple sclerosis, Argentina	2009	Three-source CRC model	Yes	Not discussed	Yes	Yes	Partly	No	No	13
Negrato *et al.*, childhood type 1 diabetes, Brasil	2009	Two-source CRC model	Yes	Not discussed	Unclear	Unclear	No	No	No	3–23
Azevedo-Silva *et al.*, child-hood acute leukemia, Brazil	2009	Two-source CRC model	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8
C										
Sanghavi *et al.*, tuberculosis, Peru	1998	Two-source CRC model	No	Not discussed	Yes	Partly	No	Yes	No	5–9
Simondon *et al.*, pertussis vaccine efficacy, Senegal	1999	Three-source CRC model	Yes	Yes	Yes	Yes	Partly	No	No	5–9
Van den Broek *et al.*, disabled leprosy patients, Nigeria	2001	Two-source CRC model	No	Yes	Yes	No	Yes	Yes	No	2
Yadon *et al.*, leishmaniasis, Argentina	2001	Two-source CRC model	Unclear	Not discussed	Unclear	Unclear	Partly	No	No	8–11
Brum *et al.*, human lepto spirosis, Brazil	2005	Three-source CRC model	Yes	No	No	No	Partly	No	Yes	3–4
Maia-Elkhoury *et al.*, visceral leishmaniasis cases, Brazil	2007	Two-source CRC model	Yes	Yes	Partly	Partly	Yes	No	Yes	4–5
Porten *et al.*, Buruli ulcer cases, Cameroon	2009	Two-source CRC model	No	Not discussed	Yes	Yes	No	Yes	No	2
Odega *et al.*, measles, Nigeria	2010	Two-source CRC model	Yes	Not discussed	Yes	No	Partly	No	Yes	2–3

HIV, human immunodeficiency virus; CRC, capture–recapture.
More details of the key assumptions and recommendations can be found in the data extraction section.
*Year, year of publication.
†L, linguistic; D, demographic.
‡Yes, ≥4 assumptions; Partly, 1–3 assumptions; No, assumptions not discussed.
§Estimating the same persons or events under study.

Table 4 Proportion of all publications using the two-source model and fulfilling recommendations for capture–recapture studies in resource-limited settings, divided over the three categories

Capture–recapture studies	Two-source model (%)	Sufficient registers (%)	Good record-linkage (%)	Sufficient capture (%)	Sufficient overlap (%)	Limitations discussed (%)	Comparison to other studies (%)	Local researchers (%)
Hidden populations, injuries and mortality (n = 19)	84	32	68	68	47	53	42	32
Non-infectious diseases (n = 25)	72	56	8	56	48	20	24	36
Infectious diseases (n = 8)	75	50	38	63	25	25	38	38

More details of the key assumptions and recommendations can be found in the data extraction section.

Table 5 Proportion of publications in low-income and lower-middle-income countries using the two-source model and fulfilling recommendations for capture–recapture studies in resource-limited settings, divided over the three categories

Capture–recapture studies	Two-source model (%)	Sufficient registers (%)	Good record-linkage (%)	Sufficient capture (%)	Sufficient overlap (%)	Limitations discussed (%)	Comparison to other studies (%)	Local researchers (%)
Hidden populations, injuries and mortality (n = 15)	93	33	73	60	53	60	33	27
Non-infectious diseases (n = 12)	77	38	0	77	62	15	31	31
Infectious diseases (n = 4)	75	50	50	100	50	25	50	25

More details of the key assumptions and recommendations can be found in the data extraction section.

Table 6 Proportion of publications in upper-middle-income countries using the two-source model and fulfilling recommendations for capture–recapture studies in resource-limited settings, divided over the three categories

Capture–recapture studies	Two-source model (%)	Sufficient registers (%)	Good record-linkage (%)	Sufficient capture (%)	Sufficient overlap (%)	Limitations discussed (%)	Comparison to other studies (%)	Local researchers (%)
Hidden populations, injuries and mortality (n = 4)	50	25	50	100	25	25	75	50
Non-infectious diseases (n = 13)	67	75	17	33	33	25	17	42
Infectious diseases (n = 4)	75	50	25	25	0	25	25	50

More details of the key assumptions and recommendations can be found in the data extraction section.

middle-income countries and upper-middle-income countries, these proportions were 46%, 23% and 50% and 50%, 24% and 25%, respectively (Tables 8 and 9). The category of hidden populations, injuries and mortality performed best and income did not affect performance considerably.

Discussion

This review examined the literature on capture–recapture studies in resource-limited countries in three categories. The performance of these methods to date has been variable; in resource-limited settings, only a few studies

Table 7 Cumulative proportion of all publications meeting numbers of assumptions or recommendations for capture–recapture studies in resource-limited settings, divided over the three categories

Country	Seven (%)	Six or more (%)	Five or more (%)	Four or more (%)	Three or more (%)	Two or more (%)	One or more (%)	None (%)
Hidden populations, injuries and mortality (n = 19)	0	5	26	42	74	90	95	5
Non- infectious diseases (n = 25)	4	4	8	20	48	72	88	12
Infectious diseasses (n = 8)	0	0	0	37	62	87	87	13

Key assumptions and recommendations: sufficient registers, good record-linkage, sufficient capture, sufficient overlap, limitations discussed, comparison to other studies and involverment of local researchers. More details of the key assumptions and recommendations can be found in the data extraction section.

