EVALUATION OF STRATEGIES TO IMPROVE CASE ASCERTAINMENT AND DATA QUALITY IN THE NATIONAL CONGENITAL ANOMALY SYSTEM

Thesis submitted to the University of London for the degree of Doctor of Philosophy

Beverley Jayne Botting

Institute of Child Health
University College London
2003
ABSTRACT

Evaluation of strategies to improve case ascertainment and data quality in the National Congenital Anomaly System

The thesis reports an evaluation of the National Congenital Anomaly System (NCAS) at a time of growing demands to use surveillance data for information on the effect of interventions (such as prenatal screening), for environmental exposures, and to provide information relevant to service provision for children with disability. In addition, recent developments in electronic data linkage have broadened the potential information that can be obtained through NCAS.

The thesis describes the original purpose of NCAS for rapid surveillance of localised and national increases in notifications of major external anomalies. Limitations of NCAS are described, together with steps taken to quantify and reduce these problems. The thesis identifies two major problems for NCAS: ascertainment and data quality. It reports a study to determine the extent of underascertainment for a range of conditions notified to NCAS. This was possible through comparisons with data held by local or condition-specific registers.

The thesis then reports a study to evaluate the effect on ascertainment of receiving notifications from local congenital anomaly registers rather than from health authorities. Effects on both overall and condition-specific notification rates are analysed. The impact of a step change in ascertainment on local surveillance and on national statistics is discussed. Notification to NCAS through four local registers resulted in substantial and significant changes in case ascertainment, which suggests that collaboration between NCAS and local and condition-specific registers should be extended.

Another study evaluated the benefits of linking individual NCAS records with birth registration records. The study demonstrated that over 95 per cent linkage was achievable. Information collected at birth registration was more complete than the same data items collected by NCAS. Therefore, those data items could be derived from linkage with birth records instead of by NCAS. Using characteristics derived from birth records (e.g. country of birth of mother and survival), new analyses were possible.

The concluding chapter discusses the impact on NCAS of these findings in the context of changing research and service demands. It looks at the strengths of NCAS, identifies priority areas for development, and presents my vision on shaping NCAS for the future.
TABLE OF CONTENTS

ABSTRACT ....................................................................................................................2
TABLE OF CONTENTS ................................................................................................3
LIST OF TABLES ............................................................................................................7
LIST OF FIGURES .......................................................................................................10
ABBREVIATIONS ........................................................................................................13
ACKNOWLEDGEMENTS ...........................................................................................15

CHAPTER 1..................................................................................................................16
Background
1.1 Summary ......................................................................................................17
1.2 Original Purpose of the system ...............................................................18
  1.2.1 Background ..........................................................................................18
  1.2.2 Description of the system ....................................................................20
  1.2.3 Congenital anomaly notification forms ................................................21
  1.2.4 Rapid notification .................................................................................22
1.3 Surveillance technique .................................................................................24
  1.3.1 Background ..........................................................................................24
  1.3.2 Description of CUSUM - principles and statistical methodology 24
  1.3.3 Application of CUSUM methodology to NCAS ..................................27
  1.3.4 Time interval used in the CUSUM analysis .......................................28
  1.3.5 Rationale for using numbers not rates in the CUSUM analysis .......29
  1.3.6 Definition of areas ................................................................................30
  1.3.7 Feedback to local areas ......................................................................31
1.4 Limitations of NCAS .....................................................................................33
  1.4.1 Underascertainment ............................................................................33
  1.4.2 Data quality ..........................................................................................40
1.5 Trends in reporting congenital anomalies ..................................................42
1.6 Evolving demands/ requirements for the system .......................................45
  1.6.1 Demands to meet policy and planning needs .....................................45
  1.6.2 Demands to investigate clusters .........................................................46
  1.6.3 Changing clinical demands .................................................................48
  1.6.4 Demands of international collaborative research ...............................49
  1.6.5 Demands of public expectations of confidentiality .............................50
1.7 Changes to meet new purposes ....................................................................52
1.8 Cost ..............................................................................................................53
1.9 Conclusion ...................................................................................................54

CHAPTER 2..................................................................................................................55
Prevalence of underascertainment of congenital anomalies by the National Congenital Anomaly System
2.1 Background ..................................................................................................56
2.2 Measurement of underascertainment .........................................................59
  2.2.1 Measuring ascertainment for congenital anomaly registers ...............59
  2.2.2 Capture - recapture methodology .......................................................61
2.2.3 Applications of capture-recapture relevant to congenital anomalies

2.3 Evidence of underascertainment

2.3.1 Anencephaly and spina bifida

2.3.1.1 Background

2.3.1.2 Methods

2.3.1.3 Results

2.3.2 Anophthalmia and microphthalmia

2.3.2.1 Background

2.3.2.2 Methods

2.3.2.3 Results

2.3.3 Congenital cataracts

2.3.3.1 Background

2.3.3.2 Methods

2.3.3.3 Results

2.3.4 Cardiac anomalies

2.3.4.1 Background

2.3.4.2 Methods

2.3.4.3 Results

2.3.5 Down syndrome

2.3.5.1 Background

2.3.5.2 Methods

2.3.5.3 Results

2.4 Summary of evidence

2.5 Discussion

2.6 Conclusions

CHAPTER 3

Effect of multi source ascertainment by local congenital anomaly registers on notification rates

3.1 Summary

3.2 Background

3.3 Description of the local registers involved in data transfer

3.3.1 Wales

3.3.2 Trent Congenital Anomaly Register (CAR)

3.3.3 North Thames West Congenital Malformation Register

3.3.4 Merseyside and Cheshire Congenital Anomaly Register (Mersey)

3.4 Methods

3.4.1 Data transfer between local registers and ONS

3.4.1.1 Preparing for data transfer

3.4.1.2 Data transfer

3.4.2 Analysis of notification rates before and after data transfer

3.4.3 Observed notification rates after data transfer

3.4.4 Expected notification rates based on rates before data transfer

3.4.5 Comparison of observed and expected rates

3.4.6 Impact of data transfer on surveillance alarms

3.5 Results

3.5.1 Observed rates before and after starting data transfer

3.5.2 Comparison with expected rates
3.5.3 Analysis of specific conditions .......................................................... 115
3.5.4 Impact on surveillance alarms ........................................................... 121

3.6 Discussion ............................................................................................... 123
3.6.1 Effect of ascertainment and data transfer by local registers on notifications to NCAS ....................................................................................... 123
3.6.2 Estimated ascertainment ................................................................... 128
3.6.3 Implications of data transfer for coverage and sustainability ............ 129
3.6.4 Implications of data transfer for surveillance .................................... 131

3.7 Conclusions ............................................................................................ 135

CHAPTER 4 ................................................................................................................. 137
An evaluation of record linkage between the National Congenital Anomaly System and Birth Registration
4.1 Summary .................................................................................................... 138
4.2 Background ................................................................................................ 139
4.2.1 Aims of linking NCAS and birth registration records ........................139
4.2.1.1 To reduce missing data ............................................................. 140
4.2.1.2 To obtain additional information on risk factors and survival ...141
4.2.2 History of record linkage in health research .....................................141
4.2.3 Previous experience of record linkage with ONS data sources ......... 143

4.3 Methods ...................................................................................................... 147
4.3.1 Choice of methodology for record linkage ........................................147
4.3.1.1 Deterministic matching ............................................................... 147
4.3.1.2 Probabilistic matching ................................................................148
4.3.1.3 Methodology used to link NCAS and birth registration records 149
4.3.2 Data sources for record linkage .......................................................... 152
4.3.3 Algorithm for linking NCAS and birth registration records ............. 155
4.3.4 Determination of linkage rate ............................................................. 159
4.3.5 Comparison of linked and unlinked records with the NCAS database .. 160
4.3.6 Evaluation of agreement between selected data items collected by both sources .................................................................................................. 161
4.3.7 Illustrative analyses enabled through linkage ................................... 164

4.4 Results ........................................................................................................ 166
4.4.1 Analysis of linkage rate ...................................................................... 166
4.4.2 Comparison of linked records with the NCAS database .................. 168
4.4.3 Evaluation of agreement between selected data items collected by both sources .................................................................................................. 175
4.4.4 Illustrative analyses enabled through linkage ................................... 182
4.4.4.1 Risk factors .................................................................................. 182
4.4.4.2 Survival ....................................................................................... 184

4.5 Discussion .................................................................................................. 187
4.5.1 Feasibility of record linkage ............................................................... 187
4.5.2 Comparison of linked records with the NCAS database .................. 188
4.5.3 Evaluation of agreement between data items collected by both sources .......................................................... 190
4.5.4 Illustrative analyses enabled through linkage ................................... 193
4.5.5 Implications for the future ................................................................. 195
CHAPTER 5 .................................................................................................................199
Looking to the future
5.1 Summary ...................................................................................................200
5.2 Summary of thesis .....................................................................................201
  5.2.1 Chapter 1 History and description of NCAS .....................................201
  5.2.2 Chapter 2 Estimation of ascertainment by NCAS .............................201
  5.2.3 Chapter 3 Evaluation of NCAS receiving data from local registers 201
  5.2.4 Chapter 4 Evaluation of linking NCAS to birth registration records 202
5.3 The way forward ........................................................................................203
  5.3.1 Scope for NCAS in the future ............................................................203
  5.3.2 Data transfer by local registers ..........................................................204
  5.3.3 Linkage of NCAS with other data sources ........................................207
5.4 Incentives for the future scope of NCAS ...................................................211
  5.4.1 The new National Service Framework (NSF) ...................................211
  5.4.2 NHS Information Strategy .................................................................211
  5.4.3 The NHS numbers for babies project ...............................................212
  5.4.4 E-government agenda .....................................................................213
5.5 Barriers to future expansion of data transfer to NCAS and record linkage
  ...........................................................................................................................214
  5.5.1 Lack of resources ...............................................................................214
  5.5.2 Untested work outside scope of this thesis .......................................214
  5.5.3 Ethics and confidentiality ...................................................................214
5.6 Conclusions ...............................................................................................219
REFERENCES ...........................................................................................................220
LIST OF TABLES

Chapter 1

Table 1.1  Notifications of congenital anomalies, England and Wales, 1964-2000  43

Chapter 2

Table 2.1  National anophthalmia and microphthalmia register by severity, whether known to NCAS and number of other reporting sources  72

Table 2.2  Congenital heart disease anomalies by whether known to NCAS and to other sources  80

Table 2.3  Summary of matching NCAS and NDSCR data, 1995 – 1998  84

Table 2.4  Ascertainment by NCAS for selected conditions  86

Chapter 3

Table 3.1  Local registers providing data to NCAS by 2000  98/9

Table 3.2  Data items required for electronic notification to NCAS  103

Table 3.3  Notification rates from registers participating in electronic data transfer, 1992-2000  112

Table 3.4  Observed and expected rates for local registers taking part in electronic data transfer with NCAS  113

Table 3.5  Observed notification rates for specific anomalies for local registers, 1992-2000  116/7
Table 3.6 Observed and expected rates per 10,000 live and stillbirths for specific conditions in the year after data transfer

Table 3.7 Number of alarms alerted by NCAS surveillance system

Table 3.8 Comparison of notification rates for four local registers and the remainder of England in 2000 with Clearinghouse Programmes in 1999.

Table 3.9 Births in areas covered by regional registers, 2000

Chapter 4

Table 4.1 Number of records extracted from NCAS and birth data files, 1997-2000

Table 4.2 Linked NCAS records by stage of algorithm, 1997-2000

Table 4.3 Unlinked records as a percentage of all notifications by year, 1997-2000

Table 4.4 Number and percentage of unlinked records by year for areas with a local congenital anomaly register participating in data transfer, 1997-2000

Table 4.5 Numbers and percentage of linked and unlinked records by birthweight groups, 1997-2000

Table 4.6 Distribution of mother's age for linked and total records, 1997-2000

Table 4.7 Distribution by multiplicity for linked and total records, 1997-2000

Table 4.8 Distribution by perinatal mortality for linked and total records, 1997-2000

Table 4.9 Unlinked records by data items with missing values, 1997-2000

Table 4.10 A comparison of mother's age reported on NCAS notifications with birth registration, 1997-2000
Table 4.11 Percentage distribution of birthweight as stated at NCAS notification according to that given at birth registration, linked records, 1997-2000

Table 4.12 Birthweight distributions of NCAS and birth records, 1997-2000

Table 4.13 Singleton/ Multiplicity status on NCAS record compared with birth registration record, 1997-2000

Table 4.14 Vital status as recorded at NCAS notification compared with that given at birth registration, 1997-2000

Table 4.15 Condition specific notification rates by mother’s country of birth per 10,000 live and stillbirths, 1997-2000
# LIST OF FIGURES

## Chapter 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Number of notifications over a 24-month period.</td>
<td>26</td>
</tr>
<tr>
<td>1.2</td>
<td>CUSUM value based on 24-month period given in Figure 1.1</td>
<td>27</td>
</tr>
<tr>
<td>1.3</td>
<td>Notification rates of anencephaly, other central nervous system anomalies, and legal abortions due to central nervous system anomalies, England and Wales, 1969-2000</td>
<td>38</td>
</tr>
<tr>
<td>1.4</td>
<td>Congenital anomaly notifications in live births, England and Wales 1964-2000</td>
<td>44</td>
</tr>
<tr>
<td>1.5</td>
<td>Congenital anomaly notifications in stillbirths, England and Wales 1964-2000</td>
<td>44</td>
</tr>
</tbody>
</table>

## Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page number</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Comparison of NTD rates for local congenital anomaly registers and NCAS, 1980-1997</td>
<td>69</td>
</tr>
<tr>
<td>2.2</td>
<td>Number of notifications of Down syndrome for births and terminations 1971-1999 before follow-up.</td>
<td>82</td>
</tr>
<tr>
<td>2.3</td>
<td>Number of notifications of Down syndrome for births and terminations 1971-1999 after follow-up.</td>
<td>85</td>
</tr>
</tbody>
</table>

## Chapter 3

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page number</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Notification rate for Wales register as published in given annual report, 1998-2000</td>
<td>107</td>
</tr>
<tr>
<td>3.2</td>
<td>Notification rates for all anomalies per 10,000 live and stillbirths, for each local register and for the remainder of England, 1992-2000</td>
<td>111</td>
</tr>
<tr>
<td>3.3</td>
<td>Observed and expected notification rates for each local register, 1992-2000</td>
<td>113/4</td>
</tr>
</tbody>
</table>
Chapter 4

Figure 4.1 Notification pathways for birthweight, mother’s age and multiplicity in NCAS and at birth registration.

Figure 4.2 Mother’s age as recorded by NCAS (pmage) and at birth registration (qmage), 1997-2000.

Figure 4.3 Birthweight as given at NCAS notification (pbwt) by that given at birth registration (qbwt) 1997-2000

Figure 4.4 Survival to age 4 years of babies born with selected conditions

Figure 4.5 Kaplan Meier survival curves for NCAS linked records with one anomaly notified compared with those with two or more, 1997-2000

Figure 4.6 Kaplan Meier survival curves for NCAS linked records with anencephaly, Down syndrome and the remainder of notifications, by whether they had one anomaly notified or two or more, 1997-2000

Chapter 5

Figure 5.1 Potential linkages to birth registration records
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>NCAS notification form – SD56</td>
<td>232</td>
</tr>
<tr>
<td>Appendix B</td>
<td>NCAS monitoring groups of anomalies affecting the same anatomical system</td>
<td>233</td>
</tr>
<tr>
<td>Appendix C</td>
<td>EUROCAT list of minor conditions that should not be notified unless they occur in combination with other major anomalies</td>
<td>235</td>
</tr>
<tr>
<td>Appendix D</td>
<td>NCAS form for notifying limb defects.</td>
<td>236</td>
</tr>
<tr>
<td>Appendix E</td>
<td>Form 309 - the standard form used for registering live birth registration</td>
<td>237</td>
</tr>
<tr>
<td>Appendix F</td>
<td>Botting B. Health The impact of more complete data from Wales on the National Congenital Anomaly System. Health Statistics Quarterly 05 2000</td>
<td>238</td>
</tr>
<tr>
<td>Appendix G</td>
<td>Botting B, Abrahams C. Linking congenital anomaly and birth records Health Statistics Quarterly 08 2000.</td>
<td>241</td>
</tr>
</tbody>
</table>
# ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHA</td>
<td>Area Health Authorities</td>
</tr>
<tr>
<td>ARL</td>
<td>Average Run Length</td>
</tr>
<tr>
<td>BDMP</td>
<td>Birth Defects Monitoring Program</td>
</tr>
<tr>
<td>BINOCAR</td>
<td>British Isles Network Of Congenital Anomaly Registers</td>
</tr>
<tr>
<td>BPCA</td>
<td>British Paediatric Cardiology Association</td>
</tr>
<tr>
<td>Bwt</td>
<td>Birthweight</td>
</tr>
<tr>
<td>CAR</td>
<td>Congenital Anomaly Register</td>
</tr>
<tr>
<td>CESDI</td>
<td>Confidential Enquiry into Stillbirths and Deaths in Infancy</td>
</tr>
<tr>
<td>CHD</td>
<td>Congenital Heart Disease</td>
</tr>
<tr>
<td>c.i.</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CMO</td>
<td>Chief Medical Officer</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>CUSUM</td>
<td>Cumulative Sum- technique for surveillance</td>
</tr>
<tr>
<td>DHA</td>
<td>District Health Authority</td>
</tr>
<tr>
<td>DOB</td>
<td>Date Of Birth</td>
</tr>
<tr>
<td>EEC</td>
<td>European Economic Community</td>
</tr>
<tr>
<td>EUROCAT</td>
<td>European Registration Of Congenital Abnormalities and Twins</td>
</tr>
<tr>
<td>Fname</td>
<td>Forename</td>
</tr>
<tr>
<td>GMC</td>
<td>General Medical Council</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>HA</td>
<td>Health Authority</td>
</tr>
<tr>
<td>HES</td>
<td>Hospital Episode Statistics</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Disease</td>
</tr>
<tr>
<td>LRD</td>
<td>Limb Reduction Defect</td>
</tr>
<tr>
<td>MACDP</td>
<td>Metropolitan Atlanta Congenital Defects Program</td>
</tr>
<tr>
<td>Mage</td>
<td>Mother's Age</td>
</tr>
<tr>
<td>MDOB</td>
<td>Mother's Date Of Birth</td>
</tr>
<tr>
<td>NCAS</td>
<td>National Congenital Anomaly System</td>
</tr>
<tr>
<td>NDSCR</td>
<td>National Down Syndrome Cytogenetic Register</td>
</tr>
</tbody>
</table>
NHS  National Health Service
NHSCR  National Health Service Central Register
NSF  National Service Framework
NTD  Neural Tube Defect
ONS  Office for National Statistics
OPCS  Office of Population Censuses and Surveys
ORLS  Oxford Record Linkage Study
Pc3  First 3 characters of postcode
PPV  Positive Predictive Value
RICHES  Regional Interactive Child Health System records
Sname  Surname
THM  Trihalomethane
UK  United Kingdom
ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible without the involvement and support of many people.

Dr Tony Ades was my supervisor in the initial years, and his support and advice formed a sound foundation to the work presented here. Dr Ruth Gilbert was my supervisor in the final years as I was writing up my research. She has provided me with inspiration as well as support and an enthusiastic interest in my work. I thank them both for being so approachable and always giving freely the time to work through the many complex issues I encountered on this journey.

I would also like to thank all my colleagues at the Office for National Statistics for their understanding and support, but in particular, Dr Peter Goldblatt for his support, Nigel Physick, for his patience in extracting datasets for my analyses, and to Carole Abrahams without whose programming skills the work presented in Chapter 4 would never have been possible. I am also grateful for the financial support given by the Office for National Statistics (ONS). However, all the views expressed in this thesis are mine and do not represent the opinion of the ONS.

My thanks also to colleagues in the local congenital anomaly registers whose collaboration was crucial to the success of the data transfer from local registers. In particular I would like to thank David Tucker from the CARIS register in Wales, Judith Budd from the Trent Congenital Anomaly Register, Lenore Abramsky from the North Thames West Congenital Malformation Register, and Grace Edwards from the Merseyside and Cheshire Congenital Anomaly Register. Their commitment to this research made the work a pleasure.

This work would never have been completed without the love and support of my friends and family. Thanks especially to Rev Peter Bradley and Liz Dighton who both believed I could complete this mammoth task when I had serious doubts. Finally my love to Charles and my parents who are always there for me, and to my children Emma and James. I promise I will have more time for you all now.
CHAPTER 1

Background

This chapter reports on the purpose, scope and operation of the National Congenital Anomaly System (NCAS) in England and Wales. I have been the Programme Director for NCAS since 1985, and during this time I identified improvements to address the quality of information provided on notification forms. Some improvements were straightforward to implement and these are reported here. Other solutions required wider co-operation and resources, and these are the subject of later chapters. This chapter also reports the findings of a collaborative study I undertook with colleagues at the University of Bergen, Norway. This established optimal methodology for surveillance of national congenital anomaly notification data.
1.1 Summary

This chapter provides the background for the thesis. It describes the original purpose of the National Congenital Anomaly System (NCAS) as a rapid system for identifying localised and national increases in notifications of major external anomalies. It also describes the surveillance technique used to identify these increases and its outputs. It aims to locate and describe the major limitations to NCAS within the context of its evolution and external developments. Over three decades there have been evolving clinical, policy, research, ethical and legal demands on the system. These are summarised, together with an overview of the operational changes made to NCAS to help meet these new requirements.
1.2 Original Purpose of the system

1.2.1 Background

The England and Wales Congenital Malformation Notification Scheme (now known as the National Congenital Anomaly System – NCAS) began in 1964. It was set up in the wake of the epidemic of limb reduction defects after maternal use of thalidomide during pregnancy in the late 1950s and early 1960s (McBride 1961, Lenz 1966).

Between 1959 and 1961 worldwide approximately 7000 births of children with absent or reduced limbs, sometimes also with gastro-intestinal and cardiac malformations, followed maternal use of thalidomide in the first trimester of pregnancy (Lenz 1988). In West Germany, where thalidomide could be bought without a prescription and was used more than in Britain, Weicker and Hungerland (1962) estimated that at least 4,000 affected children had been born before the end of 1961.

Thalidomide was first marketed in the United Kingdom (UK) in 1958 and used increasingly during the next three years (Leck 1962). As a result, the risks of exposure to thalidomide in early uterine life were practically non-existent for children born in 1957-8, but increased steadily thereafter. Since the drug was not withdrawn from sale in the UK until December 1961, some affected children were not born until the second half of 1962.

To determine the number of children affected by thalidomide in the UK, estimates were made by comparisons with previous baseline levels for the specific anomalies associated with exposure to thalidomide. A study of reduction deformities of limbs undertaken in Birmingham (Leck 1962), suggested that the rise in the incidence of reduction deformities was confined to bilateral reduction deformities. This was supported by a nationwide survey conducted by the College of General Practitioners, which concluded that the notification rate for ectromelia (absence of
a limb or limbs) affecting more than one limb rose after 1958, whereas the rate for other cases of ectromelia did not (Slater 1964). Leck and Millar estimated that bilateral reduction defects of the limbs occurred in 800 children in the United Kingdom from 1958 to the end of 1961 (Leck 1962). Smithells quoted the same figure (Smithells 1962) as the minimum number of babies born in England and Wales up to the end of August 1962 with malformations attributable to thalidomide.

