Population Based Ascertainment of Twins and their Siblings, Born in Western Australia 1980 to 1992, Through the Construction and Validation of a Maternally Linked Database of Siblings

Maxine L. Croft[1,2], Anne W. Read[2], Nicholas de Klerk[1,2], Janice Hansen[2], and Jennifer J. Kurinczuk[2,3]

[1]Department of Public Health, The University of Western Australia, Nedlands, Australia
[2]Centre for Child Health Research, The University of Western Australia, Telethon Institute for Child Health Research, Western Perth, Australia
[3]Department of Epidemiology and Public Health, University of Leicester, Leicester, United Kingdom

This paper describes the creation of a unique maternal identifier for use in the investigation of perinatal, postneonatal and child outcomes in relation to maternal characteristics. All Midwives' records of Western Australian (WA) births were routinely linked to registrations of births and deaths for infants born from 1980 to 1992 inclusive, then linked to WA hospital discharge data and to registries of birth defects and cerebral palsy to create a longitudinal health record for each infant. However, since each birth to a woman was recorded as a separate event, there was no way to identify siblings. Probabilistic record linkage, based on information about the mother, was used for this task. Logical inconsistencies within the data were used to test the validity of the linkages between birth records attributed to each mother. Information about the mother from other epidemiological studies and data abstracted from hospital case notes was also used to validate sibships. Linkage of the records of 310,255 births in WA during that period resulted in the formation of 181,133 sibships of one or more children. Pooling the results of all of the validation methods gave an error of 0.9%. Linkage identified 3678 sibships containing multiple births, and 305 sets of maternal twins. Ascertainment of twins and their siblings for an ongoing twin register, the WA Twin Child Health (WATCH) study, was a natural consequence of this process.

Perinatal mortality and morbidity in the Western Australian (WA) population from 1968 to 1975 inclusive were extensively documented by Stanley and Hobbs (1980). Partly as a consequence of their research, the Midwives' Notification of Birth System was implemented by the Health Department of Western Australia and Midwives' records of all WA births have been collected since 1980 (Gee, 1996). All births are recorded as separate events for each woman.

No unique personal identifier exists in Australia to link all births relating to each mother and hence population studies focusing on outcomes within families could not be conducted using these data. Ideally, each woman's records could have been simply linked through the use of her hospital assigned personal identifier (PID). However, on investigation, the PID was missing in over 40% of records. In addition, in some WA hospitals during that period the PID was not a reliable indicator of identity, since it may have been reassigned to a new patient at the end of a calendar year, if not in active use. Similarly, an alternate hospital record identifier, the mother's Unit Medical Record number was missing for 36% of records and may also have been reassigned to other patients. Simply linking records that agree on key fields is not practical for use in identifying sibships across a 13-year period since there may be many changes of address over time for the same woman (Bell, 1992). There is also a potential for women to change their surname following marriage, divorce and remarriage.

Selected groups of maternally linked sibships in WA were identified by Read et al. (1991) for a study of the recurrence of small-for-gestational-age births. However, this study focused only on a subset of the WA women giving birth during 1980 to 1986 inclusive. The limitations of the computer programs and equipment available to the investigators at that time precluded further large-scale research on sibship creation.

The overall aim of this project was to document the health of mothers during pregnancy for their entire reproductive life and to identify potential maternal risk factors for perinatal, infant and childhood morbidity and mortality in the infants born to those women. Thus, the specific aim of this data linkage study was to use probabilistic record linkage to link all of the Midwives' records of births in WA from 1980 to 1992 inclusive, in order to create a set of maternally linked sibships, including twins, and to validate those sibships. Ascertainment of twins and their siblings for an ongoing twin register was a natural consequence of the creation of these population data on sibships.

Address for correspondence: Maxine Croft, Telethon Institute for Child Health Research, PO Box 855, West Perth, WA 6872, Australia. Email: maxine@ichr.uwa.edu.au

This paper describes the construction and validation of the linked data set and the consequent creation of a population based register of all twins and higher order multiples and their siblings born in WA during that period.

