The final version of this paper was published in *ANZJOG* 2012; 52(1):91-95

**Title:** Prevalence of preeclampsia, pregnancy hypertension and gestational diabetes in population-based data: impact of different ascertainment methods on outcomes

**Short title:** Disease ascertainment in population-based data

**Authors:**
*Dr. Jian Sheng CHEN* – Researcher Fellow at Clinical and Population Perinatal Health Research, Kolling Institute of Medical Research, Sydney Medical School, University of Sydney, Sydney, Australia

*Prof. Christine L. ROBERTS* – Director at Clinical and Population Perinatal Health Research, Kolling Institute of Medical Research, Sydney Medical School, University of Sydney, Sydney, Australia

*Prof. Judy M. SIMPSON* – Professor of Biostatistics at Sydney School of Public Health, University of Sydney, Sydney, Australia

*Dr. Jane B. FORD* – Senior Researcher Fellow at Clinical and Population Perinatal Health Research, Kolling Institute of Medical Research, Sydney Medical School, University of Sydney, Sydney, Australia

**Address for correspondence:**
Dr. Jian Sheng CHEN, Clinical and Population Perinatal Health Research, Building 52, Royal North Shore Hospital, St Leonards 2065 NSW Australia
Tel: 61 2 9926 5877
Fax: 61 2 9906 6742
E-mail: jschen@med.usyd.edu.au

Dr. Jane B. FORD,
Fax: 61 2 9906 6742
E-mail: jane.ford@sydney.edu.au

There are no financial or other relationships that might lead to a conflict of interest with the study.
Abstract

This study investigated strategies for ascertaining preeclampsia, pregnancy hypertension and gestational diabetes mellitus from birth records and/or hospital discharge data. The results showed that ascertaining these conditions from a dataset that linked birth records to the corresponding maternal hospital record for birth was sufficient for health outcomes research. Antenatal hospital records provided few extra cases and may be necessary only for the ascertainment when a very accurate estimate of the prevalence is required.
Introduction

Pregnancy hypertension (including gestational hypertension, preeclampsia and eclampsia) and gestational diabetes mellitus (GDM) are common complications during pregnancy and have adverse consequences for both mothers and babies, including for hypertensive disorders: preterm birth, growth restricted infants, perinatal death, obstetric haemorrhage, acute renal or hepatic failure and maternal death; and for GDM: preeclampsia, macrosomic infants, induction of labour and caesarean delivery.\(^1-5\)

Accurate ascertainment of these conditions is important for maternity care service planning and health outcomes research.

Population health data such as hospital discharge and birth records have been used in studies of pregnancy hypertension and GDM.\(^2;6-8\) However, these data are not primarily collected for research. Concerns about the completeness of disease ascertainment in either dataset have been raised as a limitation of using single datasets for research.\(^9-12\) Linking birth records to the corresponding birth admission hospital records (one to one linkage) has been used to improve ascertainment.\(^7\) As only diagnoses affecting the birth admission are required to be coded in hospital discharge data, a more complex linkage adding information from antenatal hospital records is likely to further improve identification of the conditions.\(^13\) This study assessed the impact of different methods of ascertaining preeclampsia, pregnancy hypertension and GDM in population health data on the prevalence of these conditions and the statistical modelling of selected pregnancy outcomes.
Methods

**Study population and data sources**

The study population included 185,416 women who gave birth in NSW hospitals during 2007-2008 and had both a birth record in NSW Midwives Data Collection (‘birth data’) and a birth admission record in hospital data. Birth data include information on all live births or stillbirths of at least 20 weeks gestation or at least 400 grams birth weight in NSW. The collected information includes number of previous births, pregnancy complications (including preeclampsia, gestational pregnancy hypertension and GDM), pregnancy, labour, delivery and perinatal outcomes. The NSW Admitted Patient Data Collection (‘hospital data’) covers every inpatient admission in NSW, and includes demographic information and episode-related data which are coded according to the tenth revision of the International Classification of Diseases Australian Modification (ICD-10-AM). Record linkage of the birth and hospital data (by the Centre for Health Record Linkage) was approved by the NSW Population and Health Services Research Ethics Committee. As this study conforms to the standards established by the NHMRC for ethical quality review, ethics approval was not sought.