Although the epidemic led to a large number of victims worldwide, the increased numbers of such babies with obvious limb reduction defects was not picked up quickly since there were no routine congenital anomaly monitoring programmes anywhere in the world. Once the epidemic was recognised, however, less than 6 months elapsed before the etiological agent was identified and withdrawn from sale. As a result of this epidemic, systems for the surveillance of congenital anomalies were set up across the world. One of the first was the register set up in Sweden. It was estimated that had the Swedish system been in place in 1960, it would have sounded an alarm about thalidomide after the first seven infants had been born with the syndrome over a four-month period (Kallen 1968).

Another of the first congenital anomaly registers in the world was set up in England and Wales. A letter in the British Medical Journal in 1962 proposed that

'Surely one lesson to be learned from the thalidomide tragedy is that, with certain exceptions such as hernias and birth marks, congenital deformities should be notifiable, in the same way as are some infectious diseases. The onus on observing any change in incidence, whether due to a new drug, a new industrial effluent, or to radioactive fall-out, would then fall upon the Registrar General, who with his staff already performs similar functions.' (Daintree Johnson 1962).

In December 1962, it was reported in the 'Medical Notes in Parliament' of the British Medical Journal (BMJ 1962) that Lieu-Colonel JK Cordeux (Nottingham, Central, Con) had asked the Minister whether he would consider making it a statutory obligation that all congenital deformities should be officially registered at birth. The Minister, Mr Powell, replied that
I am arranging for returns of children born with recognisable abnormalities. Statutory powers should not be necessary.

In 1963 there were discussions between the Ministry of Health and Local Health Authorities (HAs) (Weatherall 1969). As a result, the Chief Medical Officers in England and Wales inaugurated NCAS in January 1964 to notify congenital anomalies and monitor their frequencies. NCAS was designed as a national surveillance system capable of detecting changes in local notification rates of congenital anomalies.

NCAS was not the only measure taken in the UK in response to the thalidomide epidemic. Subsequently the Medicines Control Agency was set up in 1972 with its primary objective being to safeguard public health by ensuring that all medicines on the UK market meet appropriate standards of safety, quality and efficacy. One way it achieves this objective is through monitoring medicines and acting on safety concerns after they have been placed on the market.

1.2.2 Description of the system

Since its inception, NCAS has been co-ordinated by the Office for National Statistics (ONS) and its predecessors, the Office of Population Censuses and Surveys (OPCS), and the General Register Office. ONS is a Government department whose Director is also the Registrar General for England and Wales.

NCAS was originally designed to detect external visible anomalies, such as the severe limb reduction defects associated with thalidomide. These anomalies were considered readily recognisable by the midwife at birth. Local HAs were asked to collect congenital anomaly details using any means they chose, although it was suggested that they might use the information collected by the midwife, since she was required by the Public Health Act 1936 to record such findings on the statutory birth notification form within 36 hours of birth. The local HA retains these forms. The birth notification form was and is the principal source of data on congenital anomalies used for notification; very rarely were other sources used.
NCAS always relied on the efforts of a large number of health service staff whose main responsibilities, however, were to provide patient care. NCAS did not specify how the data should be collected or who, within each HA, was responsible for its provision. In practice, notification of congenital anomalies to NCAS has been an additional task for a range of different, but usually junior, non-clinical clerical staff. No additional resources have ever been provided for the Health Service to supply congenital anomaly data to NCAS.

1.2.3 Congenital anomaly notification forms

Notification has always been on a standard paper notification form, known as form SD56 (Appendix A).

The original SD56 notification form had a short list of conditions on the reverse of the form, and notifiers were asked to ring the appropriate code. In January 1972 the list was removed, and a box was provided on the front of the form for a textual description of all anomalies present. The text description was coded centrally by trained coders within NCAS using the revision of the International Classification of Diseases (ICD) in use at the time. Since 1995, NCAS has used the ICD 10 International Statistical Classification of Diseases and related health problems 1992, produced by the World Health Organisation (WHO 1992). Most classifications, however, are not specific enough to be used for the level of analyses required for congenital anomalies. For example, the 4-digit ICD codes cannot separate the different types of limb reduction defects into their specific anatomical groups.

Given the large number of individual 4 digit codes within the ICD, and the small number of notifications assigned to each code, surveillance would not have been very sensitive at this level. Therefore, ICD codes were grouped to form ‘ONS monitoring groups’ of anomalies affecting the same anatomical system (Appendix B).
Other information collected on the SD56 notification form included the child’s place and date of birth, their sex, whether they were live or stillborn, and whether they were from a singleton or multiple birth. The form also included the child’s birthweight and gestation, both parents’ occupations and information about the mother, including her full address, date of birth, and number of previous pregnancies.

A designated doctor, acting on behalf of the HA, signed the covering slip that was returned to NCAS each month accompanying the SD56 forms. This slip gave the number of babies with congenital anomalies notified that month.

1.2.4 Rapid notification

NCAS was set up to allow notifications to be received in monthly batches from the notifying areas. In order to provide fast surveillance, the notification forms were sent to NCAS no later than the first day of the second month following the month in which the baby was born. For example, for an April birth, the form should have been sent to NCAS no later than 1st June.

At the inception of the system, notifications were requested within 7 days after birth. This requirement for early notification would ensure rapid notification, but it was accepted that it would also reduce the chance of detecting internal malformations or chromosomal anomalies that were unlikely to be clinically confirmed within this period.

Information collected by NCAS has been limited to live and stillbirths. Stillbirths are defined using the legal definition of 28 completed weeks gestation (24 weeks since 1 October 1992). However, malformations in fetuses aborted spontaneously earlier in pregnancy were not included. As legal abortion was not introduced into England and Wales until April 1968 following the 1967 Abortion Act, therapeutic abortions were not considered for inclusion in the original system. Prior to 1967, the Infant Life (Preservation) Act 1929 made it an offence to deliberately cause the death of a child before it had an existence independent of its mother except where the act
which caused the death of the child was done in good faith for the purpose only of preserving the life of the mother.
1.3 Surveillance technique

1.3.1 Background

Appropriate statistical tools were needed to rapidly detect increases in notification of anomalies. With minor amendments, the ‘CUSUM’ (cumulative sum) method, which originated in industry as a method for monitoring quality in production processes, was found to be a powerful tool for monitoring anomalies. In brief, this method examines the expected number of cases of each monitored condition (or group of conditions) based on the previous level of notifications for a given time period and area. If the level is higher than would be expected by chance, a warning is triggered.

The CUSUM chart for Poisson variables, since congenital anomalies follow a Poisson distribution, was one of the first methods to be proposed for surveillance (Page 1954). As the methodology is based on straightforward calculations and readily available statistical reference tables, this simplifies its application for routine large-scale surveillance. Despite the subsequent development of alternative methodologies for monitoring congenital anomalies, CUSUM has been shown more recently to be the most appropriate technique for large national monitoring programmes such as NCAS (Lie 1991).

1.3.2 Description of CUSUM – principles and statistical methodology

The CUSUM technique is a sequential method. This means that it considers the number of notifications in a given time period, and takes into account the numbers previously notified. At the end of each chosen time period (which in NCAS is the most recent month, quarter year, and year), the CUSUM statistic ‘S’ is calculated. The CUSUM statistic is denoted \( S_n \) at the \( n^{th} \) time period, and \( S_0 \) (the initial value) is set to zero.

\( S_n \) is calculated by adding the number of cases in the most recent time period \( (n) \) to the previous value of \( S \) \( (S_{n-1}) \) and subtracting a reference value \( K \). \( K \) is related to the expected number of cases, so adding the observed number of cases and
subtracting K is effectively adding to CUSUM any excess between the observed number of cases and the expected number. When S reaches a critical warning value H, which indicates a significant increase in the observed mean, an alarm is signalled. Thus the values of H and K must be set in advance and the choice depends solely on the expected number of cases in one period and the size of the increase to be detected. These values can be obtained from prepared statistical tables (Woodward 1964).

A further element of the methodology is the need to specify an 'average run length' (ARL). This is defined as the expected time to detect a true increase or a false alarm. The statistical tables used by NCAS to derive the CUSUM variables H and K were calculated assuming an ARL of 3 months (or other time period if appropriate) to detect an increase of a specified size, and an ARL of 500 time periods before a false alarm if there had been no change in the mean number of cases. In choosing the ARL there is a trade off between sensitivity and specificity: if a shorter run length is chosen there will be more false alarms but the analysis will be more sensitive to true alarms. If, however, the ARL is longer, there will be fewer false alarms but the analysis may miss some real but small increases. In the NCAS setting, each month surveillance is carried out in approximately 100 areas for 50 anomalies. Therefore an ARL of 500 would generate 10 false alarms on average each period. In 2000, each monthly surveillance run produced an average of 38 alarms.

Lie and colleagues (I was a collaborator in this work) argued that the poor specificity with an ARL of 500 months to a false alarm seemed too high for use in the surveillance of birth defects (Lie 1991). Instead an alternative reference table was recommended for the Poisson CUSUM. To detect different levels of increase, this table gave values of K and H, and the ARLs to a false alarm and to detect the chosen increase. The ARL to a false alarm was chosen to be as close to 150 as possible for the specified increases that were to be detected. With an expected value of one case per period, use of these values would allow detection of a three-fold increase in true notifications in 4.4 periods.
An example of applying the CUSUM methodology to congenital anomaly notification data is given in Figures 1.1 and 1.2 (Chen 1983).

Figure 1.1 shows the monthly number of notifications of a given anomaly over a two-year period. Visual inspection would not suggest that there had been any change in incidence over this two-year period.

![Graph showing monthly notifications](image)

Figure 1.1 Number of notifications over a 24-month period.

The average number of notifications prior to the period shown in Figure 1.1 was 4 cases per month. Therefore 4 was chosen as the expected mean value, K. For this value of K, in a setting where the CUSUM is allowed to become negative, H is set at 7. The reference value K (=4) was subtracted from the observed number of notifications each month and the results summed to give the cumulative sum (CUSUM). Figure 1.2 shows the resulting CUSUM values.
If there was no change in the mean prevalence of 4 notifications per month, the line would vary around zero. In the second 12-month period, however, the mean increased to 4.5 cases per month. This relatively small increase in mean number of cases is not easily detected from the incidence chart, but is very evident in the second 12-month period in Figure 1.2. An alarm is generated when the CUSUM reaches 7 or higher, which from Figure 1.2 is in October of the second year.

1.3.3 Application of CUSUM methodology to NCAS

The application of the CUSUM methodology to the NCAS notifications broadly follows the example given in 1.3.2 except that the value of CUSUM is not allowed to fall below zero since the interest is only in detecting increases in incidence rather than decreases. Therefore, if the resulting $S_n$ is negative it is reset to zero, and the levels of $H$ and $K$ in the calculation of the CUSUM statistic are set appropriate to a one-sided test. The restriction of substituting zero whenever the CUSUM would have fallen below zero increases the sensitivity of the technique but decreases the specificity.

Expected baseline numbers are calculated separately for each local HA, using all notifications accumulated since the beginning of the system (excluding any epidemic periods) but weighted towards the most recent period. Each January,
baselines are revised by calculating the mean of the previous baseline number and the number of notifications of the given condition in the given area during the previous 12 months.

Although it has been shown that the use of CUSUM methodology is appropriate for the type of routine surveillance carried out by NCAS, its application to NCAS data had flaws that may have resulted in missing some true increases in birth prevalence in the past. Some problems are described in the following sections.

1.3.4 Time interval used in the CUSUM analysis

Notification data are sent to NCAS in monthly batches, and the monthly CUSUM analyses are based on data received in the previous month, quarter and year. Extending the analysis beyond the previous month has two advantages. First, it allows cases that have been identified and notified after the first month of life to be included in subsequent CUSUM analyses for the previous quarter or year. This increases the sensitivity of analyses after the first monthly analysis. Second, it identifies smaller sustained increases, which would not have been significant on a monthly basis. Although local Directors of Public Health are not notified of alarms generated from just one notification from their HA, a sustained increase could be based on one notification each month for several months. In this case, the analysis of monthly data would not generate an alarm, but the quarterly or annual analysis would generate an alarm, thus increasing the sensitivity of the analysis.

In addition, the file used for CUSUM analyses was not previously updated with late returns. As a result, the CUSUM analysis was less likely to detect increases in notification rates. Also, baselines were calculated based on actual numbers of notified cases for the period, not just those that had been included in the CUSUM analysis. If late notification rates remained constant, this would artificially inflate the expected values used to calculate baselines, making the CUSUM analysis less sensitive to true increases. Improvements to the CUSUM surveillance system were incorporated into a redevelopment of NCAS in 1995 necessitated by an office wide move away from its operating system. The CUSUM methodology was changed to
run monthly, based on the number of notifications in the previous month, quarter and year, compared to the expected numbers in that area in the given time period including late notifications. Since the assumption would be that the quarterly file would be more complete than the monthly, and the annual file more complete still, the relevant baselines were increased to take into account the improved completeness. This provided a more sensitive analysis than the previous analysis based solely on notifications received in the past month, but for quarterly and annual alarms, would be less timely.

1.3.5 Rationale for using numbers rather than rates in the CUSUM analysis

The variables used in the CUSUM analysis are first, the observed number of notified cases of a given anomaly in the HA in the previous period, and second, the expected number. Thus the method requires that there are no major changes in the underlying number of births between successive time periods. Since the number of births varies from period to period, the surveillance would be made more sensitive by using notification rates in place of the number of notifications. This would adjust the number of notifications in proportion to the increase or decrease in the number of births during the same period.

Notification rates are not used, however, because the original purpose of the system was fast surveillance of notified anomalies. As births can be registered up to 42 days after birth, and further delays occur during processing before the data are entered onto the ONS births database, use of the number of births that occurred in the given period as denominators to calculate notification rates would introduce an unacceptable delay. Weekly numbers of births registered are available, but registration patterns vary, particularly around holiday periods. Therefore the number of registrations in a given period is not always a good indicator of the true number of occurrences. Hence using the number of registrations for denominators of the rates offers no improvement over using just the number of notifications directly in the analysis.
1.3.6 Definition of areas

CUSUM analyses are based on HA areas and are also performed at national level. Any increases identified nationally usually reflect increases in a small number of local areas rather than a general increase over all of England and Wales. These local increases often reflect local changes in personnel reporting congenital anomalies, but may reflect real increases in prevalence.

The major changes in the organisation of the National Health Service since NCAS began in 1964 have required NCAS to regularly revisit two questions. First, what is the most appropriate area level to provide data? Secondly, what is the most appropriate area level on which to base the CUSUM analysis?

Data were originally provided by Area Health Authorities (AHAs). In 1984, when this administrative layer was removed, District Health Authorities (DHAs) took on the role of notifying anomalies. Currently Community trusts usually provide data.

Originally the areas chosen for surveillance were AHAs, mostly consisting of groups of two or three DHAs with average populations ranging from one million upwards. When English AHAs were abolished in 1982, it was considered that the number of births in many of the 200 DHAs was too small for routine surveillance. Therefore, the old AHA boundaries were retained for surveillance purposes. Consequently, an alarm in any surveillance area could have been generated based on notifications from two or more DHAs. This presented problems when trying to pass meaningful information back to the contributing DHAs. For example, the surveillance area of Berkshire included notifications from both East and West Berkshire Health Authorities. Any significant increase detected in notifications of a specific anomaly could have resulted from an increase in East Berkshire, West Berkshire, or both DHAs. The surveillance system could not detect which DHA had notified an increased number of the specific anomaly. Therefore, both DHAs were notified of the increase despite the fact that one DHA may not have notified a single case of the specific anomaly that month.
During the 1990s groups of two or more DHAs in England merged, so that the number of DHAs fell from 200 in 1990 to 187 in 1992, 121 in 1994 and 105 in 1996. These DHAs became Health Authorities (HAs) and their number remained at this level until 2000. Thus there was a smaller number of HAs each covering a larger geographical area with a corresponding larger number of births. These new areas were therefore an appropriate size for routine monitoring, and became the current monitoring areas.

The CUSUM analysis examines only the records for a given HA without considering the number of cases in adjoining areas or any other geographical clustering. However, before signing the alarm letters, I identify by eye any geographical patterns for the same anomaly. If adjacent areas receive a warning for the same condition, both areas are informed of the increase in the adjoining area. There is no geographical software used in NCAS to automate this process.

1.3.7 Feedback to local areas

Each month NCAS provides each HA with two different sets of feedback. First, a list of cases provided by each HA for the previous month, quarter and year is returned to the data notifier and the HA. These lists can be used by the HA to check that the data provided are correct. Any additions or corrections can be subsequently notified to NCAS.

Secondly, the monthly CUSUM analysis generates each month alarms which are notified by letter to the relevant HA with details of the anomaly, the time period and the specific cases that generated the alarm. The letter also explains that the increase may be a result of a new clinician with an interest in the specific condition, a change in clinical practice that has enabled the condition to be detected more easily or duplicate notifications. If there is a further alarm for the given anomaly within a year, a different warning letter is generated which refers to the earlier letter. In 2000, 38 alarms were generated on average from the monthly
surveillance, of which 24 were first alarms and 14 related to a second or subsequent alarm within 12 months in the given area for that condition.
1.4 Limitations of NCAS

Since the inception of NCAS as the system developed and NCAS data were required for research, a number of limitations have been identified. These limitations broadly result from three problems; underascertainment, data quality, and ability of the CUSUM methodology to accurately detect increases in notifications. The problems inherent in the NCAS CUSUM system were described in sections 1.3.4 to 1.3.6. Underascertainment and data quality are discussed in this section. Investigating solutions to the problems of underascertainment and data quality are the purpose of this thesis.

1.4.1 Underascertainment

'Underascertainment' is used in this thesis to refer to babies with congenital anomalies that were not notified to NCAS. Evidence for underascertainment in NCAS will be presented in Chapter 2. Some possible explanations are outlined here, and the effect of ascertainment by local registers and data transfer to NCAS is explored in Chapter 3.

Ascertainment varies for different anomalies depending on when it could be identified prenatally or postnatally, and whether there was a need for a specific test for diagnosis. These vary by time, technology and staff. If a condition was easily identified at birth then ascertainment would be expected to be high. If the condition is an internal anomaly, not easily visible at birth, more subjective and dependent on clinical judgement or dependent on investigations such as ultrasound, then true cases may be missed.

Underascertainment can be attributed to a variety of factors. These include poor knowledge of NCAS by clinicians and clinical staff who should notify congenital anomalies, diversion of notifications to other congenital anomaly registers, changes in the NHS that lead to responsibility for notifying anomalies not being incorporated
into new job descriptions, and changes in recognition of anomalies. Some of these changes affect all HAs at the same time, some affect HAs over a given period and others just affect a given HA. These are considered in more detail below.

When monitoring anomalies, completeness is not necessarily essential to identify significant increases in notifications. Provided that levels of under-ascertainment remain constant, the surveillance techniques are able to detect changes in the number of notifications and signal if these changes are statistically significant. However, as there is no way of being certain that ascertainment is constant, it is not possible to know whether any increase is due to improved ascertainment. Underascertainment makes it more likely that statistically significant changes in notification rates are the result of changes in local congenital anomaly detection and notification practices. These include changes in compliance with the system by new clinicians or clerks, or the introduction of new diagnosis equipment and techniques, rather than real changes in birth prevalence rates. Such variation in completeness introduces bias and the CUSUM analysis will react to this bias as well as to real change. It is not possible to differentiate between these two situations if an alarm is generated by the CUSUM analysis.

Complete ascertainment is not essential for surveillance since the CUSUM variables are specific for each HA. Nevertheless, CUSUM does assume a constant level of ascertainment. However, as complete ascertainment as possible is a prerequisite for audit, research and resource planning.

Organisational factors

Several factors relating to the local organisation of NCAS contribute to underascertainment. These can be summarised as responsibility, accountability, incentives and the lack of a mandatory system.
No single individual has been made accountable for notifying congenital anomalies. NCAS is dependent on many people working in local Health Authorities, for whom notification represents an additional commitment to their work. The work involved in notifying a case must be kept to a minimum. Once a case has been identified and any relevant paperwork completed, ascertainment will then depend on the ease of transferring the information from the source to the register. If it involves extensive clerical effort it may not be completed if it is competing with other high priority work.

For any register to work there must be a commitment and belief in the system by all those actively participating and contributing data. In order for any system to achieve high ascertainment, all relevant clinicians must know of the system. To improve awareness of the need to collect good congenital anomaly data, in Utah in 1981 a congenital malformation notification form was introduced, to be completed by the infant’s physician during the infant’s hospitalisation and to be reviewed on discharge. Previously the Utah register had relied on passive reporting from information collected on the birth certificate recorded by the mother’s physician (Minton 1986). The results of these changes were analysed retrospectively for the 4949 births in 1982, and showed that completeness of reporting improved from 12 per cent using birth certificates in 1970-72 to 93 per cent in 1982 using the new form. Hence, awareness of a congenital anomaly register by clinicians can improve ascertainment. However, clinicians may have reservations about both the limited clinical utility of the register and about potential breaches of confidentiality. Experience in the Glasgow Register of Congenital Malformations showed that on occasions these concerns may have impaired the efficient notification of anomalies (Stone 1987).

Midwives have a legal responsibility to complete a birth notification form within 36 hours of delivery. There is a box on this form that asks whether an anomaly is present. The options are to leave the box blank, tick yes (and add a brief description) or tick no. Many midwives are unlikely to be aware of, or appreciate
the use that will be made of the congenital anomaly data. This problem has been appreciated recently by their profession (Coleridge 2002).

If congenital anomaly information is provided on the birth notification form, this is copied by a clerk onto the SD56 notification form. These forms are collated each month and a designated doctor in the local health area, who may not have the time to check and ensure completeness, signs the covering slip that is returned to NCAS. The sources of data available to the clerk, their knowledge of congenital anomalies and their time available for notification all have an impact on the completeness of their returns. A study of congenital anomalies among babies born in 1987 in Finland (Hemminki 1993) showed that registers using routinely collected data from patient records and discharge summaries which were linked centrally, identified more infants with congenital anomalies than a separate congenital anomaly register requiring special reporting by physicians.

NCAS is a voluntary system. The design of the NCAS was driven by its primary purpose of surveillance, to detect quickly any increases in notifications of external visible anomalies. In the 1960s it was believed (BMJ 1962) that since clinicians understood the importance of fast notification, statutory powers should not be necessary. Therefore, notification was voluntary. As memories of the thalidomide epidemic faded, however, workloads increased, and no other widespread teratogen was identified, the importance of malformation notification was very likely to diminish.