Materials and Method

Data Sources

Data from the WA Midwives' Notification of Birth System comprising birth records of all infants born in WA from 1980 to 1998 inclusive, are held in the Maternal and Child Health Research Data Base of WA (MCHRDB) (Stanley et al., 1994; Stanley et al., 1997). The data used in the linkage project described in this paper relate to births from 1980 to 1992 inclusive. In WA, a birth is defined as an infant with a birth weight of 400 g or greater or a gestational age at delivery of 20 weeks or more (Gee, 1996). Computerized linkage of the midwife's record of pregnancy and birth to the WA Registrar General's record of the registrations of birth creates a composite birth record for each infant born in WA and its mother (Stanley et al., 1994; Stanley et al., 1997). However, each birth record held in the MCHRDB is separate so that second and subsequent births, including sets of twin and higher order multiples, are recorded as isolated events.

Record Linkage Methods

Probabilistic record linkage techniques were pioneered by Newcombe and Kennedy, (1962) while the mathematical foundations underlying the algorithms in use in the computer program used, Automatch (1996), were developed by Fellegi and Sunter (1969). Since the linkage of large data sets can result in comparisons of many millions of possible pairs, subsets of birth records were selected for comparison based on agreement on certain maternal characteristics such as date of birth, family and given names and place of residence. Only records within that subset, or block, were then compared with each other using the other variables available as match variables. To allow for errors in the data in each set of blocking and matching variables, multiple passes through the data were made, each time blocking and matching on different combinations of variables. The decision as to whether to match any pair of records was made based on the value of the match weight criteria (cutoff weights). Three cutoff weights were determined separately for each pass, one for automatic matching, one for passing to a clerical review, and one for rejection. Automatch's unduplication mode (the identification of duplicates in a single file) was used to identify matching records within the MCHRDB data. Records that generated a match weight above the match cutoff in any pass were identified as linked by Automatch. The final match strategy was chosen following clerical reviews of the results of preliminary data matching strategies and the pilot study.

Pilot linkage. The pilot record linkage involved linking into maternal sibships all births in WA from 1981 to 1985 inclusive (n =113,704). The match strategy consisted of seven passes through the data set, each time blocking on a different set of maternal characteristics.

Full linkage. The final linkage strategy, used on all 310,256 births from 1980 to 1992, was designed to overcome errors which could have resulted in mismatches or failure to match through the correct records never being present in the same block for comparison. The first pass blocked on the PID alone and then records that agreed on the PID were compared for agreement on other fields. Subsequent passes were designed to allow for recording errors. For example, the full date of birth was used as a blocking variable for three passes but the day number of the date of birth was allowed to vary on the fifth pass. Similarly, errors in the name or the address were tolerated through the five pass blocking strategy and by blocking on the first two characters of the maiden name in two passes rather than on the exact name. Changes of name or spelling errors resulting in differences in the first two characters were accommodated by blocking on other fields such as the PID, the hospital code or the postcode of residence (Table 1).

Within each pair of records that agreed on the blocking fields, all of the remaining fields were compared and a composite match weight was calculated for each potentially matched pair. The fields compared in each pass comprised those listed in Table 2, with the exception of the blocking fields for that pass on which there was known agreement. In order to identify those women who had changed their surnames between births, all of the potential surnames (surname, maiden name, infant's surname and mother's birth registration surname) in each birth record were treated as an array of four possible items. Each element of that array was compared to each element of the arrays of names in the other record. An array of given names was defined using fields derived from the midwife's record and the birth registration.