**Disease ascertainment**

Preeclampsia, pregnancy hypertension or GDM can be identified from birth data and/or hospital data with very high specificities (i.e. >99%) indicating few false positives. In the birth data preeclampsia, gestational hypertension or GDM were determined if ‘Yes’ was recorded in response to the relevant question in the birth record. In hospital data, any record with a diagnosis of gestational hypertension (ICD10-AM: O13 and O16),
preeclampsia (O11 and O14) or eclampsia (O15); and diabetes mellitus arising during pregnancy (O24.4) was included.

We compared five different methods of ascertaining disease rates from population data, from: (1) birth record alone, (2) birth admission hospital record alone, (3) either the birth record or birth admission hospital record (4) antenatal hospitalisation record, and (5) any record – birth, birth admission or antenatal hospitalisation.

Outcomes for statistical modeling of these conditions included small for gestational age (SGA)\(^{17}\) (i.e. \(\leq 10^{\text{th}}\) percentile of birth weight for gestational age), large for gestational age (LGA) (i.e. \(\geq 90^{\text{th}}\) percentile of birth weight for gestational age), preterm birth (<37 weeks), caesarean section (CS) and induction.

**Data analysis**

Using the different ascertainment methods, the proportion of women with preeclampsia, pregnancy hypertension or GDM was calculated and logistic regression was used to determine the effect size (odds ratio) on the selected outcomes. The capacity of a model to predict an outcome was evaluated using the area under the receiver-operating characteristic (ROC) curve (C-statistic), with values of 1.0 representing perfect discrimination and 0.5 indicating no better than chance. Data were analysed separately for women with and without a previous birth.
Results

Of the 185,416 study subjects, 41.5% (n=76,898) were women having their first birth and 25.2% (n=46,753) had been hospitalised during pregnancy (other than the birth admission)(Table 1). These 46,753 women had a total of 75,944 antenatal hospital admissions with a median of one antenatal admission per pregnancy (range: 1 to 39).

The prevalence of preeclampsia, pregnancy hypertension and GDM based on all available data was 3.4%, 9.7% and 6.0% respectively for all women. The number and frequency of these conditions varied by the different ascertainment methods and by parity (Table 2). The number of cases ascertained and associated prevalence estimates for all three conditions were higher in the birth admission hospital records than in birth data. The prevalence estimates increased significantly when birth admission and birth records were combined, and the proportional increase from using birth admission hospital records alone ranged from 12% for GDM at first pregnancy to 50% for pregnancy hypertension at second and subsequent pregnancies. However, the additional (case) ascertainment from including antenatal hospital records was small and the largest relative increase from using combined birth admission and birth records was only 8% for preeclampsia at second and subsequent pregnancies, representing an absolute increase from 2.1% to 2.2%.

Case ascertainment based on a single data source (birth admission hospital records or birth records) gave effect sizes for SGA, LGA, preterm birth, CS and induction that could be quite different (Table 2). For example, the odds ratio for the effect of preeclampsia on induction in a first pregnancy was 2.84 (95% confidence intervals (CI): 2.63-3.06) and 4.24 (95% CI: 3.88-4.63) for cases ascertained from birth admission hospital records or
birth records respectively, whereas the odds ratio for the effect of preeclampsia on cesarean section at second and subsequent pregnancies was 2.66 (95% CI: 2.42-2.93) and 2.12 (95% CI: 1.89-2.38) respectively. In contrast, case ascertainment from combined birth admission and birth records or from all available data did not change the effect sizes of any condition for any of the outcomes, with the largest relative change being less than 5% (i.e. odds ratio for the effect of preeclampsia on preterm birth at second and subsequent pregnancies: from 5.24 to 5.55). The inclusion of additional cases from antenatal hospital records resulted in little difference in the predictive capacities of models for any outcome when compared to models using information collected at the time of birth. For example, for preeclampsia ascertained from combined birth admission and birth records compared to all data, the C-statistics for the prediction of preterm birth at first birth were 0.5670 and 0.5692 respectively; this represented the largest difference between the two C-statistics for all the analyses.

