The final version of this paper was published in *BMC Preg Childbirth* 2013, 13:37 DOI: 10.1186/1471-2393-13-37

Using hospital discharge data to identify incident pregnancy-associated cancers: a validation study

Yuen Yi (Cathy) Lee[1]

Email: calee@nsw.doh.health.gov.au

Christine L Roberts[1*]

* Corresponding author

Email: clroberts@med.usyd.edu.au

Jane Young[2]

Email: jane.young@sydney.edu.au

Timothy Dobbins[2]

Email: timonthy.dobbins@sydney.edu.au

[1]Clinical and Population Perinatal Health Research, Kolling Institute of Medical Research, University of Sydney, New South Wales, Australia

[2]Cancer Epidemiology and Services Research Group, University of Sydney, New South Wales, Australia

Authors

1. Name: Yuen Yi (Cathy) Lee

   Address: Kolling Institute of Medical Research

   Building 52, University of Sydney at Royal North Shore Hospital

   St Lenonards NSW 2006, Australia

   Email: calee@nsw.doh.health.gov.au

2. Name: Christine L Roberts

   Address: Kolling Institute of Medical Research

   Building 52, University of Sydney at Royal North Shore Hospital

   St Lenonards NSW 2006, Australia

   Email: clroberts@med.usyd.edu.au

3. Name: Jane Young

   Address: D02–QEII Research Institue for Mother and Babies

   The University of Sydney

   Sydney NSW 2006, Australia

   Email: jane.young@sydney.edu.au

4. Name: Tim Dobbins

   Address: D02–Coppleson Building

   The University of Sydney

   Sydney NSW 2006, Australia

   Email: timonthy.dobbins@sydney.edu.au

Abstract

Background

Pregnancy-associated cancer is associated with maternal morbidities and adverse pregnancy outcomes, and is reporting to be increasing. Hospital discharge data have the potential to provide timely information on cancer incidence, which is central to evaluation and improvement of clinical care for women. This study aimed to assess the accuracy and reliability of hospital data for identifying incident pregnancy-associated cancers compared with incident cancers from an Australian population-based statutory cancer registry.

Methods

Birth data from 2001–2008, comprised 470,277 women with 679,736 maternities, were linked to cancer registry and hospitalisation records to identify newly diagnosed cancers during pregnancy or within 12 months of delivery. Two hospital-identified cancer groups were examined; “all cancer hospitalisations”, which replicates a scenario where identification of individuals is not possible and hospitalisations are used as the unit of analysis, and “index cancer hospitalisation” (first cancer admission per woman per pregnancy).

Results

The incidence of pregnancy-associated cancer (according to cancer registry) was 145.4/100,000 maternities. Incidence of cancer was over- and under- estimated in “all cancer hospitalisations” and “index cancer hospitalisation” by 73.9% and 22.1% respectively. Overall, the sensitivity of “index cancer hospitalisation” was 60.4%, positive predictive value (PPV) 77.7%, specificity and negative predictive value both 100%. Melanoma ascertainment was only 36.1% and breast cancer 62.9%. For other common cancers sensitivities ranged from 72.1% to 78.6% and PPVs 56.4% to 87.3%.

Conclusion

Although hospital data provide another source of cancer identification, the ascertainment and accuracy are insufficient to obtain reliable cancer incidence estimates for the obstetric population.

Keywords

cancer; pregnancy; incidence; sensitivity; positive predictive value; validation

Background

Cancers associated with pregnancy are reported to be increasing, and this has been attributed to increasing maternal age and women's interaction with health services during pregnancy [1]. Women with cancer diagnosed during pregnancy or within 12 months of delivery are at increased risk of maternal morbidities and adverse pregnancy outcomes [1]. Current information on incidence of cancer associated with pregnancy is central to evaluation and improvement of clinical care for women.

Population-based statutory cancer registries are a reliable source for identifying incident cancers in populations [2-4]. However, the extensive quality assurance processes implemented by cancer registries can delay the timely availability of cancer data [5]. Consequently, there is increasing interest in the use of routinely collected and easily accessible administrative data, such as hospital discharge data, for identifying incident cancers and assessing health service utilisation and quality of care for cancer patients [2-4,6]. One disadvantage of using databases of hospitalisation records is that individuals with multiple hospitalisations cannot be identified. For example, several obstetric research studies in the United States (US) to date have relied solely on hospitalisation records, which means 100 records may come from 100 individuals or 10 individuals each admitted 10 times [7-9]. Further complications arise in determining pregnancy-associated cancers due to inability in identifying the duration of pregnancy in weeks of gestation, resulting in an imprecise pregnancy exposure period [7].