Table 1

Data Quality of Fields Used for Record Linkage of Births 1980 to 1992

Data Item	# Records with item missing (N = 310,255)	% Records item missing
Mother's hospital Personal Identifier (PID)	127,370	41.0
Mother's Unit Record Number (URN)	112,942	36.4
Mother's surname	199	< 1
Mother's first name	36	< 1
Maiden name	9001	2.9
Mother's middle name	208,397	67.2
Mother's full name (from the Registration of Birth)	114,321	36.5
Marriage date	156,864	50.6
Infant's surname	3733	1.2
Mother's height	3580	1.2
Mother's date of birth	338	< 1
Mother's country of birth	112,220	36.2
Mother's address	246	< 1
Mother's postcode	314	< 1
Hospital of birth	63	< 1

Table 2

Final Linkage Strategy (No. of Birth Records 1980–1992 = 310,255)

Pass	Blocking fields	# of matched pairs
1	Mother's Personal Identifier from hospital	61,716
2	Mother's date of birth First 2 characters of maiden name	62,469
3	Mother's date of birth Hospital of birth	2221
4	Mother's date of birth Postcode of residence	646
5	First 2 characters of maiden name Mother's year of birth Mother's month of birth	2010

An Automatch comparison type, which tolerates phonetic variations, character transpositions, insertions and deletion errors, and misspellings, was used whenever comparisons were made of the full text of the arrays of possible surnames and maiden name, the first names and the address. A date comparison, which recognizes differences in dates across month and year boundaries, was used to compare the maternal dates of birth, with the restriction that variation by more than one day resulted in a disagreement being recorded. All other match variables were compared as character or numeric variables.

The aim of a probabilistic record linkage strategy is the generation of match pair weights which very clearly discriminate between true matches (highly positive match weights) and non-matches (negative match weights), with as few potential pairs as possible requiring a clerical review (moderately positive match weights). Where the files being linked contain a rich source of information about each person it is possible to achieve this goal and sometimes to eliminate clerical reviews entirely. In this data set, despite the rich source of information about the mother, the time period spanned meant that many mothers had changed address and/or surname between birth events. Although a clerical review process during the pilot linkage led to the selection of the match cutoffs for each pass of the full linkage, the final match strategy still involved manual on-screen review of thousands of pairs of records.

Maternal Sisters, Sisters-in-Law and Twins

One of the first challenges encountered in determining the optimal matching strategy was the reduction of the occurrence of mismatches between sisters and between sisters-in-law. These women share personal characteristics such as maiden name and country of birth for sisters and surname for sisters-in-law. Some of the sisters were apparently twins, which increased the shared personal characteristics and thus the difficulty of assigning the correct infant(s) to the correct mother. Hence, clerical reviews were critical to the correct assignment of their births.

Validation Methods

Validation was essential in order to verify that the sibships formed did in fact reflect the true birth history for each woman, and to investigate potential false negative and false positive linkages. If errors were present, it was important to identify whether this reflected a bias in the record linkage method that could be addressed. Three methods were used to validate the linkages made, with the validation data obtained from questionnaires and medical records being regarded as the "Gold Standard" data sets (Figure 1).

Data sets from other studies. The first approach used data collected in other studies in which a complete birth history had been recorded, and was used to validate the pilot linkage. Four sub-sets of data were from studies in which women had given birth to at least two of their babies in WA from 1981 to 1985 inclusive (Figure 1). For each woman, the birth history described by her sibship data was compared to that recorded in the questionnaire data. Any discrepancies between the two sets of birth histories were recorded and reviewed to determine whether errors of linkage were present.

Medical records review. This approach used five sets of random samples of women, who were selected from those having sibships in which at least one birth had occurred at the tertiary obstetric and gynaecology hospital in WA. A manual search of the medical record(s) for each selected woman was carried out by a nurse researcher who was blind to the number of links made for that woman. Two stratified random samples were drawn from the set of linked sibships data, 50 women whose linkage indicated that they had only had one birth and 100 women whose linkage indicated they had had more than one birth. Three sets of stratified random samples (of 50 women) were chosen from the set of all women who were common to both the sibship linkage and a related record linkage study of the association between the mode of delivery and the risk of a subsequent hysterectomy (the gynaecological morbidity study). This latter set of 22,747 women was identified as the women whose data on sibships had been linked to hospital gynaecological discharge records, and for whom at least one birth and one subsequent hospitalisation had occurred at the same tertiary centre in WA.