Discussion

This study showed that simply linking birth records to the corresponding birth admission hospital records can significantly improve the ascertainment of preeclampsia, pregnancy hypertension and GDM, but additional linkage to antenatal hospital records did little to further the improvement. A validation study of 4541 women in Washington, USA, using information in medical records as the ‘gold standard’, reported that hospital discharge and birth records combined gave higher sensitivity for pregnancy hypertension and GDM (74% and 93% respectively) than did either hospital discharge (71% and 81% respectively) or birth records (49% and 64% respectively) alone.\textsuperscript{18} This was also the case
in a random sample of 1184 NSW women who gave birth in 2002\textsuperscript{9,11}; the sensitivity in birth data, hospital data and the two data combined were 63%, 68% and 82% respectively for pregnancy hypertension and 63%, 69% and 73% respectively for GDM. These results support the rationale of combining hospital data and birth data for ascertaining these conditions.

Although antenatal hospital records provided additional cases, the relative increase in the number of total cases in this study was small (less than 8% relative increase) and did not affect the estimate of the odds ratio of preeclampsia, pregnancy hypertension or GDM and the predictive capacity of any model. We previously reported that additional information on chronic diseases, such as preexisting hypertension and cardiac disease, from hospital admissions prior to pregnancy (lookback up to five years) did little to improve the modelling of risk factors for obstetric haemorrhage.\textsuperscript{19} The data manipulation required for identifying prior admissions is challenging, requiring multiple steps of data merging and development of the decision rules for resolving conflicting information from different records, but the additional effort may have little or no impact on the study results. This is reassuring for settings where antenatal data are not available, or individuals’ records are not identifiable.

In general, the odds ratios for pregnancy outcomes (Table 2) based on exposure information obtained from either birth or hospital data were lower than those obtained from hospital data alone (the more reliable source). This likely reflects that the increased ascertainment achieved using more than one data source provides a more accurate
distribution (with respect to severity) of disease in the population. Validation studies have repeatedly shown that identifying conditions from more than one data source increases accurate ascertainment, with higher sensitivity of reporting but without increasing the false positive reports.\textsuperscript{9,11,18} In addition, in any single data source, more severe disease or disease associated with adverse outcome is more likely to be reported.\textsuperscript{21,22} The combination of these facts suggests that using more than one data source would increase the proportion of women with mild disease, consistent with the lower odds ratios obtained.

In this study, pregnancy hypertension was estimated to be around 9.7%; this prevalence is similar to the estimate of 8.3\% from the audit of medical records of 1184 randomly selected women in NSW in 2002.\textsuperscript{11} The prevalence of 6.0\% for GDM in this study is higher than the NSW medical record audit (4.8\%)\textsuperscript{9} but consistent with increased GDM screening.\textsuperscript{20} For maternity care service planning, it is important to have accurate estimation of the proportion of women who are affected by these conditions as a small improvement in the ascertainment rate may translate into a large number of women at the population level and the resources required for caring for these women could be substantial.

In conclusion, whenever it is feasible, combined birth admission and birth records should be used to identify pregnancy hypertension and GDM. The resulting ascertainment rate is sufficient for health outcomes research. For assessing the burden of these conditions, additional information from the antenatal hospital records should be considered.
Acknowledgements

We thank the NSW Department of Health for access to the population health data and the NSW Centre for Health Record Linkage for linking the data sets. This work was supported by an Australian National Health and Medical Research Council (NHMRC) Capacity Building Grant (#573122). Christine Roberts is supported by a NHMRC Senior Research Fellowship (#457078).
References