The quality of estimates of hospital-ascertained cancer incidence relies on complete and accurate ascertainment of recorded diagnoses. However, the validity and completeness of hospital data for identifying incident cancers associated with pregnancy have yet to be established. Several US studies have validated Medicaid or Medicare claims data (inpatient, outpatient and physician claims) for identification of incident cancers with cancer registries as the "gold standard", with sensitivities

ranging from 68% to 97% and positive predictive values ranging from 83% to 96% [10-15]. There have also been limited efforts to link private insurer claims to cancer registries, with focus on cancer treatment information [16]. The generalisability of such findings to specific populations (e.g., obstetric populations) remains questionable, given that the majority of studies are among elderly populations. Therefore the aim of this study was to determine the accuracy and reliability of hospital diagnoses for identification of incident pregnancy-associated cancers, both overall and by cancer type, compared with incident diagnoses from a population-based statutory cancer registry.

Methods

Data sources and study population

The study population comprised 470,277 women who gave birth in New South Wales (NSW) in the period 2001 to 2008, which corresponded to 679,736 maternities. NSW is the most populous state of Australia with a resident population of approximately 7 million people. Approximately one-third of all Australian births occur in NSW public or private hospitals.

Data were obtained from three linked NSW population databases: the Perinatal Data Collection (PDC), Admitted Patient Data Collection (APDC) and Central Cancer Registry (CCR). The PDC is a statutory surveillance system that includes births of at least 20 weeks gestation or at least 400 grams birth weight. The APDC (referred to as "hospital data") is a census of all inpatient hospitalisations to NSW public and private hospitals. The CCR (referred to as "registry data") is a statutory case-based registry of all newly diagnosed cancers in NSW since 1972 with the exception of non-melanoma skin cancers. Cancer diagnoses in the CCR are coded according to the 3rd edition of the International Classification of Diseases for Oncology [17]. The registry data are validated by a rigorous procedure to ensure the completeness and quality of data are high. Over 90% of cancers are verified by pathology and are confirmed as the primary cancer diagnosis. Less than 1% of cancers are ascertained from the death certificate [18]. Based on treatment categories, cancers were categorised into 13

clinical groupings [18]. For validation purposes, we used the registry data as the 'gold standard' of identification of incident cancers associated with pregnancy.

PDC birth records from 2001 to 2008 were linked to the cancer registry and hospitalisation records to identify newly diagnosed cancers during pregnancy or within 12 months of delivery (referred to as 'pregnancy-associated cancer') [1,19]. Linkage to the PDC birth data allowed accurate identification of the duration of pregnancy in completed weeks of gestation for each woman who gave birth. Record linkage of the three databases was carried out by the NSW Centre for Health Record Linkage using probabilistic record linkage methods [20,21]. This involves a process of blocking and matching combinations of selected variables with identifying information such as name, date of birth, address and hospital. Each match was assigned with a probability weight [22]. The probabilistic record linkage for this study is highly reliable with less than three in 1,000 false positive links and less than five in 1,000 missed links [20]. The researchers were provided anonymised data. The study was approved by the NSW population and Health Services Research Ethics Committee.

For identification of pregnancy-associated cancers from hospital data we searched up to 22 diagnosis fields associated with each pregnancy and postpartum hospital admission for cancer diagnoses. The diagnosis fields were coded according to the ICD-10-AM coded (10^{th} revision of the International Classification of Disease, Australian Modification) with cancer diagnoses indicated by codes of C00–C96 [23]. As our aim was to identify incident cancers, any hospitalisation record coded as secondary, in-situ or benign cancer was excluded from the analysis. Hospital-identified pregnancy-associated cancers were examined based on two groups: i) 'all cancer hospitalisations', refers to the total number of hospitalisations with a cancer diagnosis and women could have multiple hospitalisations during pregnancy.This replicates a scenario where identification of individuals is not possible and hospitalisations (records) are used as the unit of analysis and; ii) "index cancer hospitalisation", identified by taking the record with the earliest admission date for each cancer type. The latter replicates a scenario where individuals can be identified in the hospital data and the first record is considered to be an incident cancer.