There has been discussion on many occasions, including in the Review by the Registrar General’s Working Group (OPCS 1995), on the voluntary nature of NCAS. In some other countries congenital anomaly registration is compulsory. These include the Czech Republic, Hungary, New Zealand, Denmark, Finland, Norway, and Sweden (International Clearinghouse 1991). In 1998 the Congress of the USA passed the Birth Defect Prevention Act which authorised their Centers for Disease Control and Prevention to collect, analyse, and make available data on birth defects, operate regional centres for applied epidemiologic research on the
prevention of birth defects, and inform and educate the public about the prevention of birth defects.

Notification to most registers, however, is voluntary. If notification was to be compulsory, however, there is no guarantee that completeness of notification would improve. The system would need committed resources and sanctions against non-reporting. A number of local congenital anomaly registers exist in England and Wales, and reporting to these is not compulsory, but these registers achieve high levels of completeness. Rather than whether a register has compulsory or voluntary notification, ascertainment by congenital anomaly registers has been found to be directly proportional to the number of sources of ascertainment (Leck 1963), the reliability of the sources (Ericson 1977), the length of time for which the population at risk is observed after birth (Yerushalmy 1970), and inversely proportional to the size of the population covered by the register (Leck 1963).

Lack of abortion data

One limitation of NCAS is the lack of therapeutic abortion data. Termination of pregnancy for fetal handicap was not legally possible at the inception of NCAS in 1964, therefore abortions were not included.

In the 1960s, when NCAS was set up, prenatal diagnosis techniques were largely unavailable. Therefore some anomalies that would now be detected prenatally were not diagnosed until birth or later. More recently the limitations of not including abortions in analyses of congenital anomalies have been recognised. Since prenatal diagnosis has had a major impact on the birth prevalence of anomalies, a distorted picture is given if terminations are excluded from analyses of trends in notification rates for anomalies. It is now essential to consider abortion data alongside that for live and stillbirths.
The changes to the NCAS system to include terminations have not yet been made. However, until 2002, ONS processed abortion notification forms on behalf of the Chief Medical Officer. Grounds E of the 1967 Abortion Act allows a legal termination of pregnancy if 'there is a substantial risk that if the child were born it would suffer from such physical or mental abnormalities as to be seriously handicapped'.

Data on abortions which took place under Grounds E were available in statistical form for complete analysis of congenital anomalies. As an example, Figure 1.3 shows the impact of including abortion data in an analysis of central nervous system (CNS) anomalies. This Figure shows the large fall in the number of CNS anomalies notified to NCAS. This is a group of anomalies which is thought to be well notified to NCAS, and therefore the graph reflects a true fall in prevalence rather than a change in notification practice.

Figure 1.3 Notification rates of anencephaly and other central nervous system (CNS) anomalies in live births, stillbirths and legal abortions, England and Wales, 1969-2000

Abortion data are not available for 1978
This fall has also been seen in other countries (International Clearinghouse 1991). Only part of this decrease is the result of these anomalies being detected antenatally and the parents choosing to terminate the pregnancy. Changing patterns of recognition of anomalies prenatally leads to more prenatal diagnosis, more terminations, and fewer live and stillbirths with the specific congenital anomalies.

Age restriction for notification

The original age cut-off of 7 days meant that notification would not be complete, as some anomalies would not be identified within that period, and for other notifications only partial clinical information would be provided. As diagnostic tools were introduced to detect anomalies earlier in the postnatal period, the age cut-off period was extended to 10 days in January 1990 in an attempt to enable cases identified in the immediate postnatal period to be notified to NCAS. More recently, to meet the additional research and policy uses of the NCAS data, it was felt that the data should be as complete as possible irrespective of the age at diagnosis. Therefore the age cut-off limit was abolished in January 1995 in the hope that later diagnosed cases would be notified.

Notifications can be added to the NCAS database at any time, although if the notifications are received when the child is aged more than one year the record cannot be included in the CUSUM surveillance analysis. Despite these changes there has been no significant change in the number of notifications to NCAS. This suggests that information collected locally, at neonatal discharge or at the 6-week postnatal check, is not being picked up by those who notify congenital anomalies. The implications for NCAS are that there are additional data locally that are not being notified to the system.
1.4.2 Data quality

Non-anomalies being notified

Prior to 1990, a large proportion of notifications to NCAS related to minor anomalies. In order to try to reduce the burden on those notifying and to try to improve compliance with reporting, in January 1990 an exclusion list was introduced to help advise notifiers on which conditions should be notified (Appendix C). This list, also based on that used by EUROCAT (a European collaboration, European Registration Of Congenital Abnormalities and Twins), gives a number of minor conditions that should not be notified unless they occur in combination with other major anomalies.

As a result of introducing this list, the overall number of notifications to NCAS fell sharply from 12,391 in 1989 to 8,154 in 1990. The number of notifications continued to fall as notifiers became familiar with the exclusion list, reaching 6,054 in 1992. The fall in notifications for live births was a direct result of falls in the number of notifications of conditions on the exclusion list. Since the conditions on the list were largely minor, non-lethal conditions, its introduction did not have such a noticeable impact on the number of notifications of stillbirths.

Inadequate description

The format of the SD56 notification form has changed only slightly over time. The main improvement, in January 1972, was to collect from notifiers a full textual description of the anomalies present in the child. Previously, on the reverse of the form there had been only a list of major and grouped anomalies to choose between. The textual descriptions are coded centrally by NCAS. Having a text description enables a more detailed coding to the International Classification of Disease (ICD) than prior to 1972. It also allows an analysis of the text information when the ICD code is not sufficient. One routine use of the text information is an annual analysis of limb reduction defects according to whether they are postaxial,
preaxial, transverse or intercalary. The ICD is not sufficient to identify these different categories.

A separate additional form for recording limb defects was introduced in 1990 (Appendix D). Based on a form designed by EUROCAT, it asks for a pictorial description of the identified anomalies by allowing the notifiers to mark on an anatomical drawing how the limb deviated from the norm. When used in combination with the text on the SD56 notification form, this is helpful in identifying the specific types of limb defects described above. Despite NCAS including these forms when sending supplies of blank SD56 notification forms, unfortunately, the limb form is only completed for approximately 50 per cent of all limb defect notifications. This is likely to be a result of those notifying having insufficient information about the defect to complete the form.
1.5 Trends in reporting congenital anomalies

Although the primary purpose of NCAS is to monitor changes in the notification of anomalies, it also provides the most extensive data on the birth prevalence of congenital anomalies in England and Wales. Table 1.1 shows the number and rate of notifications, by live or stillbirth since 1964 when the system first began. The notification rate per 10,000 live births remained steady between 1964 and 1975 at around 160-170. Therefore, just fewer than 2 per cent of live births were notified as having one or more anomalies.

The notification rate rose between 1975 and 1984 reaching 210 per 10,000, and then fell to 180 between 1984 and 1989. Notifications fell sharply between 1990 and 1994 after the introduction of an exclusion list, reaching a low for live births of 81 per 10,000 in 1994. The number of notifications began to increase again in 1998, reaching almost 8,000 in 2000. This latter improvement was largely due to data transfer to NCAS by a small number of local congenital anomaly registers. This improvement in ascertainment is the subject of the study reported in Chapter 3.
<table>
<thead>
<tr>
<th>Year</th>
<th>Notifications of congenital anomalies</th>
<th>Births</th>
<th>Rates per 10,000 births</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Live</td>
<td>Still</td>
<td>Live</td>
</tr>
<tr>
<td>1964</td>
<td>14729</td>
<td>2971</td>
<td>875972</td>
</tr>
<tr>
<td>1965</td>
<td>13610</td>
<td>2832</td>
<td>862725</td>
</tr>
<tr>
<td>1966</td>
<td>13479</td>
<td>2681</td>
<td>849823</td>
</tr>
<tr>
<td>1967</td>
<td>14112</td>
<td>2640</td>
<td>832164</td>
</tr>
<tr>
<td>1968</td>
<td>14011</td>
<td>2474</td>
<td>819272</td>
</tr>
<tr>
<td>1969</td>
<td>12072</td>
<td>1830</td>
<td>797538</td>
</tr>
<tr>
<td>1970</td>
<td>12246</td>
<td>1712</td>
<td>784486</td>
</tr>
<tr>
<td>1971</td>
<td>12567</td>
<td>1840</td>
<td>783155</td>
</tr>
<tr>
<td>1972</td>
<td>12782</td>
<td>1629</td>
<td>724827</td>
</tr>
<tr>
<td>1973</td>
<td>11920</td>
<td>1431</td>
<td>675394</td>
</tr>
<tr>
<td>1974</td>
<td>10315</td>
<td>2222</td>
<td>639359</td>
</tr>
<tr>
<td>1975</td>
<td>10968</td>
<td>1209</td>
<td>602872</td>
</tr>
<tr>
<td>1976</td>
<td>11337</td>
<td>1013</td>
<td>583761</td>
</tr>
<tr>
<td>1977</td>
<td>11380</td>
<td>990</td>
<td>568718</td>
</tr>
<tr>
<td>1978</td>
<td>11827</td>
<td>918</td>
<td>595897</td>
</tr>
<tr>
<td>1979</td>
<td>12667</td>
<td>825</td>
<td>637490</td>
</tr>
<tr>
<td>1980</td>
<td>13349</td>
<td>723</td>
<td>655728</td>
</tr>
<tr>
<td>1981</td>
<td>12774</td>
<td>600</td>
<td>634005</td>
</tr>
<tr>
<td>1982</td>
<td>12732</td>
<td>497</td>
<td>625431</td>
</tr>
<tr>
<td>1983</td>
<td>13444</td>
<td>452</td>
<td>628749</td>
</tr>
<tr>
<td>1984</td>
<td>13546</td>
<td>391</td>
<td>636434</td>
</tr>
<tr>
<td>1985</td>
<td>12949</td>
<td>342</td>
<td>656072</td>
</tr>
<tr>
<td>1986</td>
<td>12758</td>
<td>265</td>
<td>606047</td>
</tr>
<tr>
<td>1987</td>
<td>13237</td>
<td>275</td>
<td>681146</td>
</tr>
<tr>
<td>1988</td>
<td>12723</td>
<td>239</td>
<td>693187</td>
</tr>
<tr>
<td>1989</td>
<td>12191</td>
<td>200</td>
<td>687376</td>
</tr>
<tr>
<td>1990</td>
<td>7941</td>
<td>213</td>
<td>705786</td>
</tr>
<tr>
<td>1991</td>
<td>6889</td>
<td>186</td>
<td>698885</td>
</tr>
<tr>
<td>1992</td>
<td>5909</td>
<td>145</td>
<td>689307</td>
</tr>
<tr>
<td>1993</td>
<td>5565</td>
<td>156</td>
<td>673051</td>
</tr>
<tr>
<td>1994</td>
<td>5356</td>
<td>181</td>
<td>664322</td>
</tr>
<tr>
<td>1995</td>
<td>5365</td>
<td>169</td>
<td>647734</td>
</tr>
<tr>
<td>1996</td>
<td>5280</td>
<td>158</td>
<td>649078</td>
</tr>
<tr>
<td>1997</td>
<td>5319</td>
<td>166</td>
<td>641720</td>
</tr>
<tr>
<td>1998</td>
<td>5393</td>
<td>199</td>
<td>635549</td>
</tr>
<tr>
<td>1999</td>
<td>6541</td>
<td>227</td>
<td>621579</td>
</tr>
<tr>
<td>2000</td>
<td>6998</td>
<td>286</td>
<td>604130</td>
</tr>
</tbody>
</table>

Table 1.1 Notifications of congenital anomalies, England and Wales: 1964-2000
Figure 1.4 Congenital anomaly notifications in live births, England and Wales, 1964-2000

Figure 1.5 Congenital anomaly notifications in stillbirths, England and Wales 1964-2000
1.6 Evolving demands/requirements for the system

The limitations of ascertainment and data quality inherent in NCAS have been described in this chapter. There have also been a number of other external changes which impact on the future requirements for NCAS data. Whilst surveillance remains the primary purpose of the NCAS, the system needs to evolve to be more sensitive to changing policy, clinical and research demands. It needs to be able to react quickly to local and national concerns about increases (real or perceived) in prevalence rates of congenital anomalies. In the past these increases have largely been investigated using expensive retrospective separate studies.

1.6.1 Demands to meet policy and planning needs

NCAS data are used by Government, and Health departments for evaluating and monitoring the impact of new and changing policy on, for example, antenatal screening, immunisation policies (for example rubella), and health education campaigns (for example folic acid). The data are also used to assist in Government enquiries and subsequent research following public concerns on suspected local clusters.

In addition, for predicting need and planning services, HAs and Social Services increasingly need to know the absolute number of surviving disabled people in the area they cover. This requires the available data to be as complete as possible for those still surviving. NCAS could help meet these needs if its ascertainment were improved. Most conditions, however, vary in severity, so to use NCAS data for this purpose would require additional information on severity. NCAS does not include several conditions required for resource planning, such as blindness and cerebral palsy. Therefore, there would need to be a link with other 'disability' registers, so NCAS could not be a universal panacea.
1.6.2 Demands to investigate clusters

Surveillance through NCAS was intended to identify local increases, but because of poor ascertainment, these data could not be used to investigate specific clusters. Instead separate expensive studies have been required to investigate local concerns resulting from apparent clusters of specific anomalies where the public concern is that toxins from a given location may have a teratogenic effect on the developing fetus. Usually there has been a local hypothesis about the mutagenic agent involved, which has required good quality local data to test the hypothesis. Recent examples include the possible adverse effects of disinfectant by-products in drinking water, of which trihalomethanes (THMs) are the most common and routinely measured compounds. This has been largely fuelled by studies that have shown an increase risk of NTDs in areas which use THMs to disinfect water (Bove 1995, Klotz 1999). Another example is the EUROHAZCON study (Dolk 1998) which found that living within 3KM of selected European landfill sites was associated with a significantly raised odds ratio for NTDs (odds ratio 1.86; 95 per cent confidence intervals 1.24-2.79).

Within England and Wales, therefore, there is a need to examine smaller geographical areas to investigate hypotheses of iatrogenic, environmental or occupational hazards. Testing these hypotheses or investigating identified clusters requires data at a small geographical area. Methodology for geographical epidemiology has become much more sophisticated in recent years. Postcode can be used to place cases within the boundaries of different local areas. Therefore, postcode has been collected by NCAS and held on the computer record since the early 1980s.

The Black Report (Black 1984) stated that

‘encouragement should be given to an organisation such as the Office of Population Censuses and Surveys (now part of ONS) or the Medical Research Council, to co-ordinate centrally the monitoring of small area statistics around major installations producing discharges which might present a carcinogenic or mutagenic hazard to the public.'
Ideally the data required for these purposes would be as complete as possible, requiring good ascertainment not only in live and stillbirths, but also, for many anomalies, in abortions and spontaneous miscarriages. In the past the albeit limited NCAS data have been used to help allay media driven fears, but better quality and more complete data could reduce the need for separate more intensive studies. To be more sensitive to changes the data need to be more complete, requiring both more records and more complete data items within records.

For example, between February 1989 and May 1990 four children with transverse limb reduction defects (LRD) were born in the same town on the Isle of Wight. The only common characteristic identified by the mothers was that they had swum in the sea during pregnancy. This raised the possibility that an environmental factor associated with living near to the coast and swimming in the sea might be implicated for these anomalies. Studies were undertaken in several countries to compare the prevalence of LRD in babies born in coastal areas with those born inland. The results from three different national congenital anomaly registers were published as letters in the Lancet on 23 April 1994. (Botting 1994, Castilla 1994a, Mastroiacovo 1994) These reports, from England and Wales, Latin America and Italy all found that there was no difference in the reported rates of LRD for babies born in coastal areas compared with those living inland.

Another example was the local fear that the pesticide Benomyl was associated with anophthalmia following a media report of a small cluster in Lincolnshire (Paduano 1993). NCAS data showed no trend in the notification of anophthalmia and other related eye defects associated with the introduction of this pesticide, despite a one thousand-fold increase in the use of Benomyl. This strengthens the position of NCAS to investigate such alerts despite its underascertainment. A wide scale investigation by the Clearinghouse in other countries similarly found no association (Castilla 1994).

More complete data for specific data items collected by NCAS, and for analysis of risk factors not collected by NCAS could be obtained by linking NCAS data with
other sources of data. For example, birth registration records collect information on the parents' countries of birth and occupation, but these data items are not collected by NCAS. A study to evaluate this linkage is reported in Chapter 4.

1.6.3 Changing clinical demands

Complete NCAS data will be required increasingly for clinical governance and feedback to units, requiring both more records and better information for each record. Congenital anomalies will be detected in different ways and at different times compared with in the past. NCAS must be able to respond with data pertaining to new techniques for prenatal diagnosis, and improved understanding of the etiology of congenital anomalies. As a result, NCAS will need to include both prenatally and postnatally detected anomalies.

Teratogens, to date, have often resulted in more than one anomaly. For example, exposure to rubella during pregnancy can result in a range of anomalies in the fetus (Gregg 1941). Therefore it is important that NCAS monitors children with multiple anomalies. Syndrome recognition is very important but also a complex process. NCAS has not had the resources to date to develop a procedure to use standard syndrome recognition software.

It is important to ensure that for each child notified, the clinical diagnosis is as accurate and complete as possible. This is not necessarily easy. Whilst some anomalies are straightforward to diagnose, others are more subjective. Absence of the external ear (anotia) may be straightforward, but it is more subjective at what point the external ear is considered small (microtia). Differences resulting from different practices of notifying these more subjective conditions will also affect completeness of notification. Chapter 2 reports a series of studies which estimate completeness of ascertainment by NCAS for a range of conditions, both those which are easily diagnosed and those which are more subjective.
1.6.4 Demands of international collaborative research

Some congenital anomalies are so rare as to require large scale global monitoring to have sufficient numbers to enable analyses. In 1974 NCAS became a founder member of the International Clearinghouse for Birth Defects Monitoring Systems (the ‘Clearinghouse’). The Clearinghouse is a non-governmental organisation recognised by the World Health Organisation. In 1999, the Clearinghouse had 35 participating programmes from 32 countries and collected information on over 3 million births annually worldwide (International Clearinghouse 2001). Congenital anomaly monitoring systems vary; some are nationally based, some cover residents of smaller areas within a country, while others collect data from local hospitals. NCAS contributes statistical summary data to the Clearinghouse quarterly and annually.

Whilst the primary function of the Clearinghouse is to exchange data and detect changes in notification rates, these data are also used for other large scale epidemiological studies and monitoring the implementation of preventative policies. International studies are usually set up when larger numbers of cases are required than would be available in any local programme or to evaluate different policies within the different member programmes. The data must be as complete as possible to be able to evaluate the impact of different environmental or occupational exposure on the birth prevalence of different anomalies in different countries. Recent collaborative projects by the Clearinghouse include studies of limb deficiencies among children with multiple anomalies (Rosano 2000, Stoll 2001), of the teratogenic effects of antiepileptic drugs (Arpino 2000), NTDs (Rosano 1999), and a comparison of national policies on periconceptional use of folic acid to prevent spina bifida and anencephaly (Cornel 1997).

The other international congenital anomaly collaborative group is EUROCAT. The EUROCAT project was originally conceived in 1974 at a workshop convened by the EEC to improve the methodology for population studies throughout the community. Monitoring congenital anomalies was selected as the focus of the
project, since it met their criteria of having great public health importance, was not regarded at the time as controversial, had scientific opportunities for the future, was amenable to preventive measures and concrete action, had educational impact, needed the active participation of medical practitioners and promoted standards internationally (Weatherall 1985). Some of the local registers within England, and the Congenital Anomaly Register Information System (CARIS) in Wales, are members of EUROCAT.

EUROCAT has also set up a number of collaborative studies including the recent work on the risk of congenital anomalies in the vicinity of landfill sites (Dolk 1998, Vrijheid 2002), and evaluations of prenatal diagnosis in relation to congenital diaphragmatic hernia (Garne 2002), and congenital heart disease (Garne 2001).

1.6.5 Demands of public expectations of confidentiality

There is a public expectation that personal privacy will be respected whilst collecting data to investigate public concerns. These conflicting demands would require that to ensure confidentiality, personal details were not retained on data bases, yet congenital anomaly registers need those identifying details to match information from different sources, to avoid duplicate records and to investigate increases in notification rates.

Parents whose children are born with a congenital anomaly are usually concerned to identify the ‘cause’ of their child’s anomaly, and understand the role of epidemiological research to find these causal associations. Many parents want to be assured that their children are included in any register. When there is an apparent cluster of babies born with a given malformation, an investigation usually follows. In 1994, following the cluster of babies born without a left hand on the Isle of Wight and the investigation of coastal areas that followed (Botting 1994), there
were many telephone calls from parents who wanted to check that their children **had** been reported to NCAS.

This needs to be viewed against potential confidentiality issues for the child with anomalies and their parents. There is increasing public and institutional concern about the confidentiality of patient data. Currently parents are not asked to give their consent for their child’s details to be added to NCAS. This is being examined by NCAS, together with local congenital anomaly registers, in the light of the Data Protection Act 1998, introduced on 1 March 2000, the guidelines from the General Medical Council in September 2000 stating that General Practitioners (GPs) do not need to pass information about their patients to registers, and the Health and Social Care Act 2001 which provides legal cover for the Secretary of State for Health to require that certain information is collected without informed consent if this is in the public interest. Confidentiality issues are discussed in Chapter 5 when considering the future scope and role of NCAS.
1.7 Changes to meet new purposes

NCAS must change to meet the new demands for complete, good quality data. In later chapters I discuss the evidence for the shortfall in ascertainment and data quality by NCAS and evaluate two methods that may resolve these deficiencies in the future.

Linkage with other data sources might improve data quality. The linkage would be straightforward if each person had a unique identifier. Everyone registered with a GP in England and Wales has a unique NHS number. However, there are problems in using NHS number, which are discussed in more detail in Chapter 4. This chapter discusses a feasibility study to evaluate record linkage between NCAS records and birth registration records. Therefore other operational changes were made within NCAS to assist linkage. One small change in the SD56 notification form was the introduction of name information in 1994. The first three characters of first and surname are collected. Once linkage had taken place named information could be removed to preserve confidentiality.
1.8 Cost

Currently NCAS operates at a cost to ONS of approximately £170 k per year. This is therefore the cost of a basic system with suboptimal ascertainment. In reviewing NCAS to design a better system more fitted for the purposes described in this chapter, the true current cost needs to include the cost of providing the data to NCAS; the cost of those in the HAs who complete the SD56 notification forms, the costs of the local and condition-specific congenital anomaly registers, and the costs of in-depth studies that are necessary because data from NCAS are insufficient to meet many policy and research purposes. Whenever the NCAS data are insufficient to investigate a given hypothesis, additional resources are required by other research bodies to run an in-depth investigation. These usually require the collection of extensive additional information, and seek retrospective notification of cases by clinicians. As a result, these studies are expensive.

For surveillance and research purposes, it is important in the future that NCAS data are as complete as possible and that information is held for a range of risk factors. Thus, the total costs of the various different agencies currently collecting information on congenital anomalies must be considered against that investment in improved continuous monitoring by NCAS. If more complete NCAS data were available, it would be easier to produce information about a topical concern quickly and inexpensively directly from the NCAS data.
1.9 Conclusion

The purpose of this thesis is to evaluate the contribution NCAS can make to the changing demands for congenital anomaly data, and the implications for NCAS' future role.

