RESEARCH COMMUNICATION

Searching for Cancer Deaths in Australia: National Death Index vs. Cancer Registries

Christina M Nagle[1]*, David M Purdie[2], Penelope M Webb[2], Adèle C Green[2], Christopher J Bain[1,2]

Abstract

    <u>Objective</u>: To compare the accuracy, costs and utility of using the National Death Index (NDI) and state-based cancer registries in determining the mortality status of a cohort of women diagnosed with ovarian cancer in the early 1990s. <u>Methods</u>: As part of a large prognostic study, identifying information on 822 women diagnosed with ovarian cancer between 1990 and 1993, was simultaneously submitted to the NDI and three state-based cancer registries to identify deceased women as of June 30, 1999. This was compared to the gold standard of 'definite deaths'. A comparative evaluation was also made of the time and costs associated with the two methods. <u>Results</u>: Of the 450 definite deaths in our cohort the NDI correctly identified 417 and all of the 372 women known to be alive (sensitivity 93%, specificity 100%). Inconsistencies in identifiers recorded in our cohort files, particularly names, were responsible for the majority of known deaths not matching with the NDI, and if eliminated would increase the sensitivity to 98%. The cancer registries correctly identified 431 of the 450 definite deaths (sensitivity 96%). The costs associated with the NDI search were the same as the cancer registry searches, but the cancer registries took two months longer to conduct the searches. <u>Conclusions and Implications</u>: This study indicates that the cancer registries are valuable, cost effective agencies for follow-up of mortality outcome in cancer cohorts, particularly where cohort members were residents of those states. For following large national cohorts the NDI provides additional information and flexibility when searching for deaths in Australia. This study also shows that women can be followed up for mortality with a high degree of accuracy using either service. Because each service makes a valuable contribution to the identification of deceased cancer subjects, both should be considered for optimal mortality follow-up in studies of cancer patients.

Key Words: Cancer mortality statistics - National Death Index - cancer registries

Introduction

    In Australia, the registration of death has been compulsory since about the mid 1850s (Australian Institute of Health and Welfare, 2005a). Information about death is recorded on a death certificate and, along with accompanying demographic and administrative information is registered with the Registrar General of Births, Deaths and Marriages in each state and territory. These data are then passed periodically to other repositories of vital records for, among other things, use in public health research and planning. One such repository, the National Death Index (NDI), is a particularly useful resource because it registers all deaths recorded in Australia since 1980 and can use up to 19 identifying criteria for matching purposes (Sadkowsky, 2001). It has consistently been shown to have high sensitivity and specificity for death ascertainment (Magliano et al., 2003; Kelman, 2000; Powers et al., 2000). Death data are

also passed periodically to other specialised repositories such as individual state and territory cancer registries.

    The completeness and quality of mortality data are crucial as a basis for good epidemiological research. It is therefore desirable that data repositories are regularly assessed for the quality of their data collection and retrieval procedures. We report here the comparative cost and utility of using the NDI and three state-based cancer registries during a follow-up study of women diagnosed with ovarian cancer in the early 1990s.

Materials and Methods

    The women in this study were participants in a large Australian case-control study of ovarian cancer conducted in the early 1990s (Purdie et al., 1995). The study population comprised 822 women (cases) with histologically confirmed incident epithelial ovarian cancer and 855 aged-matched

1. School of Population Health, University of Queensland 2. Queensland Institute of Medical Research *CORRESPONDING AUTHOR: University of Queensland, Medical School, Herston, Queensland 4006 Australia Fax +61-7-3365-5130 Email c.nagle@sph.uq.edu.au

controls. Cases were women treated in major gynaecology-oncology centres in New South Wales and Victoria between 1991 and 1992, and in Queensland from 1991 to 1993. As part of a follow-up study we tracked all the women who were diagnosed with ovarian cancer (cases) for mortality using both the NDI and state-based cancer registries.