Logical inconsistency checks. The third method involved using programs written in SAS to identify logical inconsistencies within all of the sibships in the pilot linkage. For example, identifying those women who appeared to have given birth within 180 days or less of a prior birth or whose height increased or decreased by more than five cms between births. A further potential indicator of errors was a discrepancy of more than one year in the period between consecutive births defined by the maternal age recorded at consecutive births, when compared to the actual number of elapsed years as measured by the dates of birth. The sibships of every woman who had given birth to more than five children during the 13-year period were checked. Similarly, those sibships where the woman's parity compared with the number of previous issue recorded at consecutive births was inappropriate for a set of correctly matched births were reviewed. Following linkage of the data on sibships to hospital admissions for the subset of women in the gynaecological study, a SAS program was used to identify further logical inconsistencies. These included identifying: women who delivered an infant or were admitted to hospital after their death; any birth, dilation and curettage or hysterectomy occurring after a hysterectomy; and any admission to a hospital occurring while the woman was still a patient in a different hospital. Further validation of the 1980 to 1992 sibships by two independent programmers identified those where the woman's given names and date of birth differed within the sibship.

Missing data rates. Missing data rates were calculated separately for variables used in the linkage of records of multiple births and for women with poor pregnancy outcomes such as preterm birth (less than 37 weeks gestation), low birth weight (less than 2500 gms) and stillbirth. These missing data rates were then compared with those present in the records of all other births (Table 3).

Twin and Higher-order Multiple Infants

All women, for whom more than one birth had occurred within an unusually short interval, were identified as potentially being the mothers of multiples. Extensive clerical reviews were conducted to ensure that these sibships contained true multiple births and not linkage errors. This enabled the creation of a population-based register of all multiple births in WA. The current postal address for each mother of multiples was obtained through probabilistic record linkage of the maternal demographic information from up to 18 years ago to the most recent WA electoral