Statistical analysis

The incidence estimated from 'all cancer hospitalisations' and the 'index cancer hospitalisation' were compared with the incidence estimated from registry data (both overall and by cancer type). The percentage differences of the numbers of incident pregnancy-associated cancers between registry data and hospital data were calculated. To assess the accuracy and reliability of the hospital identification of incident pregnancy-associated cancers only 'index cancer hospitalisation' data were compared with registry data. The following reporting characteristics with 95% exact binomial confidence interval (95%CI) were computed: sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Sensitivity (sometimes referred to as true positive fraction) is the proportion of maternities with an associated cancer, as ascertained by the registry gold standard, that were also identified in the hospital data, thus measuring completeness of identification. Specificity (equivalent to one minus false positive fraction) is the proportion of maternities without an associated cancer that were also not identified in the hospital data. PPV is the proportion of maternities with an associated cancer identified in the hospital data that in fact had an associated cancer, denoting accuracy. NPV is the proportion of maternities without an associated cancer identified in the hospital data that in fact did not have an associated cancer. We examined the false-positive and false-negative pregnancy-associated cancers in the hospital data (compared with registry data) to determine if some of these were prevalent cancers or had an inconsistent cancer type. To differentiate truly incident from false-positive cancers, we compared the hospital-identified cancers with registry data from 1994 onwards and regarded those previously notified cancers of the same type as prevalent cancers. Analysis was carried out in SAS, Version 9.2 (SAS Institute, Cary NC, USA) [24].

Results

Incidence of pregnancy-associated cancer from registry and hospital data

Between 2001 and 2008, a total of 988 pregnancy-associated cancers were identified from the registry data, corresponding to an overall crude incidence of 145.4 per 100,000 maternities (Table 1). The number and resultant incidence estimates of hospital-identified pregnancy-associated cancer were

notably higher among “all cancer hospitalisations” (252.7 per 100,000 maternities) and lower among the “index cancer hospitalisation” (113.3 per 100,000 maternities) (Table 1). Breast and lymphohaematopoeitic cancers resulted in large number of hospitalisations. The most common cancer in the registry data was melanoma with an incidence of 47.4 per 100,000 maternities, which was 55.9% higher than that reported in the “index cancer hospitalisation” (20.9 per 100,000 maternities). The next most common cancers were breast (32.1 per 100,000 maternities in registry data) and thyroid and other endocrine cancers (19.6 per 100,000 maternities), and these rates were also comparatively higher than that reported in the “index cancer hospitalisation” (22.4 and 17.5 per 100,000 maternities respectively) (Table 1).

Reporting characteristics of pregnancy-associated cancer

As the majority of women did not have a cancer associated with pregnancy, the specificities and negative predictive values of all cancers were close to 100%. Overall the sensitivity of hospital record identification of cancer was 60.4% and PPV was 77.7% (Table 2). The sensitivities for the majority of pregnancy-associated cancers were over 70.0%, except for melanoma (36.1%), breast (62.9%), colorectal (64.3%) and head and neck cancers (69.2%) (Table 2). Respiratory and urogenital cancers had 100% sensitivities, although their sample sizes were small. There was some variation in positive predictive values by cancer type. Breast, thyroid and other endocrine, gynaecological, neurological and head and neck cancers had PPVs ranging from 81.8% to 87.3%. For the remaining cancers, PPV ranged from 56.4% for lymphohaematopoeitic cancer to 79.9% for melanoma (Table 2).

Investigation of false-positive and false-negative pregnancy-associated cancers

Compared with registry data, there were 166 false-positive and 380 false-negative pregnancy-associated cancers among “index cancer hospitalisation” (Table 2). False-positive cancers were predominately lymphohaematopoeitic cancer (56.4%, n = 48) and false-negative cancers were predominately melanoma (53.7%, n = 204) (Table 2). Of the 166 false positives, registry data indicated that 86 (51.8%) were prevalent cancers and 22 (13.3%) were inaccurate/inconsistent identification of the cancer type. The cancers that were most frequently misclassified in the hospital

data were colorectal cancer (n = 7, misclassified as upper gastrointestinal cancer), followed by melanoma (n = 3, misclassified as breast or bone and other connective tissue cancers) and gynecological cancer (n = 4, misclassified as upper gastrointestinal cancer). False-positive prevalent cancers were primarily lymphohaematopoeitic (38 out of 86, 44.2%), melanoma (13, 15.1%), breast (14, 16.3%) and thyroid and other endocrine cancers (10, 11.6%).