<table>
<thead>
<tr>
<th></th>
<th>1&lt;sup&gt;st&lt;/sup&gt; pregnancy</th>
<th>≥2&lt;sup&gt;nd&lt;/sup&gt; pregnancy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=76,898†</td>
<td>N=108,518†</td>
<td>N=185,416†</td>
<td></td>
</tr>
<tr>
<td>N (%)</td>
<td>N (%)</td>
<td>N (%)</td>
<td></td>
</tr>
<tr>
<td><strong>Women with any antenatal hospital record</strong></td>
<td>21,358 (27.8)‡</td>
<td>25,395 (23.4)‡</td>
<td>46,753 (25.2) ‡</td>
</tr>
<tr>
<td><strong>Number of antenatal hospital record/s</strong></td>
<td>13,374 (62.6)</td>
<td>16,577 (65.3)</td>
<td>29,951 (64.1)</td>
</tr>
<tr>
<td>1</td>
<td>5,101 (23.9)</td>
<td>5,315 (20.9)</td>
<td>10,416 (22.3)</td>
</tr>
<tr>
<td>2</td>
<td>1,704 (8.0)</td>
<td>1,942 (7.6)</td>
<td>3,646 (7.8)</td>
</tr>
<tr>
<td>3</td>
<td>625 (2.9)</td>
<td>798 (3.1)</td>
<td>1,423 (3.0)</td>
</tr>
<tr>
<td>≥5</td>
<td>554 (2.6)</td>
<td>763 (3.0)</td>
<td>1,317 (2.8)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21,358 (100)</td>
<td>25,395 (100)</td>
<td>46,753 (100)</td>
</tr>
</tbody>
</table>

† Number of women with linked birth and birth admission hospital records
‡ Number and percentage of women who were hospitalised during the pregnancy, other than the birth admission.
Table 2. Prevalence of preeclampsia, pregnancy hypertension and gestational diabetes and odds ratio (OR) of these conditions for various maternal and neonatal outcomes by different methods of ascertainment, New South Wales, 2007-2008

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1st pregnancy</th>
<th>2nd pregnancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preeclampsia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth record</td>
<td>2187 (2.8%)</td>
<td>1172 (1.1%)</td>
</tr>
<tr>
<td>Birth admission hospital record</td>
<td>2854 (3.7%)</td>
<td>1712 (1.6%)</td>
</tr>
<tr>
<td>Birth record + birth admission hospital record</td>
<td>3687 (4.8%)</td>
<td>2255 (2.1%)</td>
</tr>
<tr>
<td>Antenatal hospital record/s</td>
<td>780 (1.0%)</td>
<td>501 (0.5%)</td>
</tr>
<tr>
<td>Birth record + all hospital records</td>
<td>3919 (5.1%)</td>
<td>2430 (2.2%)</td>
</tr>
<tr>
<td>Pregnancy hypertension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth record</td>
<td>6215 (8.1%)</td>
<td>5171 (4.8%)</td>
</tr>
<tr>
<td>Birth admission hospital record</td>
<td>7226 (9.4%)</td>
<td>5211 (4.8%)</td>
</tr>
<tr>
<td>Birth record + birth admission hospital record</td>
<td>9178 (11.9%)</td>
<td>7809 (7.2%)</td>
</tr>
<tr>
<td>Antenatal hospital record/s</td>
<td>2696 (3.5%)</td>
<td>2033 (1.9%)</td>
</tr>
<tr>
<td>Birth record + all hospital records</td>
<td>9718 (12.6%)</td>
<td>8335 (7.7%)</td>
</tr>
<tr>
<td>Gestational diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth record</td>
<td>3420 (4.4%)</td>
<td>5029 (4.6%)</td>
</tr>
<tr>
<td>Birth admission hospital record</td>
<td>3860 (5.0%)</td>
<td>5754 (5.3%)</td>
</tr>
<tr>
<td>Birth record + birth admission hospital record</td>
<td>4339 (5.6%)</td>
<td>6633 (6.1%)</td>
</tr>
<tr>
<td>Antenatal hospital record/s</td>
<td>580 (0.8%)</td>
<td>948 (0.9%)</td>
</tr>
<tr>
<td>Birth record + all hospital records</td>
<td>4359 (5.7%)</td>
<td>6705 (6.2%)</td>
</tr>
</tbody>
</table>

† SGA - Small for gestational age is the only independent variable in the analysis ‡ SGA - Large for gestational age is the only independent variable in the analysis