Table 8 Cumulative proportion of all publications in low-income and lower-middle-income countries meeting numbers of assumptions or recommendations for capture–recapture studies in resource-limited settings, divided over the three categories

Country	Seven (%)	Six or more (%)	Five or more (%)	Four or more (%)	Three or more (%)	Two or more (%)	One or more (%)	None (%)
Hidden populations, injuries and mortality (n = 15)	0	7	20	46	73	93	93	7
Non-infectious diseases (n = 12)	0	0	0	23	54	85	93	7
Infectious diseasses (n = 4)	0	0	0	50	100	100	100	0

Key assumptions and recommendations: sufficient registers, good record-linkage, sufficient capture, sufficient overlap, limitations discussed, comparison to other studies and involverment of local researchers. More details of the key assumptions and recommendations can be found in the data extraction section.

Table 9 Cumulative proportion of all publications in upper-middle-income meeting numbers of assumptions or recommendations for capture–recapture studies in resource-limited settings, divided over the three categories

Country	Seven (%)	Six or more (%)	Five or more (%)	Four or more (%)	Three or more (%)	Two or more (%)	One or more (%)	None (%)
Hidden populations, injuries and mortality (n = 4)	0	0	50	50	75	75	100	0
Non-infectious diseases (n = 13)	8	8	16	24	41	58	83	17
Infectious diseasses (n = 4)	0	0	0	25	25	75	75	25

Key assumptions and recommendations: sufficient registers, good record-linkage, sufficient capture, sufficient overlap, limitations discussed, comparison to other studies and involverment of local researchers. More details of the key assumptions and recommendations can be found in the data extraction section.

achieved good individual quality criteria and only a minority met a minimum quality criterion per category. Capture–recapture approaches provide a potential tool to assess the burden of human attributes and disease, also in resource-limited countries, but their application needs to be rigorous. In particular, studies should ensure that, as far as possible, the basic underlying assumptions and other recommendations are met.

Limitations and strengths of this review

Only PubMed/MEDLINE indexed studies were included in this review and therefore those indexed elsewhere and non-indexed national or local studies were not considered. We used the search term 'recapture' but some studies may have used the term 'multiple system estimation' or 'mark-release' for their methodology. It cannot be excluded that a limited number of relevant studies have not been identified. Although standardised by a World Bank 2008 country income list, conditions can vary over time and within countries, for example when foreign research institutes are involved. Therefore, the World Bank income index may not always be a comparative indicator for each of the studies. The individual study quality score only applied to 30 studies (58%), leaving 22 studies with an indeterminate result. All studies identified in our search could be retrieved.

Limitations of capture–recapture analysis in epidemiology

Most of the limitations of capture–recapture analysis pertain to the possibility of violation of some of the underlying assumptions, which in contrast to animal population studies are unlikely to be satisfied in epidemiological applications (Desenclos & Hubert 1994; Brenner 1995; Hook & Regal 1995, 2000a; International Working Group for Disease Monitoring and Forecasting 1995a,b, Papoz *et al.* 1996; Chang *et al.* 1999; Cormack 1999; Cormack *et al.* 2000; Tilling 2001a). (i) Sufficient reliable identifiers for perfect record-linkage, preferably unique identification numbers, e.g. social security number, are not always available, for example owing to privacy regulations, or prone to linguistic or clerical errors, and sometimes relaxed or probabilistic record-linkage has to be performed on a combination of proxy-identifiers. (ii) Human populations are rarely closed, for example owing to health care–seeking behaviour. (iii) The probability of ascertainment by any particular source should be equal but in epidemiological settings often it is not, because of the intrinsic nature of human variation, e.g. socioeconomic differences or variation of severity of disease. (iv) Dependence of sources is often a problem in epidemiological capture–recapture applications. 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 because of referral. In two-source capture–recapture analysis, this assumption is crucial because it is difficult to check independence mathematically and researchers must make the judgement intuitively. Dependencies can cause under-estimation (in case of positive dependence) or over-estimation (in case of negative dependence) (Brenner 1995). Violation of the other assumptions can also cause bias in both directions. (5) In epidemiology, capture–recapture analysis often uses existing administrative registers, not designed for this methodology, instead of random surveys of the population according to a common protocol as in ecology. The accuracy of the registers used, such as correct diagnosis and coding, is important. In capture–recapture analysis, errors are highly likely to have a more than additive effect on estimates (van Hest *et al.* 2007). Registers containing poor quality data lead to poor capture–recapture outcomes (Hook & Regal 1995; Cormack 1999), and a low positive predictive value results in over-estimation.

The use of simple two-source capture–recapture models for epidemiological data is often limited by violation of the underlying capture–recapture assumptions, resulting in biased estimates (Brenner 1995; Tilling 2001a), and more complicated three-source log-linear models were developed to partly address this problem (Wittes & Sidel 1968; Hook & Regal 1995; International Working Group for Disease Monitoring and Forecasting 1995a,b). Although many apparently successful capture–recapture studies have been published, only few did reportedly fail and confidence in the validity of capture–recapture results may reflect publication bias in favour of successful capture–recapture studies rather than the inherent strength of this method (Hay 1997).

