BMJ Open

Using probabilistic record linkage methods to identify Australian Indigenous women on the Queensland Pap Smear Register: the National Indigenous Cervical Screening Project

Lisa J Whop,[1] Abbey Diaz,[1] Peter Baade,[2] Gail Garvey,[1] Joan Cunningham,[1] Julia M L Brotherton,[3,4] Karen Canfell,[5,6] Patricia C Valery,[1,7] Dianne L O'Connell,[5] Catherine Taylor,[8] Suzanne P Moore,[1] John R Condon[1]

ABSTRACT

Objective: To evaluate the feasibility and reliability of record linkage of existing population-based data sets to determine Indigenous status among women receiving Pap smears. This method may allow for the first ever population measure of Australian Indigenous women's cervical screening participation rates.

Setting/participants: A linked data set of women aged 20–69 in the Queensland Pap Smear Register (PSR; 1999–2011) and Queensland Cancer Registry (QCR; 1997–2010) formed the Initial Study Cohort. Two extracts (1995–2011) were taken from Queensland public hospitals data (Queensland Hospital Admitted Patient Data Collection, QHAPDC) for women, aged 20–69, who had ever been identified as Indigenous (extract 1) and had a diagnosis or procedure code relating to cervical cancer (extract 2). The Initial Study Cohort was linked to extract 1, and women with cervical cancer in the initial cohort were linked to extract 2.

Outcome measures: The proportion of women in the Initial Cohort who linked with the extracts (true -pairs) is reported, as well as the proportion of potential pairs that required clerical review. After assigning Indigenous status from QHAPDC to the PSR, the proportion of women identified as Indigenous was calculated using 4 algorithms, and compared.

Results: There were 28 872 women (2.1%) from the Initial Study Cohort who matched to an ever Indigenous record in extract 1 (n=76 831). Women with cervical cancer in the Initial Study Cohort linked to 1385 (71%) records in extract 2. The proportion of Indigenous women ranged from 2.00% to 2.08% when using different algorithms to define Indigenous status. The Final Study Cohort included 1 372 823 women (PSR n=1 374 401; QCR n=1955), and 5 062 118 records.

Conclusions: Indigenous status in Queensland cervical screening data was successfully ascertained through record linkage, allowing for the crucial assessment of the current cervical screening programme for Indigenous women. Our study

Strengths and limitations of this study

- This study provides evidence that record linkage methodology can be used to identify Indigenous Australian women on Pap Smear Registers (PSRs).
- The validated linked data set obtained allows for evaluation of the National Cervical Screening Program for Indigenous Australian women for the first time since the inception a quarter of a century ago.
- While the linkage was deemed successful, it is inevitable that a small proportion of PSR records would not have successfully linked to the hospital data collection. Thus, some Indigenous women on the PSR may not have been identified as such. Interpretation of future results should bare this in mind.

highlights the need to include Indigenous status on Pap smear request and report forms in any renewed and redesigned cervical screening programme in Australia.

BACKGROUND

In 1991, Australia established an organised approach to cervical screening called the National Cervical Screening Program (NCSP).[1] The programme currently recommends that all women aged 20–69 years who have ever been sexually active are screened using the Papanicalou test (commonly abbreviated to Pap smear or Pap test) every 2 years to detect abnormal cell changes in the cervix. Each State and Territory implemented a Pap Smear Register (PSR) to record women's screening history and to provide a reminder function for women and their healthcare providers. Since the programme's inception, there has

been a 50% reduction in cervical cancer incidence and mortality,[2] resulting in one of the lowest incidence rates in the world.[3][4]

However, Australian Aboriginal and Torres Strait Islander women (hereafter respectfully referred to as Indigenous Australians) have a disproportionately higher burden of cervical cancer, with incidence nearly three times that of non-Indigenous women and mortality over four times higher.[2][5] It is not clear whether incidence has decreased nationally for Indigenous women in recent years. Despite their higher disease burden, cervical screening participation and outcomes are not known for Indigenous women. Two regional studies indicated that screening participation was considerably lower than national rates for Indigenous women in remote areas of Queensland and the Northern Territory,[6][7] but information using the standard performance measures of the NCSP for Indigenous women, such as participation or adequate follow-up of abnormalities, is lacking.

These outcomes cannot be measured directly by the NCSP because pathology forms that inform the PSRs do not record Indigenous status.[2] The issue of adding Indigenous status to pathology forms has been under discussion for some time, not only for the NCSP but for other areas in which pathology report forms are the main source of notification (eg, communicable diseases).[8][9] Despite recent progress,[9] change is still pending and it is likely to take many years before a change in the data collection mechanisms would produce high-quality data on Indigenous identification in PSRs.[10]

Record linkage may be a feasible and more immediate solution to the lack of Indigenous status in PSRs, as has been discussed in other contexts.[11] Record linkage, also referred to as data linkage, is the process of combining information within or across multiple sources relating to an individual.[12][13] Linking the PSR to a population-based data source that includes an accurate measure of Indigenous status has the potential to identify which women in the PSR are Indigenous.[10] In Australia, hospital inpatients data (also known as 'hospital separations data') are known to be reasonably accurate sources of Indigenous status for most jurisdictions; the most recent national data quality assessment in 2011–2012 found 88% agreement between hospital records and self-report, nationally.[14]

The *National Indigenous Cervical Screening Project (NICSP)* is utilising linkage to obtain a data set to assess participation in cervical screening and follow-up of abnormal Pap smear results among Indigenous Australian women compared with non-Indigenous women. Another aim is to examine how screening participation and follow-up of detected abnormalities are associated with survival among Indigenous and non-Indigenous Australian women diagnosed with cervical cancer.