Chapter 2 analyses underascertainment for a range of different congenital anomalies. The thesis then evaluates two feasibility studies; the first, reported in Chapter 3, aims to improve ascertainment through data transfer to NCAS by local congenital anomaly registers. The second, presented in Chapter 4, aims to improve completeness of data items and enable new analyses through record linkage. Both studies aim that NCAS will be better fitted to meet its new purposes, whilst retaining its primary purpose of surveillance and the benefit of being low cost.
CHAPTER 2
Prevalence of underascertainment of congenital anomalies by the National Congenital Anomaly System

The purpose of this Chapter is to report on the extent of underascertainment in five conditions notified to the National Congenital Anomaly System (NCAS); anencephaly, anophthalmia, congenital cataracts, cardiovascular anomalies and Down syndrome.

In each of the five studies reported here, I was involved in the study design and responsible for specifying the NCAS data required for analysis, which were subsequently extracted by an ONS programmer. I played an active role in the analysis, and assisted in the preparation of papers for publication. I was the lead researcher in the study on Down syndrome.
2.1 Background

This chapter reports on five recent studies undertaken by the National Congenital Anomaly System (NCAS) in collaboration with other researchers, which aimed to measure the prevalence of underascertainment for selected conditions. Each study was a comparison of notifications to NCAS with a register which used multi-source ascertainment. All the registers used in these comparisons had been initiated in response to policy or public health questions when NCAS data had been judged to be inadequate.

The main outcome from this work is an estimate of underascertainment by NCAS for each condition. In this chapter 'underascertainment' is defined as the existence of cases not notified to NCAS. This chapter also describes the 'capture-recapture' methodology used in subsequent sections of the chapter to estimate the true prevalence of a condition, based on ascertainment by independent sources.

It is well established that NCAS fails to ascertain all cases of congenital anomalies, and the extent of underascertainment varies with condition. Such studies were either conducted nationally (Swerdlow 1988) or for a smaller geographical area (Dutton 1991, Knox 1984, Payne 1992).

Information on the prevalence of underascertainment is important for interpreting condition-specific notification rates. This chapter reports on five studies which aimed to measure the prevalence of underascertainment in anencephaly and spina bifida, anophthalmia and microphthalmia, congenital cataracts, cardiovascular anomalies and Down syndrome. These conditions were selected as marker conditions, using examples of external major anomalies (anencephaly and spina bifida), external anomalies manifesting in a spectrum of severity (anophthalmia, microphthalmia), major anomalies requiring a specialised clinical diagnosis (congenital cataracts), internal anomalies (cardiac anomalies) and congenital anomalies requiring a test for confirmation (Down syndrome). It was not possible
with the resources available to evaluate ascertainment for every condition included in NCAS. Nevertheless, from known characteristics of the different conditions presented here it was possible to make generalisations about ascertainment by NCAS of other similar conditions.

Anencephaly and spina bifida were selected as marker conditions of a single external anomaly that is very obvious at birth and should be diagnosed immediately at or before birth. In contrast, anophthalmia and microphthalmia represent a spectrum disorder. Anophthalmia is a single severe anomaly that should be obvious at birth, but mild microphthalmia may be difficult to distinguish from normal eyes. Both anophthalmia and microphthalmia require specialist diagnosis, which may affect notification rates to NCAS.

Congenital cataracts may not be obvious at birth, and require detection by a paediatrician and confirmation by an ophthalmologist, many of whom are unlikely to be familiar with NCAS. This section therefore explores issues about diagnosis of single less-obvious conditions that vary in severity and age at detection. This study allowed comparisons between severe conditions requiring skilled examination and specialist diagnoses and those presented in other studies that did not.

Cardiac anomalies were selected to consider serious internal anomalies. These anomalies affect one system, but are not visible at birth, although the symptoms may be. Therefore, these anomalies may not be detected at birth unless previously identified during prenatal screening. Hence this section explores ascertainment for single anomalies not visible at birth, requiring a test for diagnosis, and which may not be detected until some time after birth.

Finally, Down syndrome was selected since it is not a single anomaly but manifests as a recognised pattern of external signs. Identification of Down syndrome may initially be through recognition of syndromic signs, but it requires cytogenetic confirmation to make a diagnosis. This may be a short time after birth, unless it had
been detected during prenatal screening. If the condition was known prior to birth, the midwife attending the birth would be aware of the diagnosis and hence more likely to record the condition at birth notification, which in turn would increase the chance of the condition being notified to NCAS.
2.2 Measurement of underascertainment

2.2.1 Measuring ascertainment for congenital anomaly registers

Measuring underascertainment by comparison with a single complete data source is rarely possible. In an ideal world tracing all babies with a given condition would use as many different sources as possible to try to achieve as complete ascertainment as possible. This is very resource intensive, however, and funds rarely allow such an exercise at more than local level or for a specific national study of short duration.

One way to measure ascertainment is to use a local register or a condition-specific register that may use multiple source ascertainment as a standard against which to measure a national register. However, comparisons of notification rates in different populations are easily confounded. The analysis can attempt to take account of different confounders, for example, maternal age, but this would only be possible if they could be measured in the population at risk (usually all births). Direct or indirect standardisation could then be applied to the national rates using local condition specific rates. There may also be other unrecognised confounders, however, which have not been recorded on the congenital anomaly record.

Underascertainment can be determined by comparing ascertainment by different sources for the same population. When comparing data from a number of different sources, each known case needs to be compared with those held by the other data sources, to be able to merge the databases and eliminate duplicate cases. This final merged list, however, is biased downwards, because it misses those in the population who were not ascertained by any of the sources (Chao 2001).

All sources of data are incomplete; therefore it is important to use a methodology that estimates the prevalence of cases including those that failed to be detected by any source. If just one source of data was used to notify a condition it would not be possible to estimate completeness of ascertainment. With two sources it would be
possible to measure ascertainment in terms of the cases known to both sources, those known to one source and not to the other, and vice versa. Each additional source of information that was included would increase the number of known cases, usually by a smaller and smaller percentage, and might approach 100 per cent notification. There would usually be other cases that avoid notification to any system.

Some previous studies fitted statistical models to data from several sources to adjust for possible bias of ascertainment of affected births. These methods, however, assume that there are no errors of recording or diagnosis. An analysis of data from Aberdeen and Belfast (Little 1984) showed that there may be discrepancies in recorded diagnoses between sources which make this assumption untenable. Also, these methods assumed that the model which is the best fit to the data on ascertained cases is also the best model for the cases which were not ascertained. The work using Aberdeen and Belfast data also showed that there was a social process in ascertainment which renders these methods of adjusting for bias of ascertainment at best very complicated and at worst inapplicable. Within the known populations of cases ascertained by any source in Aberdeen and Belfast, the relative probability of ascertainment by any particular source was found to depend on variables such as the presence of an anomaly of another type, or parity; these variables differed from source to source. Thus it was concluded (Little 1992) that there was no satisfactory mathematical method of adjusting for bias of ascertainment. Ideally registries should aim for complete ascertainment. Data on the prevalence of abnormalities and on associations with other factors must be interpreted with a consideration of the possible effects of incomplete or biased recording. Consistency of results from several centres is therefore important in making causal inferences.

Despite these known problems of estimating the true prevalence of anomalies, researchers still require estimates to be made. Capture-recapture methodology evolved over 40 years to estimate the true total population using two or more incomplete sources of data, each of whose completeness was limited by the
resources available to detect the given conditions. This methodology was developed by ecologists in order to assess the size of animal populations in the wild (Seber 1982). Estimates of various animal populations using capture-recapture showed that the degree of under-counting could be estimated precisely and used to measure ascertainment. Therefore it was no longer necessary to identify every relevant case to obtain reliable estimates of total prevalence.

In recent years the method has been used to estimate the size of human populations and, increasingly, to determine prevalence estimates for a variety of health related conditions. An editorial in the British Medical Journal (Laporte 1994) suggested that using capture-recapture as a primary means of monitoring the human condition could bring substantial benefits. Capture-recapture estimation of the number of 'hidden' cases has been applied to a number of difficult to reach populations such as estimating the prevalence of drug misuse in Dundee, Scotland (Hay 1996) where complete ascertainment of cases was impractical. This methodology has been frequently used to determine the true prevalence of rare conditions and can be used to measure the completeness of congenital anomaly registers. Section 2.2.2 presents the methodology for capture-recapture, and applications of the methodology relevant to congenital anomalies are given in section 2.2.3.

2.2.2 Capture-recapture methodology

The capture-recapture methodology is based on the concept that the true population size can be estimated from the degree of overlap between two or more samples from the same population.

The simplest method is based on taking a sample of size 'a' from a closed population, marking this sample and replacing it back into the population. A second sample is then taken. This sample would comprise 'b' unmarked
individuals and 'c' marked ones. Assuming that every individual had the same probability 'p' of being a member of the second sample, then 'p' is estimated by

\[ p = \frac{c}{a} \quad \text{(i)} \]

If 'n' was the number of unmarked individuals in the total population at the time of the second sample, then 'n' was estimated by

\[ n = \frac{b}{p} = \frac{ab}{c} \quad \text{from (i) above} \quad \text{(ii)} \]

The estimate, N, of the total population size, is therefore given by N = the number of unmarked individuals plus the number of marked individuals, so

\[ N = n + a = \frac{a(b+c)}{c} \quad \text{from (ii) above} \]

The method is based on two assumptions. Firstly, that the samples are independent, meaning that their chance of being selected for one sample is not related to their chance of being selected for another sample. Secondly, the method assumes that the population is closed, which means that between samples being taken there is no migration into or out of the population. The technique can be applied to samples from the same population taken concurrently or at different times. The latter requires the additional assumption that the true prevalence does not change over the time period from which the samples were drawn.

In practice the assumption of independence, i.e. that every individual had the same probability of being selected for the second sample, may not be true because population members can differ in their inherent catchability. Some members may be more likely to be selected on either the first or second sample or both. For example, using the wildlife example, an injured animal may be more likely to be captured in both samples as it is less able to flee from capture. The probability of being selected may vary across the area being covered by the population, in that individuals may be more likely to be selected in some areas, or in some areas more effort may be put into selecting the sample (Cormack 1989). Also, the chance
of an individual being selected may change after being selected for the first sample.

An additional difficulty of applying the methodology to human populations was that it might also not be possible to confirm that the same subject had been selected for both samples. In animal censuses, those caught in the first selection were marked in some way (for example using paint) so that it was easy to identify whether the second or subsequent selection includes previously selected animals. This approach is clearly not appropriate for humans, although the use of unique personal identification, such as fingerprints, has been used to identify whether an individual had been previously known to the given system.

The usual application of capture-recapture to congenital anomaly populations is that of comparison with data from different registers. However, if the data sources are dependent, then capture-recapture underestimates true prevalence. Adaptations of the methodology have been suggested to help overcome these problems. One adaptation of the method was to use a third or subsequent selection to quantify failures of the earlier selections and therefore adjust for the lack of independence. Multiple pooling of all but one source was suggested to provide a generally efficient method for overcoming the problem of likely variable catchability (Hook 1993).

To overcome variability in the probability of being selected, a stratified adaptation has been suggested (Darroch 1961). Later work introduced a logistic regression model. This model allowed different capture probabilities across individuals and across capture times. It uses the characteristics of the captured individuals to explain their probabilities of capture. This analysis was illustrated by applying the model to occupational disease data from Finland in 1981 (Hemminki 1981). More complex models allow for the non-independence of samples with individuals observed in earlier samples having either an increased or reduced chance of being selected again compared to the remainder. These models also allow for the
population being open, with birth, death and migration within the population (Cormack 1989).

2.2.3 Applications of capture-recapture relevant to congenital anomalies

Knowledge of congenital anomaly prevalence is important locally to help estimate the prevalence of disability and assessing their needs. Two studies suggested that capture-recapture might be helpful to meet these needs by providing a quick and cheap alternative to population surveys for estimating the prevalence of physical disability (Laporte 1994) and as part of a needs assessment for a new rehabilitation service in East Berkshire (Gutteridge 1994). Other applications have used the methodology specifically to estimate ascertainment of congenital anomalies. The Metropolitan Atlanta Congenital Defects Program (MACDP) used capture-recapture using a newly available delayed source of notifications (birth certificates) with the result that MACDP was estimated to have underestimated defects by 3 per cent at one year after birth and 5 per cent at two years after birth (Honein 1999). Hook and Regal compared capture-recapture methods applied to the estimation of the true prevalence of Down syndrome (Hook 1982).

Studies using capture-recapture to estimate total human populations highlighted problems which are relevant in an assessment of the suitability of this methodology for congenital anomaly registers. These include the population not being closed as people move into or out of the area being studied, the quality of health information varying between sources, people with more severe problems being inherently more catchable by the different sources, and duplicate cases being missed if people used different names when presenting at different sources.

The study of homelessness in Westminster (Fisher 1994) identified several of these problems. The population was not closed as people could have moved and become or stopped being homeless. Their mental health data relied on the use of information recorded by different sources, cross-referencing between sources, and
the quality of this information varied between sources. The study found that mental health problems were significantly less common in the unobserved or hidden homeless, compared with those who were easily surveyed suggesting that those with more severe problems were inherently more catchable.

Duplicate notifications to a congenital anomaly register may not be detected if a child is notified using different names. This problem arises if the child was known by a different name to that registered at birth, or if their name had changed between birth and birth registration. Matching for recapture based on name can have severe limitations, particularly in developing countries. A study of causes and rates of death in three areas of Tanzania attempted to use capture-recapture techniques (Black 1994). Names were used as identifiers but problems were encountered with variations in the spelling of names at different times, together with people using completely different names at different times and for different purposes.

Capture-recapture has been used in all the five studies presented in this chapter to estimate the true birth prevalence of these conditions. Chronologically, the order of the five studies were; the study of anencephaly and spina bifida, then cardiac anomalies, then anophthalmia / microphthalmia, then congenital cataracts, and finally Down syndrome. The first three studies use capture-recapture to compare NCAS notifications with pooled data from other sources. The latter two studies used the simple application of capture-recapture, to compare NCAS notifications with one other source. The studies are presented in Section 2.3 in the order they were discussed in Section 2.1, since this follows the spectrum from the external major anomalies obvious at birth, to internal anomalies, to those requiring a test for confirmation.
2.3 Evidence of underascertainment

2.3.1 Anencephaly and spina bifida

2.3.1.1 Background

Anencephaly is a severe neural tube defect (NTD) where a major portion of the brain, skull and scalp is absent. It is incompatible with survival for more than a few days. Spina bifida is another NTD caused by the failure of the fetus’s spine to close properly during the first month of pregnancy, and will result in varying degrees of paralysis of the lower limbs. Both conditions are unlikely to be overlooked by midwife or doctor and are usually unequivocally the cause of the infant’s stillbirth or death. Therefore local areas using birth notifications as their sole source of information for notification to NCAS were expected to have detected most cases. These conditions were selected for analysis here since they are single severe conditions which are easily identifiable at birth.

An analysis in 1966 of ascertainment of anencephaly by NCAS matched 1233 notifications of anencephaly against the 1584 stillbirths and infant deaths registered from this condition in the same period (Weatherall 1969). Problems were resolved through correspondence with the Medical Officer of Health. There were 1155 cases of anencephaly which matched stillbirths and death information against notifications. Another 77 babies notified to NCAS with anencephaly were recorded on their stillbirth or death certificate with a cause of ‘gross central nervous system abnormalities’ or ‘multiple gross congenital abnormalities’. There remained 429 babies certified at death or stillbirth with anencephaly, which had not been notified to NCAS (27 per cent). The proportion varied for different geographical areas. However, areas that reported efficiently were distributed all over England and Wales, as were those that were poorly reported. The analysis showed that two factors favoured efficient notification, first that all information required to complete the notification form should be collected at one time, and second that routine follow-up procedures for collecting the necessary information needed to be
established. It also suggested that the competing interests of local surveys mitigated against the efficient notification to a central body since those reporting to a local register tended not to report their data also to a national register.

Ascertainment in NCAS is estimated here from two studies through collaboration with the Oxford Record Linkage Study (ORLS). In order to undertake a historical and prospective study of the relationship between successful infertility treatment and the subsequent outcome of the pregnancy as a neural tube defect (NTD), a register of NTD affected babies or terminations in Oxfordshire and West Berkshire was established in the ORLS. The ORLS is assembled from computerised extracts of hospital inpatient and day case records, together with birth registrations of all babies born in Oxfordshire and West Berkshire and death certificates for them and for all current residents of the area (Murphy 1996). This analysis, therefore only evaluates notifications for one geographical area and results cannot necessarily be extended to all of England and Wales.

2.3.1.2 Methods

Cases identified locally to build the register of NTD affected babies within the ORLS were distinguished from the NCAS data available for the locality. Voluntary notification to NCAS of live and stillbirths, together with the statutory notification to the Department of Health of terminations and the grounds on which they were performed, provided an independent count of the prevalence at birth or at termination for NTD in the ORLS area. Independence could reasonably be assumed in this comparison because birth notification via the district community health offices and the physicians undertaking the abortion were the usual source for NCAS and abortion data, whilst the ORLS was built directly from notifications from clinicians and not from the district offices. Therefore it was possible to use capture-recapture to estimate the true prevalence of the condition based on ascertainment by two sources.
Terminations were included in the analysis because most NTD pregnancies, if prenatally diagnosed, ended in termination rather than birth. From 1979 onwards there was only an average of 10 NTD cases each year in live and stillbirths in the ORLS area. Therefore restricting the analysis to anencephaly cases in live and stillbirths in the ORLS area resulted in too few cases to provide a meaningful analysis.

2.3.1.3 Results

The ORLS comprised 880 cases of which 372 were anencephaly (where more than one NTD was present, anencephaly was counted as the definitive condition in combination with other NTDs). Using a capture-recapture analysis on the data from the local register and from NCAS and abortion data between 1968 and 1990 (Hey 1994) showed that the ORLS ascertained 96 per cent and NCAS 66 per cent of the true number of local NTD cases (in live births, stillbirths and legal terminations), and suggested that combining data from NCAS and the local register would produce data that were 99 per cent complete.

This analysis was subsequently repeated for data between 1974 and 1994 (Murphy 1996). Again the ORLS was estimated to be more than 90 per cent complete compared with ascertainment by NCAS of 60 per cent.

Figure 2.1 shows NTD notification rates in 1980 to 1997 for NCAS and other local congenital anomaly registers in Great Britain. If the prevalence of NTDs in Oxfordshire and West Berkshire were typical of England and Wales, the national rate (based on notifications of NTDs in births and terminations combined) was underestimated by about one third.
As shown in Figure 2.1, the under-reporting by NCAS compared to the local registers seemed to be reasonably consistent over time. The statistical stability of national data was felt to be greater, however, and it was judged that the NCAS data for England and Wales could be used to monitor trends if not absolute risks (Murphy 1996).

2.3.2 Anophthalmia and microphthalmia

2.3.2.1 Background

Anophthalmia (absent eyes) is an external severe single anomaly. This example was selected to estimate underascertainment in NCAS as it represented anomalies that, despite being obvious, require specialist diagnosis. The study that enabled the analysis was set up in response to a policy demand that could only be met to a
limited extent by NCAS, and so necessitated funding for a separate retrospective identification and register of affected babies.

In early 1993 there had been media reports of an alleged cluster of cases following exposure to the pesticide Benomyl (Paduano 1993). An early analysis of the cases known to NCAS showed that there was serious underascertainment. Although analyses of NCAS over 30 years, a period in which there was a 1000 fold increase in use of Benomyl (Castilla 1994), showed no change in birth prevalence, funding was granted for a national register to be compiled of babies in England with anophthalmia and microphthalmia born in 1988-94. This used multi-source ascertainment, to achieve the maximum number of children for cluster analysis.

2.3.2.2. Methods

The National Anophthalmia/ Microphthalmia Register included terminations following prenatal diagnosis of congenital anomaly, stillbirths, or live births. Initially cases were identified from 21 different sources, including NCAS, local congenital anomaly registers, Moorfields Hospital, child health surveillance systems, and special needs registers. Questionnaires were sent to the clinicians who had responsibility for the care of the children. The register, covering 4.5 million births over the time period of 1988-94, was chosen to overlap as far as possible with the period of concern but not to go back too far and risk major underascertainment of cases in earlier years (Dolk 1998a). Although Trisomy 13 is the most common chromosomai syndrome associated with these eye defects, notification for these cases were subsequently excluded.

For the purposes of the analysis of ascertainment by NCAS presented here, since NCAS only included anomalies recorded for live and stillbirths, termination data from the National Anophthalmia/ Microphthalmia Register were excluded.
2.3.2.3. Results

The register identified a total of 490 affected babies from 21 different sources (Busby 1998). An analysis undertaken by the National Anophthalmia Register (Busby 1998) had shown that no one source had identified more than 26 per cent of affected babies nationally, with paediatricians (26 per cent), district sources (25 per cent) and NCAS (22 per cent) notifying the greatest number of cases. The two local registers which had been collecting data for the entire time period knew of 55 per cent (Oxford region) and 63 per cent (Northern Region) of cases in their regions.

Using the National Anophthalmia Register, the prevalence of non-chromosomal anophthalmia/microphthalmia was estimated to be 0.96 per 10,000 births.

Once terminations and Trisomy 13 cases were excluded, 414 confirmed records remained for the analysis presented here. Table 2.1 shows the number of cases of anophthalmia and microphthalmia by severity, whether they were known to NCAS and the number of other sources which had reported the case.

Of the 414 records in the analysis, only 12 cases (3 per cent) had been notified solely from NCAS. A further 108 (26 per cent) had been notified from just one source but not NCAS, 180 (43 per cent) from two sources and 114 (28 per cent) from 3 or more sources. In total, NCAS had known of 14 per cent of confirmed cases on the register. Twenty five per cent of severely affected babies had been notified to the register by NCAS, compared with only 6 per cent of mild cases. Severely affected babies were more likely to be notified from two or more sources; 96 per cent were notified by more than one source compared with only 55 per cent of mild cases.
Table 2.1 National Anophthalmia and Microphthalmia Register data by severity, whether known to NCAS, and number of other reporting sources

<table>
<thead>
<tr>
<th>Number of other sources</th>
<th>Anophthalmia</th>
<th>Severe microphthalmia</th>
<th>Moderate microphthalmia</th>
<th>Mild microphthalmia</th>
<th>Microphthalmia severity unknown</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Known to NCAS No</td>
<td>Yes</td>
<td>Total</td>
<td>Known to NCAS No</td>
<td>Yes</td>
<td>Total</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>36</td>
<td>3</td>
<td>39</td>
<td>26</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>5</td>
<td>17</td>
<td>12</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>6</td>
<td>14</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>61</td>
<td>20</td>
<td>81</td>
<td>58</td>
<td>14</td>
<td>72</td>
</tr>
</tbody>
</table>

Estimated true total 84.2 76.5 107.9 406.0 104.0 506.9
Using a capture-recapture analysis on the data presented in Table 2.1, comparing cases known to NCAS with those known to the pooled other sources, the total number of anophthalmia cases was estimated to be 84, of which NCAS ascertained 20 (24 per cent). For cases of mild microphthalmia the total was estimated to be 406 cases of which NCAS knew of only 7 (2 per cent). Of the cases known to NCAS, only two had been reported from other sources.