Discussion

This record linkage-based validation study determined the accuracy of ICD-10-AM hospital diagnoses in estimating the incidence of pregnancy-associated cancer. We have demonstrated that using hospitalisations as the unit of analysis, rather than the individual, substantially over-estimated the incidence of pregnancy-associated cancer. In contrast, among individual-level hospital data the overall incidence was under-estimated by 22.1%. Specifically, the incidence of melanoma and breast cancer was under-estimated by 55.9% and 30.3% respectively. Pregnancy-associated melanoma was frequently missed in the hospital data, and was the major contributor to the false negatives. Other common cancers (breast, thyroid, gynaecological and lymphohaematopoeitic cancers) achieved only moderate levels of ascertainment (72.1–78.6%) and accuracy (56.4–87.3%).

There is no literature based on obstetric populations available for direct comparison of our reporting characteristics. Published studies have predominately assessed the use of Medicare claims data compared with medical records or cancer registries for ascertainment of selected cancers (mainly breast cancer) with varying sensitivities and positive predictive values, depending on the definitions used, the study timeframe and identification algorithms (first diagnostic code only, all diagnostic codes or a combination of diagnostic or surgical procedure codes) [10-15]. Studies using ICD-9 coded hospital discharge data to identify incident breast, colorectal and lung cancers with cancer registries as the “gold standard” achieved better reporting characteristics than in our obstetric population [2,15,25]. For example, for breast cancer the sensitivity was 77–85% PPV 57–91%, colorectal cancer: sensitivity 72% PPV 60–88% and lung cancer: sensitivity 81% PPV 59–79%. A study comparing ICD-10 coded

diagnoses in hospital discharge data with medical records as the "gold standard" reported a much greater ascertainment and accuracy (breast cancer: sensitivity 96% PPV 94%, colon cancer: sensitivity 93% PPV 96% and lung cancer: sensitivity 97% PPV 94%) than we report using ICD-10 for an obstetric population [26]. This is likely due to an older, non-specific hospitalised study population, which in general has a different pattern of hospital activity from an obstetric population. In NSW between 2001 and 2008, about 16% women with singleton pregnancy and 64% women with multifetal pregnancies were admitted to hospital at least once during 20–36 gestation weeks [27,28].

The evaluation of false-positive and false-negative pregnancy-associated cancers was insightful in considering the poorer identification of cancer from the hospital records of pregnant women. In accordance with literature, the common reasons for false positives were prevalent cancers or misclassification of cancer type [3]. Unfortunately, we were unable to evaluate the underlying causes of false-negative cancers. Others have speculated that the false-negative (missed) cancers may be due to coding errors, cancers notified from death certificate or other non hospital-based institutions [3]. In our obstetric population we believe outpatient cancer management (e.g., for melanoma) to be an important reason for false negatives. Furthermore, cancer may not be the reason for hospitalisation of a pregnant woman and diagnoses that are not relevant to the current admission are not required to be coded [29].

The strength of our study is the use of ICD-10 for coding hospital diagnoses. ICD-10 codes, based on surgical speciality, allow more detailed diagnoses coding than ICD-9 codes [30]. Importantly, there is a complete registration of cancers in the NSW statutory data collection, which provided the gold standard of identification of incident pregnancy-associated cancers without the need of an independent validation source. The linkage to cancer registry dating back to 1994 provided a unique opportunity to assess extensive history of cancer for identification of prevalent cancers, and to mitigate their impact on the positive predictive values [3,4]. Unlike other studies which focused on specific cancers, our comparison of reporting characteristics was done overall and by cancer type, and accounted for women with multiple primary cancers in a pregnancy. However, we were unable to

perform comparison by stratifying age or stage of cancer because of the limited sample size. Several limitations of our study also deserve consideration. The hospital data represent only inpatient stays; it may be less complete in capturing cancers where outpatient diagnosis and treatment are more common, e.g., melanoma. However, the data also provide a general representation of inpatient stays regardless of age or insurance status [15], and they are more assessable than the primary health utilisation data. Finally, as the reporting characteristics are sensitive to the sample size, careful interpretation is needed for cancers with small sample sizes.