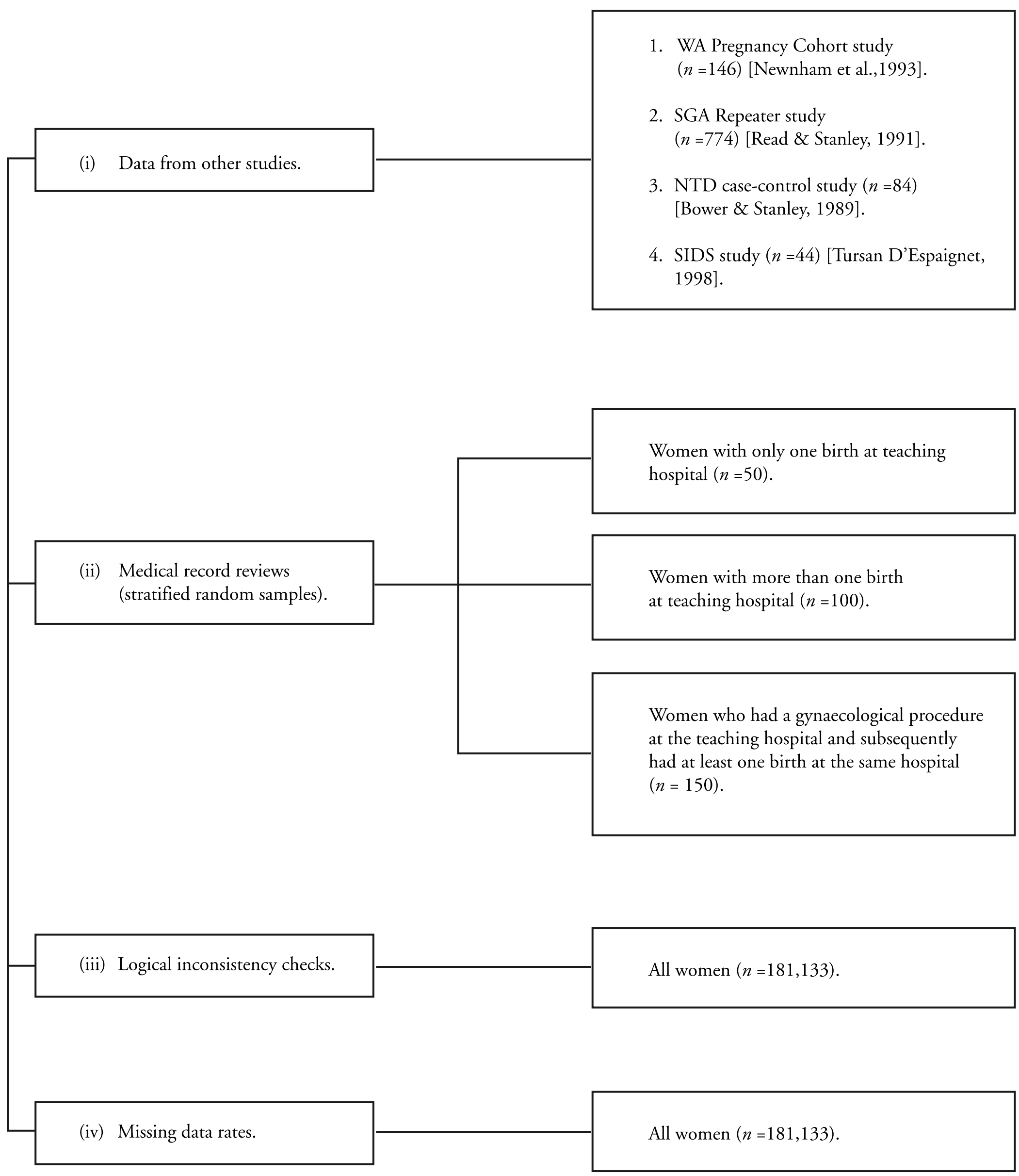

Figure 1
Validation Processes

https://doi.org/10.1375/twin.5.5.317 Published online by Cambridge University Press

Table 3

The Number of Sibships Formed and the Size of the Sibships

Number of births per sibship	Number of sibships	% of total sibships
One	89,892	49.6
Two	61,759	34.1
Three	22,716	12.5
Four	5,456	3.0
Five	1,026	< 1
Six	223	< 1
Seven	50	< 1
Eight	11	< 1
Total sibships	181,133	100

roll. The linkage techniques used to link birth information to the electoral roll were modelled on those used to identify siblings. Women who had given birth to one or more sets of multiples during 1980 to 1991 inclusive, were then sent a postal invitation to join a study of multiple births, the WA Twin Child Health (WATCH) Study (Hansen et al., 2000). Data in the MCHRDB enabled us to identify women whose infants or children had died and no attempt was made to contact those women either for this register or for a subsequent study into asthma. Those families were invited to participate in a separate study into bereavement in multiple birth families (Swanson et al., 2002).

It was possible that the responders to the WATCH study questionnaire differed systematically from the women who either did not respond or who did respond but declined to participate in the study. Hence responder mothers were compared with those who did not respond and with those who declined to participate in the study. These groups were compared in terms of the: index of social and economic adversity code (1998) for the mother's area of residence at the first birth; maternal and paternal age at the first birth; rural or metropolitan area of residence at the first birth; and whether the mother was recorded as being Indigenous or non-Indigenous. The mothers of the twins and their siblings were asked to complete questionnaires relating to the prevalence of asthma and atopy. The exposure to passive smoking of children in WA multiple birth families was measured and related to the reported prevalence of asthma.

The current project is recruiting families with seven to ten year old twins into the WATCH for Asthma study in order to measure the asthma phenotype. Specifically, via measures of atopy, lung function, expired nitric oxide and airway responsiveness to methacholine challenge. To assess the role of hereditary factors in determining the development of asthma, DNA samples will be collected and members of families in which one twin has asthma will be compared to families in which neither or both twins have asthma.

Ethical Issues

Record linkage to create the sibships was conducted during a doctoral program of research (Croft, thesis in preparation) undertaken as part of a major program of Epidemiological Studies in Maternal and Child Health. This study has the approval of the Confidentiality of Health Information Committee of the WA Health Department. The WATCH study and the WATCH for Asthma Study were approved by the joint Ethics committee of the Princess Margaret Hospital for Children and the King Edward Memorial Hospital.