National Death Index

The NDI is a computerised list of death records in Australia compiled by the Australian Institute of Health and Welfare (AIHW) (Australian Institute of Health and Welfare, 2005b). The NDI database includes all deaths since 1980 and is updated every two to three months from records of the Registrar-General, Births, Deaths and Marriages in the six Australian States and two Territories. At the time of this study, the NDI death records were complete through to the end of June 1999. After obtaining ethical clearance, NDI users submit information on as many as 19 potential matching variables (e.g. names, date of birth, sex, alternate names, marital status) from which returned records are ranked on the basis of a probabilistic scoring mechanism to determine the likelihood of a true match. The user must then decide which NDI records are associated with the subjects in question.

Cancer registries

Each Australian State and Territory maintains a cancer registry which provides information on incidence of and survival from cancer (Australian Institute of Health and Welfare, 2005c). Notification of malignant neoplasms is a statutory requirement in all Australian States and Territories for public and private hospitals, departments of radiation oncology, nursing homes, pathology laboratories, outpatients departments and day procedure centres (Australian Institute of Health and Welfare, 2005c). Mortality data are obtained periodically by abstraction from the files of the Registrar General, Births, Deaths and Marriages in each State and Territory. This is supplemented with information from hospital inpatient notifications and, in some States, on an ad hoc basis via links to other cancer registries and the NDI if it is known or suspected that the person with cancer has moved interstate. After obtaining ethical clearance from individual cancer registries, users submit information on matching variables (e.g. names, date of birth, sex, date of diagnosis/last contact) from which records are returned if a possible match is made. The user must then decide which records represent the subjects in question.

Comparison study

After obtaining ethical clearance from the AIHW and cancer registries, we searched for deaths in our cohort of ovarian cancer patients up to June 30, 1999. Using the NDI we searched for all deaths (n=822), and at each cancer registry we searched for deaths in the state where cases had been initially diagnosed e.g. we searched the Queensland Cancer Registry for the cases diagnosed in Queensland. The following identifiers were available from our cohort for

matching: surname, first name or initial, full date of birth, sex, date of diagnosis/last contact and address (none missing), hospital attended (2.5% missing), hospital identification number (26% missing), middle name or initial (79% missing), occupation code (6.2% missing) and country of birth (<1% missing).

The NDI used a probabilistic record linkage-matching package (Automatch) followed by clerical review to match our cohort to their records (Australian Institute of Health and Welfare, 2005b; Sadkowsky, 2001). The matching software followed predefined algorithms each with slightly different constraints by varying the variables that had to match (Sadkowsky, 2001). In the first pass surname, gender, and date of birth had to match precisely, with successive passes being more liberal (and matches less certain). Individual passes are not mutually exclusive so women with common names may be identified in several passes. The cancer registries use algorithms based on a combination of standard identifying information (names, sex, and date of birth) and matches were reviewed clerically by registry staff.

We tracked all the information on each search including numbers of hours worked, costs and total time. We classified NDI matches on all identifiers as 'probable matches'. 'Doubtful matches' were matches identified by the NDI using the less stringent algorithm that allowed for significant variation in identifying information (those with very low probabilistic scores). Matches with the cancer registries were all considered 'probable matches' because we did not have a probabilistic score to make an evaluation of the likelihood a match. We presumed that the remaining subjects without probable or doubtful matches were alive.

In our analyses we compared results from the NDI and cancer registry searches with our presumed gold standard 'definite deaths'. Definite deaths comprised deaths ascertained from multiple sources (NDI, cancer registries, hospital records and physicians reports), traced, reviewed and verified with a death certificate by two investigators (CN, CB).

Results

It took approximately 6 months to obtain ethics approval from the NDI and three state-based cancer registries. The NDI search took two months from ethics approval being granted to receiving results, compared to two to four months for the cancer registry searches. The total costs, excluding personnel time for the NDI to match the 822 women were $890 ($250 for the ethics approval and $640 for the matching) in 1999. One cancer registry charged a matching fee ($250) in 1999.