Conclusion

The timely availability of population level data is a key factor in surveillance [31]. In 2012, NSW cancer registry data were available for linkage to the end of 2008, while hospital data were available through June 2011 [20]. Unfortunately, we do not consider the ascertainment and accuracy of pregnancy-associated cancers as determined by our hospital data of sufficient to be relied on for contemporary cancer incidence estimates.

Our study shows that the use of hospital data for identifying incident pregnancy-associated cancers achieved only moderate levels of ascertainment and accuracy. Although hospital data may provide another source of cancer identification for a cancer registry there will still need to be rigorous assurance to confirm cancers in order to obtain reliable estimates of incidence.

Competing interests

The authors declare no conflicts of interest.

Author's contributions

YYL took responsibility for the integrity of data, the design of study, the accuracy of data analysis and drafting of the manuscript. CLR participated in conception and study design, analysis and

interpretation of data, and drafting of the manuscript. TD and JY were involved in study design, acquisition of data and critical revision of the manuscript for important intellectual content.

Acknowledgements

The study was supported by the Australian Government National Collaborative Research Infrastructure Strategy's Population Health Research Network. Christine Roberts is supported by a NHMRC Senior Research Fellowship (#1021025). We wish to acknowledge the NSW Ministry of Health and NSW Central Cancer Registry in maintaining the population health data and the NSW Centre for Health Record Linkage for linking the datasets. The study was approved by the NSW Population and Health Services Research Ethics Committee on 06 October 2009 (Ref: 2009/08/172).

References

1. Lee YY, Roberts CL, Dobbins T, Starvou E, Black K, Morris JM, Young J: **Incidence and outcomes of pregnancy-associated cancer in Australia, 1994–2008: a population-based linkage study.** *An International Journal of Obstetrics & Gynaecology* 2012. DOI: 10.1111/j.1471-0528.2012.03475.x
2. Yuen E, Louis D, Cisbani L, Rabinowitz C, De Palma R, Maio V, Leoni M, Grilli R: **Using administrative data to identify and stage breast cancer cases: implications for assessing quality of care**. *Tumori* 2011, **97**(4):428-435.
3. Bernal-Delgado EE, Martos C, Martinez N, Chirlaque MD, Marquez M, Navarro C, Hernando L, Palomar J, Izarzugaza I, Larranaga N *et al*: **Is hospital discharge administrative data an appropriate source of information for cancer registries purposes? Some insights from four Spanish registries**. *BMC health services research* 2010, **10**:9.
4. Wang PS, Walker AM, Tsuang MT, Orav EJ, Levin R, Avorn J: **Finding incident breast cancer cases through US claims data and a state cancer registry**. *Cancer causes & control : CCC* 2001, **12**(3):257-265.

5. Izquierdo JN, Schoenbach VJ: **The potential and limitations of data from population-based state cancer registries**. *Am J Public Health* 2000, **90**(5):695-698.

6. Stavrou EP, Pesa N, Pearson SA: **Hospital discharge diagnostic and procedure codes for upper gastro-intestinal cancer: how accurate are they?** *BMC health services research* 2012, **12**(1):331.

7. Kuklina EV, Whiteman MK, Hillis SD, Jamieson DJ, Meikle SF, Posner SF, Marchbanks PA: **An enhanced method for identifying obstetric deliveries: implications for estimating maternal morbidity**. *Maternal and child health journal* 2008, **12**(4):469-477.

8. Chakravarty EF, Khanna D, Chung L: **Pregnancy outcomes in systemic sclerosis, primary pulmonary hypertension, and sickle cell disease**. *Obstet Gynecol* 2008, **111**(4):927-934.

9. Chakravarty EF, Nelson L, Krishnan E: **Obstetric hospitalizations in the United States for women with systemic lupus erythematosus and rheumatoid arthritis**. *Arthritis Rheum* 2006, **54**(3):899-907.

10. Nattinger AB, Laud PW, Bajorunaite R, Sparapani RA, Freeman JL: **An algorithm for the use of Medicare claims data to identify women with incident breast cancer**. *Health Serv Res* 2004, **39**(6 Pt 1):1733-1749.