Results

Linkage

Linkage of the 310,255 birth records of infants born from 1980 to 1992 inclusive, using the match strategy shown in Table 2, resulted in 129,062 unique sibships being automatically identified. Agreement of many items of data with low match weights (such as postcode, hospital, race, country of birth) may, in combination, create a match weight high enough to result in automatic matching of records belonging to different women. This was addressed by clerical review processes, which reduced the number of unique sibships with two or more children to 91,241, with family structure as shown in Table 4. The remaining 89,892 unlinked birth records belonged to women who had apparently delivered only one child in WA between 1980 and 1992. These women may have only ever had one child, may have had one or more previous children born outside WA, or in WA before 1980, or had subsequent children who were born after 1992. It is also possible that some of the women may have had another birth in WA, but the two (or more) records of those births were not linked.

Table 4

Percentage of Records Where Record Linkage Fields Are Missing, in Pregnancies with High Risk or Poor Outcomes Compared to Those With Low Risk and Good Outcomes

Data Item	High risk or poor outcome pregnancies*	Low risk and good outcome pregnancies**
Mother's hospital Personal Identifier (PID)	34.6	33.5
Mother's Unit Record Number (URN)	35.9	36.4
Mother's surname	0.1	0
Mother's first name	0	0
Mother's middle name	61.2	60.0
Mother's full name from the Registration of Birth	36.8	36.8
Marriage date	51.6	48.0
Infant's surname	1.8	1.1
Mother's height	3.2	0.7
Mother's date of birth	0.2	0
Mother's country of birth	34.6	36.1
Mother's postcode	0.2	0

Note: *High risk or poor outcome pregnancies = stillborn or < 2500 g or < 37 weeks gestation or multiple birth

**Low risk and good outcome pregnancies = live born singletons of 2500+ g and 37+ weeks gestation

https://doi.org/10.1375/twin.5.5.317 Published online by Cambridge University Press

Validation

Data sets from other studies. All of the birth histories derived from the pilot linkage for the 1048 women in the validation study data sets agreed fully with data from the questionnaires. Three of the Gold Standard data sets (902 sibships) resulted in full agreement for the full linkage. Of the 146 remaining sibships from the WA Pregnancy Cohort study, 35 women had reported births that did not apparently exist in the Midwives' records of birth held in the MCHRDB. An independent midwife conducted further investigations of the birth histories of these women. The missing births for 22 of the 35 women were confirmed to have occurred outside of WA and ten women could not be traced at all. This left three women for whom more births should have been found by the pilot record linkage. Further searches of the final corrected data base of sibships were then made. Of those three women, the sibships of two had been completely corrected via the logical consistency checks and that of the third partially corrected. The birth records of the latter woman's remaining two infants could not be found in the MCHRDB despite exhaustive searches using the mother's surname, maiden name, date of birth and father's country of birth.

Medical records review. Two different types of samples of linked birth records were compared with the women's hospital case notes. The first type consisted of two stratified random samples of 50 and 100 women who, according to the sibship linkage, had either only one birth or more than one birth respectively, during 1980 to 1992 inclusive. All of those 150 women's hospital records agreed completely with their sibship records constructed through linkage. The second type consisted of three random samples, each of 50 women, whose records of births had been linked to hospital discharges for the gynaecological morbidity study. Of that group of 150 women, 138 sets of sibships (92%) were found to agree fully with the case notes. For eight women the linked sets under-ascertained the true number of births at that hospital and, for four women, the linked sets included births that were not recorded in the case notes.