The NICSP was designed as a national study, but due to the differing legislative and data custodian requirements across the country, the data are being obtained separately from each State and Territory. The purpose of this paper is to describe and evaluate the methods used for the Queensland component of the study (the first jurisdiction completed), which linked records in three existing Queensland Health data sets, and to compare and evaluate algorithms for defining Indigenous status.

METHODS
Data sources
The Queensland component of the NICSP involved the collection of data from three existing administrative data sets: the Queensland PSR, Queensland Cancer Registry (QCR) and the Queensland Hospital Admitted Patient Data Collection (QHAPDC).

The PSR, in operation since February 1999, collects information on Pap smears (ie, demographic details of the woman and the provider, date and results of each test) for all women who screen in Queensland (including interstate residents), and the date and results of follow-up diagnostic tests conducted after an abnormal Pap smear result. Although the PSR includes a data item for Indigenous status, these data are missing for the majority of records because they are not available from pathology reports.[15]

The QCR records information on all invasive cancer cases diagnosed among Queensland residents. Cancer notification and registration in Queensland has been a statutory requirement for all public and private hospitals, nursing homes and pathology services since 1982.[16] Data recorded in the QCR include date of cancer diagnosis, site and type of cancer, date and cause of death and Indigenous status. Indigenous status in the QCR is sourced from hospital notifications, death certificates and pathology when available. Completeness of Indigenous status in the QCR has been reported as 83.3%.[17]

The QHAPDC contains information on hospital episodes of care for patients admitted to Queensland public and private hospitals. However, only public hospital data were included in this study as the majority of Indigenous women in Queensland (approximately 96%) seek hospital care through the public system.[18] The QHAPDC does not include data on emergency department or outpatient clinic episodes. Information recorded in this collection includes clinical characteristics (eg, primary and other diagnoses, procedures occurring during the hospital admission), and demographic characteristics (eg, marital status, residential location, health insurance status and Indigenous status). The collection of Indigenous status is known to be reasonably accurate in Queensland public hospitals. In the most recent audit of Indigenous status data in 2011–2012 in Queensland, 87% (95% CI 84% to 91%) of

people who self-reported as being Indigenous were recorded as Indigenous in hospital records.[14]

Record linkage

Specifications for extracting records from each data source were developed and tested in consultation with the relevant data managers and custodians, who then extracted the data from the PSR, QCR and QHAPDC, and provided the extractions to the Queensland Record Linkage Group (QRLG). The QRLG utilised probabilistic record linkage as implemented by the LinkageWiz Data Matching Software (LinkageWiz Inc, Adelaide) with full name, sex, date of birth and address as matching variables to identify potential matching records of women between data sets. Potential matches on each variable were allocated a 'linkage weight' to indicate the probability that the match was a 'true match'. The linkage weights were predefined as the natural logarithm of the ratio of the frequency of agreement in linked pairs to the frequency of agreement in unlinked pairs. The LinkageWiz software classifies potential pairs weighted 11 and lower to be non-matches. While this can be adjusted on a study-by-study basis, the QRLG found this cut-off to be satisfactory and did not clerically review potential pairs with a weighting of 11 or under. Those with a weighting 12 and above were reviewed.

Given the number of records in each of the QHAPDC and the PSR and the resources available to the QRLG at the time, the QRLG was not able to conduct a probabilistic matching of the two complete data sets. For this reason, two specific QHAPDC extracts were used in the linkage (see figure 1, stage 1) instead of the entire QHAPDC data set. Extract 1 contained all episodes for women aged 20–69 years, within hospitals, who had ever been identified as Indigenous between 1995 and 2011 in the QHAPDC. For these women, records from hospitals where the woman never identified as Indigenous are not included. Extract 2 contained all episodes for women aged 20–69 years (regardless of Indigenous status) who were admitted to hospital between 1995 and 2011 with a diagnosis or procedure code related to cervical cancer. After these extracts were obtained, the second stage was implemented.

The second stage was to establish our Initial Study Cohort by linking records in the extracts from the PSR and the QCR (figure 1, stage 2). The Initial Study Cohort was defined as any woman aged 20–69 years who had at least one Pap smear recorded in the PSR between 1999 and 2011 and/or any woman registered in the QCR diagnosed with cervical cancer between 1997 and 2010, the latest available year at the time of record linkage. Once established, the Initial Study Cohort was linked to QHAPDC extract 1 to assign Indigenous status. We assumed women on the PSR who did not match to at least one QHAPDC record were not Indigenous and were assigned as non-Indigenous. This was then linked to extract 2 from the QHAPDC. Individual women were assigned a unique cohort identification number,

consistent across data sets. Personal identifiers were removed by the QRLG before the linked data sets were sent to the research team for analysis (figure 1, stage 3).