2.3.3. Congenital cataracts

2.3.3.1 Background

Congenital cataracts were selected as a condition to evaluate underascertainment in a single system anomaly that varies in severity and requires a test to confirm the diagnosis. Congenital cataracts may be detected by paediatricians at the birth examination or by GPs at the 6 week surveillance check, but the condition requires confirmation by ophthalmologists, many of whom may be unfamiliar with NCAS. Therefore, congenital cataracts may not be diagnosed until after discharge from hospital, and so not be notified to NCAS. Nevertheless, the optimal age for operation is by 3 months, therefore it is important that it is detected as quickly as possible.

The purpose of the study reported here was to assess the level of ascertainment by NCAS of children with congenital cataract. This study was undertaken in conjunction with an independent population-based active surveillance study of this disorder from paediatricians and ophthalmologists (Rahi 1999, Rahi 2000). Despite being the main source for estimation of birth prevalence of congenital ocular anomalies, the proportion of eligible children actually notified to NCAS had not previously been evaluated at national level.

In the UK, the reports of two national joint working parties of the Royal College of Paediatrics and Child Health and the Royal College of Ophthalmologists (Hall
1996, Royal College of Ophthalmologists and British Paediatric Association. 1994) recommended inspection of the eyes and evaluation of the papillary red reflux of all infants during the newborn period and again at 6 to 8 weeks. Despite these recommendations, a substantial proportion of children with congenital and infantile cataract had not been diagnosed by 3 months of age. A national study of congenital and infantile cataract found that 35 per cent of babies with cataracts present at birth were detected at the routine newborn examination and a further 12 per cent at the 6-8 weeks examination leaving half of all children with the condition being undetected (Rahi 1999).

Non-ophthalmic health professionals detected most cases: 41 per cent by a hospital paediatrician and 27 per cent by a general practitioner. Routine ocular examination of young infants requires specific knowledge and skills but with appropriate training can be performed by clinical staff with limited previous experience (Ruttum 1987). Thus, given that only 35 per cent of cases were identified in the newborn period, the principal source of information used for notification to NCAS, it would be expected that this condition would be under-notified. This study aimed to measure this underascertainment.

### 2.3.3.2 Methods.

NCAS data of children newly diagnosed with congenital cataract in England and Wales were compared with paediatric surveillance data by age at detection and severity. Independent ophthalmic and paediatric national active surveillance schemes were used to identify all infants (up to 1 year of age) newly diagnosed with congenital and infantile cataract in England and Wales between 1 September 1995 and 31 August 1996. These notifications were compared with those made independently to NCAS during the same period to determine the proportion of cases identified by both sources.

The ophthalmic surveillance scheme was established for this study through the British Congenital Cataract Interest Group (Rahi 2000). The long-established
paediatric scheme, run by the British Paediatric Surveillance Unit of the Royal College of Paediatrics and Child Health, had successfully facilitated incidence studies of a number of uncommon childhood conditions (British Paediatric Surveillance Unit 2000). For the duration of the study reported here, paediatricians were sent reporting cards monthly, and ophthalmologists every two months, and asked to notify new cases or to confirm that no new cases had been seen, thus making the surveillance active rather than passive. Detailed information was collected about all notified cases using standard forms developed specifically for the study. A separate capture-recapture analysis indicated that 92 per cent of eligible infants had been identified through these two sources (Rahi 1999a).

All new notifications to NCAS in the two years between January 1995 and December 1996 that included any ICD 10 codes (World Health Organisation 1992) applicable to congenital or infantile cataract, isolated or in conjunction with systemic or other ocular disease, were extracted. The search was restricted to those aged 12 months or less at notification as a review of a random sample of previous notifications had indicated that children diagnosed outside infancy were unlikely to be notified to NCAS, despite the abolition of the previous upper age limit for notification of ten days after birth. To ensure late notifications were identified, a two-year period was used to straddle the one-year case ascertainment period of the surveillance study. Cases identified in England and Wales through the active surveillance schemes described earlier were matched manually with notifications to NCAS using initials, gender, date of birth, partial postcode, laterality of cataract and the presence of other anomalies. A successful match required agreement on at least four criteria. This procedure was conducted twice during the course of the study and again at the end of the case ascertainment period. The number of cases identified by active surveillance and also notified to the NCAS was determined.

2.3.3.3 Results

In 12 months from October 1995, 149 children, born in 1995 or 1996, with congenital cataract diagnosed by their first birthday were identified in England and
Wales through the ophthalmic and paediatric surveillance schemes. Of these, 85 (57 per cent) children were diagnosed in the first month of life.

In two years from January 1995, there were 21 notifications to NCAS of congenital cataract in children aged one year or less. Three (14 per cent) children who had also been notified through the paediatric surveillance scheme were subsequently confirmed by their managing ophthalmologists as not having cataract. One child with multiple anomalies died shortly after birth, precluding further verification, and together with two other cases with insufficient information for matching, had to be excluded. The remaining 15 (71 per cent) children, 14 aged one month or less at notification, were matched with cases identified through the ophthalmic and paediatric surveillance schemes.

Thus, 10 per cent (15/149) of eligible children with newly diagnosed congenital or infantile cataract were notified to the NCAS. A higher proportion of those diagnosed as neonates (16 per cent, 14/85) than in later infancy (2 per cent, 1/64) was ascertained through the NCAS. Given the very small number of cases notified to the NCAS, analysis by factors that might influence notification, such as laterality of cataract, geographic region or associated systemic disease, was considered inappropriate, being subject to a high level of random error. Since there were no additional cases notified to NCAS which had not been detected through either the ophthalmic or paediatric surveillance schemes, a capture-recapture analysis of NCAS data and pooled data from the surveillance schemes, would suggest that complete ascertainment had been achieved, whereas the earlier analysis based on the two surveillance schemes had suggested that the schemes had achieved 92 per cent ascertainment (Rahi 1999a).
2.3.4 Cardiac anomalies

2.3.4.1 Background

The next example selected was cardiac anomalies since these might not be detected immediately at birth given that they are internal anomalies. Thus, these anomalies would be expected to be less well notified. This was shown previously in an analysis of congenital malformations of the cardiovascular system notified from Trent region in 1984-87. Notification rates were shown to vary considerably within Health Districts in the Trent Region, whilst a review of other data sources suggested that there might be no substantial differences in the underlying rates of these anomalies (Payne 1992).

In the study presented here (Smeeton 1999), the prevalence of congenital heart disease (CHD) in the first year of life was estimated using five different datasets based on two English Health Regions covering a population of 6,872,000. The aim of the study was to assess the prevalence of CHD in South East Thames Regional Health Authority (South East Thames) and Wessex Regional Health Authority (Wessex). These two areas were of interest because in the late 1980s an active training scheme to detect severe malformation of the heart had taken place in several hospitals in South East Thames. Such formal training did not take place in Wessex. Data for this study were also used here to estimate ascertainment of cardiac anomalies by NCAS.

per 1,000 had been reported, but they were based on hospitalisation, and some children were admitted more than once in a given year (Gillum 1994). There were large variations in rates between the different regions and districts in England and Wales (Payne 1992). Regional registers were able to search their records intensively for cases and hence reported higher rates. For example, the Register in Birmingham reported a prevalence of CHD between 3 and 5 per thousand in the period 1964 to 1982 (Knox 1991). Other smaller studies using data from the 1960s and 1970s have reported a prevalence of between 5.5 and 7.2 per thousand (Bound 1977, Dickenson 1981, Mitchell 1971), but these studies covered a relatively small population and were carried out over a 10 year period. In 1994, based on notifications of cases identified within 10 days of birth, NCAS reported the prevalence of CHD and anomalies of the circulatory system to be less than 1 per 1000 (ONS 1997).

2.3.4.2 Methods

Five different datasets were used comprising data on cardiac anomalies for South East Thames and Wessex. Each data set had a varying degree of completeness and covered the period June 1993 to August 1994. The first source of data was NCAS data for the period reported from the two regions. Secondly ONS infant death and stillbirth data for the same geographical areas were used in the analysis. As capture-recapture depends on overlap between groups, and stillbirths and infant deaths are mutually exclusive, these two groups were combined to provide a single source of data.

The third source of data was that obtained by the research team at United Medical and Dental Schools. All paediatricians and pathologists in South East Thames and Wessex were asked to complete a form for each patient with severe malformation of the heart diagnosed before the age of 6 months. The Department of Foetal Cardiology at Guy’s and the Departments of Paediatric Cardiology at Guy’s and Southampton Hospitals also notified the researchers of appropriate cases.
The fourth source of data was from a national audit of heart malformations during the antenatal and postnatal period carried out by the British Paediatric Cardiology Association (BCPA). Patients were notified to the audit only if they had CHD of sufficient severity to require operation or to need catheter intervention in the first year of life. The fifth source was based on the South East Thames Congenital Anomaly Register, which began in 1992 and ran until funding ceased in 1998. This register had requested notification of all cases of major congenital anomalies from obstetric, fetal medicine, genetic, paediatric and pathology units in the region. Data were provided to the study by the co-ordinator.

A database of all cases known to the five different sources was compiled. Children who appeared in more than one source were manually recognised from various combinations of identical postcode, child’s date of birth, mother’s date of birth, child’s name and mother’s name. Records from different sources relating to the same child were merged and their record indicated presence or absence in each source. Terminations of pregnancy were excluded, as were cases for residents outside the Wessex or South East Thames areas, because NCAS only records anomalies recorded in live and stillbirths for residents of the given health areas.

Two separate analyses were carried out using these data. The first analysis (Smeeton 1999) was carried out using capture-recapture, including log-linear modelling. It compared the overlap between different registers relating to the same period of time. The models were applied to each Region separately, starting with the assumption that the sources were independent. Interaction terms were introduced into the model as appropriate and goodness-of-fit was assessed using the deviance, which is based on the chi-square statistic, rejecting the model if the p value was less than 5 per cent. The analysis used the Generalised Linear Interactive Modelling (GLIM) statistical package (Francis 1994) applying the program described by Cormack (Cormack 1989). The second capture-recapture analysis, presented here, includes NCAS records, and pooled data for cases known to the other sources.
2.3.4.3 Results

In total 770 records were derived from the five sources. These reduced to 366 individuals when information from one or more sources was brought together and the cases were restricted to those resident at birth in Wessex or South East Thames.

<table>
<thead>
<tr>
<th>Known to other sources</th>
<th>Wessex</th>
<th>South East Thames</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Known to NCAS</td>
<td>Known to NCAS</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>12</td>
<td>129</td>
</tr>
<tr>
<td>No</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2 Congenital heart disease anomalies by whether known to NCAS and to other sources

Table 2.2 shows the number of cases of CHD according to whether they were known to NCAS and whether they were known to other sources.

A capture-recapture analysis of NCAS and the pooled data estimated that the number of cases in Wessex was 306 and the number of cases in South East Thames was 491, suggesting that NCAS ascertained 8 per cent of cases in Wessex and 5 per cent of cases in South East Thames.

The capture-recapture models based on all sources presented in the published analysis (Smeeton 1999) showed that ascertainment of CHD by one or more of the five sources was estimated to be between one third and one half. It was estimated that for South East Thames the prevalence rate was between 5.5 and 9.0 per 1000, and between 4.3 and 5.1 per thousand in Wessex.
2.3.5 Down syndrome

2.3.5.1 Background

Down syndrome was selected since it would usually require cytogenetic confirmation that might not be available at the time of notification. It is not a single anomaly but manifests in a recognised pattern of external signs.

Information about most Down syndrome cases was sent to the National Down Syndrome Cytogenetic register (NDSCR) based at St Bartholomew's Hospital Medical School, London. This register was established in 1989 and was originally funded by the MRC. It was a register of all diagnoses of Trisomy 21 and its variants, reported from all the Clinical Cytogenetic Centres in England and Wales. The study presented here reviewed NCAS data on Down syndrome (Trisomy 21) in live and stillbirths, in comparison with data from the NDSCR, and used these results to undertake a follow-up study with the aim of improving ascertainment of Down syndrome by NCAS.

The balance between two conflicting trends determines the birth prevalence of Down syndrome in a given year. First, Down syndrome is a condition where, if detected antenatally, parents may choose a termination. Assuming no change in background incidence, over time the birth prevalence of Down syndrome would fall if more cases were prenatally detected and the pregnancy was terminated. Secondly, in contrast, the average age at childbirth has been increasing (ONS 1999) as women delayed childbearing. As the risk of Down syndrome increases with maternal age, increasing average maternal age would result in an increase in prevalence.

Down syndrome was estimated from epidemiological studies to have a birth prevalence of 10.8 per 10,000 births, and an analysis of NCAS data in 1974-87 showed that only 67 per cent of the estimated affected births had been notified (Cuckle 1991). Subsequently the NDSCR found a decreasing trend in birth
prevalence from 11.1 cases per 10,000 live births in 1989 to 9.2 in 1993 (Alberman 1995). Using capture-recapture methodology, a more recent comparison of NCAS data with that from the NDSCR based on data for 1990-93 (Huang 1997) suggested that while the NDSCR was estimated to be 94 per cent complete, only 48 per cent of Down syndrome live births had been notified to NCAS. Given this fall in the completeness of Down syndrome notifications to NCAS, a further capture-recapture exercise was undertaken using more recent data.

In addition, over the period 1971 to 1999, the number of notifications of Down syndrome had decreased from 582 to 336. Over the same period, the number of abortion notifications, which gave Down syndrome as the principal medical condition, increased from 80 to 329 (Figure 2.2). In 1997, for the first time ever, the number of terminations for Down syndrome was higher than the number of notifications for births. Given this increase in the number of terminations, some decrease in the number of Down syndrome notifications would be expected in births. Nevertheless, the number of Down syndrome notifications in births was at the low level which previous analysis had shown missed 50 per cent of cases (Huang 1997). Therefore, it was important to follow-up notifiers to try to improve completeness.

Figure 2.2 Down Syndrome Births before follow-up, and Abortions; England & Wales, 1987-2000
2.3.5.2 Methods

The analysis reported here is a capture-recapture analysis of NCAS and NDSCR data for 1995 to 1998. An earlier survey postal of all HAs by ONS (unpublished) showed that only one HA had used cytogenetic laboratories as a source of congenital anomaly information. Therefore, since NCAS and the NDSCR use independent sources of data, capture-recapture techniques can be used to estimate completeness of notification to both sources.

Following the capture-recapture analysis, NCAS wrote to data suppliers in the NHS Trusts in April 1999 giving details of babies resident in their HA notified as having Down syndrome. The Trusts were asked to check against their known cases to see whether they could find additional cases of Down syndrome and to supply completed notification forms for any additional cases identified.

For cases in Trent and Wales, however, where NCAS already received notifications from their local Register (CARIS in Wales and the Trent CAR), and in the Northern Region, NCAS worked with the local register to improve completeness. These registers were asked to provide electronic details of cases not previously notified.

This request for additional cases was restricted to 1997 and 1998 births because it was felt that given recent changes in Trusts, earlier cases might be difficult for them to trace. Comparisons were made between the NCAS database before and after the additional cases had been notified.

2.3.5.3 Results

Table 2.3 shows the number of NCAS and NDSCR records which were matched for each year 1995 to 1998, together with the estimated total number from the capture-recapture analysis.
<table>
<thead>
<tr>
<th>Year</th>
<th>Known to NDSCR</th>
<th>Known to NCAS</th>
<th>Estimated total cases</th>
<th>Percentage ascertainment by NCAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes No Total</td>
<td>Yes No Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>Yes 297</td>
<td>No 307</td>
<td>Total 604</td>
<td>641</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>Yes 300</td>
<td>No 328</td>
<td>Total 628</td>
<td>670</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>Yes 264</td>
<td>No 421</td>
<td>Total 685</td>
<td>732</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>Yes 265</td>
<td>No 376</td>
<td>Total 641</td>
<td>683</td>
</tr>
</tbody>
</table>

Table 2.3 Summary of matching NCAS and NDSCR data, 1995 - 1998

Ascertainment by NCAS was between 39 per cent for 1997 and 49 per cent in 1995, lower than the earlier analysis (Huang 1997).

The follow-up exercise resulted in an increase in the number of notifications for 1997 and 1998 from 282 to 419 in 1997 and from 283 to 396 in 1998, an overall increase of 44 per cent (Figure 2.3). This represents on average two extra cases per HA per year, with the resulting birth prevalence of 6.3 cases per 10,000 births. The overall rate for England and Wales increased from 4.4 to 6.3 per 10,000 births. This reflects a good improvement in notifications and is closer to levels in registers known to be more complete. However, subsequent to the follow-up period, the number of notifications fell slightly in 1999 and 2000 (Figure 2.3).
Figure 2.3 Down Syndrome Births after follow-up and Abortions; England & Wales, 1987-2000
2.3 Summary of evidence

Table 2.4 shows a summary of levels of underascertainment for each of the conditions reported in this chapter.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Year</th>
<th>Area</th>
<th>Severity</th>
<th>Estimated % notified to NCAS</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anencephaly and spina bifida</td>
<td>1968-90</td>
<td>national</td>
<td></td>
<td>66%</td>
<td>61% 71%</td>
</tr>
<tr>
<td></td>
<td>1970-94</td>
<td>Oxon &amp; W Berks</td>
<td></td>
<td>60%</td>
<td>54% 66%</td>
</tr>
<tr>
<td>Anophthalmia and microphthalmia</td>
<td>1988-94</td>
<td>national</td>
<td>severe</td>
<td>24%</td>
<td>8% 40%</td>
</tr>
<tr>
<td></td>
<td>1988-94</td>
<td>national</td>
<td>mild</td>
<td>2%</td>
<td>0% 4%</td>
</tr>
<tr>
<td>Congenital cataracts</td>
<td>1995-96</td>
<td>national</td>
<td>Diagnosed as neonates Later diagnosis</td>
<td>16%</td>
<td>3% 29%</td>
</tr>
<tr>
<td></td>
<td>1995-96</td>
<td>national</td>
<td></td>
<td>2%</td>
<td>0% 8%</td>
</tr>
<tr>
<td>Cardiac anomalies</td>
<td>June 1993 to August 1994</td>
<td>SE Thames Wesssex</td>
<td></td>
<td>8%</td>
<td>3% 13%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5%</td>
<td>2% 8%</td>
</tr>
<tr>
<td>Down syndrome</td>
<td>1974-87</td>
<td>national</td>
<td></td>
<td>67%</td>
<td>61% 73%</td>
</tr>
<tr>
<td></td>
<td>1990-93</td>
<td>national</td>
<td></td>
<td>48%</td>
<td>41% 55%</td>
</tr>
<tr>
<td></td>
<td>1995-98</td>
<td>national</td>
<td></td>
<td>39-49%</td>
<td>33% 56%</td>
</tr>
</tbody>
</table>

Table 2.4 Ascertainment by NCAS for selected conditions

Easily visible conditions, such as anencephaly tend to be better notified to NCAS than those such as Down syndrome and congenital cataracts which require expert clinical confirmation or laboratory tests. Severe manifestations of the condition tend to be better notified than mild cases.
2.4 Discussion

Anencephaly and spina bifida

The capture-recapture analysis of NTDs estimated that the ORLS was more than 90 per cent complete compared with ascertainment by NCAS of 60 per cent. This difference is likely to be a result of the ORLS actively seeking new cases, compared to NCAS which is a passive system, receiving data from a network of notifiers.

A study in the United States compared data on the prevalence of Spina bifida from two surveillance systems, one passive and one active (Lary 1996). Data from 16 state congenital anomaly programmes with active collection of congenital anomaly information were analysed for live and stillborn infants born between 1983 and 1990 with spina bifida. The programmes differed in size, racial/ethnic composition, surveillance methods and ascertainment. These findings were compared with CDC’s Birth Defects Monitoring Program (BDMP) for the same period. BDMP is a passive case ascertainment surveillance system that obtains data from participating hospitals in 50 states. The birth prevalence rate for spina bifida from the 16 states was 4.6 cases per 10,000 births; the BDMP rate was nearly identical (4.4 cases). The report concluded, therefore, that active and passive surveillance systems could provide reliable information concerning national trends in the birth prevalence of spina bifida. Of course, ascertainment varies for different conditions, and spina bifida as a major anomaly visible at birth and capable of prenatal detection would be expected to be well reported. It may not be the size of the national system that affects ascertainment but the number, type and awareness of reporting sources.

The birth prevalence of anencephaly and spina bifida was estimated to have been under notified to NCAS by 19 and 13 per cent respectively based on data for 1964-72 (Rogers 1976). Adjusting for underreporting using this estimate, Morris and Wald estimated that the birth prevalence of NTDs had fallen by 96 per cent
between 1970 and 1997, of which 40 per cent was due to antenatal screening and termination of pregnancy and 56 per cent to a decline in incidence.

**Anophthalmia and microphthalmia**

The various sources of data used to create the anophthalmia register, other than clinicians, could only identify half of all cases eventually registered. NCAS identified only 22 per cent of all cases and 29 per cent of severe cases. This lack of duplication between the different sources shows the difficulty in retrospective reporting and the need for multiple sources when establishing a register.

A large international epidemiological study on anophthalmia and microphthalmia found a birth prevalence rate of 1.50 per 10,000 births, varying between 0.92 and 2.29 between registers (Kallen 1996). This study excluded chromosomal cases, as did the study reported in this chapter. The birth prevalence rate of 0.96 shown in Section 2.3.2.2 was within this range of rates.

**Congenital cataracts**

Ascertainment of children with congenital cataract through passive reporting to NCAS was shown to be low. As a result, NCAS was likely to be insensitive to small but important changes in risk factors for congenital cataract, and large changes in frequency were likely to be detected with limited precision. This restricts its use in monitoring secular and other trends of congenital and infantile cataract, and other ocular anomalies with low ascertainment. This is an important omission from NCAS since congenital ocular anomalies contribute significantly to childhood visual morbidity, with congenital cataract being a major cause of visual impairment throughout the world (Hornby 2000). In England and Wales isolated congenital cataract accounted for 3 per cent of all new blindness or partial sight certifications amongst children in 1990, with other ocular or central nervous system anomalies accounting for a further 19 per cent (Evans 1995).
Congenital heart disease
The estimated number of children with cardiac anomalies was generally large and in all capture-recapture analyses greatly exceeded the number of known cases. In addition, the ability of a single source to know of all cases was less than adequate. The most comprehensive source, the project of the BPCA, contained only 70 per cent of the known Wessex cases and only 40 per cent of the South East Thames cases. Additional cases were identified after the analysis had been completed, but even these additional cases would have still fallen short of the number of estimated cases based on capture-recapture methods.