Application and limitations of capture–recapture analysis in resource-limited countries

In the year 1981, Karel Styblo and Annik Rouillon wrote as follows: 'For Africa, Asia and Latin America the reported tuberculosis incidence figures were, with a few exceptions, totally unreliable and incomplete, and should not be extrapolated to areas with no notification of tuberculosis and, most important, should not be taken into consideration to assess the trend of the incidence of tuberculosis in the world' (Styblo & Rouillon 1981). Almost 30 years after this statement, despite progress in a limited number of countries, epidemiological surveillance in resource-limited countries often remains weak and unreliable. Cost-effective and accurate human conditions and health indices are greatly needed, not only for infectious diseases but also for emerging non-infectious diseases (International Working Group for Disease Monitoring and Forecasting 1995b; Trébucq *et al.* 1998). Community-based studies, such as surveys, are expensive and difficult to perform but may be justifiable in high-burden countries where many cases are missed by routine surveillance systems (Frerichs 1991; Dye *et al.* 2008).

As an alternative, capture–recapture techniques were proposed for epidemiological surveillance in resource-limited countries in the mid-1990s (Chiu *et al.*1993; Watts *et al.* 1994, 1995; International Working Group for Disease Monitoring and Forecasting 1995b), with early indications of limitations (Black *et al.* 1994).

In contrast to resource-rich countries, only a limited number of capture–recapture studies have since been published from resource-limited countries estimating completeness of road traffic accident, injury or surgery registrations (Razzak & Luby 1998; Spichler *et al.* 2001; Tercero & Anderson 2004; Van *et al.* 2006; Zavareh *et al.* 2008; Sivamani *et al.* 2009); numbers of sex workers and drug users (Mastro *et al.* 1994; Watts *et al.* 1994, 1995; Caiaffa *et al.* 2003; Kruse *et al.* 2003; Böhning *et al.* 2004; Khan *et al.* 2004; Minh *et al.* 2004; Platt *et al.* 2004; Luan *et al.* 2005; Mingoti *et al.* 2006; Chiang *et al.* 2007; Geibel *et al.* 2007), or other hidden populations (Gurgel *et al.* 2004); the completeness of demographic surveillance (Diallo *et al.* 1996; Eisele *et al.* 2003; Ronsmans *et al.* 2009); the completeness of non-infectious disease registration (Xia *et al.* 1995; Pal *et al.* 1998; Argollo & Lessa 1999; Debrock *et al.* 2000; Sánchez *et al.* 2000, 2002, 2004; Gill *et al.* 2001; Chen *et al.* 2004; Vázquez *et al.* 2008; Azevedo-Silva *et al.* 2009; Cristiano *et al.* 2009), mainly (juvenile) diabetes (Elamin *et al.* 1992; Fu *et al.* 1994; Ramachandran *et al.* 1996; Shen *et al.* 1996; Tull *et al.* 1998; Yang *et al.* 1998; Tuchinda *et al.* 2002; Coeli *et al.* 2004; Ogbera *et al.* 2006; Negrato *et al.* 2010), especially after the WHO Diamond project (WHO Diamond Project Group 1990); and vaccine efficacy (Simondon & Khodja 1999). Regarding infectious diseases, four capture–recapture studies on tuberculosis, leishmaniasis and leptospirosis were reported in upper-middle-income countries in South-America (Sanghavi *et al.* 1998; Yadon *et al.* 2001; Brum & Kupek 2005; Maia-Elkhoury *et al.* 2007), and only three such studies, estimating leprosy, Buruli ulcer and measles patients, in a low-income (Porten *et al.* 2009) or a lower-middle-income country (van den Broek *et al.* 2001; Odega *et al.* 2010), all in Africa, were found, despite the huge burden of disease.

Various difficulties with capture–recapture analysis in resource-limited settings have been described, possibly violating the underlying assumptions (Black *et al.* 1994; Mastro *et al.*1994; Watts *et al.* 1995; Razzak & Luby 1998; Debrock *et al.* 2000; van den Broek *et al.* 2001; Gill *et al.* 2001; Coeli *et al.* 2004; Fleury *et al.* 2004; Tercero & Anderson 2004; Brum & Kupek 2005).

Firstly, for perfect record-linkage conventional, often existing, non-aggregated data sources, containing sufficient and reliable patient identifiers, used in resource-rich countries, such as notification, laboratory and hospital admission registers for disease incidence estimation, are rarely available in resource-limited countries. Linguistic and demographic problems, such as variability in spelling or registration in local language and English, lacking precise date of birth or age and unavailability of street names and numerical addresses, hinder perfect record-linkage and use of computer software (Razzak & Luby 1998; Debrock *et al.* 2000; Gill *et al.* 2001; Eisele *et al.* 2003; Gurgel *et al.* 2004). The single-source approach with truncated models to reduce record-linkage problems can only be used in specific situations rarely met in resource-limited countries (Böhning *et al.* 2004; van Hest *et al.* 2008c).

Second, high levels of migration in many resource-limited settings violate the closed population assumption. Third, inaccurate and incomplete data prevent proper investigation of violation of the homogeneity assumption and possible correction through stratified or co-variate analysis (Tilling & Sterne 1999; Tilling *et al.* 2001b; van Hest *et al.* 2008b). Fourth, insufficient available and suitable registers hinder three-source log-linear modelling to investigate and handle possible interdependencies. The majority of the capture–recapture studies in resource-poor settings use the simple two-source methodology, making it difficult to investigate potential interdependencies, possibly resulting in bias in the estimates, limited (van Hest 2008) or substantial (Bassili *et al.* 2010). Some studies *a priori* assume independent registers (Elamin *et al.* 1992; Shen *et al.* 1996; Yang *et al.* 1998; Martinucci *et al.* 2002; Chiang *et al.* 2007; Maia-Elkhoury *et al.* 2007; Sivamani *et al.* 2009).

