

## Invasive pneumococcal disease in Victoria: a better measurement of the true incidence

H. J. CLOTHIER<sup>1,2\*</sup>, T. VU<sup>1</sup>, V. SUNDARARAJAN<sup>3</sup>, R. M. ANDREWS<sup>4</sup>,  
M. COUNAHAN<sup>1</sup>, G. F. TALLIS<sup>1</sup> AND S. B. LAMBERT<sup>2,5</sup>

<sup>1</sup> Communicable Diseases Section, Rural and Regional Health and Aged Care Services Division, Department of Human Services, Melbourne, Victoria, Australia

<sup>2</sup> Master of Applied Epidemiology Program, National Centre for Epidemiology and Population Health, Australian National University, Canberra, Australia

<sup>3</sup> Health Surveillance and Evaluation Section, Rural and Regional Health and Aged Care Services Division, Department of Human Services, Melbourne, Victoria, Australia

<sup>4</sup> Clinical Epidemiology and Biostatistics Unit, Research Institute & Department of Paediatrics, University of Melbourne, Australia

<sup>5</sup> Vaccine and Immunisation Research Group, Research Institute and the School of Population Health, University of Melbourne, Australia

(Accepted 24 January 2007; first published online 15 March 2007)

### SUMMARY

Invasive pneumococcal disease (IPD) notifications are used to monitor IPD vaccination programmes. We conducted sequential deterministic data-linkage between IPD notifications and hospitalization data in Victoria, Australia, in order to determine whether all diagnosed cases were being reported. The proportion of each relevant hospital admission ICD-10-AM code that could be linked to notified cases was calculated. Total and age-specific annual rates were calculated and compared for notified and non-notified cases. Total incidence was estimated using data-linkage results and application of a two-source capture–recapture method. The first 2 years of IPD surveillance in Victoria missed at least one-sixth of laboratory-confirmed IPD cases. Estimated annual IPD rate increased from 9·0 to 10·7/100 000 and rose even higher, to 11·5/100 000, with age-specific rates possibly reaching 90·0/100 000 children aged <2 years, when using capture–recapture. Strategies to improve notification and coding of hospitalized cases of IPD are required.

### BACKGROUND

Invasive pneumococcal disease (IPD) is a serious and potentially vaccine-preventable disease. IPD commonly manifests as septicaemia or meningitis but is occasionally isolated from other normally sterile sites. Mucosal pneumococcal disease can be associated with IPD if the pneumococcus is also isolated

from a normally sterile site, such as is the case with bacteraemic pneumonia.

In Victoria, Australia, IPD became a notifiable disease under the Health (Infectious Diseases) Regulations 2001 on 16 May 2001 [1]. These Regulations made it mandatory for both the diagnosing laboratory and doctor to notify the Department of Human Services (DHS) upon identification of IPD. Under-reporting impacts on the ability of the surveillance system to inform preventative policy decisions around vaccine use in high-risk populations and the general community.

\* Author for correspondence: Ms. H. J. Clothier, Epidemiology Unit, Victorian Infectious Diseases Reference Laboratory, 10 Wreckyn Street, North Melbourne 3051, Victoria, Australia.  
(Email: hazelclothier@optusnet.com.au)

In addition to notifications, IPD cases can be found in hospital discharge abstracts recorded using codes from the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian modification* (ICD-10-AM) [2] in the Victorian Admitted Episodes Dataset (VAED). The VAED records up to 25 diagnoses, operations or procedures for each episode of care occurring in acute hospitals in Victoria [3].

We sought to identify non-notified cases of IPD by data-linkage between the surveillance and hospitalization datasets, and attempted to estimate cases potentially missed by both datasets using a capture–recapture process.

## METHODS

We performed data-linkage on two datasets that routinely collected information on IPD cases: (1) notifications of IPD received by the DHS with onset of illness between 1 July 2001 and 30 June 2003; and (2) all admissions with any of five identified pneumococcal-related illness ICD-10-AM codes in the VAED during the same period (hereafter called VAED admissions). Two of the five ICD-10-AM codes used were specific for IPD: G00.1 (meningitis due to *S. pneumoniae*) and A40.3 (septicaemia due to *S. pneumoniae*). The other three codes, included to maximize the identification of miscoded IPD cases, referred either to mucosal disease that may be associated with invasive disease: J13 (pneumonia due to *S. pneumoniae*) and streptococcal disease that may potentially include IPD (G00.2, streptococcal meningitis; B95.3, *S. pneumoniae* as the cause of disease classified to other ICD-10 chapters).