Each of the different sources of data on cardiac anomalies operated under different objectives, and hence, the level of detail recorded differed between sources. South East Thames rates were consistently higher than those in Wessex. This is surprising for two reasons. First, given the extra training in South East Thames to detect these anomalies, more cases would have been expected to be detected antenatally and termination offered if appropriate. Since terminations were not included in this analysis, this should have resulted in higher rates in stillbirths and infant deaths in Wessex. Also, other studies have suggested that rates of CHD are consistent across different population groups (Benson 1989). Possible reasons for this opposite finding may be that the levels of notification were systematically lower in Wessex, there is a different population mix in the two areas or that the difference was due to random variation in the rates over time. It is unlikely that the magnitude of the difference between the two areas is due to random variation.

Down syndrome
The NCAS notification rate for Down syndrome was initially 4.1 per 10,000 births, increasing to 5.4 after follow-up. Yet published data from the West Midlands Register give 3-year incidence rate for live and stillborn Down syndrome cases in the West Midlands to be 10.2 per 10,000 registered births, almost double the rate after follow-up (West Midlands Congenital Anomaly Register 2000). This illustrates the deficiencies in reporting to NCAS by the local trusts. Nevertheless, the identification of new cases of Down syndrome shows the high level of participation...
in the follow-up exercise reported in this Chapter. Despite the fact that NHS trusts can notify NCAS at any time, most focus on congenital anomalies given on birth notification forms. Late diagnoses are seldom notified to NCAS. Given that NCAS was thought to miss over 50 per cent of all Down syndrome notifications, however, the additional cases identified represent only half of the estimated shortfall. Sustained effort will be needed to maintain these levels, as shown by the subsequent fall in notification rates in 1999-2000.

Limitation of these studies
The findings of these various studies are subject to the limitation that the prevalence estimates derived using the capture-recapture method probably is an underestimate because of some unquantified passive dependence between the two or more data systems. For example, the children identified by one data source may have an increased likelihood of being identified on another as supported by higher estimated rates of completeness on NCAS for severely affected children than for less severely affected children.

Completeness by condition
Completeness of reporting to NCAS varies with the nature and severity of the anomaly, some severe, life-threatening anomalies being better ascertained (Payne 1992, Knox 1984, Hey 1994, Dutton 1991, Knox 1991), as are children with multiple congenital anomalies (Calle 1991). Thus better ascertainment might be expected of anencephaly which is readily identifiable at birth, than other conditions which require a clinical or laboratory test. Ocular anomalies, such as anophthalmia, may be more readily diagnosed than cataract, but nevertheless ascertainment is unlikely to be high. Indeed, from NCAS the combined birth prevalence of all ocular anomalies, at 1 per 10,000 total births (ONS 1999) was one sixth of that reported by the European Registration of Congenital Anomalies (EUROCAT) (EUROCAT 1993) and similar to the prevalence of anophthalmia and microphthalmia in England reported recently from the National Anophthalmia Register (Dolk 1998). In contrast to the latter sources, notification to NCAS is largely passive and previously relied mainly on a single source.
In common with other malformation reporting systems elsewhere, under-ascertainment and inaccuracies in NCAS were attributed to aspects of both design and implementation (Knox 1984). Specifically, a major review of the NCAS (OPCS 1995) advocated new approaches should be adopted to improve early ascertainment together with further promotion of notification of later diagnosed anomalies. The need for better validation of reported data was also emphasised but the difficulties in undertaking this, given the limited information currently recorded about notified cases, were illustrated by the studies presented in this chapter.
2.5 Conclusions

Ascertainment by NCAS varies with condition. Those conditions which are easily visible at birth being better ascertained than those conditions which are internal or require specialist clinical diagnosis or a laboratory test.

Some conditions required diagnosis from specialist health professionals who might have been unaware of NCAS and that they could report conditions to NCAS when affected babies were identified. Effective monitoring of anomalies remains important to improving the health of children: those responsible for the care of affected children remain best placed to strengthen the ability to achieve this. Strategies to improve NCAS data must include enhancing the awareness and participation in NCAS by all clinical specialties involved in managing children with congenital anomalies. In some specialities, however, there are condition-specific registers in which the clinicians may be active participants. Linkage between these registers and NCAS should also be actively pursued.

Collaboration with condition specific registers, where these exist, and active follow up, is an effective way to improve notification to NCAS. Its main disadvantage is that it does not change the regular notification practices to NCAS and so the improved levels of notification are difficult to sustain. Regular follow-up exercises would therefore be necessary.
CHAPTER 3

Effect of multi source ascertainment by local congenital anomaly registers on notification rates

In the work presented in this chapter, the original ideas were mine, and I performed the statistical analyses. My aim was to test the feasibility of electronic data transfer by local congenital anomaly registers and to determine the effect on notification rates. I co-ordinated the collaboration with the local registers, and colleagues in ONS made the necessary changes to the NCAS computer programmes to allow electronic receipt of data from local registers.
3.1 Summary

The purpose of the study reported in this chapter was to determine the effect on notification rates to NCAS of ascertainment by local congenital anomaly registers with subsequent transfer of information to NCAS. Previously, as described in Chapter 1, all notifications to NCAS had been reported directly from Health Authorities (HAs).

The analysis is based on notifications to NCAS from the areas covered by four local registers between 1992 and 2000, and compares notification levels before and after these registers took over responsibility for ascertaining congenital anomalies and data transfer to NCAS. The step changes in overall notification rates and condition specific rates were evaluated. Notification rates for the three local registers in England were compared with those for the remainder of England. Finally, the impact of the change on surveillance alarms was analysed.
3.2 Background

The role of NCAS is to monitor trends in the prevalence of congenital anomalies, to contribute to epidemiological studies of the major causes of anomalies, to evaluate the impact of medical interventions and public health policies, and to contribute to investigations about changes in the prevalence of congenital anomalies (see Chapter 1). These activities are more likely to be achieved if ascertainment of congenital anomalies is as complete as possible since it gives an indication of the true burden of disease and reduces the risk that changes in prevalence are due to changes in ascertainment.

National surveillance systems operate in many countries including England and Wales (International Clearinghouse 2002), but, as discussed in Chapter 1, have several limitations. Due to the need to minimise costs, national systems often rely on cases being notified, known as passive reporting, rather than actively seeking new cases. The most serious limitations with such data are incomplete ascertainment of cases and little or no follow-up information on the survival of the notified child (Brookmeyer 1995). Both limitations have always been a problem for NCAS.

In the early 1960s, prior to establishing NCAS, recognition of the problems of underascertainment inherent in a passive notification system lead Leek and Smithells to advocate the collection of congenital anomaly data at local rather than national level (Leek 1963). It was believed that physicians would be more supportive to a visible local register. In Britain, following the pioneering example of Birmingham (Knox 1984), local registers were created in several locations including Liverpool (closed early 1990s), Merseyside and Cheshire (started in 1995), South Wales (now covering all of Wales), Trent, Northern Region, West Midlands and Glasgow (Stone 1987).

NCAS was set up in 1964, independently of the local registers in existence at the time. Indeed, an analysis of 1966 NCAS data suggested that the competing
interests of local surveys mitigated against notification to NCAS (Weatherall 1969), as cases notified to a local register were less likely to be notified to NCAS. This may in part have been due to the incorrect assumption by those notifying locally that the local register worked together with NCAS.

Local registers continued to operate in parallel to NCAS until the mid-1990s. One advantage of local registers was that they used multiple sources of information to identify children with congenital anomalies which was also shown to improve ascertainment (Leck 1963). They also followed-up affected children to verify diagnoses, an approach known to improve specificity (Weatherall 1969). Consequently, the local registers held data that were more complete and accurate. As a result, a review of NCAS in 1993 (OPCS 1995) recommended that

‘where good congenital malformation registers exist outside OPCS (now ONS), information should be exchanged with these to improve the completeness and validity of both local and national data’.

It was accepted, however, that implementation of this recommendation would be difficult in the absence of a good working relationship between the registers and NCAS. In 1998, I collaborated with Dr David Stone from the Glasgow register of congenital anomalies to create BINOCAR (the British Isles Network Of Congenital Anomaly Registers). It was set up as an informal network with aims which included acting as a forum for communication and collaboration between all congenital anomaly registers in the British Isles, pool experience and expertise, facilitate the exchange of ideas and information, foster collaborative activities, and provide a framework for mutual support.

All local congenital anomaly registers in the United Kingdom (UK) and the Republic of Ireland are affiliated to BINOCAR. Members meet annually to discuss issues of concern relating to the collection and monitoring of data about congenital anomalies. Scientific papers based on recent research are also presented. Through BINOCAR, a number of collaborative studies began. For example, the results from a study comparing rates of neural tube defects for different registers were published as a letter in the Lancet (Abramsky 1999).
Most importantly, BINOCAR established a collaborative infrastructure to facilitate data transfer, by helping to build a good working relationship and trust between the local registers and NCAS. It was through this network that several local registers agreed to transfer data to NCAS.

Four registers volunteered to take part in electronic data transfer for the study reported in this chapter. One register began data transfer in 1998, one in 1999, and the remaining two in 2000. In 2000 the combination of these four registers covered 27 per cent of all births in England and Wales.

In this Chapter I describe the local registers participating in this study, and report a ‘before and after’ study to investigate the effect of local ascertainment and data transfer on NCAS notification rates. I also discuss whether such improvements could be sustained and extended to other local and condition specific registers.
### 3.3 Description of the local registers involved in data transfer

Table 3.1 shows the start date, population base, sources of notification and aims of each of the four local registers included in this study.

<table>
<thead>
<tr>
<th></th>
<th>Wales</th>
<th>Trent</th>
<th>Mersey</th>
<th>North Thames West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date local register</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>began data supply to</td>
<td>1 Jan 1998</td>
<td>1 Jan 1999</td>
<td>1 Jan 2000</td>
<td>1 Jan 2000</td>
</tr>
<tr>
<td>NCAS:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Date register began:</td>
<td>1 Jan 1998</td>
<td>1 Jan 1997</td>
<td>1 Jan 1995</td>
<td>1 Jan 1990</td>
</tr>
<tr>
<td>Funding:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCAS:</td>
<td>National</td>
<td>Trent NHS</td>
<td>Supra district</td>
<td>Regional genetics</td>
</tr>
<tr>
<td></td>
<td>Assembly</td>
<td>Executive</td>
<td>Audit</td>
<td>purchasing forum</td>
</tr>
<tr>
<td></td>
<td>of Wales</td>
<td></td>
<td>Commissioning</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Group</td>
<td></td>
</tr>
<tr>
<td>Funding level in 2001:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCAS:</td>
<td>£100,000</td>
<td>£65,000</td>
<td>£39,000</td>
<td>£70,000</td>
</tr>
<tr>
<td>Hosp/residence based:</td>
<td>Residence</td>
<td>Residence</td>
<td>Residence</td>
<td>Hospital</td>
</tr>
<tr>
<td>Number of births in</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000:</td>
<td>31449</td>
<td>55541</td>
<td>49113</td>
<td>25878</td>
</tr>
<tr>
<td>Number of different</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sources of notification:</td>
<td>23</td>
<td>22 hospitals + Several groups</td>
<td>Several groups in 11 hospitals</td>
<td>10</td>
</tr>
<tr>
<td>AIMS:*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitor anomalies</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Provide a database for research</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Assess interventions</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan and co-ordinate information for health services</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Assess clusters and their causes
Influence public policy & improve public health
Audit prenatal screening and diagnostic programmes
Forge links with other surveys

Age at reporting:
Termination of pregnancy
Stillbirths
Live births
Aged under 12 months
Aged over 12 months

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audit prenatal screening and diagnostic programmes</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 3.1 Local registers providing data to NCAS in 2000

* as published in the register’s most recent annual report

As shown in Table 3.1, all four local registers used multiple sources for ascertaining cases. All included anomalies detected antenatally as well as postnatally, and collected more data items than were required by NCAS. Three of these registers were based on a geographically defined population, but the fourth register (North Thames West) was based on the population of several hospitals.

3.3.1 Wales

The Congenital Anomaly Register Information Service in Wales (CARIS) is funded by the Welsh Office (now the National Assembly of Wales). It is based at the Singleton Hospital in Swansea. One condition of their funding was that they were to take over responsibility for notifying congenital anomalies from Wales to NCAS.
Data collection commenced on 1 January 1998 and included any baby with congenital anomalies where the pregnancy ended after this date (CARIS 2000 Annual report).

Any fetus or baby who has, or is suspected of having, a congenital anomaly and whose mother is normally resident in Wales at the time of birth should be notified to the Wales register. Notifiable anomalies can be diagnosed at any time from conception to the end of the first year of life. There are 23 different disciplines who notify the Wales register of anomalies, including obstetricians, paediatricians and midwives. Wales has co-ordinators in all 16 main delivery hospitals, plus a further 9 covering community units, antenatal clinics and small cottage hospitals (25 co-ordinators in total), who are able to supply notification forms for the local register. They can also help with filling in the form and retrieving notes. The Wales register provides details of reported cases to all co-ordinators on a regular basis so that they are aware of cases the register may have received from other sources. These data are then available for local audit.

To ensure notification for all residents of Wales, the register includes notifications for the 2 per cent of births in Wales which occur at home. In addition, a significant number of births to Welsh mothers take place in hospitals across the English border. The register has therefore established links with these hospitals, and with the Mersey and West Midlands congenital anomaly registers since they also border Wales and may receive details of congenital anomalies affecting Welsh residents.

3.3.2 Trent Congenital Anomaly Register (CAR)

The Trent CAR began data collection in 1997, from the outset intending to provide notifications to NCAS. It is co-ordinated from the Department of Epidemiology and Public Health at the University of Leicester, and is a collaboration between the
universities of Leicester, Sheffield and Nottingham. The register includes all anomalies detected antenatally and up to age 16. This is the only one of the four local registers included in this study which includes anomalies detected after one year of age. Funding for the register is from the Trent NHS Executive through the Trent Institute of Health Services Research. The Trent CAR was established in response to local concern and confusion over a perceived rise in the occurrence of anomalies, a complete absence of local data held by the Region for research or audit, to investigate possible causes of anomalies, and the urgent need to improve reporting to NCAS. It relies on an extensive network of 19 different types of notifier including clinicians, midwives, sonographers, community child health departments, neonatal screening, pathology and cytogenetic services, together with other databases and colleagues working in other registers (Trent CAR Annual Report 1999).

3.3.3 North Thames West Congenital Malformation Register

The North Thames West Congenital Malformation Register began data collection on 1 January 1990. The Register is essentially hospital based rather than residence based, covering all deliveries in hospitals within a geographically defined region. Residents near the edges of the region may deliver outside and not be ascertained by the register. However, the register does include babies delivered out of region to women originally booked for delivery at a hospital in the region. Its funding is from the Regional Genetics Purchasing Forum through the Genetics Services Contract. It includes information about all births, miscarriages and terminations in contributing hospitals that had or were thought to have a significant structural malformation diagnosed antenatally or within one year of birth, or who had a chromosome abnormality diagnosed at any time antenatally or postnatally. The register is entirely dependent on information provided by 11 different sources of notification, including sonographers, midwives, doctors, system supervisors and secretaries from contributing obstetric units, together with cytogeneticists and
those from other registers and referring centres (North Thames West Congenital Malformation Register Annual Report 2000).

3.3.4 Merseyside and Cheshire Congenital Anomaly Register (Mersey)

The Merseyside and Cheshire Congenital Anomaly Register (Mersey) was set up in 1995, covering the former Mersey Region. It is linked to the Confidential Enquiry into Stillbirths and Deaths in Infancy (CESDI). The Mersey register records all anomalies which are detected antenatally, at the time of birth, at termination of pregnancy, or during the first year of life, or which involve a structural, metabolic, endocrine or genetic defect in the child or foetus apart from an internationally accepted list of minor defects. Its funding, until 2003, is through the Supra District Audit Commissioning Group. The CAR relies for its notifications on an extensive network including health professionals, obstetricians, paediatricians, midwives, neonatal nurses, pathologists and ultrasonographers based at the 11 hospitals within their area, together with other databases, hospital records, and local birth and death records (Merseyside and Cheshire Congenital Anomaly Survey Annual Report).
3.4 Methods

3.4.1 Data transfer between local registers and ONS

This chapter reports on a before and after study to evaluate the change from paper notification to NCAS by HAs, to ascertainment of congenital anomalies and data transfer to NCAS by local registers.

3.4.1.1 Preparing for data transfer

The procedures for electronic data transfer were developed *ab initio* with the first registers that volunteered to transfer data. NCAS supplied the local register with a list of data items, indicating which were mandatory for the record to be included in NCAS, and which were desirable but not essential. These are shown in Table 3.2.

<table>
<thead>
<tr>
<th>Data item</th>
<th>Type</th>
<th>Data item</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record ID</td>
<td>C</td>
<td>Home postcode</td>
<td>C</td>
</tr>
<tr>
<td>DHA residence of mother</td>
<td>C</td>
<td>Home address</td>
<td>C</td>
</tr>
<tr>
<td>DHA of birth</td>
<td>C</td>
<td>Mother's occupation</td>
<td>D</td>
</tr>
<tr>
<td>Date of birth of child</td>
<td>C</td>
<td>Father's occupation</td>
<td>D</td>
</tr>
<tr>
<td>Surname of child</td>
<td>D</td>
<td>Mother's date of birth</td>
<td>C</td>
</tr>
<tr>
<td>Forename of child</td>
<td>D</td>
<td>Mother's age</td>
<td>C</td>
</tr>
<tr>
<td>Sex of child</td>
<td>C</td>
<td>Number previous live births</td>
<td>D</td>
</tr>
<tr>
<td>Place of birth</td>
<td>C</td>
<td>Number previous stillbirths</td>
<td>D</td>
</tr>
<tr>
<td>Stillbirth / death in first week</td>
<td>C</td>
<td>Number previous other</td>
<td>D</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>C</td>
<td>informant</td>
<td>C</td>
</tr>
<tr>
<td>Gestation</td>
<td>C</td>
<td>Text description of anomalies</td>
<td>C</td>
</tr>
<tr>
<td>Birthweight</td>
<td>C</td>
<td>Record type</td>
<td>C</td>
</tr>
</tbody>
</table>

Table 3.2 Data items required for electronic notification to NCAS

*Where type C=compulsory, D=desirable*
Local registers were required to provide this pre-designated minimum data set to NCAS. For each of these data items, NCAS and the local register agreed on the specific format of the data item, its coding and how missing values would be handled. Coding by the local registers to existing NCAS classifications for data items such as sex and multiplicity, ensured that NCAS could add new records from the local registers directly into the NCAS system. NCAS worked in close collaboration with the local registers to understand the limitations of both the NCAS and the local computer systems. This ensured that the developing specification was achievable by the local registers whilst meeting NCAS' requirements for data and its quality.

3.4.1.2 Data transfer

In 1998, Wales began data transfer to NCAS of all congenital anomalies reported in live births or stillbirths known to them from any source. Terminations were not included since, as described in Chapter 1, NCAS only includes anomalies detected in live and stillbirths. Establishing the mechanism for electronic data exchange with the Wales register proved to be a lengthy, iterative procedure that required several trials before implementation was possible. Wales provided test disks containing data for one month during 1997. These data were not added to the NCAS database, but were tested to determine whether individual data items had been provided in the format requested and with the agreed coding for missing data. Problems were identified by NCAS and there was a series of trials as the Wales register revised their submitted data. This experience meant that many of these problems could be pre-empted for subsequent registers that began data transfer.

Although the Wales register formally commenced in January 1998, it also notified NCAS of additional babies with congenital anomalies known to the register but born in years prior to the register being set up. This included a substantial number of babies ascertained by an earlier register based in South Wales. The analyses presented here excluded these retrospective notifications.
The Trent register began data transfer to NCAS in 1998. Initially they transferred to NCAS only the information they had received from HAs on the NCAS SD56 notification forms. In 1999, however, they began complete electronic data transfer, based on multiple source ascertainment. The analysis presented here is based on data transferred by Trent CAR for births in 1999 onwards.

Two other registers, North Thames West and Mersey, began data transfer of prospectively ascertained congenital anomalies in 2000.

3.4.2 Analysis of notification rates before and after data transfer

The principal objective of the analysis was to determine the effect on total and condition-specific notification rates of receiving notifications directly from local registers. This is a before and after comparison of an intervention at a defined point in time. Observed and expected notification rates for the first year of data transfer were calculated for each of the four local registers. These rates were used to test the null hypothesis that data transfer from local registers had no statistically significant effect on notification rates for the area covered by the local register.

The specific anomalies included were neural tube defects, eye anomalies, cleft lip and palate, heart and circulatory system anomalies, respiratory anomalies, limb reduction defects, abdominal defects and Down syndrome. These anomalies were selected since they are well defined, universally collected, easily detected at birth or subject to minimal clinical variation in diagnosis. As a result, these conditions are among the standard conditions analysed and published annually by NCAS.

These conditions were analysed in groups according to whether they should be easily visible at birth (neural tube defects, cleft lip and palate, and limb reduction defects), internal anomalies that might not be detected at birth unless prenatally
detected (heart and circulatory system anomalies, respiratory anomalies and abdominal defects), and those requiring a test before clinical diagnosis (eye anomalies and Down syndrome). These categories are fairly crude as specific anomalies within each group might be better placed in another group. However, the purpose of these groupings was to evaluate whether any changes in notification rate on starting data transfer could be explained by previous levels of ascertainment by NCAS. The results presented in Chapter 2 showed that levels of ascertainment by NCAS were higher for anencephaly, representing major anomalies visible at birth, intermediate for Down syndrome, representing anomalies requiring a diagnostic test, and very low for cardiac defects, representing internal anomalies which might not be diagnosed at birth unless prenatally identified.

A secondary aim of the analysis was to determine the effect of data transfer on surveillance and alarms. This analysis was based on the number of alarms generated by the NCAS surveillance programmes before and after data transfer with the local register.

All analyses were based on notifications of cases born between 1992 and 2000 inclusive. 1992 was selected as the start date as this was the first year for which NCAS had full text and coded information for each case recorded on the computer database. 2000 was the most recent complete year of data available for analysis at the time the study was undertaken.

3.4.3 Observed notification rates after data transfer

The best estimate of the observed notification rate after data transfer was chosen to be the observed notification rate for the first year of data transfer. Figure 3.1 shows the rate of notification of all congenital anomalies to the Wales register for births in 1998 to 2000.
Figure 3.1 Rate of notification of all congenital anomalies for the Wales register as published in the annual report

The notification rates for 1998, 1999 and 2000 published for the first time in their respective annual reports are very similar at around 250 notifications per 10,000 total births. However, by the publication of the 1999 report, additional notifications had been received for 1998 births. Further notifications for 1998 births were received in time for inclusion in the 2000 report. As a result, the notification rate for Wales for 1998 increased by 24 per cent between that published in the 1998 report and that published in the 2000 report due to the later inclusion of notifications of children diagnosed in the first year of life but notified later. A similar trend is seen for the 1999 notification rate. Late notifications for children born in 1998, the first year after data transfer, had had longest to accumulate and so provided the best estimate of the true birth prevalence rate.