Fifth, inaccurate and incomplete data prevent proper elimination of duplicate cases within registers and insufficient registers prevent identification of false-positive cases and cross-validation of true cases between registers. Sixth, insufficient registers, even after additional surveys, also limit opportunities for adequate source-selection, possibly resulting in mutually exclusive registers (Gill *et al.* 2001), or poor overlap (Mastro *et al.* 1994; Gurgel *et al.* 2004; Tercero & Anderson 2004; Chiang *et al.* 2007). Our criterion for sufficient overlap (≥15%) is rather conservative as others indicate ≥30% as sufficient (Kruse *et al.* 2003).

Despite previous discussion and recommendations on the application, interpretation and presentation of capture–recapture studies (International Working Group for Disease Monitoring and Forecasting 1995a,b; Hook & Regal 1995, 1999, 2000b), only a few studies adequately address feasibility and assumptions of the capture–recapture approach in resource-limited settings. No infectious diseases study belonged to this group and only one non-infectious disease study. The presentation of the data, models and outcomes varied considerably between the studies reviewed. Capture–recapture studies in resource-limited countries often involve international experts and possibly foreign funding, leaving the question whether local authorities would have been able to apply the methodology independently and consistently. In contrast to the outcomes of this review, a considerable number of capture–recapture studies in resource-limited settings mention the approach as low-cost, relatively easy to use and appropriate (Xia *et al.*1995; Tull *et al.* 1998; Argollo & Lessa 1999; Sánchez *et al.* 2000; van den Broek *et al.* 2001; Spichler *et al.* 2001; Tuchinda *et al.* 2002; Kruse *et al.* 2003; Khan *et al.* 2004; Van *et al.* 2006; Maia-Elkhoury *et al.* 2007; Azevedo-Silva *et al.* 2009; Cristiano *et al.* 2009).

Recommendations for the application of the capture–recapture method

Several steps are important in planning, applying, presenting and evaluating capture–recapture techniques in epidemiological studies(Hook & Regal 1999, 2000b; van Hest 2007). Briefly, countries or regions must be well defined and without substantial, for example health care purpose, migration. The aim of the study and the required accuracy of the data should be described and appropriate sources for capturing cases should be selected. Sufficient data should be collected, perhaps by covering a longer period, in three fairly independent registers, approximately equal in number, each capturing more than 15% of cases and having sufficient overlap. Possible relationships between the selected sources should be investigated and described as well as their influence on the capture–recapture results. An unambiguous and uniform case-definition for the various sources should be used. In morbidity and mortality studies, the accuracy of diagnosis and disease classification in each of the sources should be examined. The accuracy of record-linkage should be described. There must be a reliable means of identifying cases across registers, e.g. by distinctive and accurately recorded patient names, universally quoted personal identity numbers, or highly specific combinations of personal and clinical conditions. Case-ascertainment, distribution of cases over various registers, the examined capture–recapture models and selected model, and the estimate of the number of missing cases and thus the total number of cases from which prevalence or incidence rates can be calculated, should be given. The choice of model for analysis should take into account known features of the country and systems and not rely wholly on the counts of cases by sources. The three-sample capture–recapture approach can handle pair-wise dependencies between registers, identified as interactions in the model, analytically. If heterogeneity exists, it can be allowed for to some extent by stratifying the data and analysing each stratum separately or by using co-variate analysis (Tilling & Sterne 1999; Tilling *et al.* 2001b; van Hest *et al.* 2008b). Models other than log-linear, such as the truncated binomial model, should be considered (van Hest *et al.* 2008a,c). The limitations of the capture–recapture method in epidemiology should be addressed. Preferably preparatory fieldwork and explorative research should be performed, when possible including a small-scale pilot capture–recapture study investigating feasibility, limitations and costs, ideally with the possibility of validation of the estimates against known data from a census or survey and, when necessary, international capture–recapture experts should be consulted. Bayesian model averaging, i.e. models are weighted according to their goodness of fit (posterior probability), with a weighted average presented as the best estimate, should be considered when in three-source modelling one of the models fits the data only marginally better than the others.

Conclusion

Most resource-limited countries, despite the burden of human attributes and disease, do not yet have reliable surveillance systems to measure, monitor and report accurately on progress of control programmes, identification of priorities and implementation and evaluation of interventions. Capture–recapture approaches can provide a alternative tool but this review indicates that in resource-limited settings only few achieved good individual quality criteria and a minority met a minimum quality criterion per category. Careful consideration and reconsideration are therefore needed regarding capture–recapture techniques for surveillance and trend analysis in resource-limited countries, and their application needs to be rigorous. Although recently some innovative capture–recapture approaches are being implemented in resource-limited settings, for global, standardised and reliable use this methodology must be further developed. Increasing expertise and ability of local researchers to conduct high-quality capture–recapture studies in resource-limited countries should be promoted, in collaboration with international expert panels. However, capture–recapture analysis is not a panacea for the ultimate goal of global surveillance, namely that all countries should strengthen their routine national surveillance systems.

Acknowledgements

Engelien de Ruiter and David Rodriguez of the Bibliographic and Documentation Department of the Municipal

Public Health Service Rotterdam-Rijnmond assisted in the literature search.

References

Aptel I, Salmi LR, Masson F, Bourdé A, Henrion G & Erny P (1999) Road accident statistics: discrepancies between police and hospital data in a French island. *Accident, Analysis and Prevention* **31**, 101–108.