Various components of the linked data sets are being utilised to achieve the objectives of the NICSP. The PSR data with Indigenous status derived from QHAPDC are being used to calculate performance indicators for cervical screening for Indigenous women. QCR and PSR data with Indigenous status assigned are being used to examine incidence and survival of cervical cancer. QHAPDC data also provides information on comorbidities recorded in the hospital records. Factors associated with screening participation/outcomes and cancer incidence/survival, such as comorbidities (derived from the QHAPDC) and remoteness and socioeconomic status (derived from information in the PSR, QCR or QHAPDC), are also being investigated.

Assessment of the linkage quality

Linkage quality was assessed initially by the QRLG which required discussions with data custodians and the research team. First, key variables were checked for authenticity (eg, plausible dates of birth, females only, dates of Pap smears, address details, etc). Possible matches were either accepted or rejected after clerical review, and the total number of possible matches accepted as true matches or rejected matches were calculated at each probability score (see online supplementary file 1a–c). We would expect to see relatively few matches between the PSR and the QCR (as cervical cancer is a relatively rare outcome for women who have Pap smears), and a high number of matches between the QCR and QHAPDC extract 2, as most of these women diagnosed with cervical cancer would have been admitted to hospital for their cancer.

Definitions of the Indigenous status algorithms

There is a national standardised method of ascertaining and recording Indigenous status so as to maintain consistency across and within administrative data sets.[19] Algorithms for defining Indigenous status using linked data sets, including hospital inpatient data, have been developed by the Australian Institute of Health and Welfare (AIHW).[13] Such algorithms are necessary because Indigenous status for an individual may vary across records for the same individual.[11 20] Reasons for this have previously been articulated, including legitimate changes in identity, changes in reporting procedures over time and changes in perceived acceptance of identifying as an Indigenous person in mainstream institutions.[20 21]

We compared four algorithms to determine Indigenous status from multiple records, based on the AIHW guidelines and published evidence about the performance of different algorithms[13 22 23]:

Ever indigenous: a woman was coded as Indigenous if at least one of her QHAPDC records within the study period identifies her as Indigenous.
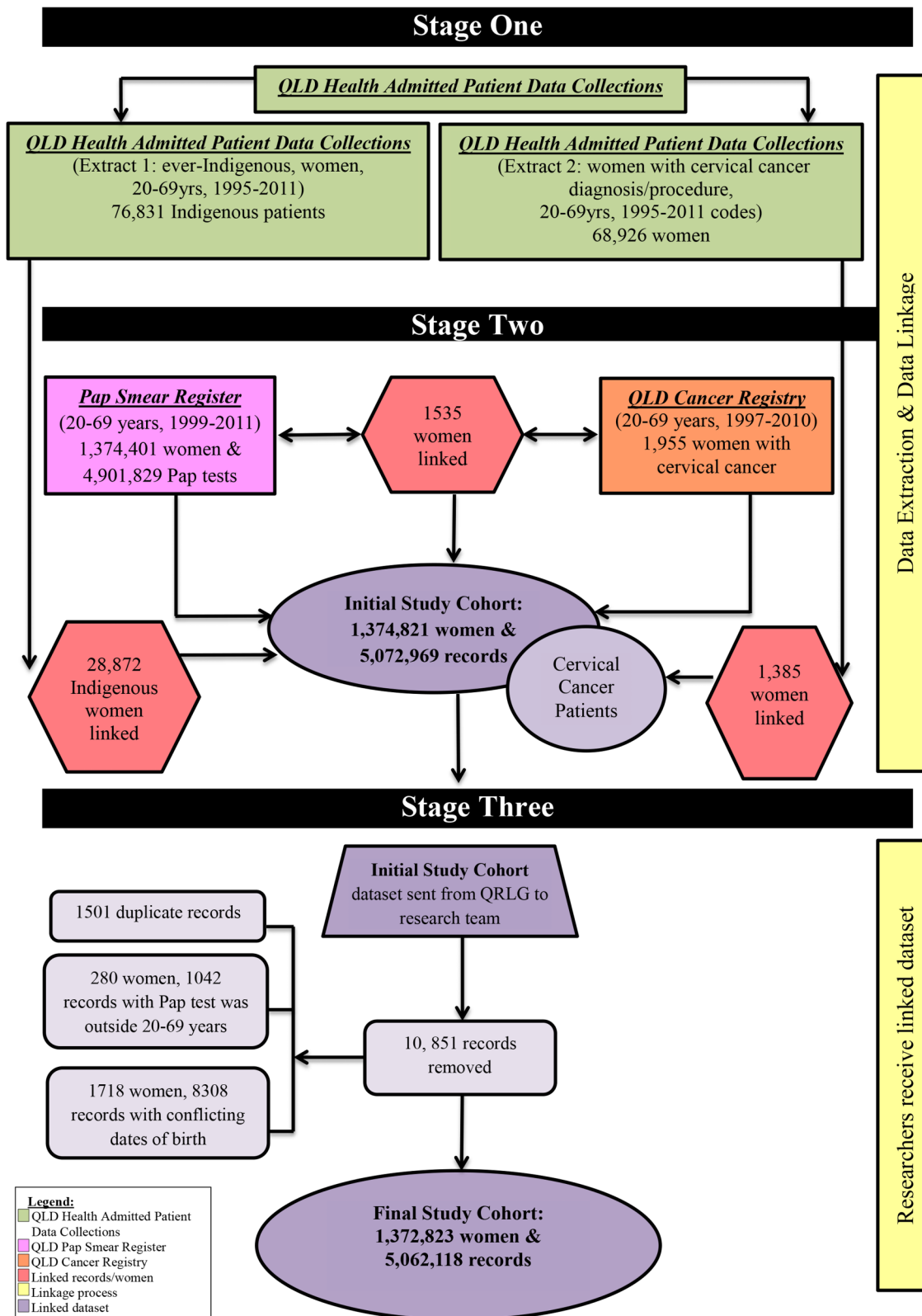
Figure 1 Queensland record linkage process. QLD, Queensland; QRLG, Queensland Record Linkage Group.