The NDI provided a list of matches for each study subject, a probabilistic score for each match and an ascribed ICD cause of death code. Each cancer registry returned a list of the names of matches. Death certificates were obtained directly from two of the cancer registries and from the Registrar, Births, Deaths and Marriages in the other state.

At June 30 1999 there had been 450 'definite deaths' in

Table 1. Comparison of 'Definite Deaths' and NDI and Cancer Registries Probable Deaths

		Definite deaths		
		Dead	Alive	
NDI	Dead	417	0	417
	Alive	33	372	405
		450	372	822

Sensitivity = 417/450 = 93%, Specificity 372/372 = 100%

		Definite deaths		
		Dead	Alive	
CANCER	Dead	431	0	431
REGISTRIES	Alive	19	372	391
		450	372	822

Sensitivity =431 /450 = 96%, Specificity 372 /372 = 100 %

our case group and 372 women remained alive. The NDI identified 439 potential matches with the 822 cases in our cohort. Of these 14 were repeat records, that is, individual women were matched with two (and in one case three) different deaths on the NDI and 1 match was with a male. The remaining 424 matches were reviewed clerically (CN, DP). 417 were classified as 'probable matches' and 7 as 'doubtful matches' (matches with very low probability scores and non ovarian cancer cause of death). Probable matches with the NDI, cancer registries and 'definite deaths' are compared in Table 1. The sensitivity of the NDI was 93% (417 matches of 450) and the specificity 100%. The cancer registries identified a total of 431 women from the cohort as deceased and after clerical review (CN, DP) all were classified as 'probable matches'. This equated to a sensitivity of 96% for the cancer registries.

Manual review of all deaths missed by the NDI and cancer registries was undertaken, including re-submission of records to each of the agencies for re-assessment. Review of the 33 deaths that failed to match to the NDI, revealed that 18 had not been identified by the NDI because names recorded in our cohort files were significantly different from those on the NDI (the use of aliases, surname changes and significant spelling variations) and 3 had differences in date of birth. A further 4 deaths had not been registered on the NDI at the time of matching (these women had all died in the 6 months prior to matching). One woman had died overseas and her death had not been registered in Australia and for the remaining 7 non-matches no obvious reasons were found. Recalculation of the sensitivity, removing the woman that had died overseas and the 4 deaths that had not been registered with the NDI from the denominator increased the sensitivity to 94% (417/445). Elimination of the non-matches resulting from variations in names and dates of birth in our cohort files could potentially increase the sensitivity to 98% (438/445).

Of the 19 deaths missed by the cancer registries, 10 had died in a state/territory other than that in which they were originally diagnosed. Three deaths were not matched because of discrepancies in date of birth and the remaining 6 non-matches were unexplained.

Figure 1 shows the overlap between the two sources of decedent information. Of all 'probable matches' the NDI and cancer registries agreed on 398 deaths (89% of deaths). The NDI identified 19 deaths which the cancer registries did not and the cancer registries identified 33 deaths that the NDI did not find.