Argollo N & Lessa I (1999) [Estimated prevalence of cerebral neoplasm of childhood from the capture–recapture method]. *Arquivos de Neuropsiquiatica* **57**, 435–441.

Azevedo-Silva F, De Souza Reis R, De Oliveira Sanos M, Raggio Luiz R & Pombo-De Oliveira MS (2009) Evaluation of childhood acute leukemia incidence and underreporting in Brazil by capture–recapture methodology. *Cancer Epidemiology* **33**, 401–402.

Bassili A, Grant AD, El-Mohgazy E *et al.* (2010) Estimating tuberculosis case-detection rate in resource-limited countries: a capture–recapture study in Egypt. *International Journal of Tuberculosis and Lung Disease* **14**, 727–732.

Bernillon P, Lievre L, Pillonel J, Laporte A & Costagliola D (2000) 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. *International Journal of Epidemiology* **29**, 168–174.

Black JF, McLarty D & Mtasiwa DE (1994) Capture–recapture techniques. Difficult to use in developing countries. *British Medical Journal* **308**, 531.

Böhning D, Suppawattanabodee B, Kusolvisitkul W & Viva-twongkasem C (2004) Estimating the number of drug users in Bangkok 2001: a capture–recapture approach using repeated entries in one list. *European Journal of Epidemiology* **19**, 1075–1083.

Brenner H (1995) Use and limitations of the capture–recapture method in disease monitoring with two dependent sources. *Epidemiology* **6**, 42–48.

van den Broek J, van Jaarsveld T, de Rijk A, Samson K & Patrobas P (2001) Capture–recapture method to assess the prevalence of disabled leprosy patients. *Leprosy Review* **72**, 292–301.

Brum L & Kupek E (2005) Record linkage and capture–recapture estimates for underreporting of human leptospirosis in a Brazilian health district. *Brazilian Journal of Infectious Diseases* **9**, 515–520.

Caiaffa WT, Mingoti SA, Proietti FA *et al.* (2003) Estimation of the number of injecting drug users attending an outreach syringe-exchange program and infection with human immunodeficiency virus (HIV) and hepatitis C virus: the AjUDE-Brasil project. *Journal of Urban Health* **80**, 106–114.

Cathcart SJ, Lawrence J, Grant A *et al.* (2009) Estimating unreported malaria cases in England: a capture–recapture study. *Epidemiology and Infection* **18**, 1052–1058.

Chang YF, LaPorte RE, Aaron DJ & Songer TJ (1999) The importance of source selection and pilot study in the capture–recapture application. *Journal of Clinical Epidemiology* **52**, 927–928.

Chen CC, Chen TF, Lin HC *et al.* (2004) Estimation of prevalence and incidence of infantile spasms in Taiwan using capture–recapture method. *Epilepsy Research* **58**, 37–42.

Chiang SC, Chen CY, Chang YY, Sun HJ & Chen WJ (2007) Prevalence of heroin and methamphetamine male users in the northern Taiwan, 1999–2002: capture–recapture estimates. *BMC Public Health* **7**, 292.

Chiu WT, Dearwater SR, McCarty DJ, Songer TJ & LaPorte RE (1993) Establishment of accurate incidence rates for head and spinal cord injuries in developing and developed countries: a capture–recapture approach. *Journal of Trauma* **35**, 206–211.

Coeli CM, Coutinho ES & Veras RP (2004) The challenge of applying capture–recapture methodology to diabetes mellitus surveillance in the elderly: lessons from a Brazilian experience. *Cadernos de Saude Publica* **20**, 1709–1720.

Cojocaru C, Van Hest NA, Mihaescu T & Davies PD (2009) Completeness of notification of adult tuberculosis in Iaşi county, Romania: a capture–recapture analysis. *International Journal of Tuberculosis and Lung Disease* **13**, 1094–1099.

Cormack RM (1999) Problems with using capture–recapture in epidemiology: an example of a measles epidemic. *Journal of Clinical Epidemiology* **52**, 909–914.

Cormack RM, Chang YF & Smith GS (2000) Estimating deaths from industrial injury by capture–recapture: a cautionary tale. *International Journal of Epidemiology* **29**, 1053–1059.

Cristiano E, Patrucco L, Rojas JI *et al.* (2009) Prevalence of multiple sclerosis in Buenos Aires, Argentina using the capture–recapture method. *European Journal of Neurology* **16**, 183–187.

Debrock C, Preux PM, Houinato D *et al.* (2000) Estimation of the prevalence of epilepsy in the Benin region of Zinvie using the capture–recapture method. *International Journal of Epidemiology* **29**, 330–335.

Desenclos JC & Hubert B (1994) Limitations to the universal use of capture–recapture methods. *International Journal of Epidemiology* **23**, 1322–1323.

Diallo DA, Habluetzel A, Esposito F & Cousens SN (1996) Comparison of two methods for assessing child mortality in areas without comprehensive registration systems. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**, 610–613.

Dye C, Bassili A, Bierrenbach AL *et al.* (2008) Measuring tuberculosis burden, trends, and the impact of control programmes. *Lancet Infectious Dieases* **8**, 233–243.

Eisele TP, Lindblade KA, Rosen DH, Odhiambo F, Vulule JM & Slutsker L (2003) Evaluating the completeness of demographic surveillance of children less than five years old in western Kenya: a capture–recapture approach. *American Journal of Tropical Medicine and Hygiene* **69**, 92–97.