Prior to linkage, multiple episodes of care for each individual were collapsed into a single record with all admission date information preserved unless it was clear they related to a single episode (such as admission dates on consecutive days). If more than one of the identified ICD-10-AM codes (excluding combined A40.3 and J13) was recorded per individual record, we maintained a single ICD-10-AM code, using the hierarchical order: G00.1, A40.3, J13, G00.2 and B95.3.

Three rounds of deterministic data-linkage were performed, with each round requiring an exact match of a three-variable algorithm using combinations of four variables: name-code (first three letters of the first name, the name-code recorded in the VAED if the patient provides a Medicare card), date of birth,

residential postcode and hospital admission date. The rounds were applied sequentially, with only non-matched records included in subsequent rounds (Fig. 1). A link was only made where all three fields were in exact agreement. Where the name-code was missing the case was excluded from the linkage rounds using name-code, except in Algorithm 3, which allowed for either name-code or admission date to match. The sequential matching algorithms used were:

- Algorithm 1: date of birth AND sex AND postcode;
- Algorithm 2: name-code AND sex AND postcode;
- Algorithm 3: date of birth AND sex AND [name-code OR admission date].

All cases notified had been previously verified as laboratory-confirmed cases of IPD by DHS surveillance staff. To identify which of the non-linked VAED cases met the case definition for laboratory-confirmed IPD, a copy of any laboratory report confirming isolation of *S. pneumoniae* from a sterile site was requested from the admitting hospital.

Post-linkage matching, using all additional information obtained from the admitting hospital and diagnosing laboratory, was performed to ensure that all possible linked pairs were identified.

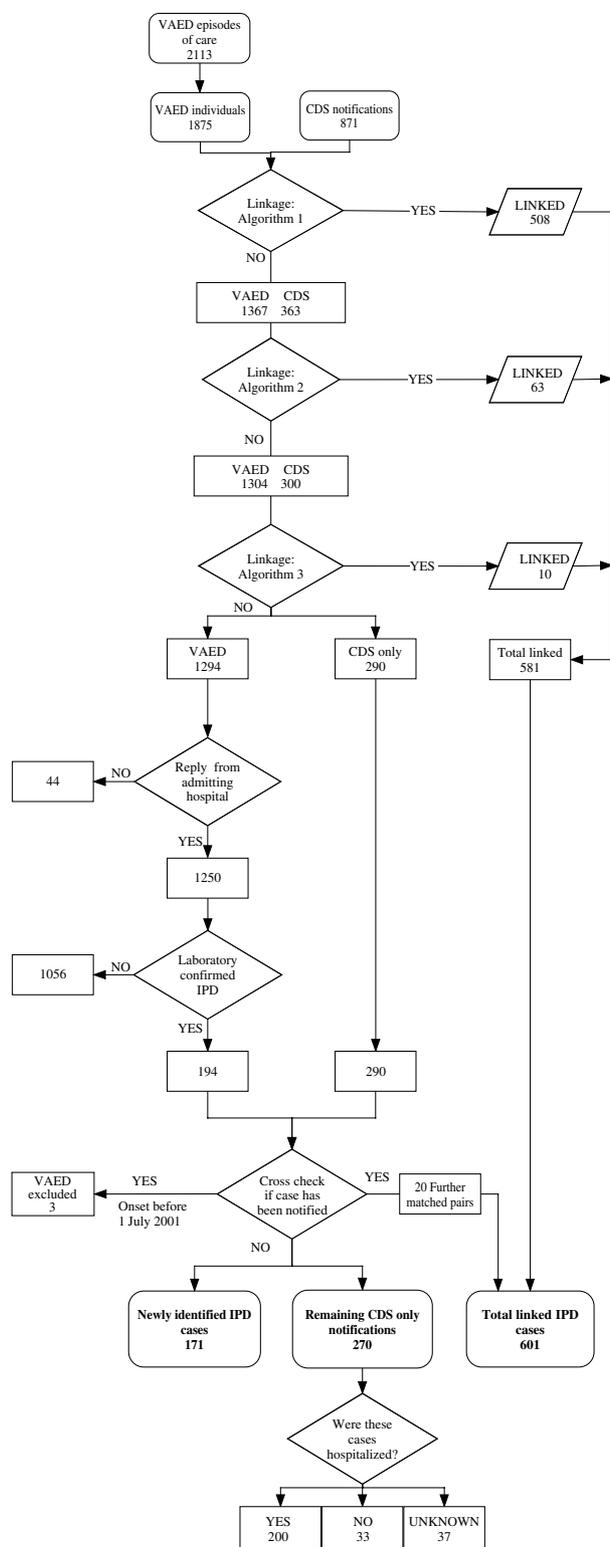
The new total number of cases and age-specific population rates were calculated to include the additional IPD cases identified following the data-linkage exercise. A two-source capture–recapture technique [4, 5] was then used to provide a projected estimate of the total incidence of IPD in Victoria. Population rates were calculated using the Australian Bureau of Statistics (ABS) estimated resident population for Victoria [6]. Average annual rates for the period were calculated using the average of the two mid-year (total and by age) estimated resident populations for 2001 and 2002 [6].