Most recent admission: a woman was coded as Indigenous if her most recent QHAPDC record in the study period identifies her as Indigenous.

Majority-based: a woman was coded as Indigenous if at least 50% of her QHAPDC records within the study period identify her as Indigenous.

Combination: a woman was counted as Indigenous if she was identified as Indigenous on *either* her most recent QHAPDC record or on at least 50% of her QHAPDC records within the study period.

Indigenous status in QHAPDC is classified as 'Aboriginal but not Torres Strait Islander origin', 'Torres Strait Islander but not Aboriginal origin', 'both Aboriginal and Torres Strait Islander origin', 'neither Aboriginal nor Torres Strait Islander origin' and 'not stated'. Prior to June 1998 Indigenous status was recorded with a different coding system. Therefore, Indigenous status was recoded as 'Indigenous' (defined as Aboriginal and/or Torres Strait Islander origin) and 'non-Indigenous' (neither Aboriginal nor Torres Strait Islander origin or not stated/unknown). Each woman in the PSR was assigned Indigenous status using each of the four algorithms. Women who did not link to extract 1 were assumed to be non-Indigenous. These variables were then merged into the PSR by cohort ID. The proportion of women identified as Indigenous on the PSR was calculated using each of these four algorithms and compared overall, and for 5-year age groups and remoteness groups based on place of residence. For context, the proportion of Indigenous women within the averaged estimated resident population (ERP) was also reported for 5-year age groups and remoteness categories. The proportion of women within our cohort identified as Indigenous was expected to be lower than the ERP regardless of algorithm use given that screening rates are not 100%. Records across all years were mapped to 2011 statistical local areas (SLAs) boundaries based on suburb and postcode. SLAs were then grouped according to level of geographic remoteness based on the Accessibility/Remoteness Index of Australia (ARIA+).[24]

Approvals

Approvals to access and link records were obtained from the Queensland Research Linkage Group (QRLG), data custodians of the included data sets and the Director General of Queensland Health. To facilitate the calculation of time-dependent Pap smear participation rates,[2] subsequent ethics and data custodian amendments were made to request additional information for screening history prior to age 20 for women within our cohort.

RESULTS
Approvals
From the initial ethics approval (Queensland Health) it took 19 months to obtain all relevant ethics and data custodian approvals, 5 months for the data to be extracted, and 3 months for the records to be linked and reviewed. Changes in the State government following a general election in March 2012 reduced the QRLG's resources, which halted progress for several months and prevented linkage of the entire QHAPDC data set. Consequently,

the linkage method was revised to use the two extracts. QRLG did not charge for record linkage.

Linked data set
The Initial Study Cohort was established by linking records from the PSR (n=1 374 401) and the QCR (n=1955), where 1535 women existed in both the PSR and QCR data sets. There were 28 872 women (2.1%) from the Initial Study Cohort who matched to an ever Indigenous record in extract 1 (n= 76 831). Women with cervical cancer from the Initial Study Cohort were then linked to extract 2 (n=68 926), resulting in 1385 (71%) women in the Initial Study Cohort being linked to a record relating to a cervical cancer diagnosis or procedure code. Researchers received the non-identifiable Initial Study Cohort data set containing 1 374 821 women and 5 072 969 records. After receiving the linked file, further records were removed by the research team (n=10 951). There were 693 records identified with the same cohort identification number, test date, provider number and result; these were deemed to be duplicates. A further 808 records were identified with the same cohort identification number, Pap smear date and, in most cases, the same provider number but with different results. These were unable to be verified through the PSR manager or laboratory and, therefore, were all excluded. A further 1042 records were excluded as the age at Pap smear was outside target group (20–69 years). Some women were found to have conflicting dates of birth across their records (n=8308); these were unable to be verified by the PSR and were removed. The Final Study Cohort included 1 372 823 women and 5 062 118 records.

Linkage quality
Linkage of PSR, QCR and QHAPDC
Potential matches which scored 12 or higher were deemed potential true matches and underwent clerical review by QRLG. The linking of the initial cohort (PSR to QCR) identified 47 432 potential matches of records, with 85% (n=40 449) rejected as not being a true match. The initial cohort linked to the QHAPDC extract 1 (women ever identified as Indigenous) had 102 342 potential matches of records identified; 95% were accepted as a true match. As expected, a high proportion of matches (3189 of 3249 potential matches) between women diagnosed with cervical cancer and women with cervical cancer-related diagnosis or procedure codes recorded in the QHAPDC extract 2 were accepted as true matches. The number of potential matches rejected or accepted at each weighting score for each part of the linkage is detailed in online supplementary file 1a–c.