11. Rolnick SJ, Hart G, Barton MB, Herrinton L, Flores SK, Paulsen KJ, Husson G, Harris EL, Geiger AM, Elmore JG *et al*: **Comparing breast cancer case identification using HMO computerized diagnostic data and SEER data**. *Am J Manag Care* 2004, **10**(4):257-262.

12. Koroukian SM, Cooper GS, Rimm AA: **Ability of Medicaid claims data to identify incident cases of breast cancer in the Ohio Medicaid population**. *Health Serv Res* 2003, **38**(3):947-960.

13. Cooper GS, Yuan Z, Stange KC, Dennis LK, Amini SB, Rimm AA: **The sensitivity of Medicare claims data for case ascertainment of six common cancers**. *Med Care* 1999, **37**(5):436-444.

14. Warren JL, Feuer E, Potosky AL, Riley GF, Lynch CF: **Use of Medicare hospital and physician data to assess breast cancer incidence**. *Med Care* 1999, **37**(5):445-456.

15. Penberthy L, McClish D, Pugh A, Smith W, Manning C, Retchin S: **Using hospital discharge files to enhance cancer surveillance**. *Am J Epidemiol* 2003, **158**(1):27-34.

16. Meguerditchian AN, Stewart A, Roistacher J, Watroba N, Cropp M, Edge SB: **Claims data linked to hospital registry data enhance evaluation of the quality of care of breast cancer**. *J Surg Oncol* 2010, **101**(7):593-599.

17. Percy C, Van Holten V, Muir Ce: **ICD-O- International Classification of Diseases for Oncology, 3ed edition**. Geneva: WHO; 1990.

18. Tracey E, Kerr T, Dobrovic A, Currow D: **Cancer In NSW: Incidence and Mortality Report 2008**. In. Sydney, NSW: Cancer Institute; 2010.

19. Smith LH, Danielsen B, Allen ME, Cress R: **Cancer associated with obstetric delivery: results of linkage with the California cancer registry**. *Am J Obstet Gynecol* 2003, **189**(4):1128-1135.

20. **Centre for Health Record Linkage** [http://www.cherel.org.au]

21. **Open Source ChoiceMaker Technology** [http://oscmt.sourceforge.net]

22. Bentley JP, Ford JB, Taylor LK, Irvine KA, Roberts CL: **Investigating linkage rates among probabilistically linked birth and hospitalization records**. *BMC medical research methodology* 2012, **12**(1):149.

23. **The International Statistical Classification of Diseases and Related Health Problems, Australian Modification – Tabular List of Diseases and Alphabetic Index of Diseases** [nccc.uow.edu.au/icd10am/icd10am/index.html]

24. **SAS (2010) SAS/STATA**. In., 9.2 edn. Cary, NC, USA: SAS International

25. Baldi I, Vicari P, Di Cuonzo D, Zanetti R, Pagano E, Rosato R, Sacerdote C, Segnan N, Merletti F, Ciccone G: **A high positive predictive value algorithm using hospital administrative data identified incident cancer cases**. *J Clin Epidemiol* 2008, **61**(4):373-379.

26. Henderson T, Shepheard J, Sundararajan V: **Quality of diagnosis and procedure coding in ICD-10 administrative data**. *Med Care* 2006, **44**(11):1011-1019.

27. Badgery-Parker T, Ford JB, Jenkins MG, Morris JM, Roberts CL: **Patterns and outcomes of preterm hospital admissions during pregnancy in NSW, 2001-2008**. *Med J Aust* 2012, **196**(4):261-265.

28. Badgery-Parker T, W. SA, Ford JB, Jenkins MG, Morris JM, Roberts CL: **Multifetal pregnancies: preterm admissions and outcomes.** *Aust Health Rev* 2012. http://doi.dx.org/10.1071/AH11106.

29. Lain SJ, Hadfield RM, Raynes-Greenow CH, Ford JB, Mealing NM, Algert CS, Roberts CL: **Quality of Data in Perinatal Population Health Databases: A Systematic Review**. *Med Care* 2011.

30. Roberts CL, Bell JC, Ford JB, Hadfield RM, Algert CS, Morris JM: **The accuracy of reporting of the hypertensive disorders of pregnancy in population health data**. *Hypertension and Pregnancy* 2008, **27**(3):285-297.