Elamin A, Omer MI, Zein K & Tuvemo T (1992) Epidemiology of childhood type I diabetes in Sudan 1987–1990. *Diabetes Care* **15**, 1556–1559.

Fleury A, Preux PM & Fragroso G (2004) Use of capture–recapture method for determining the prevalence of neurological parasitic diseases: a reply. *Neuroepidemiology* **23**, 99–100.

Frerichs RR (1991) Epidemiologic surveillance in developing countries. *Annual Review of Public Health* **12**, 257–580.

Fu H, Shen SX, Chen ZW *et al.* (1994) Shanghai, China, has the lowest confirmed incidence of childhood diabetes in the world. *Diabetes Care* **17**, 1206–1208.

Geibel S, van der Elst M, King'ola N *et al.* (2007) 'Are you on the market?': a capture–recapture enumeration of me who sell sex to men in and around Mombassa, Kenya. *AIDS* **21**, 1349–1354.

Gill GV, Scott B, Beeching NJ, Wilkinson D & Ismail AA (2001) Enumeration of non-communicable disease in rural South Africa by electronic data linkage and capture–recapture techniques. *Tropical Medicine and International Health* **6**, 435–441.

Glaziou P, van der Werf MJ, Onozaki I *et al.* (2008) Tuberculosis prevalence surveys: rationale and cost. *International Journal of Tuberculosis and Lung Disease* **12**, 1003–1008.

Guernier V, Guégan JF & Deparis X (2006) An evaluation of the actual incidence of tuberculosis in French Guyana using a capture–recapture model. *Microbes and Infection* **8**, 721–727.

Gurgel RQ, Da Fonseca JD, Neyra-Castañeda D, Gill GV & Cuevas LE (2004) Capture–recapture to estimate the number of street children in a city in Brazil. *Archives of Disease in Childhood* **89**, 222–224.

Gutteridge W & Collin C (1994) Capture–recapture techniques. Quick and cheap. *British Medical Journal* **308**, 531.

Hay G (1997) The selection from multiple data sources in epidemiological capture–recapture studies. *Statistician* **46**, 515–520.

van Hest NA (2007) Capture–recapture methods in surveillance of tuberculosis and other infectious diseases. Academic Thesis. Erasmus University, Rotterdam, http://publishing.eur.nl/ir/repub/asset/10506/070914_Hest,%20Norbertus%20Alphonsus%20Henricus%20van.pdf

van Hest R (2008) Use of capture–recapture analysis in data sources. *Revista de Saude Publica* **42**, 377–379.

van Hest NA, Smit F & Verhave JP (2002) Improving malaria notification in the Netherlands: results from a capture–recapture study. *Epidemiology and Infection* **129**, 371–377.

van Hest NA, Smit F, Baars HW *et al.* (2007) Completeness of registration of tuberculosis in the Netherlands: how reliable is record-linkage and capture–recapture analysis? *Epidemiology and Infection* **135**, 1021–1029.

van Hest NA, Grant D, Smit F, Story A & Richardus JH (2008a) Estimating infectious diseases incidence: validity of capture–recapture analysis and truncated models for incomplete count data. *Epidemiology and Infection* **136**, 14–22.

van Hest NA, Hoebe CJ, Den Boer JW *et al.* (2008b) Incidence and completeness of notification of Legionnaires' disease in the Netherlands: covariate capture–recapture analysis acknowledging geographical differences. *Epidemiology and Infection* **136**, 540–550.

van Hest NA, de Vries G, Smit F, Grant AG & Richardus JH (2008c) Estimating the coverage of tuberculosis screening among drug users and homeless persons with truncated models. *Epidemiology and Infection* **136**, 628–635.

Hook EB & Regal RR (1992) 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. *American Journal of Epidemiology* **135**, 1060–1067.

Hook EB & Regal RR (1995) Capture–recapture methods in epidemiology: methods and limitations. *Epidemiologic Review* **17**, 243–264.

Hook EB & Regal RR (1999) Recommendations for presentation and evaluation of capture–recapture estimates in epidemiology. *Journal of Clinical Epidemiology* **52**, 917–926.

Hook EB & Regal RR (2000a) Accuracy of alternative approaches to capture–recapture estimates of disease frequency: internal validity analysis of data from five sources. *American Journal of Epidemiology* **152**, 771–779.

Hook EB & Regal RR (2000b) On the need for a 16th and 17th recommendations for capture recapture analysis. *Journal of Clinical Epidemiology* **53**, 1275–1277.

Iňigo J, Arce A, Martín-Moreno JM, Herruzo R, Palenque E & Chaves F (2003) Recent transmission of tuberculosis in Madrid: application of capture–recapture analysis to conventional and molecular epidemiology. *International Journal of Epidemiology* **32**, 763–769.

International Working Group for Disease Monitoring and Forecasting (1995a) Capture–recapture and multiple-record estimation I: history and theoretical development. *American Journal of Epidemiology* **142**, 1047–1058.

International Working Group for Disease Monitoring and Forecasting (1995b) Capture–recapture and multiple-record estimation II: applications in human diseases. *American Journal of Epidemiology* **142**, 1059–1068.

Khan SI, Bhuiya A & Uddin AS (2004) Application of the capture–recapture method for estimating number of mobile male sex workers in a port city of Bangladesh. *Journal of Health, Population and Nutrition* **22**, 19–26.