Indigenous status algorithms
Depending on the algorithm, the proportion of Indigenous women in QHAPDC varied only slightly. There was an absolute difference of 0.08%, ranging from 2.00% (majority-based) to 2.08% (ever Indigenous;

Table 1 Indigenous status algorithms derived from applying Queensland Health Admitted Patient Data Collection's Indigenous status to the Queensland Pap Smear Register, 1999–2011

Algorithm	Indigenous*	Non-Indigenous†
Ever Indigenous	28 565 (2.08%)	1 343 838 (97.92%)
Most recent admission	27 677 (2.02%)	1 344 726 (97.98%)
Majority-based	27 444 (2.00%)	1 344 959 (98.00%)
Combination	28 124 (2.05%)	1 344 279 (97.95)

*Indigenous' (defined as Aboriginal and/or Torres Strait Islander origin).
†'Non-Indigenous' (neither Aboriginal nor Torres Strait Islander origin or not stated/unknown).

table 1). The algorithms produced similar proportions across 5-year age groups (figure 2) and remoteness of place of residence (figure 3) for the women's first Pap smear recorded in the past 5 years of the study period (2007–2011).

DISCUSSION
Cervical cancer is more common and more fatal for Australian Indigenous women than for the rest of the population.[2][5][25][26] Given that it is largely preventable through cervical screening, insight into participation by Indigenous women is of paramount public health importance. This study, through the linkage of three existing public health data collections, has been able to determine, with reasonable confidence, which women in the Queensland PSR are Indigenous. This will enable

state-wide and regional-specific screening participation rates for Australian Indigenous women to be estimated and reported in a forthcoming paper.

In Queensland, the record linkage approach provides a cost-effective way to overcome the shortcomings of other epidemiological studies, such as small sample size, attrition and difficulty in ascertaining data for vulnerable and marginalised groups.[27][28] One limitation of this study was that the linkage was not ongoing, and so only provides Indigenous identification information for the cohort of women included in our study and not for subsequent cohorts of women who screen for cervical cancer. Consideration should be given to determining the feasibility of ongoing record linkage in lieu of changes to the pathology forms.

As previously discussed,[10] our experience has been that the application and approval process for record linkage is complex, time consuming and often out of the researchers' control. Despite the challenges associated with record linkage, this project is critical given two decades of national reporting of Pap smear participation rates without any measurement of the performance of the national programme for Indigenous women. The benefits of utilising record linkage (despite complex administrative processes) have meant that: first, we were able to achieve whole of population data; and second, the individual's anonymity was able to be truly preserved, as is also reported in other record linkage studies,[27] thus overcoming the need for obtaining informed consent which was not feasible.
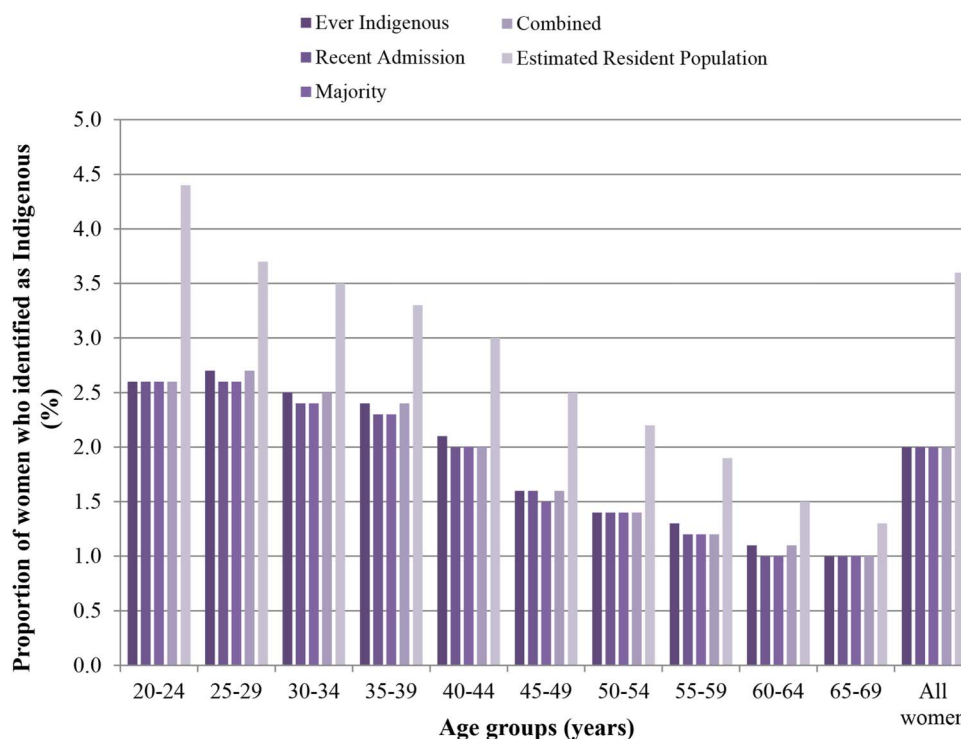