For registers that began data transfer after 1998 it was not possible to document similar increases in notification rates as the study only included at maximum one additional year of observations after data transfer. If the other registers subsequently received late notifications, these analyses would provide a minimum estimate of any increase in notification rates after data transfer and a maximum estimate of any decrease in notification rates.
The observed notification rate $p$ for babies born in a given calendar year with an anomaly, was calculated using the number, $r$, of notifications for babies born in the given calendar year, divided by the total number of person-years at risk, $n$ – here estimated by the number of live and stillbirths in the same year.

So $p = \frac{r}{n}$, expressed per 10,000 live and stillbirths.

### 3.4.4 Expected notification rates based on rates before data transfer

The expected notification rate in the first year of data transfer is based on an extrapolation using linear regression of the observed rates in the years from 1992 to the year prior to data transfer i.e. six years for Wales which began data transfer in 1998, seven years for Trent which began data transfer in 1999, and eight years for the North Thames West and Mersey registers which began data transfer in 2000.

The analysis assumed that the notifications follow a Poisson distribution as cases occur independently of each other and randomly in time. Its use was justified as there had been no surveillance alarms that might provide evidence of clustering during the period covered by this study.

The regression line was calculated using $y = a + bx$ where $y$ was the notification rate for the year $x$, $a$ was the intercept and $b$ was the regression coefficient.

Linear regression was used to estimate the expected notification rate, $\bar{p}$, in the first year after data transfer.
3.4.5 Comparison of observed and expected rates

The observed rate was compared with the expected rate using the normal significance test for a single proportion for large samples:

\[ z = \frac{p - \bar{p}}{s.e. (p)} = \frac{p - \bar{p}}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \]

where \( p \) is the observed notification rate calculated as in section 3.4.3, \( \bar{p} \) is the expected notification rate in the first year of data transfer, \( n \) is the number of live and stillbirths in the first year of data transfer, and \( z \) is the standard normal deviate.

Confidence limits around the expected rate (\( \bar{p} \)) were calculated using the formula for large samples:

\[ C.L. = \bar{p} \pm 1.96 \times \text{standard error (s.e.)} \] where \( \text{s.e.} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \)

Significance testing based on the normal distribution was used since the Poisson distribution approximates to a normal distribution if the average number of cases in the time period is large (10 or more).

3.4.6 Impact of data transfer on surveillance alarms

This study was concerned with the impact of data transfer from local registers on the number of alarms generated by the CUSUM programme. The CUSUM methodology used for NCAS surveillance and its handling of alarms is described in Chapter 1 sections 1.3.2 and 1.3.3. In summary, each month NCAS compares current notification levels for selected conditions in each HA against expected ‘baseline’ values based on previous notification rates. Any statistically significant increases are notified to the local Director of Public Health. Therefore, any changes in reporting practice, such as more complete notification from a local register,
would only affect the surveillance outputs from that area. All other areas would be unaffected.

The number of alarms each month for the combined health authorities covered by the local register were analysed before and after data transfer to test for significant change.
3.5 Results

3.5.1 Observed rates before and after starting data transfer

Figure 3.2 and Table 3.3 show the notification rates between 1992 and 2000 for the four local registers, together with the total notification rate for the remainder of England excluding the three English local registers.

Figure 3.2 Notification rates for all anomalies per 10,000 live and stillbirths, for each local register and for the remainder of England, 1992-2000

Between 1992 and 1995 the notification rates for each of the local registers in England were higher than those for the rest of England. Despite Trent and Mersey having falling notification rates in the years prior to data exchange, their notification rates were consistently higher than those for the rest of England. The rates for the rest of England, however, were higher than those for North Thames West from 1996 until 2000 when the register began data transfer.
Table 3.3 Notification rates from registers participating in electronic data transfer, 1992-2000

Note: Rates in bold are those after electronic data transfer

The data in Figure 3.2 and Table 3.3 show the step increase in notification rates to NCAS on beginning data transfer with each local register (1998 in Wales, 1999 in Trent, and 2000 in North Thames West and Mersey). In every case, on beginning data transfer, NCAS notification rates for the areas covered by local registers were higher than those for the rest of England. For Wales and Trent, there was a lower rate in the second year of data transfer compared with the first year. This apparently lower rate replicates the analysis by the Wales register given in Figure 3.1, which showed this to be a temporary difference until late notifications had accumulated.

3.5.2 Comparison with expected rates

Table 3.4 shows for each local register the observed notification rate to NCAS in the first year of data transfer, the coefficients from the linear regression, and the expected notification rate after data transfer. Figures 3.3 show for each local register the observed notification rates and the fitted line from the regression analysis giving the expected rates after the start of data transfer.
Table 3.4 Observed and expected rates for local registers taking part in electronic data transfer with NCAS

*per 10,000 live and stillbirth

Figures 3.3 Observed and expected notification rates for each local register, 1992-2000

![Graph showing observed and expected rates for Wales](image-url)
Figures 3.3 Observed and expected notification rates for each local register, 1992-2000
Notification rates for the areas covered by all four local registers had been falling in the years before data transfer, shown by the negative regression coefficient. Rates for Wales experienced the largest fall, whereas those for Trent only fell slightly over time. For the rest of England, when notification rates from the three English local registers engaged in electronic notification had been excluded, notification rates increased slightly between 1992 and 1998.

Table 3.4 shows that all observed rates in the first year of data transfer were significantly higher than those expected from the linear regression, p<0.001.

The increase in the NCAS total notification rate after data transfer was six-fold for Wales and two-fold for Trent and Mersey. For these three registers, after data transfer with NCAS, 2 to 3 per cent of all children born in these areas were notified as having at least one congenital anomaly. The notification rate for North Thames West increased by 69 per cent after data transfer. The highest observed rate in the first year of data transfer was for Trent, which more than doubled between 1998 and 1999, from 119 to 280 per 10,000 live and stillbirths. The notification rate for Wales increased from 41 in 1997 to 260 in 1998, and the notification rate for Mersey increased from 87 to 251 between 1999 and 2000. In North Thames West the notification rate increased from 68 to 115 per 10,000 live and stillbirths which was statistically significant, although their notification rate after data transfer was considerably lower than the other local registers. Nevertheless, the notification rate for North Thames West in 2000 was significantly higher than that for the remainder of England, (p < 0.01).

3.5.3 Analysis of specific conditions

Table 3.5 shows the observed notification rates for specific congenital anomalies for the whole study period. Notification rates after data transfer are shown in bold.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All Babies</td>
<td>Wales</td>
<td>74.3</td>
<td>77.2</td>
<td>71.7</td>
<td>68.1</td>
<td>45.6</td>
<td>40.9</td>
<td>260.0</td>
<td>223.1</td>
<td>215.9</td>
</tr>
<tr>
<td>All ICD codes</td>
<td>Trent</td>
<td>132.6</td>
<td>122.6</td>
<td>111.0</td>
<td>114.0</td>
<td>116.8</td>
<td>139.2</td>
<td>119.3</td>
<td>280.1</td>
<td>266.1</td>
</tr>
<tr>
<td></td>
<td>N Thames West</td>
<td>97.5</td>
<td>81.5</td>
<td>77.0</td>
<td>66.5</td>
<td>56.4</td>
<td>56.6</td>
<td>63.8</td>
<td>68.0</td>
<td>114.6</td>
</tr>
<tr>
<td></td>
<td>Mersey</td>
<td>89.3</td>
<td>86.7</td>
<td>80.9</td>
<td>97.3</td>
<td>82.7</td>
<td>78.3</td>
<td>76.0</td>
<td>86.6</td>
<td>251.2</td>
</tr>
<tr>
<td></td>
<td>Rest of England</td>
<td>63.1</td>
<td>62.2</td>
<td>63.3</td>
<td>64.6</td>
<td>67.0</td>
<td>66.7</td>
<td>65.4</td>
<td>72.6</td>
<td>68.0</td>
</tr>
<tr>
<td>Neural tube defects</td>
<td>Wales</td>
<td>2.9</td>
<td>1.4</td>
<td>1.4</td>
<td>0.9</td>
<td>0.3</td>
<td>0.3</td>
<td>1.2</td>
<td>3.1</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Trent</td>
<td>1.6</td>
<td>1.4</td>
<td>2.0</td>
<td>1.9</td>
<td>1.3</td>
<td>2.2</td>
<td>1.5</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>N Thames West</td>
<td>2.1</td>
<td>1.8</td>
<td>1.0</td>
<td>2.4</td>
<td>0.8</td>
<td>0.8</td>
<td>1.2</td>
<td>0.6</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td>Mersey</td>
<td>2.5</td>
<td>2.0</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
<td>1.4</td>
<td>1.8</td>
<td>0.8</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Rest of England</td>
<td>1.6</td>
<td>1.4</td>
<td>1.3</td>
<td>1.4</td>
<td>1.1</td>
<td>1.0</td>
<td>1.2</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>Eye anomalies</td>
<td>Wales</td>
<td>0.8</td>
<td>0.8</td>
<td>1.4</td>
<td>2.6</td>
<td>0.9</td>
<td>0.9</td>
<td>13.4</td>
<td>12.4</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>Trent</td>
<td>1.5</td>
<td>1.1</td>
<td>2.2</td>
<td>1.4</td>
<td>2.3</td>
<td>2.0</td>
<td>1.8</td>
<td>4.5</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>360.0-379.9</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.4</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>743.0-743.9</td>
<td>0.3</td>
<td>0.7</td>
<td>0.3</td>
<td>0.7</td>
<td>1.0</td>
<td>1.4</td>
<td>-</td>
<td>1.1</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>756.1</td>
<td>0.7</td>
<td>0.8</td>
<td>1.1</td>
<td>0.6</td>
<td>1.0</td>
<td>0.8</td>
<td>0.7</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Cleft lip and palate</td>
<td>Wales</td>
<td>9.8</td>
<td>9.2</td>
<td>11.0</td>
<td>8.1</td>
<td>5.7</td>
<td>4.9</td>
<td>17.0</td>
<td>13.6</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>Trent</td>
<td>13.2</td>
<td>12.9</td>
<td>9.6</td>
<td>11.2</td>
<td>9.1</td>
<td>9.7</td>
<td>8.4</td>
<td>15.7</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td>N Thames West</td>
<td>11.7</td>
<td>10.1</td>
<td>7.5</td>
<td>6.5</td>
<td>7.2</td>
<td>5.2</td>
<td>5.8</td>
<td>7.7</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>Mersey</td>
<td>11.3</td>
<td>11.1</td>
<td>7.4</td>
<td>8.7</td>
<td>5.9</td>
<td>11.1</td>
<td>6.2</td>
<td>6.4</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>Rest of England</td>
<td>7.8</td>
<td>7.1</td>
<td>7.4</td>
<td>6.7</td>
<td>6.9</td>
<td>7.1</td>
<td>7.3</td>
<td>6.7</td>
<td>7.0</td>
</tr>
<tr>
<td>Heart and circulatory</td>
<td>Wales</td>
<td>9.0</td>
<td>8.7</td>
<td>10.4</td>
<td>8.4</td>
<td>4.6</td>
<td>3.2</td>
<td>71.4</td>
<td>47.7</td>
<td>52.5</td>
</tr>
<tr>
<td></td>
<td>Trent</td>
<td>10.8</td>
<td>11.0</td>
<td>10.4</td>
<td>9.8</td>
<td>14.0</td>
<td>18.0</td>
<td>17.8</td>
<td>43.2</td>
<td>42.9</td>
</tr>
<tr>
<td></td>
<td>N Thames West</td>
<td>5.2</td>
<td>6.0</td>
<td>7.3</td>
<td>3.0</td>
<td>5.6</td>
<td>5.8</td>
<td>6.2</td>
<td>6.9</td>
<td>21.6</td>
</tr>
<tr>
<td></td>
<td>Mersey</td>
<td>7.5</td>
<td>6.8</td>
<td>4.3</td>
<td>3.5</td>
<td>4.8</td>
<td>2.5</td>
<td>4.4</td>
<td>7.5</td>
<td>85.0</td>
</tr>
<tr>
<td></td>
<td>Rest of England</td>
<td>4.4</td>
<td>4.3</td>
<td>4.6</td>
<td>5.6</td>
<td>5.6</td>
<td>6.0</td>
<td>6.0</td>
<td>6.7</td>
<td>7.8</td>
</tr>
<tr>
<td>Respiratory anomalies</td>
<td>Wales</td>
<td>0.8</td>
<td>0.5</td>
<td>1.1</td>
<td>1.7</td>
<td>0.3</td>
<td>0.3</td>
<td>8.9</td>
<td>9.9</td>
<td>5.7</td>
</tr>
<tr>
<td></td>
<td>Trent</td>
<td>1.2</td>
<td>1.4</td>
<td>0.3</td>
<td>1.4</td>
<td>2.9</td>
<td>2.7</td>
<td>1.0</td>
<td>4.5</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>N Thames West</td>
<td>1.0</td>
<td>0.2</td>
<td>1.2</td>
<td>0.8</td>
<td>0.4</td>
<td>1.0</td>
<td>0.2</td>
<td>0.4</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>Mersey</td>
<td>0.6</td>
<td>0.3</td>
<td>0.7</td>
<td>0.3</td>
<td>1.0</td>
<td>1.1</td>
<td>1.1</td>
<td>1.5</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Rest of England</td>
<td>0.4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Table 3.5 Observed notification rates for specific anomalies for local registers, 1992-2000

*Boundaries are those at 1 April 2000
NCAS data as at 22 January 2002*

Between 1992 and 1997, notification rates for the selected conditions in the rest of England were amongst the lowest rates when compared with the local registers. In 1992 the notification rates for the rest of England were the lowest for each condition except eye anomalies and limb reduction defects. In 1993 the rest of England had the lowest rates for five of the nine conditions presented, and in 1994 and 1995 their rates were the lowest for only two conditions each year. In 1996 and 1997 the area covered by at least one local register had a lower rate than the rest of England for each condition presented in Table 3.5.
When the local registers began data transfer, however, the observed notification rates for specific conditions increased for these areas and most were significantly higher than those expected from a linear regression, as shown in Table 3.6. This table compares the observed and expected rates for the four local registers for the year after data transfer, based on a linear regression of the rates for each specific condition from 1992 to the year prior to data transfer.

Neural tube defects, cleft lip and palate, and limb reduction defects are usually obvious at birth. All four local registers saw a significant increase (p<0.01) in notification rates for cleft lip and palate compared to expected rates. Wales had the highest notification rate for cleft lip and palate. All registers except Trent experienced a significant increase in notification rate for neural tube defects after data transfer, and the rates were at least four times higher than expected. North Thames West had the highest rate for neural tube defects. Wales and Trent had significant increases in notification rates for limb reduction defects on beginning data transfer, but only for Wales, however, was the notification rates more than double that expected for this condition. The highest notification rate for limb reduction defects, however, was in Trent.

Heart and circulatory, respiratory, and abdominal defects are largely internal anomalies and may not be diagnosed quickly unless prenatally detected. All four local registers saw a significant increase (p<0.01) in notification rates for heart and circulatory anomalies and abdominal defects, compared to expected rates. For heart and circulatory anomalies, the rate was 23 times higher in Wales and 20 times higher in Mersey than expected. Mersey had the highest notification rate for heart and circulatory conditions. The North Thames West notification rate for heart and circulatory conditions was noticeably lower than those of the other local
<table>
<thead>
<tr>
<th>Year</th>
<th>Wales</th>
<th>Trent</th>
<th>North Thames West</th>
<th>Mersey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>observed</td>
<td>expected</td>
<td>observed</td>
<td>expected</td>
</tr>
<tr>
<td></td>
<td>1998</td>
<td>1999</td>
<td>2000</td>
<td>2000</td>
</tr>
<tr>
<td>All Babies</td>
<td>260.0</td>
<td>36.5(30.42.9)</td>
<td>280.1</td>
<td>121.7(112.6,133.5)</td>
</tr>
<tr>
<td>Neural tube defects</td>
<td>1.2</td>
<td>0.0 (0.0,0.1)</td>
<td>2.4</td>
<td>1.8 (0.7,2.9)</td>
</tr>
<tr>
<td>Eye anomalies</td>
<td>13.4</td>
<td>1.4 (0.1,2.6)</td>
<td>4.5</td>
<td>2.2 (1.0,3.4)</td>
</tr>
<tr>
<td>Cleft lip and palate</td>
<td>17.0</td>
<td>4.3 (2.1,6.5)</td>
<td>15.7</td>
<td>7.5 (5.3,9.8)</td>
</tr>
<tr>
<td>Heart and circulatory anomalies</td>
<td>71.4</td>
<td>3.0 (1.1,4.9)</td>
<td>43.2</td>
<td>18.7 (15.1,22.2)</td>
</tr>
<tr>
<td>Respiratory anomalies</td>
<td>8.9</td>
<td>0.5 (0.0,1.3)</td>
<td>4.5</td>
<td>2.2 (1.0,3.5)</td>
</tr>
<tr>
<td>Limb reductions</td>
<td>5.4</td>
<td>1.5 (0.2,2.7)</td>
<td>5.7</td>
<td>2.9 (1.5,4.3)</td>
</tr>
<tr>
<td>Abdominal defects</td>
<td>8.6</td>
<td>1.6 (0.3,3.0)</td>
<td>8.0</td>
<td>1.1 (0.3,2.0)</td>
</tr>
<tr>
<td>Down syndrome</td>
<td>11.3</td>
<td>2.8 (1.0,4.6)</td>
<td>9.9</td>
<td>8.4 (6.0,10.7)</td>
</tr>
</tbody>
</table>

Table 3.6 Observed and expected rates per 10,000 live and stillbirths for specific conditions in the year after data transfer
registers, but still had a significant increase in notification rate on starting data transfer. In all registers except Mersey, the rates for abdominal wall defects were over 5 times higher than expected. Wales had the highest notification rates for abdominal anomalies.

All registers except Mersey experienced a significant increase in notification rates for respiratory anomalies on beginning data transfer. The notification rate was double that expected for Trent and over five times that expected for Wales and North West Thames in these conditions. Particularly noticeable was the Wales notification rates for respiratory conditions, which at 8.9 per 10,000, was twice that of the next highest rate for this condition. In contrast, notification rates for respiratory anomalies in Mersey were noticeably lower than those for the other local registers.

Eye anomalies and Down syndrome require a test for clinical diagnosis. Notification rates from Wales were the highest for both Down syndrome and eye anomalies. Particularly noticeable was the notification rates for eye anomalies from Wales, which at 13.4 per 10,000 was three times higher than the next highest rate for this condition. All registers except Mersey experienced a significant increase in notification rates for eye anomalies on beginning data transfer. The notification rate was double that expected for Trent, five times that expected by North West Thames and over nine times that expected by Wales. Notification rates for eye anomalies in Mersey were noticeably lower than those for the other local registers. Only Wales and Mersey had a significant increase in notification rate for Down syndrome on starting data transfer, with the notification rate for Wales being four times that expected.
3.5.4 Impact on surveillance alarms

Table 3.7 shows the number of alarms generated by the CUSUM surveillance programmes for 1997 for the HAs covered by the four local registers. Alarms after starting data transfer are shown in bold.

<table>
<thead>
<tr>
<th></th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wales</td>
<td>5</td>
<td>647</td>
<td>425</td>
<td>10</td>
</tr>
<tr>
<td>Trent</td>
<td>230</td>
<td>165</td>
<td>778</td>
<td>721</td>
</tr>
<tr>
<td>North Thames West</td>
<td>56</td>
<td>112</td>
<td>61</td>
<td>267</td>
</tr>
<tr>
<td>Mersey</td>
<td>108</td>
<td>99</td>
<td>67</td>
<td>114</td>
</tr>
</tbody>
</table>

Table 3.7 Number of alarms alerted by NCAS surveillance system

Source: NCAS as at 31/7/01

Notification rates for all four local registers had been falling in the years from 1992 until starting data transfer, as shown in Figures 3.3. On starting data transfer each of the local registers experienced a significant increase in notification rate. Simultaneously, the surveillance programmes for each area generated an increased number of alarms, as shown in Table 3.7.

Prior to data transfer the number of notifications to NCAS from Wales had been falling; therefore fewer alarms had been generated. In 1998 the surveillance programme for Wales generated 647 alarms compared with just five in 1997. These were assumed to be result of the increase in completeness of notification, because in 2000, once the new reporting levels had stabilised and the CUSUM baseline levels had been adjusted to reflect the new reporting levels, the number of alarms fell to 10, back to the previous level.
In Trent, however, the number of notifications had fallen only slightly prior to data transfer, resulting in a high number of alarms prior to data transfer. On starting data transfer in 1999 the number of alarms for Trent rose to 778 from 165 in the previous year. A similar number were generated in 2000, but I anticipated that once these new reporting levels were incorporated into the baselines the number of alarms relating to 2001 notifications would fall. Up to September 2001 (based on 9 months of the year) there were 130 alarms. This would suggest a similar number of alarms for 2001 as received in 1998. In 2000 the number of alarms from notifications for North Thames West rose to 267 from 61 in the previous year, and for Mersey to 114 from 67 in 1999.
3.6 Discussion

3.6.1 Effect of ascertainment and data transfer by local registers on notifications to NCAS

The initial data transfer stages had implications for the study, since the ability and the ease with which data transfer was possible was important to identify whether results could be extended to other registers and could be sustained. The series of iterations working with the Wales register to transfer data in the correct format was important to resolve problems of providing data in the required format. Once these were resolved, data transfer with the subsequent registers was much faster, suggesting that the general problems in providing data in the required format had been resolved, thus facilitating data transfer from other local registers.

Electronic data transfer resulted in significant increases in total notification rates from all four local registers, by 67 per cent to 600 per cent, with the largest increase for Wales which had had the lowest notification rate before data transfer. Those local registers with higher notification rates prior to data transfer would be expected to increase less than those with lower rates. In addition, the Wales register had been transferring data for longer than the other registers at the time of the study reported here, so had had more time to accumulate late notifications for babies born in the first year of data transfer.

In the absence of any reason why true prevalence should be lower in Wales than in England, the lower notification rates for Wales in 1992-1997 could have been due to less complete notification from Wales. This might have been due to problems when transferring responsibility for notifying NCAS of congenital anomalies as a result of administrative changes in Welsh Health Authorities in the early and mid 1990s. For Wales and Trent there was a lower rate in the second year of data transfer compared with the first year, but this may be due to fewer late notifications
for that year having accumulated by the time of the analysis for this study, as shown for the Wales register in Figure 3.1.

Levels of ascertainment varied for different conditions in different registers. Neural tube defects, cleft lip and palate, and limb reduction defects are usually obvious at birth. All four local registers saw a significant increase ($p<0.01$) in notification rates for cleft lip and palate compared to expected rates. Heart and circulatory, respiratory, and abdominal defects are largely internal anomalies and may not be diagnosed quickly unless prenatally detected. All four local registers saw a significant increase ($p<0.01$) in notification rates for heart and circulatory anomalies and abdominal defects, compared to expected rates. For heart and circulatory anomalies, the rate was 23 times higher in Wales and 20 times higher in Mersey than expected. All registers except Mersey experienced a significant increase in notification rates for respiratory anomalies on beginning data transfer. The notification rate was double that expected for Trent and over five times that expected for Wales and North West Thames in these conditions. Eye anomalies and Down syndrome require a test for clinical diagnosis. Notification rates from Wales were the highest for both Down syndrome and eye anomalies. All registers except Mersey experienced a significant increase in notification rates for eye anomalies on beginning data transfer.