Discussion and Conclusions

This study has demonstrated that the NDI and cancer registries were both able to provide reliable and accurate information regarding deaths for a cohort of female cancer patients. We found that the NDI correctly identified 93% of definite deaths and the cancer registries 96%. These results
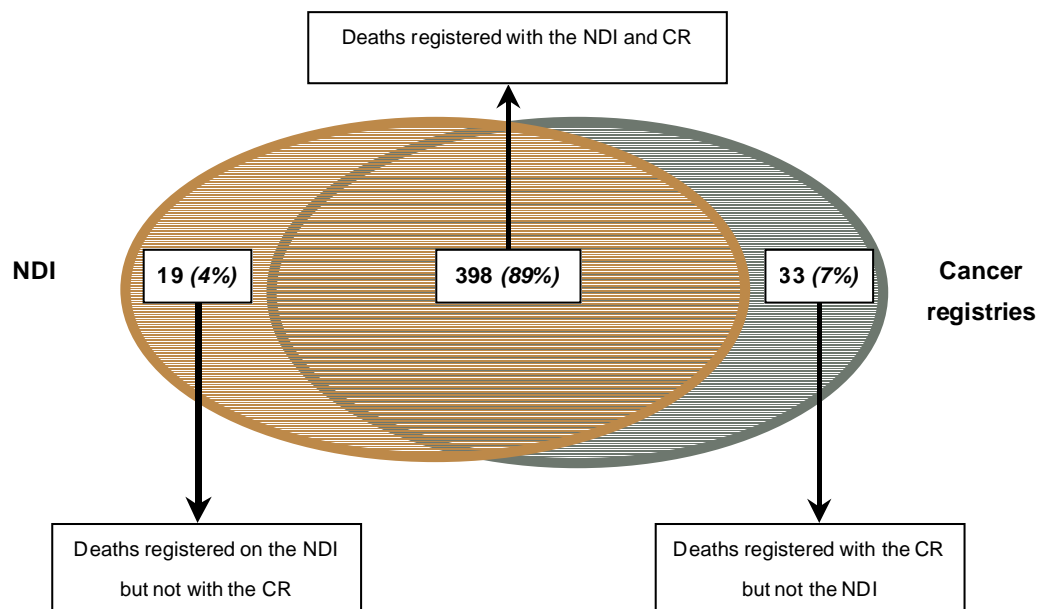
Figure 1. Comparison of Deaths Identified Using the NDI and Cancer Registries

are comparable to Powers et al. (2000) who compared mortality data from the Australian Longitudinal Study on Women's Health to the NDI in 12,432 women and reported a sensitivity of 95%. Our results and those of Power et al. (2000) provide evidence that mortality studies of women do not necessarily mean poorer ascertainment of death. Our results for the NDI are also as good as other Australian validation studies of men and women reported by Kelman (2000) and Magliano et al. (2003) who found sensitivities of 89% and 94% respectively. Kelman (2000) compared mortality information obtained from families with that of the NDI in a cohort of patients (n=2,990) who had received medical implants. Magliano and colleagues (2003) in their study of 5,821 Australian participants in the long-term intervention with Pravastatin in ischaemic disease study reported that the sensitivity increased from 94% to 98% when subjects with mis-recorded identifiers were excluded. Currently we have no validation studies to compare our cancer registry result with, but we assume the slightly better sensitivity for the cancer registries may be the result of rigorous manual rexaew of all matches by registry staff.

In determining the accuracy of mortality services such as the NDI, researchers must consider both the method of matching (algorithm) as well as the information in the databases. A matching algorithm may use single (i.e. social security number) or multiple (i.e. names and date of birth) matching criteria (Blakely and Salmond, 2002). Both the NDI and cancer registries used probabilistic record linkage and a multiple matching algorithm that allowed for minor discrepancy in names (limited to common variations) and date of birth. Name changes disproportionately involve women, and indeed patients whose names in the NDI and cancer registry files differed substantially from the names recorded previously in our cohort files were not successfully identified. This source of inaccuracy is potentially avoidable by using unique identifiers, such as the personal identification number used in Scandinavian countries or a de facto national personal identification number such as the social security number in the US (currently not available in Australia). A number of US studies have shown that the social security number is the most important determinant of sensitivity, and when used, the sensitivity of the NDI exceeds 95% (Cowper et al., 2002). Maintaining accurate and up to date personal identification information and using additional matching algorithms will also help, but the latter will likely be at the expense of specificity (Magliano et al, 2003).