Figure 2 Proportion of women at first Pap smear during 2007–2011 who were identified as Indigenous using.
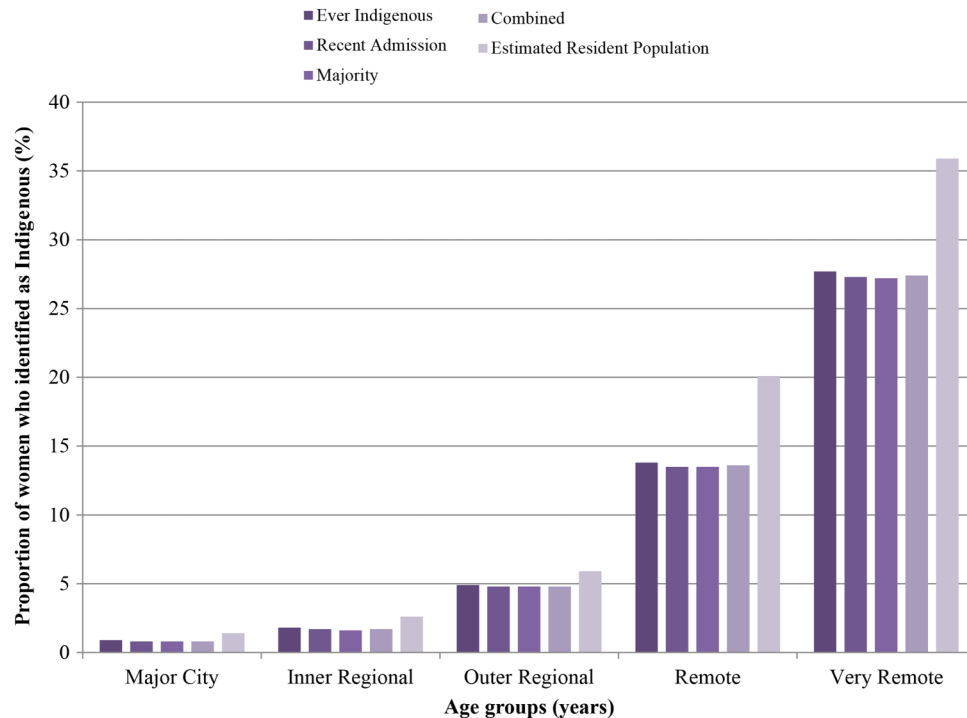
Figure 3 Proportion of women at first Pap smear during 2007–2011 who were identified as Indigenous using.

The success of this project heavily relied on a high match rate between the PSR and the QHAPDC. It is likely though that some Indigenous women in the PSR have not been identified as such—either through misclassification in the hospital records, failure to link to a QHAPDC record, or because they did not attend a public hospital during the study period. Consequently, this will lead to some outcome measures (such as participation rates) being underestimated for Indigenous women.

The known, reasonably high, accuracy of the Indigenous identifier contained in the QHAPDC (87% accuracy) was a major advantage of this study.[14 29] The accuracy in the QHAPDC, however, also varied by remoteness areas. The accuracy of the Indigenous identifier improved with increasing remoteness where major cities reported 72% (95% CI 62% to 80%) accuracy and remote/very remote reported 100% (95% CI 88% to 100%).[14] This means up to 13% of Indigenous women in our cohort overall or up to 28% Indigenous women in major cities may have been incorrectly identified as non-Indigenous or of unknown Indigenous status.[14] While we are unable to quantify the exact extent of misclassification bias in our study, sensitivity analyses using correction factors devised by the AIHW, will be performed for certain outcome measures to account for potential underidentification of Indigenous women for both overall Queensland estimates and by remoteness.[14]

In addition, there remains some uncertainty regarding how many Indigenous women in the PSR were not identified because their PSR record failed to link to their QHAPDC record. It is impossible to quantify these false negatives, because the QRLG did not review potential matches with a weighting lower than 12, as these are deemed too low to be a true match (LinkageWiz Data Matching Software, LinkageWiz Inc, Adelaide). There were many potential pairs with weights equal to 12 which were rejected, but, as the weights increased, the number of rejections decreased. Probabilistic matching based on weighted variables coupled with clerical review has been previously reported as a robust method and, therefore, we expect minimal false-negative matches within this study.[27]

Not all Indigenous women in the PSR would have been admitted to a public hospital during our study period and, as such, would not have been assigned Indigenous status through the record linkage process. In 2011, there was an estimated 50 189 Indigenous women who may have been eligible for inclusion in our study, but that does not include women who died, moved interstate or exclude women who may have had a hysterectomy before 2011.[30] Extract 1 contained 76 831 Indigenous women who were resident in Queensland and aged 20–69 at any time between June 1995 and December 2011. While we cannot estimate the number of women who were eligible for screening at any time between 1995 and 2011, the large excess of women in the extract indicates that a high proportion of eligible Indigenous women were available to be included in the linkage.

The high proportion of eligible women included in extract 1 is plausible given the 15-year time frame of

QHAPDC data. In addition, from the most recent national reports hospital separation rates are 2.3 times higher for Indigenous than non-Indigenous Australians (896 and 384 per 1000 population).[31] Of the separations for Indigenous Australians, 58% were for women, 91% were from public hospitals, and 74% were for those aged between 15 and 64 years.[31] Further, the hospital inpatient data collection in Western Australia, which has a state-wide unique client identifier for all hospitals, includes approximately 90% of the Western Australian adult female population (personal communication, D Rosman, 2012. Record linkage Unit Manager, WA Department of Health).