Kruse N, Frieda M-T, Behets F *et al.* (2003) Participatory mapping of sex trade and enumeration of sex workers using capture–recapture methodology in Diego-Suarez, Madagascar. *Sexually Transmitted Diseases* **30**, 664–670.

Laporte RE (1994) Assessing the human condition: capture–recapture techniques. *British Medical Journal* **308**, 5–6.

Liberati A, Altman DG, Tetzlaff J *et al.* (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of Internal Medicine* **151**, W65–W94.

Luan R, Zeng G, Zhang D *et al.* (2005) A study on methods of estimating the population size of men who have sex with men in Southwest China. *European Journal of Epidemiology* **20**, 581–585.

Maia-Elkhoury AN, Carmo EH, Sousa-Gomes ML & Mota E (2007) Analysis of visceral leishmaniasis reports by the capture–recapture method. *Revista de Saude Publica* **41**, 931–937.

Martinucci ME, Curradi G, Fasulo A *et al.* (2002) Incidence of childhood type 1 diabetes mellitus in Gomel, Belarus. *Journal of Pediatric Endocrinology and Metabolism* **15**, 53–57.

Masjedi MR, Fadaizadeh L & Taghizadeh Asl R (2007) Notification of patients with tuberculosis detected in the private sector. *International Journal of Tuberculosis and Lung Disease* **11**, 882–886.

Mastro TD, Kitayaporn D, Weniger BG *et al.* (1994) Estimating the number of HIV-infected injection drug users in Bangkok:

a capture–recapture method. *American Journal of Public Health* **84**, 1094–1099.

Mingoti SA, Caiaffa WT & Projeto AjUDE-Brasil II (2006) Capture–recapture technique to estimate the size of the injecting drug user population attending syringe exchange programs: AjUDE-Brasil II Project. *Cadernos de Saude Publica* **22**, 783–789.

Minh TT, Nhan DT, West GR *et al.* (2004) Sex workers in Vietnam: how many, how risky. *AIDS Education and Prevention* **16**, 389–404.

Negrato CA, Dias JP, Teixeira MF *et al.* (2010) Temporal trends in incidence of type 1 diabetes between 1986 and 2006 in Brazil. *Journal of Endocrinological Investigation* **33**, 373–377.

Odega CC, Fatiregun AA & Osagbemi GK (2010) Completeness of suspected measles reporting in a southern district of Nigeria. *Public Health* **124**, 24–27.

Ogbera AO, Fasanmade O, Ohwovoriole AE & Adediran O (2006) An assessment of the disease burden of foot ulcers in patients with diabetes mellitus attending a teaching hospital in Lagos, Nigeria. *International Journal of Lower Extremity Wounds* **5**, 244–249.

Pal DK, Das T & Sengupta S (1998) Comparison of key informant and survey methods for ascertainment of childhood epilepsy in West Bengal, India. *International Journal of Epidemiology* **27**, 672–676.

Papoz L, Balkau B & Lellouch J (1996) Case counting in epidemiology: limitations of methods based on multiple data sources. *International Journal of Epidemiology* **25**, 474–478.

Pezzoti P, Piovesan C, Michieletto F, Zanella F, Rezza G & Gallo G (2003) Estimating the cumulative number of human immunodeficiency virus diagnoses by cross-linking from four different sources. *International Journal of Epidemiology* **32**, 778–783.

Platt L, Hickman M, Rhodes T *et al.* (2004) The prevalence of injecting drug use in a Russian city: implications for harm reduction and coverage. *Addiction* **99**, 1430–1438.

Porten K, Sailor K, Comte E *et al.* (2009) Prevalence of Buruli ulcer in Akonolinga health district, Cameroon: results of a cross sectional survey. *PLoS Neglected Tropical Diseases* **3**, e466.

Ramachandran A, Snehalatha C & Krishnaswamy CV (1996) Incidence of IDDM in children in urban population in southern India. *Diabetes Research and Clinical Practice* **34**, 79–82.

Razzak JA & Luby SP (1998) Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture–recapture method. *International Journal of Epidemiology* **27**, 866–870.

Ronsmans C, Scott S, Qomariyah SN *et al.* (2009) Professional assistance during birth and maternal mortality in two Indonesian districts. *Bulletin of the World Health Organization* **87**, 416–423.

Sánchez JL, Aguirre C, Arcos-Burgos OM *et al.* (2000) [Prevalence of multiple sclerosis in Colombia]. *Revista de Neurologica* **31**, 1101–1103.

Sánchez JL, Uribe CS, Franco A, Jiménez ME, Arcos-Burgos M & Palacio LG (2002) [Prevalence of myasthenia gravis in Antioquia, Colombia]. *Revista de Neurologica* **34**, 1010–1012.

Sánchez JL, Buriticá O, Pineda D, Uribe CS & Palacio LG (2004) Prevalence of Parkinson's disease and parkinsonism in a Colombian population using the capture–recapture method. *International Journal of Neuroscience* **114**, 175–182.

Sanghavi DM, Gilman RH, Lescano-Guevara AG, Checkley W, Cabrera LZ & Cardenas V (1998) Hyperendemic pulmonary tuberculosis in a Peruvian shantytown. *American Journal of Epidemiology* **148**, 384–389.

Shen SX, Wang HB, Chen ZW *et al.* (1996) The incidence of insulin-dependent diabetes mellitus in urban districts of Shanghai (1989–1993). *Journal of Pediatric Endocrinology and Metabolism* **9**, 469–473.