The sensitivity of the DHS IPD surveillance system for all identified IPD cases was calculated using both the data-linkage and capture–recapture estimated total IPD cases:

$$N = (a + c)(a + b) / a,$$

where  $a + c$  is the total number of cases ascertained from the primary data source,  $a + b$  is the total number of cases from the secondary data source, and  $a$  is the number of cases common to both sources.

The sensitivity and positive predictive value of the IPD-specific VAED ICD-10-AM codes (G00.1, A40.3) were calculated in relation to all identified IPD cases.



**Fig. 1.** Data-linkage and diagnosis verification process for IPD notifications and hospital admissions, Victoria, June 2001 to July 2003. VAED, Victorian Admitted Episodes Dataset; CDS, Communicable Diseases Section, Department of Human Services; IPD, invasive pneumococcal disease.

The DHS Human Research Ethics Committee granted approval for this study (reference number: 69/03). Original data extractions were performed in Excel. Data record linkage was conducted using STATA version 7, and statistical analyses were performed using STATA version 7 (StataCorp., College Station, TX, USA) and Epi-Info 6.04d (CDC, Atlanta, GA, USA).

### RESULTS

The DHS dataset recorded notifications for 871 individuals for the 2-year period; six cases, with onset dates between 1 July 2001 and 30 June 2003 were notified after 30 June 2003 and were included in this analysis. Of these 871 cases, 721 were recorded as being hospitalized for their IPD. In the VAED, there were 2113 episodes of care recorded with a relevant ICD-10-AM code in 1922 individuals. This reduced to 1875 individuals after 47 people with non-Victorian residential postcodes were removed.

Of the variables used in the matching algorithms, date of birth and gender were recorded for all records in both datasets. The first name name-code was missing for one individual in the DHS dataset and was unavailable for 286 (15%) records in the VAED. Admission date, consistently recorded in the VAED, was complete for 83% of the notifications received by the DHS, recorded as having been hospitalized.

Of the 1875 individuals with at least one VAED pneumococcal-related admitted episode, 31% (581/1875) could be linked to a DHS notification using the algorithms described (Fig. 1).

In seeking confirmation of IPD for the remaining 1294 non-linked VAED cases, 87% (104/120) of admitting hospitals responded, providing data for 96.6% (1250/1294) of cases. Diagnosis of IPD was verified by a laboratory report confirming isolation of *S. pneumoniae* from a sterile site for 15% of these (194/1294) cases (Fig. 1).

Of the 1056 hospitalizations excluded as IPD, isolation of *S. pneumoniae* from a non-sterile site (including sputum) was the main reason ( $n=796$ , 75%), followed by the isolated pathogen not being *S. pneumoniae* ( $n=91$ , 9%), clinical diagnosis only ( $n=79$ , 7%), no relevant pathology request found ( $n=30$ , 3%), culture results negative ( $n=25$ , 2%), a coding error identified by the responding hospital ( $n=19$ , 2%) and no record of the case found in the hospital medical records ( $n=16$ , 2%).

Table. VAED admission results for each stage; data-linkage, verification and post-linkage matching, by ICD-10-AM code

ICD-10-AM code	Data source	Data-linkage	Verification of IPD diagnosis for non-linked cases		After post-linkage matching		
	Individuals with a relevant VAED admission	Linked to a notification using algorithms 1–3	Response rate from admitting hospital		Confirmed IPD, not previously notified (% of non-linked, verified VAED admissions)		
G00.1	64	78.1% (50/64)	92.9%	(13/14)	1	7.7%	(1/13)
A40.3 only	286	66.8% (191/286)	94.7%	(90/95)	55	61.1%	(55/90)
A40.3+J13*	118	66.1% (78/118)	100.0%	(40/40)	30	75.0%	(30/40)
J13 only	938	22.0% (206/938)	96.3%	(705/732)	60	8.5%	(60/705)
G00.2	47	31.9% (15/47)	96.9%	(31/32)	5	16.1%	(5/31)
B95.3	422	9.7% (41/422)	97.4%	(371/381)	20	5.4%	(20/371)
Total	1875	31.0% (581/1875)	96.6%	(1250/1294)	171	13.7%	(171/1250)

VAED, Victorian Admitted Episodes Dataset; IPD, invasive pneumococcal disease.