The time and effort that would have been required to obtain approvals for private hospital data far outweighed the benefit of including the few Indigenous women who would have been identified through this added process. We may have missed some women from our cohort because we only collected public hospital data; however, this would be small as 96% of hospital separations for Indigenous people in Queensland occur in the public sector.[18] Given that 70% of all women and 98% of Indigenous women who gave birth in Queensland during 2000–2009 did so in a public hospital,[32] we are confident that over the 17 years of QHAPDC collected, we will have ascertained as close to a population-based sample as possible. While we acknowledge that we may not have been able to capture all Indigenous women, we believe we have captured most using the best data and method that is currently available.

Indigenous identification using QHAPDC data, which may include multiple admission records for each patient, also relies on the algorithm used to define Indigenous status. There were minor differences in the proportion of women in the PSR who were identified as Indigenous using the different QHAPDC-based algorithms for Indigenous status (0.08% difference). Using a combination of the 'most recent admission' and the 'majority-based' algorithms, 2.05% of women in the PSR were identified as Indigenous. While this is lower than that for the Queensland Indigenous female population aged 20–69 years (2.95% of the Queensland female population aged 20–69 years[33]), this was expected based on our hypothesis of lower participation rates for Indigenous women.[6 7] Given that we may have missed some hospital records for some Indigenous women, the 'most recent' and 'majority-based' algorithms may overestimate the proportion of Indigenous women in the cohort.

It is possible that some eligible women may have been excluded from our study. For example, women who request to 'opt-off' are deleted from the Queensland PSR and are not included here or in any population statistics derived from the PSR. The proportion of screened women who opt-off the PSR has been reported as less than 1% in other states, but has not been reported for the Queensland PSR.[34 35] Similarly, women resident in Queensland but who had all their Pap smears during

the study period outside of Queensland will not be included in this study. Queensland resident women who were screened interstate and those whose screening history was not retained in the register will be included in the population denominator but not in the numerator of women who have screened, thus resulting in an underestimate of participation. In contrast, women who are not residents, yet were screened in Queensland will be counted in the numerator but not the population denominator (eg, women who live in border towns, such as Tweed Heads), which would overestimate the screening participation rate. Consideration of how these effects impact on the outcomes will be made when reporting the final results, including assessment of relevant sensitivity analyses.

In 2014, after recommendation by the Medical Services Advisory Committee (MSAC), the Australian Government announced a renewed cervical screening programme (known as the 'Renewal') will be implemented by May 2017.[1 36] The current implementation stage is concerned with, among other things, implementing a national data collection and register system. As a component of Renewal implementation the aim is to establish a national cervical screening register (real or virtual). Despite the difficulties in collecting information on Indigenous status at an individual level, we recommend that the work programme for the national screening register considers this important issue, which will ultimately facilitate better delivery of care to Indigenous women.

CONCLUSION

This study provides a proof of concept that record linkage can be used to identify Indigenous women in PSR data. The lack of an existing reliable and complete Indigenous identifier in the PSR to date has meant that the performance of the NCSP in Queensland, as in other Australian States and Territories, cannot be evaluated for Indigenous women using the PSR alone. Linkage of the PSR to the QHAPDC, which contains a reasonably accurate Indigenous identifier, does allow for such evaluation for the first time since the inception of the NCSP a quarter of a century ago. While this method can be used to produce reasonable estimates, it may not be a suitable long-term solution. Developing and implementing ongoing culturally safe and accurate ways to capture Indigenous identification in cervical screening registers must be a priority in the Renewal of the NCSP. The assessment of screening participation and outcomes for Indigenous women over the first two decades of the programme, which will be facilitated by the current linkage project, will thus provide a baseline for ongoing assessment of participation and outcomes for Indigenous women if the opportunity to consider the issues around possible collection of information on Indigenous status is implemented for the Renewed NCSP.

Author affiliations
[1]Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia
[2]Cancer Council Queensland, Fortitude Valley, Queensland, Australia
[3]Victorian Cytology Service, East Melbourne, Victoria, Australia
[4]School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
[5]Cancer Research Division, Cancer Council NSW, Sydney, New South Wales, Australia
[6]School of Public Health, Sydney Medical School, Sydney University, Sydney, New South Wales, Australia
[7]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
[8]Queensland Record Linkage Group, Queensland Health, Brisbane, Queensland, Australia