Researchers must also consider the quality of the data in the databases. There were a number of reasons why deaths were not matched to the NDI or cancer registries and they do not necessarily reflect deficiencies of the databases themselves. The accuracy of death identification services vary with respect to time, marital status, race and, importantly, the data that each agency is supplied with (Wong, 2001). For example, inaccuracies in the reporting of age on death certificates are common (Hill et al., 2000). Currently the Registrars General of Births, Deaths and Marriages are trialling an electronic death certificate to facilitate the registration of deaths in Australia (Magliano et al., 2003). It is thought that this will improve the quality and accuracy of death data.

Although the amount of the time taken from approval of ethics applications to receiving the death data was a little longer for the cancer registries compared to the NDI, the costs associated with the searches were equivalent. Currently (in 2005) the NDI charges an Ethics Committee review fee of $250 (regardless of success of the application) and the service charges for manipulation and dealing with the data are $160 per hour or $1000 per day (Australian Institute of Health and welfare, 2005b). The pro rata charge for matching by one registries was $250 in 1999.

The NDI does however have several advantages over the cancer registries, including greater flexibility when conducting searches. It allows matching on many identifiers. It also provides the coded cause of death for matches and deals with interstate migration by identification of the state or territory in which the death occurred and the corresponding certificate number. Importantly it provides a probabilistic score to assist researchers in determining the likelihood of a true match.

When funding levels permit, both the NDI and cancer registries can be used to enhance the certainty of identifying deaths. For example, researchers relying solely on the cancer registries for mortality ascertainment will invariably fail to identify some proportion of actual decedents because of interstate migration. Because of the higher costs associated with searches of the NDI compared with cancer registries, cost-effective two stage search protocols might be developed that save more expensive NDI searches for study subjects not located during initial searches of the cancer registries.

This research indicates that local sources such as the cancer registries remain valuable agencies for short term follow-up of cancer cohorts, particularly if the cohort was state-based, but for those with more extensive budgets the NDI provides additional flexibility and information that the cancer registries do not.

Acknowledgments

References

Australian Institute of Health and Welfare. The classification of death data in Australia. Retrieved on 14.2.2005a from http://www.aihw.gov.au/mortality/history.cfm

Australian Institute of Health and Welfare. National Death Index. Retrieved on 7.2.2005b from http://www.aihw.gov.au/cancer/ndi/index.htm.l

Australian Institute of Health and Welfare. Australasian Association of Cancer Registries. Retrieved on 7.2.2005c from http://www.aihw.gov.au/cancer/accr/accr.html

Blakely T, Salmond C (2002). Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol*, **31**, 1246-52.

Cowper D, Kubal J, Maynard C, Hynes D (2002). A primer and comparative review of major U.S. mortality databases. *AEP*, **12**, 462-8.

Hill M, Preston S, Rosenwaike I (2000). Age reporting among white Americans aged 85+: results of a record linkage study. *Demography,* **37**, 175-86.

Kelman C (2000). The Australian National Death Index: an assessment of accuracy. *Aust N Z J Public Health*, **24**, 201-3.

Magliano D, Liew D, Pater H et al (2003). A. Accuracy of the Australian National Death Index: comparison with adjudicated fatal outcomes aogng Australian participants in the Long-term Intervention with Pravastatin in Ischaemic Disease (LIPID) study. *Aust N Z J Public Health*, **27**, 649-53.

Powers J, Ball J, Adamson L, Dobson A (2000). Effectiveness of the National Death Index for establishing the vital status of older women in the Australian Longitudinal Study on Women's Health. *Aust N Z J Public Health*, **24**, 526-8.

Purdie D, Green A, Bain C et al (1995). Reproductive and other factors and risk of epithelial ovarian cancer: an Australian case-control study. Survey of Women's Health Study Group. *Int J Cancer*, **62**, 678-84.

Sadkowsky K (2001). Epidemiological research using the National Cancer Statistics Clearing House and the National Death Index databases. *Australas Epidemiol*, **8**, 29-33.

Wong O (2001). Use of administrative databases in occupational and environmental epidemiologic research. *Annals Epidemiol*, **11**, 281-5.