Post-linkage matching, using additional information obtained from the admitting hospital and laboratory report, identified 23 of the 194 non-linked VAED IPD cases as previously notified (Fig. 1). Three of these cases had onset of illness prior to 1 June 2001 and were excluded from the analysis at this point. Thus, the complete data-linkage and verification process identified 171 IPD cases that had not previously been notified (Fig. 1).

All notifications to the DHS had been confirmed as having IPD, meaning that all linked cases were also confirmed as IPD, regardless of their VAED ICD-10-AM code. A high proportion of the VAED admissions with IPD-specific ICD-10-AM coded cases (G00.1 and A40.3) were directly linked to notifications (66.1% and 78.1%, respectively). In contrast, a much lower proportion of the non-specific codes were matched to notifications (between 9.7% and 31.9%) and yet these codes contributed half ( $n=85$ ) of the newly identified IPD cases (Table).

In total we identified 1042 IPD cases with onset between 1 July 2001 and 30 June 2003; 871 (84%) cases had been notified to the DHS and 171 (16%) cases, recorded as hospitalized with a pneumococcal-related illness, had not been notified. There were 270 cases notified to the DHS that were not linked to a VAED record, 200 (74%) were reported as being hospitalized, 33 (12%) reported as not hospitalized, and hospitalization status was unknown for 37 (14%) cases (Fig. 1).

The largest absolute increase in average annual age-specific IPD rate was seen in infants aged <1 year (75.0–82.5/100 000) and 1-year-old age groups

(83.4–88.4/100 000) (Fig. 2). The largest proportional increase (31%) in number of cases was seen among people in the 35–39 years age group; an increase in average annual rate from 4.5 to 6.5/100 000 over the 2 years.

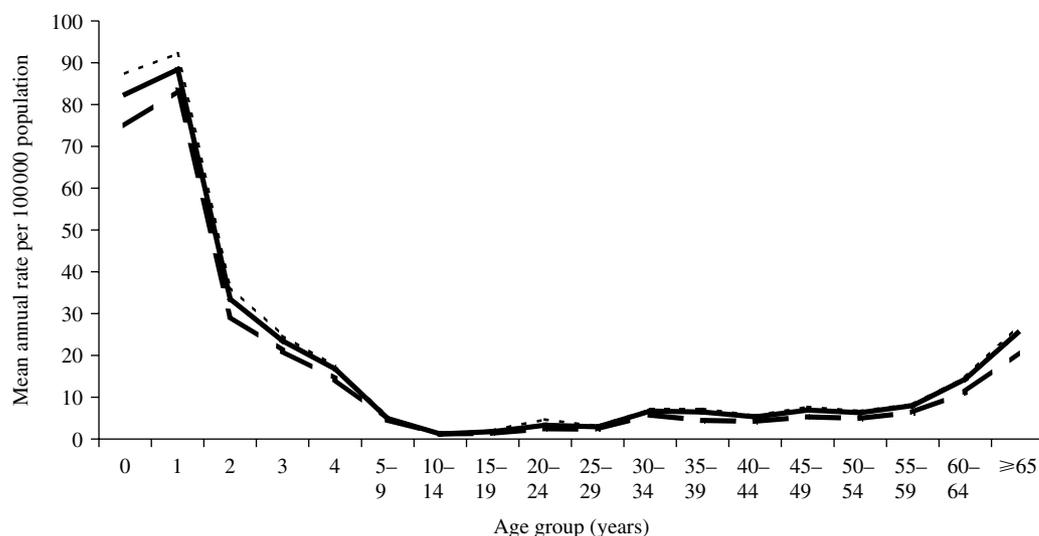
Capture–recapture estimated that an additional 77 IPD cases were missed by both data sources bringing the total estimate of IPD incidence for the 2-year review period to 1119 cases, the overall rate to 11.5/100 000 population, and age-specific rates possibly reaching 90.0/100 000 in children aged <2 years (Fig. 2).