REFERENCES

1. Australian Government Department of Health. National Cervical Screening Program. Secondary National Cervical Screening Program 18 November 2013. 2013. http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/cervical-screening-1
2. Australian Institute of Health and Welfare. *Cervical screening in Australia 2012–2013*. Cancer series 93 Cat no CAN 91. Canberra: AIHW, 2015.
3. Ferlay J, Soerjomataram I, Dikshit R, *et al*. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 2015;136:E359–86.
4. Arbyn M, Castellsagué X, de Sanjosé S, *et al*. Worldwide burden of cervical cancer in 2008. *Ann Oncol* 2011;22:2675–86.
5. Zhang X, Condon JR, Rumbold AR, *et al*. Estimating cancer incidence in Indigenous Australians. *Aust N Z J Public Health* 2011;35:477–85.
6. Binns PL, Condon JR. Participation in cervical screening by Indigenous women in the Northern Territory: a longitudinal study. *Med J Aust* 2006;185:490–4.
7. Coory MD, Fagan PS, Muller JM, *et al*. Participation in cervical cancer screening by women in rural and remote Aboriginal and Torres Strait Islander communities in Queensland. *Med J Aust* 2002;177:544–7.
8. Acumen Alliance. *Aboriginal & Torres Strait Islander identifier on pathology forms: feasibility study into increasing the completeness of the Aboriginal and Torres Strait Islander identifier in ACT government registries*. Canberra: ACT Health, 2007.
9. Australian Institute of Health and Welfare. *The inclusion of Indigenous status on pathology request forms*. Cat. no. IHW 103. Canberra: AIHW, 2013.
10. Whop LJ, Cunningham J, Condon JR. How well is The National Cervical Screening Program performing for Indigenous Australian women? Why we don't really know, and what we can and should do about it. *Eur J Cancer Care (Engl)* 2014;23:716–20.
11. Thompson SC, Woods JA, Katzenellenbogen JM. The quality of indigenous identification in administrative health data in Australia: insights from studies using data linkage. *BMC Med Inform Decis Mak* 2012;12:133.
12. Queensland Health. *Queensland Data Linkage Framework*. Brisbane: Queensland Health, 2014.
13. Australian Institute of Health and Welfare and Australian Bureau of Statistics. *National best practice guidelines for data linkage activities relating to Aboriginal and Torres Strait Islander people*. Cat No IHW 74. Canberra: AIHW, 2012.
14. Australian Institute of Health and Welfare. *Indigenous identification in hospital separations data-Quality report*. Cat. no. IHW 90. Canberra: AIHW, 2013.
15. Australian Institute of Health and Welfare. *National summary of the 2001 and 2002 jurisdictional reports against the Aboriginal and Torres Strait Islander health performance indicators*. AIHW cat. no. IHW 12. Canberra: AIHW, 2004.
16. Parliamentary Counsel. *Public Health Act 2005*. Queensland.
17. Cramb SM, Garvey G, Valery PC, *et al*. The first year counts: cancer survival among Indigenous and non-Indigenous Queenslanders, 1997–2006. *Med J Aust* 2012;196:270–4.
18. Australian Institute of Health and Welfare. *Australian hospital statistics 2011–12*. Health services series no. 50. Cat. no. HSE 134. Canberra: AIHW, 2013.
19. Australian Institute of Health and Welfare. *National best practice guidelines for collecting Indigenous status in health data sets*. Cat. no. IHW 29. Canberra: AIHW, 2010.
20. Kennedy B, Howell S, Breckell C. *Indigenous identification in administrative data collections and the implications for reporting*

Indigenous health status. Brisbane: Health Statistics Centre, Queensland Health, 2009.

21. Australian Bureau of Statistics. *Perspectives on Aboriginal and Torres Strait Islander Identification in Selected Data Collection Contexts*. Australia: ABS, 2012.

22. Katzenellenbogen JM, Sanfilippo FM, Hobbs MS, *et al*. Incidence of and case fatality following acute myocardial infarction in Aboriginal and non-Aboriginal Western Australians (2000–2004): a linked data study. *Heart Lung Circ* 2010;19:717–25.

23. Neville SE, Taylor LK, Moore H, *et al*. Using linkage between hospital and ABS mortality data to enhance reporting of deaths among Aboriginal and Torres Strait Islander peoples. *Aust N Z J Public Health* 2011;35:543–8.

24. Australian Institute of Health and Welfare. *Rural, regional and remote health, a guide to remoteness classifications*. Canberra: AIHW, 2004.

25. Condon JR, Zhang X, Baade P, *et al*. Cancer survival for Aboriginal and Torres Strait Islander Australians: a national study of survival rates and excess mortality. *Popul Health Metr* 2014;12:1.

26. Australian Institute of Health and Welfare & Cancer Australia. *Cancer in Aboriginal and Torres Strait Islander peoples of Australia: an overview*. Cancer series no.78. Cat. no. CAN 75. Canberra: AIHW, 2013.

27. Holman CD, Bass AJ, Rosman DL, *et al*. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev* 2008;32: 766–77.

28. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol* 1996;25:435–42.

29. Australian Institute of Health and Welfare. *Indigenous identification in hospital separations data—quality report*. Health Services Series no 35 Cat no HSE 85. Canberra: AIHW, 2010.

30. Australian Bureau of Statistics. *3238.0—Estimates and projections, Aboriginal and Torres Strait Islander Australians 2001 to 2026*. Canberra: ABS, 2014.

31. Australian Institute of Health and Welfare. *Admitted patient care 2013–14: Australian hospital statistics*. Health services series no. 60. Cat. no. HSE 156. Canberra: AIHW, 2015.

32. Humphrey M. *Maternal and Perinatal Mortality and Morbidity in Queensland: Queensland Maternal and Perinatal Quality Council Report 2011*. Brisbane: Queensland Health, 2011.

33. Office of Economic and Statistical Research. *Census 2011: women in Queensland*. Brisbane: Queensland Treasury and Trade, 2012.

34. Victorian Cervical Cytology Registry. *Statistical report 2013*. Melbourne: Victorian Cervical Cytology Registry, 2013.

35. Cancer Institute NSW. Cervical cancer screening in New South Wales: annual statistical report 2009–2010. Sydney: Cancer Institute NSW, 2013.

36. MSAC (Medical Services Advisory Committee). *MSAC application no.1276: National Cervical Screening Program renewal*. Canberra: MSAC, 2014.