Simondon F & Khodja H (1999) Capture–recapture method for estimating misclassification errors: application to the measurement of vaccine efficacy in randomized controlled trails. *International Journal of Epidemiology* **28**, 113–116.

Sivamani M, Balraj V & Muliyil JP (2009) Validity of a surveillance system for childhood injuries in a rural block of Tamilnadu. *Indian Journal of Community Medicine* **34**, 43–47.

Spichler ER, Spichler D, Lessa I *et al.* (2001) Capture–recapture method to estimate lower extremity amputation rates in Rio de Janeiro, Brazil. *Pan American Journal of Public Health* **10**, 334–340.

Styblo K & Rouillon A (1981) Estimated global incidence of smear-positive tuberculosis. Unreliability of officially reported figures on tuberculosis. *Bulletin of the International Union against Tuberculosis* **56**, 118–126.

Tercero F & Anderson R (2004) Measuring transport injuries in a developing country: an application of the capture–recapture method. *Accident, Analysis and Prevention* **36**, 13–20.

Thompson-Cerna AR & Medina MT (2007) [The prevalence of cerebrovascular disease in the rural community of Salamá, Honduras, using the capture–recapture epidemiological method]. *Revista de Neurologica* **44**, 460–464.

Tilling K (2001a) Capture–recapture methods-useful or misleading? *International Journal of Epidemiology* **30**, 12–14.

Tilling K & Sterne JA (1999) Capture–recapture models including covariate effects. *American Journal of Epidemiology* **149**, 392–400.

Tilling K, Sterne JA & Wolfe CD (2001b) Estimation of the incidence of stroke using a capture–recapture model including covariates. *International Journal of Epidemiology* **30**, 1351–1359.

Trébucq A, Aït-Khaled N, Gninafon M & Louis J-P (1998) Information management in national tuberculosis control programmes and national health information systems. *International Journal of Tuberculosis and Lung Disease* **2**, 852–856.

Tuchinda C, Likitmaskul S, Unachak K, Panamonta O, Patarakijavanich N & Chetthakul T (2002) The epidemiology of type 1 diabetes in Thai children. *Journal of the Medical Association of Thailand* **85**, 648–652.

Tull E, Butler C, Gumbs L & Williams S (1998) The use of capture–recapture methods to monitor diabetes in Dominica, West Indies. *Revista Panamericana de Salud Publica* **3**, 303–307.

Van HT, Singhasivanon P, Kaewkungwal J, Suriyawongpaisal P & Khai LH (2006) Estimation of non-fatal road traffic injuries in

Thai Nguyen, Vietnam using capture–recapture method. *Southeast Asian Journal of Tropical Medicine and Public Health* **37**, 405–411.

Vázquez MC, Ketzoián C, Legnani C *et al.* (2008) Incidence and prevalence of amyotrophic lateral sclerosis in Uruguay: a population-based study. *Neuroepidemiology* **30**, 105–111.

Waters WE (1994) More unreliable in humans than in birds. *British Medical Journal* **308**, 531.

Watts C, Zwi A, Wilson D, Mashababe S & Foster G (1994) Capture–recapture as a tool for programme evaluation. *British Medical Journal* **308**, 859.

Watts CH, Zwi A & Foster G (1995) Using capture–recapture in promoting public health. *Health Policy and Planning* **10**, 198–203.

WHO Diamond Project Group (1990) WHO Multinational Project for childhood diabetes. *Diabetes Care* **13**, 1062–1068.

Wittes J & Sidel VW (1968) A generalization of the simple capture–recapture model with applications to epidemiological research. *Journal of Chronic Diseases* **21**, 287–301.

Woodward JA (1990) *Design of a Study to Estimate the Size of a Heroin Using Population in Thailand*. University of California, Los Angeles, CA.

World Bank (2009) *Data and Statistics: Country Classification* World Bank, Washington, http://siteresources.worldbank.org/DATASTATISTICS/Resources/CLASS.XLS, Accessed 27 November 2010.

Xia ZL, Jin XP, Lu PL, Gu XQ, LaPorte RE & Tajima N (1995) Ascertainment corrected prevalence rate (ACPR) of leukopenia in workers exposed to benzene in small-scale industries calculated with capture–recapture methods. *Biomedical and Environmental Sciences* **8**, 30–34.

Yadon ZE, Quigley MA, Davies CR, Rodrigues LC & Segura EL (2001) Assessment of Leishmaniasis notification system in Santiago del Estero, Argentina, 1990–1993. *American Journal of Tropical Medicine and Hygiene* **65**, 27–30.

Yang Z, Wang K, Li T *et al.* (1998) Childhood diabetes in China. Enormous variation by place and ethnic group. *Diabetes Care* **21**, 525–529.

Zavareh DK, Mohammadi R, Laflamme L, Naghavi M, Zarei A & Haglund BJ (2008) Estimating road traffic mortality more accurately: use of the capture–recapture method in the West Azarbaijan province of Iran. *International Journal of Injury Control and Safety Promotion* **15**, 9–17.

Corresponding Author Rob van Hest, Tuberculosis Control Section, Infectious Disease Control Cluster, Rotterdam Public Health Service, P.O. Box 70032, 3000 LP Rotterdam, The Netherlands. Tel.: +31 10 4339 338; Fax: +31 10 4339 950; E-mail: r.vanhest@ggd.rotterdam.nl