In relation to all IPD cases identified in this study, the DHS surveillance system had a sensitivity of 84% (871/1042). When the further cases estimated by capture–recapture were included sensitivity fell to 78% (871/1119).

The IPD-specific ICD-10-AM codes (G00.1, A40.3) had a sensitivity for detecting all confirmed IPD cases identified in the VAED of 54%, with the positive predictive value of these codes being 90%.

## DISCUSSION

This study has shown current and ongoing deficits in the IPD surveillance system in Victoria. Case ascertainment via passive notification of IPD cases missed at least one-sixth (16.9%) of IPD cases for the first 2 years of surveillance. Using data-linkage we found the mean annual incidence rate for IPD in Victoria during the review period rose from 9.0 to 10.7/100 000 as a result of the additional hospitalized cases identified and rose even higher, to 11.5/100 000,



**Fig. 2.** Invasive pneumococcal disease incidence by age, Victoria June 2001 to July 2003. ---, Notifications; —, notifications + newly identified cases; -.-., notifications + newly identified cases + capture-recapture.

when the incidence was adjusted using capture-recapture.

Rates in children aged <2 years, using capture-recapture, rose to 90/100 000 population; a rate consistent with those reported in the neighbouring state of New South Wales [7]. This finding suggests that the lower rates previously reported in Victoria for this age group may be a function of case ascertainment rather than truly lower incidence [8]. In contrast to the increased rates seen in children, rates increased only marginally in persons aged  $\geq 65$  years: providing further supportive evidence of the success of Victoria's publicly funded pneumococcal vaccine programme in this age group [9].

Previous reports using hospitalization data to review the incidence of IPD or pneumococcal disease in Victoria have used the three pneumococcal-specific ICD-10-AM codes: G00.1 (pneumococcal meningitis), A40.3 (pneumococcal septicaemia) and J13 (pneumococcal pneumonia) without verifying the diagnosis or confirming J13 coded cases as invasive disease [10]. Our study included two additional non-pneumococcal-specific ICD-10-AM codes and diagnosis verification in order to identify miscoding of true IPD cases, and to validate hospital coding for hospitalized cases. Of the 171 newly identified IPD cases, 85 (50%) were only identified through use of non-IPD-specific codes. Most of these were found by investigating the code for pneumococcal pneumonia, with many non-notified cases coded for pneumococcal pneumonia having bacteraemic pneumonia.

It is important that hospital coding for pneumococcal disease clearly delineates when the organism has been isolated from a normally sterile site. This could be readily achieved by ensuring that when there is a laboratory report confirming isolation of *S. pneumoniae* from blood the ICD-10-AM code A40.3 is used instead of, or in addition to, the J13 code.

It has been estimated that only a small proportion of all IPD cases get diagnosed. Our study found that even when diagnosed, it is not always reported [11]. Proportionally, young adults were less likely to be notified as having IPD compared to children, however, there were larger absolute numbers of children, particularly aged <2 years, with IPD who were not notified. These cases are of particular interest given the increased risk of disease in this age group, and the recent availability of a conjugate vaccine against IPD-causing serotypes in Australian infants and children [12]. Given the often severe nature of IPD, non-notification of even small numbers of cases in children could impact on health economic and other assessments of a publicly funded vaccine programme. In order to adequately monitor the recent expansion of the pneumococcal vaccination programme to cover all Australian children aged <2 years with the seven-valent conjugate vaccine [13], it is important this deficiency in case ascertainment be addressed; studies to determine the reasons for the lack of notification may be warranted. Potential missed opportunities for notification should be identified, and improvements

could include: a system to automatically remind the treating doctor when a notifiable disease ICD-10-AM code is assigned to one of his/her patients upon hospital discharge, or provision of computer software, capable of being updated, to all laboratories that will identify whenever a notifiable disease is diagnosed and automatically forward a notification to the DHS. In the interim, notification may be improved by encouraging hospital coders to seek evidence of documented notification when they identify a notifiable condition.

Appropriate use of capture–recapture methods requires that several underlying assumptions be met [5, 14]. Diagnosis of IPD is by laboratory confirmation and is therefore both accurate and consistent. As diagnosis was verified for all DHS-notified cases as part of surveillance and we conducted follow-up to verify diagnosis of all non-notified hospitalized cases, we are confident that IPD cases have been correctly identified. Our use of conservative data-linkage, where we required exact matches and the algorithms used were not exhaustive of all possible combinations, minimized incorrect linkage; and the post-linkage matching ensured that all matched pairs were correctly assigned to the linked dataset. Apart from the possibility of movement of cases across state borders the two data sources sampled the same essentially closed population. Finally, both datasets required laboratory-confirmed diagnosis and, therefore, were not entirely independent. However, as non-independence causes an underestimation of the total incidence it is considered that use of the two-source capture–recapture method is still useful [15].