Logical inconsistency checks. Of the 22,747 women, who were part of the gynaecological morbidity study, 17 had been admitted to hospital following an apparent death. Fifteen had apparently undergone a delivery, dilation and curettage or a hysterectomy after a hysterectomy; seven infants had been delivered within 126 days of a prior dilation and curettage and 162 women had apparently been admitted to more than one hospital at the same time. Clerical reviews of these records found that all of the deaths and hysterectomies followed by births or dilation and curettage procedures resulted from apparent errors in transcription or coding of the gynaecological procedures. Similarly, the seven dilation and curettage procedures where infants were delivered soon afterwards appeared to be correctly matched birth records. Of the 162 multiple admissions to hospital within the same time period, 161 sets of birth records were correctly linked and one set of birth records may have been those belonging to a set of wrongly linked maternal twins. Of those 1310 sibships with more than five children, 82 sibships were found to contain mismatched records.

Further logical inconsistency checks by independent programmers found that there were 2535 women whose first names and dates of birth differed within their sibships. Many of these apparent linkage errors resulted from reversals of first and second names and errors in the date of birth. The remainder, which were true linkage errors, affected 786 infants and resulted in the formation of 657 new single infant sibships and the reassignment of 129 infants to different sibships. Overall, the validation processes found that only 0.9% of sibships contained errors.

Missing data rates. Missing data rates were calculated for maternal variables used in the linkage of multiple birth sibships and the sibships containing birth records of infants who were preterm (less than 37 weeks gestation), low birth weight (less than 2500 g) or stillborn. Comparison of those missing data rates with missing data rates for all other births has shown that the birth records for the latter contained very similar percentages of missing values to those for births which resulted in high risk or poor outcomes. The only linkage field to show a large increase in the missing data rate for women whose infants had high risk or had poor outcomes was maternal height.

Twins and Higher-order Multiples

Through the construction of the sibships, 3678 women were identified as the mothers of at least one set of twins or higher order multiples.

Maternal Twins

It was not possible to accurately determine what proportion of the WA population of same sex twin females of childbearing age would be represented in the sibships through having given birth in WA. Three hundred and five sets of apparent maternal twins were identified. Information about maternal twins in WA was obtained from the WATCH study questionnaires. Of the 2848 WATCH study responder mothers, 60 (2%) are twins, with 26 (43%) monozygotic, 20 (33%) dizygotic same sex and 14 (23%) dizygotic male-female twins.

Discussion

This project has used computerized record linkage to create a set of maternally linked sibships for all WA births from 1980 to 1992 inclusive. Minimization of the number of mismatched records resulting from linkage procedures was a major objective of this research into the optimal methods of record linkage for the creation of sibships. Probabilistic methods of record linkage may compromise the data base quality through mismatches, hence this study included several different approaches to sibship creation. This was complemented by the use of multiple types of validation processes.

In the absence of a unique personal identifier, probabilistic record linkage has been demonstrated to be a viable method for creating such sibships. However, further research is needed to measure the linkage errors in other high and low risk groups within these sibships to ascertain if there are differences in the occurrence of errors between such groups. In this study, any errors identified in the linkage were immediately corrected. Further research is also required to measure the magnitude and direction of any bias that may have been introduced into epidemiological studies of perinatal outcomes if these rigorous validation and correction processes had not taken place.

Research into the health of the maternal twins identified through this project, and of their infants, could aid in our understanding of the etiology of poor perinatal outcomes such as stillbirth, low birth weight, preterm birth and birth defects. This type of perinatal research would be enhanced by the collection of data in future study questionnaires relating to the multiple birth status of all study participants.

A population-based registry of all WA born twins and higher-order multiples and their siblings was created as a result of the creation of these sibships. The linkage identified the siblings of all multiple birth infants and enabled us to compare the mothers who responded to postal invitations with non-responder mothers and with those who declined to take part in the WATCH study. Other studies into aspects of maternal and child health using the sibship data are being conducted. In particular, studies of the association between the health of WA mothers and the perinatal, infant and childhood outcomes of their infants. Outcomes of interest include the risk factors for, and the recurrence rates of, Sudden Infant Death Syndrome, birth defects, and childhood deaths. A study of the outcomes of vaginal births after cesarean sections is also being undertaken (personal communication, Jan de Groot). The creation of sibships described in this paper has demonstrated the usefulness of probabilistic record linkage in linking records about women collected over a considerable time period. Ascertainment of twins and their siblings for an ongoing twin register was a natural consequence of this process.