31. Teutsch SM, Thacker SB: **Planning a public health surveillance system**. *Epidemiol Bull* 1995, **16**(1):1-6.

Table 1. Pregnancy-associated cancer rates in the registry and hospital discharge datasets

Clinical group of cancer	Incidence of pregnancy-associated cancer by data source										
	Cancer Register[a]			All hospitalisations[b]				Index hospitalisation[b]			
	N	*%*	*Rate*	*N*	*%*	*Rate*	*%diff*	*N*	*%*	*Rate*	*%diff*
Melanoma	322	33	47.4	168	9.8	24.7	-47.8	142	18	20.9	-55.9
Breast	218	22	32.1	491	29	72.2	125.2	152	20	22.4	-30.3
Thyroid and other endocrine	133	14	19.6	198	12	29.1	48.9	119	16	17.5	-10.5
Gynaecological	94	9.5	13.8	161	9.4	23.7	71.3	81	11	11.9	-13.8
Lymphohaematopoeitic	87	8.8	12.8	340	20	50	290.8	110	14	16.2	26.4
Colorectal	35	3.5	5.1	44	2.6	6.5	25.7	22	2.9	3.2	-37.1
Neurological	22	2.2	3.2	57	3.3	8.4	159.1	24	3.1	3.5	9.1
Bone and other connective tissue	19	1.9	2.8	64	3.7	9.4	236.8	25	3.2	3.7	31.6
Head and Neck	14	1.4	2.1	13	0.8	1.9	-7.1	11	1.4	1.6	-21.4
Upper Gastrointestinal	18	1.8	2.6	124	7.2	18.2	588.9	52	6.8	7.7	188.9
Respiratory	7	0.7	1	17	1	2.5	142.9	9	1.2	1.3	28.6
Ill-defined and unknown primary sites	11	1.1	1.6	22	1.3	3.2	100	12	1.6	1.8	9.1
Urogenital	8	0.8	1.2	19	1.1	2.8	137.5	11	1.4	1.6	37.5
Any	988	100	145	1,718	100	253	73.9	770	100	113	-22.1

Rates are expressed per 100,000 pregnancies.

%diff , percentage difference of the number of incident pregnancy-associated cancers between registry and hospital data.

[a]Two women had multiple cancer notifications of different type in a pregnancy.

[b]Twenty-three women had multiple cancer hospitalisations of different type in a pregnancy.

Table 2. Reporting characteristics of incident pregnancy-associated cancer identified in hospital data compared with cancer registry data

Clinical group of cancer	Reporting characteristics of pregnancy-associated cancer						
	TP (*N*)	Sn (%)	95% CI (%)	FP (*N*)	PPV (%)	95% CI (%)	FN (*N*)
Melanoma	115	36.1	[30.8,41.6]	29	79.9	[72.4,86.1]	204
Breast	134	62.9	[56.0,69.4]	20	87.0	[80.7,91.9]	79
Thyroid and other endocrine	103	78.6	[70.6,85.3]	15	87.3	[79.9,92.7]	28
Gynaecological	70	77.8	[67.8,85.9]	15	82.4	[72.6,89.8]	20
Lymphohaematopoeitic	62	72.1	[61.4,81.2]	48	56.4	[46.6,65.8]	24
Colorectal	18	64.3	[44.1,81.4]	11	62.1	[42.3,79.3]	10
Neurological	20	90.9	[70.8,98.9]	4	83.3	[62.6,95.3]	2
Bone and other connective tissue	16	88.9	[65.3,98.6]	7	69.6	[47.1,86.8]	2
Head and Neck	9	69.2	[38.6,90.9]	2	81.8	[48.2,97.7]	4
Upper Gastrointestinal	15	83.3	[58.6,96.4]	5	75.0	[50.9,91.3]	3
Respiratory	6	100	[54.1,100]	3	66.7	[29.9,92.5]	0
Ill-defined and unknown primary sites	4	50.0	[15.7,84.3]	4	50.0	[15.7,84.3]	4
Urogenital	8	100	[63.1,100]	3	72.7	[39.0,94.0]	0
Any	580	60.4	[57.2,63.5]	166	77.7	[74.6,80.7]	380

TP: True positive; Sn: Sensitivity; TN: True negative; Sp: Specificity; FP: False positive; PPV: Positive predictive value; FN: False negative NPV: Negative predictive value.

Note: The specificities and negative predictive values of all cancers were close to 100%.