This study attempted to identify all hospitalized IPD cases. Given that in the 2 years under review there were 200 cases of confirmed IPD notified to the DHS as having been hospitalized but who were not identified on the VAED, there may be substantial numbers of other IPD cases not notified and not coded as a pneumococcal-related illness. Further options for research in this area should include obtaining the ICD-10-AM codes assigned to the 200 notified cases reported as hospitalized but who were not identified in the VAED in order to better understand coding issues and to perform an audit of selected laboratories to identify other non-notified, non-hospitalized IPD cases.

With the recently announced public funding of pneumococcal vaccine programmes for all children (conjugate vaccine) and the elderly (polysaccharide

vaccine) in Australia, robust and accurate surveillance will be important to better monitor changes in IPD epidemiology.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the health information units of each admitting hospital for their assistance in confirming the diagnosis of IPD. At the time of this study Hazel Clothier was a Masters of Applied Epidemiology Scholar at the National Centre for Epidemiology and Population Health, Australian National University. Her scholarship for this programme was provided by the Australian Commonwealth Department of Health and Ageing.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **Health (Infectious Diseases) Regulations. Statutory Rule 41/2001.** Victorian Legislation and Parliamentary Documents, Victoria ([www.dms.dpc.vic.gov.au](http://www.dms.dpc.vic.gov.au)). Accessed 13 October 2005.
2. **National Centre for Classification in Health.** *The International Statistical Classification of Disease and Related Health Problems, Tenth revision, Australian modification*, 3rd edn. Sydney: National Centre for Classification in Health, 2002.
3. **Acute Health Division.** The Victorian Admitted Episodes Dataset: An Overview, 2001. Victorian Government Department of Human Services, Melbourne (<http://www.health.vic.gov.au/hdss/archive/vaed/vaedover.pdf>). Accessed 13 October 2005.
4. **Hook EB, Regal RR.** Capture-recapture methods in epidemiology: methods and limitations. *Epidemiological Review* 1995; **17**: 243–264.
5. **Whitfield K, Kelly H.** Using the two-source capture-recapture method to estimate the incidence of acute flaccid paralysis in Victoria, Australia. *Bulletin of the World Health Organization* 2002; **80**: 846–851.
6. **Australian Bureau of Statistics.** Victoria in the future, estimated resident population by single year of age and sex, 2001 to 2051. Canberra, Australian Bureau of Statistics, 2004.
7. **McIntyre P, Gilmour R, Watson M.** Differences in the epidemiology of invasive pneumococcal disease, Metropolitan NSW, 1997–2001. *NSW Public Health Bulletin* 2003; **14**: 85–89.
8. **Liu M, et al.** Invasive pneumococcal disease among children in Victoria. *Communicable Disease Intelligence* 2003; **27**: 362–366.

9. **Andrews RM, et al.** Effectiveness of a publicly funded pneumococcal vaccination program against invasive pneumococcal disease among the elderly in Victoria, Australia. *Vaccine* 2004; **23**: 132–138.
10. **Hogg GG, Strachen JE, Lester RA.** Invasive pneumococcal disease in the population of Victoria. *Medical Journal of Australia* 2000; **173** (Suppl.): S32–S35.
11. **Fedson DS, Scott JA.** The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known. *Vaccine* 1999; **17** (Suppl. 1): S11–S18.
12. **Roche P, et al.** Invasive pneumococcal disease in Australia, 2003. *Communicable Disease Intelligence* 2004; **28**: 441–454.
13. **The Hon Tony Abbott MHR.** Media release: Commonwealth Government to provide pneumococcal vaccine for both young and old. 2004. Commonwealth Department of Health and Ageing (<http://www.health.gov.au/internet/ministers/publishing.nsf/content/health-mediarel-yr2004-ta-abb078.htm>). Accessed 13 October 2004.
14. **Tilling K.** Capture-recapture methods – useful or misleading? *International Journal of Epidemiology* 2001; **30**: 12–14.
15. **Brenner H.** Application of capture-recapture methods for disease monitoring: potential effects of imperfect record linkage. *Methods of Information in Medicine* 1994; **33**: 502–506.