Acknowledgments

The pilot phase of this record linkage project was supported by a grant from the Public Health Research and Development Committee of the National Health and Medical Research Council (NH&MRC) (grant #95/4028). Jennifer Kurinczuk and Anne Read were supported by a program grant from NH&MRC (grant #96/3209). Nick de Klerk and Anne Read are currently supported by a program grant from NH&MRC (grant #00/3209). Jennifer Kurinczuk is currently funded by a National Public Health Career Scientist Award from the DH and NHSR&D (#PHCS022). The WATCH study was funded by a grant from WA Health Promotion Foundation (Healthways grant #4920) and the WATCH for Asthma Study is currently funded by a grant from the NH&MRC (grant #110320). The authors wish to thank Kim Wong for her assistance in extracting the data sets for validation and Sue Forbes, Helen Galea, Carol Garfield, Raywin Tuohy and Karen Moonen for their assistance in the independent validations of the sibships. Gail Reading assisted in conducting the linkage of the birth records to the WA electoral roll. Phyllis Alessandri was responsible for clerical reviews of multiple birth sibships and telephone follow-up of the families for the WATCH study. Helen Ewart is the Clinical Coordinator of the WATCH for Asthma Study.

References

Australian Bureau of Statistics. (1998). *1996 Census of population and housing — Socioeconomic indexes for areas.* Canberra: Australian Government Publishing Service.

Automatch. (1996). *AUTOMATCH: Generalized record linkage system.* Burtonville, MD: Matchware Technologies Inc.

Bell, M. (1992). *Internal migration in Australia 1981–1986.* Canberra: Australian Government Publishing Service.

Bower, C. & Stanley F. J. (1989). Dietary folate as a risk factor for neural-tube defects: Evidence from a case-control study in Western Australia. *Medical Journal of Australia, 150*(11), 613–619.

Fellegi, I., & Sunter, A. (1969). A theory for record linkage. *Journal of the American Statistical Association, 64,* 1183–1210.

Gee, V. (1996). *Perinatal statistics in Western Australia: Thirteenth annual report of the Western Australian midwives' notification system for 1994.* Perth, WA: Health Department of Western Australia.

Hansen, J., de Klerk, N., et al. (2000). The Western Australian twin child health (WATCH) study: Work in progress. *Australasian Epidemiologist, 7*(2).

Newcombe, H., & Kennedy, J. (1962). Record linkage. *Communications of the Association for Computing Machinery, 5,* 563–566.

Newnham, J. P., Evans, S. F., et al. (1993). Effects of frequent ultrasound during pregnancy: A randomised controlled trial. *Lancet, 342*, 887–891.

Read, A. W., & Stanley, F. J. (1991). A comparison of recurrent and isolated small-for-gestational-age term births. *Paediatric & Perinatal Epidemiology, 5*(2), 138–156.

SAS Institute Inc. (1997). *SAS Release 6.12.* Cary, NC: SAS Institute Inc.

Stanley, F., & Hobbs, M. (1980). Neonatal mortality and cerebral palsy: The impact of neonatal intensive care. *Australian Paediatric Journal, 16*(1), 35–39.

Stanley, F. J., Croft, M. L., et al. (1994). A population database for maternal and child health research in Western Australia using record linkage. *Paediatric & Perinatal Epidemiology, 8*(4), 433–447.

Stanley, F. J., Read, A. W., et al. (1997). A population maternal and child health research database for research and policy evaluation in Western Australia. *Seminars in Neonatalogy, 2,* 195–201.

Swanson, P. B. Pearsall-Jones, J. G., et al. (2002). How mothers cope with the death of a twin or higher multiple. *Twin Research, 5*(3), 156-164.

Tursan d'Espaignet, G. D. E. (1998). *Sudden Infant Death Syndrome. Following up on the 1991 Reduce the Risk Campaign.* Doctoral dissertation, Canberra, Australian National University.