The various studies reported in Chapter 2 showed that underascertainment in NCAS was largely a result of NCAS only being notified of the most severe cases and those easily visible at birth, whilst multi-source ascertainment by the local registers picked up more minor anomalies and internal anomalies not detected at birth. The increases for the selected conditions reported in this Chapter varied by condition, with heart defects which are known to be least well notified to NCAS showing the largest increases compared with the expected rates.

Limb reduction defects and cleft lip and palate are visible at birth and so thought to be among the better notified conditions to NCAS. Only two local registers
experienced significant increases in notification rates for limb reduction defects on the change to data transfer, and only in Wales was the notification rate more than double that expected. Whilst all the local registers experienced a significant increase in notification rates for cleft lip and palate on beginning data transfer, the increases were much smaller than increases seen for other conditions. Both these conditions which were visible at birth and specifically checked at the baby's delivery and therefore likely to have been well recorded on the birth notification form, which was the primary or sole source used for notification by the majority of HAs prior to data transfer by local registers.

Down syndrome had already been the subject of an earlier follow-up exercise for notifications in 1997 and 1998 which had resulted in an increase in the number of notifications by 44 per cent (Botting 2000a). Therefore, it was not surprising that data transfer had not lead to a significantly higher rate for Down syndrome than expected for two of the three local registers that began data transfer in 1999 and 2000.

In 2000, after the four local registers had begun data transfer, notification rates for the registers showed close agreement for cleft lip and palate, abdominal wall defects and Down syndrome. To test whether a maximum detection level had been reached, notification rates for these conditions were compared with international notification rates, on the assumption that there are no real differences between different countries for the prevalence of these conditions.

Table 3.8 compares a selection of these notification rates with the range of rates for participating programmes in the International Clearinghouse for Birth Defects Monitoring Systems (International Clearinghouse 2002). The chosen conditions were those available from the annual published reports.
<table>
<thead>
<tr>
<th>Condition (based on ICD codes as shown in Table 3.5)</th>
<th>Range of rates from local registers, 2000</th>
<th>Range of rates from Clearinghouse Programmes, 1999</th>
<th>Rate for Rest of England, 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Babies</td>
<td>114.6-266.1</td>
<td>68.0</td>
<td></td>
</tr>
<tr>
<td>Cleft lip and palate</td>
<td>10.4-13.7</td>
<td>8.24-23.62</td>
<td>7.0</td>
</tr>
<tr>
<td>Abdominal defects</td>
<td>5.4-7.3</td>
<td>0.18-7.54</td>
<td>2.0</td>
</tr>
<tr>
<td>Down syndrome</td>
<td>8.6-11.2</td>
<td>4.34-25.35</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Table 3.8 Comparison of notification rates for 4 local registers and the remainder of England in 2000 with Clearinghouse Programmes in 1999.

The range of rates for Clearinghouse programmes was wide reflecting the diversity of programmes and their different prevalence and ascertainment rates. Compared with international data for 1999, which was the latest published data from the International Clearinghouse, the notification rates from the local registers for these conditions were consistent with rates from Clearinghouse programmes, showing that either their rates were closer to true prevalence or that some maximum ascertainment been reached. The rates for the rest of England were lower, and, for cleft lip and palate and for Down syndrome, were below the range of international data.

Although the notification rates from local registers were higher than those received by NCAS from HAs, the rates varied between the local registers. Another study in the USA found similar large differences in prevalence rates between different registries (Hobbs 2001). The authors concluded that factors which influence prevalence rates include case ascertainment sources, case inclusion criteria and inclusion of elective terminations and stillbirths. The impact of each of these factors was thought to be defect specific. Levels of notification for the specific conditions analysed here also varied between the four local registers. This reflects real differences in prevalence levels, the different ways the local registers operated, and the level of difficulty in detecting and notifying different conditions. These
differences needed to be explored further to explain the differences in terms of the strengths and weaknesses of the registers. This could help identify whether there were common factors and core requirements for an ideal local register to ensure the sustainability of local registers as the core data providers for NCAS. This work was outside the scope of the study reported here.

There were several differences in the ways registers operated. These included the location of the register (for example in a hospital or child health department), the range of sources notifying to the local register, and the support and interests of local clinicians. Other factors also included the local registers' use of prenatal diagnosis information, and later notification of older children with anomalies. It also reflected different policies on which conditions should be recorded by the local register. Some registers excluded more minor cases of certain anomalies, for example, North Thames West does not collect isolated hypospadias.

The total notification rate for North Thames West after beginning data transfer was noticeably lower than the total rate for the other two local registers in England. This was likely to be due to a range of factors including being a hospital-based register, and from the use of exclusion criteria by the register. Notification rates for North Thames West suggest that hospital based registers find it harder to achieve good coverage, but there are a number of other local factors that need to be taken into account. There are inherent problems of hospital-based registers as cases are harder to follow up in the long term as they may have been referred from a wide geographical area. The differences reported here, however, could also be a function of this register being in London where other registers have experienced problems in achieving good coverage (Roderick 1992), or their different IT or data collection systems.
3.6.2 Estimated ascertainment

To estimate true birth prevalence of congenital anomalies in England using data from just one or two registers it is necessary to make assumptions that total notification rates are consistent across England, which assumes that the rates for individual conditions are also consistent across England. Tables 3.5 showed that prior to electronic data transfer, notification rates from the areas covered by the four local registers, except for Trent, were comparable with those from the remainder of England. Overall notification rates from Trent were noticeably higher. Therefore, given the similar levels of notification prior to electronic data transfer, this suggested that the improvements in local notification seen on beginning data transfer could be replicated from other areas if they had a local register. This assumption was used to estimate the level of ascertainment to NCAS in 2000 for births in the rest of England. Notifications from Wales were not included in this estimate as only local registers in England were used to estimate true prevalence in England. The notification rate for North Thames West was also excluded from this analysis since this register was the only known hospital based register in England, and had its own specific issues of ascertainment as described earlier in this chapter. All other local registers currently operating were population based.

The total notification rates for the two local registers in their first year of data exchange were 280.1 per 10,000 live and stillbirths for Trent, and 251.2 for Mersey. Based on the assumptions that levels of notification from these two local registers were complete, and that notification rates would be the same for all health areas in the rest of England if complete ascertainment was achieved, the true birth prevalence of congenital anomalies for the rest of England in 2000 was estimated to be in the range of 251.2 to 280.1 per 10,000 live and stillbirths. This would suggest that between 8188 and 9589 congenital anomalies in the rest of England were not notified to NCAS.
3.6.3 Implications of data transfer for coverage and sustainability

Given the shortfall of notifications in England estimated in section 3.6.2, it is important to aim for complete coverage of congenital anomalies in England and Wales by local registers; local snapshots are not sufficient. Incomplete coverage would not enable investigations into hypothesised environmental hazards in different geographical areas. Given the very low prevalence of some conditions, complete geographical coverage is also needed to identify as many cases as possible to provide sufficient numbers for analysis.

There are other local congenital anomaly registers in England in addition to those included in this study. Table 3.9 shows the local registers in existence in 2000, both those who had volunteered to take part in data transfer with NCAS and those who did not participate at that time, together with the number of births covered by each register in 2000.

Table 3.9 shows that if all the other local congenital anomaly registers began data transfer to NCAS, the coverage of births by local registers, at just under one half of all births in England in 2000, would not be able to replace notification from Health Authorities. A new congenital anomaly register in the South West had been funded from 2002 which would increase to an estimated 54 per cent the coverage of births by a congenital anomaly register in England and Wales.

The Oxford and Wessex local registers began data transfer to NCAS on 1 January 2002. At that stage, based on the number of births in 2000, 33 per cent of all births in England and Wales were covered by a local congenital anomaly register that participated in local ascertainment and data transfer to NCAS. It was hoped that BINOCAR would facilitate the extension of electronic data transfer to other local registers.
<table>
<thead>
<tr>
<th>Registers providing data to NCAS January 2000</th>
<th>Number of births 2000</th>
<th>Percentage of total births in England and Wales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wales</td>
<td>31449</td>
<td></td>
</tr>
<tr>
<td>Trent</td>
<td>55541</td>
<td></td>
</tr>
<tr>
<td>North Thames West</td>
<td>49113</td>
<td></td>
</tr>
<tr>
<td>Mersey and Cheshire</td>
<td>25878</td>
<td></td>
</tr>
<tr>
<td><strong>Total births</strong></td>
<td><strong>161981</strong></td>
<td><strong>27</strong></td>
</tr>
<tr>
<td>Other Regional registers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORCAS (Northern region)</td>
<td>29785</td>
<td></td>
</tr>
<tr>
<td>West Midlands</td>
<td>61845</td>
<td></td>
</tr>
<tr>
<td>Wessex (WANDA)*</td>
<td>32631</td>
<td></td>
</tr>
<tr>
<td>Oxford*</td>
<td>7240</td>
<td></td>
</tr>
<tr>
<td><strong>Total Births</strong></td>
<td><strong>131501</strong></td>
<td><strong>22</strong></td>
</tr>
<tr>
<td>No registers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yorkshire and Humberside</td>
<td>38716</td>
<td></td>
</tr>
<tr>
<td>North West (excl Mersey and Cheshire)</td>
<td>48411</td>
<td></td>
</tr>
<tr>
<td>Anglia and Oxford (excl Oxford)</td>
<td>41110</td>
<td></td>
</tr>
<tr>
<td>South West (excl Wessex)</td>
<td>36923</td>
<td></td>
</tr>
<tr>
<td>South Thames</td>
<td>72045</td>
<td></td>
</tr>
<tr>
<td>North Thames (excl North Thames West)</td>
<td>76617</td>
<td></td>
</tr>
<tr>
<td><strong>Total Births</strong></td>
<td><strong>313822</strong></td>
<td><strong>52</strong></td>
</tr>
<tr>
<td>England and Wales total births</td>
<td>607304</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.9 Births in areas covered by regional registers, 2000

*Estimated

One major concern for most local registers is that they were often precariously and poorly funded which threatened their existence. Some local registers had closed through lack of continued funding. A congenital anomaly register had existed in South East Thames but had lost its funding, and funding for the Trent register was due to end in 2002. Ideally registers would be funded consistently and comparably across England and Wales.
In summary, an ideal register would have the following attributes: To achieve more complete ascertainment multiple sources of information are essential. With their active follow-up, the local registers are more likely to pick up cases. A core set of data items should be collected for every case, although local needs may require further data items. It is important to be able to return to case notes to test future new hypotheses. Prenatal diagnosis should be included, first to ensure that terminations for fetal anomaly are included, and second to flag the pregnancies to follow up the child when born. Given complete coverage, and consistent and comparable funding, different disciplines should be encouraged to work with the registers, since diversity of interest could result in a wider range of research using the local database.

The study reported here only evaluated the improvements achieved by data transfer with local registers. This showed that different increases were achieved for different conditions, depending on the timing of diagnosis and the involvement of different specialisms in notifying the local register. There are also condition-specific registers within England and Wales which cover a wider geographical area but for just one or a limited range of conditions. One example is the National Down Syndrome Cytogenetic Register which is based at the Wolfson Institute of Preventive Medicine and has collected reports of trisomy 21 or its cytological variants from cytogenetic laboratories in England and Wales since January 1989. A further possible approach is to test the feasibility of data transfer from this and other condition-specific registers to NCAS. This work was outside the scope of the study reported in this chapter, and further pilot work would be necessary to test whether the results reported here for local congenital anomaly registers could be extended to condition-specific registers.

3.6.4 Implications of data transfer for surveillance

The methodology used by NCAS to revise CUSUM baselines each year, assuming no epidemic had taken place, used a weighted average of previous notifications
plus those in the most recent year. Given the large increase in notification rates seen initially from Wales, I anticipated that alarms reflecting the higher than expected notification would occur regularly in the initial year of data transfer. It was therefore important that the surveillance programmes were quickly adjusted to expect these new raised levels of notification. Then any subsequent increase could be identified as a potential true increase in prevalence.

The current methodology for amending the baselines would have lead to slow stepwise increases in baselines over several years until current notification levels were reached. This would lead to continued false alarms over coming years. Therefore I decided to calculate new baselines for the areas covered by the registers engaged in data transfer using the first two years data received after starting data transfer. Surveillance baselines change in January each year, so baselines for Wales in 2000 were recalculated based on the new reporting levels for 1998 and 1999. Similarly, the baseline levels were revised for Trent 2001 data based on 1999 and 2000 data, and those for Mersey and North Thames West would be adjusted for 2002 using the 2000 and 2001 data.

As a result of the large increases in notification rates for the four local registers, the number of alarms increased for the HAs covered by the local registers. These were assumed to be a result of improvements in completeness of notification, because once the new reporting levels had stabilised and the CUSUM baseline levels had been recalculated to reflect the new reporting levels, the number of alarms fell substantially.

Data transfer leads to a large increase in the number of alarms generated by the surveillance programmes. Table 3.7 showed that the number of alarms generated by the CUSUM programmes after data transfer reflected the size of the increase in notification rates and the prior decline in notification rates. I recognised that given so many false alarms, any true alarms would be easily overlooked. During the first
year of data transfer with Wales, therefore, I compared all increases in specific conditions in areas of Wales with alarms in areas of England, to see if any patterns in the same conditions arose simultaneously. None were detected.

In the future the increase in the number of alarms should be anticipated when beginning data transfer with other local registers. Ideally the surveillance baselines would be reset at the outset of data transfer. For existing registers this would be possible using their data for previous years to generate appropriate baselines. This would not be possible, however, for new registers, as there are no historic data from which to calculate baselines. For new registers, therefore, there would be three options. First, existing NCAS baselines for the health areas covered by the register could be used. This was the option used in the study reported here, and had the disadvantage of baselines being too low for the new reporting rates, which lead to a large increase in the number of alarms. In turn this generated an increase in workload for the local register co-ordinators involved. Secondly, the baselines of a ‘similar’ register could be adopted until the new register had generated sufficient data to calculate local baselines. The problem with this option would be to identify an appropriate ‘similar’ register, since all the registers reported here had different notification rates for the selected conditions due to their different sources of data and operation of their register. Thirdly, there could be a ‘blind’ period when surveillance was suspended, until sufficient local data had been generated to calculate appropriate baselines. This could lead to the major disadvantages of missing an epidemic in that local area.

In the future, when another new register begins data transfer, of the three options given above, my choice would be to use baselines of a similar register for the first year of data transfer. However, this register would need to be chosen with care to match as closely as possible with another register which has the same population basis (hospital or residence), and similar aims. I would then revise the baselines to those of the register over its first year of data transfer. I would not suspend surveillance since this is the primary purpose of NCAS. In addition, I would not
continue to use existing baselines, as used in the study presented here, due to the large number of alarms generated.
3.7 Conclusions

In the study reported here, data transfer by all four local registers resulted in improvements in ascertainment. Therefore this approach for NCAS to obtain data, where possible, from local congenital anomaly registers is to be advocated. Data transfer by each local register improved ascertainment in total and for most of the specific conditions presented. Given that NCAS was estimated to have ascertained only one quarter of all babies with congenital anomalies born in the rest of England, it is essential to aim to extend data transfer to as many other local and condition-specific congenital anomaly registers as possible. Such an extension would not cover all of England, however, since in 2000, there were eight local congenital anomaly registers in England which in total only covered 50 per cent of all births in England.

Collaboration with the register in Wales and the three local registers in England demonstrated immediate benefits in improving the completeness of data. The initial years of electronic data transfer with the four local registers showed that highly significant increases in notification rates could be achieved. These improvements should be sustainable and therefore this would provide a solution to the problem of improving ascertainment. Given the presence of other local and condition specific registers, this solution could also be extended to include other registers, if they supported this initiative. BINOCAR would have a key role in facilitating the opportunities for networking and collaborative work to develop a relationship between the registers and NCAS.

In each of the registers included in the study presented in this chapter, however, there were enthusiastic register leaders and colleagues who were all dedicated to achieving the highest levels of ascertainment. Similar commitment would be essential to ensure the success of any new register. In addition, local registers were funded in different and sometimes precarious ways. If a local register was to lose its funding, there must be a mechanism to restore local notification from HAs to NCAS. BINOCAR would also have a role to lead the lobby for a consistent
approach to establishing and funding local congenital anomaly registers across the UK and the Republic of Ireland.

One incentive for local registers to work with NCAS was that NCAS provided a surveillance function for those registers engaged in data transfer. Whilst NCAS is in the transition stage of improving the completeness of notification there would be an impact on the local surveillance outputs and on national statistics. It is important that the fluctuations in the number of false alarms generated is managed by quantifying the expected increases due to improvements in notification for given anomalies. This would enable trends in the underlying number of anomalies to be monitored.

As electronic data transfer was extended to cover more of England and Wales, NCAS data would become more complete and therefore more sensitive to real changes in prevalence. This would improve the quality of the surveillance alarms. National data would also be closer to true prevalence rates so better able to be used to inform Government policy, assist in geographical comparisons, and assist in medical research.
CHAPTER 4

An evaluation of record linkage between the National Congenital Anomaly System and Birth Registration

My contribution:

The original ideas for the work presented in this chapter were mine, and I designed the algorithm used in this chapter based on my experience with other linkage and matching exercises. A senior research analyst within ONS programmed the algorithm using SAS. I performed the statistical analyses using Stata.

I have long-standing experience in developing record linkage algorithms, having instigated a number of record matching exercises in ONS, linking different data sources to birth and death registration records. For example, I have developed algorithms to link ONS births data with ONS child death records, with RICHS (Regional Interactive Child Health System) records, with HES (Hospital Episode Statistics) records and with CESDI (Confidential Enquiry into Stillbirths and Deaths in Infancy) records.
4.1 Summary

The feasibility of linking NCAS notifications with birth registration records is evaluated in this chapter. The aims of linkage were to reduce missing data and to add new data items to the existing database. The chapter reports on the selection of appropriate linkage methodology, and evaluates the percentage of NCAS records linked, potential bias in the linked records, and differences in data items collected by both sources. New data items allowed analyses of NCAS records according to covariates collected at birth registration. The survival of babies with congenital anomalies was also analysed, being made possible through the existing linkage of death registration records to birth records.

The study showed that linkage with birth registration records was feasible, and identified some systematic differences between the linked data and the NCAS database. Analyses of congenital anomaly rates by mother’s country of birth, and survival of babies with selected anomalies are presented to illustrate the use of linked data in the evaluation of risk factors for congenital anomalies and for mortality rates.
4.2 Background

Record linkage has been used elsewhere in the UK and abroad to bring together existing datasets to utilise all information available for individuals in order to analyse associations between characteristics and outcomes. The advantages of population-based record linkage include the avoidance of selection bias which can occur in epidemiological case-control and cohort studies (Stanley 1994, Acheson 1987). Record linked data assembled routinely can address many different hypotheses and can avoid the need for costly prospective cohort studies to address specific hypotheses. Another advantage of linked data is rapidity of analyses as the data are already assembled.

The four objectives of the study were to:

- Review two methodologies for record linkage - deterministic and probabilistic matching, and select appropriate methodology for linking NCAS records with birth registration records.
- Determine whether there were any systematic differences between linked records and those on the NCAS database.
- Assess the extent of missing data on the linked records, and evaluate agreement between NCAS and birth records for data items recorded by both sources.
- Illustrate new analyses enabled through linkage with birth records. The examples presented here are (a) a comparison of rates of selected congenital anomalies by the mother's country of birth, and (b) the survival of babies with congenital anomalies, made possible from the death registration information already linked to the birth record.

4.2.1 Aims of linking NCAS and birth registration records

The aims of the evaluation presented in this chapter were to evaluate whether linkage of NCAS records with birth registration records would be feasible, whether
it could be used to improve the quality of NCAS data by reducing missing data, and whether it would provide additional data items for new analyses of covariates and outcomes.

4.2.1.1 To reduce missing data

For all linked records it is possible to compare values for data items that are collected by both NCAS and birth registration. These comparisons were used to evaluate the proportion of linked records where the value on the NCAS record agreed with the birth record, to assess characteristics of linked records where the same data item had different values on the birth and NCAS records, and, if the data item was missing, whether it was more likely to be missing from one rather than the other source. This information would enable future decisions on replacement and supplementation of data items on the NCAS record.

In the longer term, if the quality of information on the birth record was at least equal to that on the congenital anomaly record, record linkage between the NCAS and the birth registration records would allow some items to be derived directly from the birth record, and eliminate the need to collect them again at NCAS notification. This would reduce the burden on the NCAS data providers, which in turn might improve the completeness of the remaining information needed by NCAS. However, there is currently little empirical evidence to support a relationship between questionnaire length and response rates (Bogen 1996).

A further aim of linkage, if successful, is to use birth registration information to replace that collected at NCAS notification, in order to reduce numerator-denominator bias. Birth records are used as denominators to calculate rates for congenital anomaly prevalence. Using data items collected at NCAS notifications where the information is incomplete, introduces numerator-denominator bias, as the numerator is based on a less complete population with possibly a different distribution of values to those on the birth record. Linkage of birth and NCAS records, and the transfer of information collected at birth record to the NCAS
notification could reduce numerator-denominator bias, since these data items on both records were from the same single source.

4.2.1.2 To obtain additional information on risk factors and survival

A principal objective of record linkage is to extend the range of data items available for research and analysis. Data items collected at birth registration but not by NCAS included mother’s and father’s country of birth, and the parent’s marital status at the child’s birth. Consequently, linkage can provide additional covariates from the birth record for research into factors associated with congenital anomalies.

Linkage would also allow analyses of survival in babies with congenital anomalies. This used an additional linkage of NCAS records with the existing links between a child’s birth and death record which had been made routinely since 1993 (section 4.2.3). This linkage between three sources of data would enable survival rates and causes of death to be obtained for children notified to NCAS with particular anomalies.

4.2.2 History of record linkage in health research

The main uses of record linkage have been in the field of health. Farr in 1861 (Farr 1861) and Florence Nightingale in 1890 (Nightingale 1997) described the advantages of record linkage for health care. Nevertheless it was not until the advent of computers in the 1950s that it became feasible to store medical records electronically for a large number of people and bring together records from different sources.

Dunn gave the first definition of record linkage in 1946:
...Each person in the world creates a book of life. The book starts with birth and ends with death. Its pages are made up of all the principal events in life. Record linkage is the name given to the process of assembling the pages of this book into one volume. The person retains the same identity throughout the book. Except for advancing age, he is the same person...... (Dunn 1946)

Record linkage has continued to be an aspiration over the intervening years. The value of record linkage for seamless health care was also recognised: ....

If we are to have seamless care and the ability to track individuals, whether in hospitals, or between the health care providers, then the Management Executive recognises that we must have, throughout the NHS, a unique person identifier. Even within one organisation, such as a hospital, we are likely to find several identifiers for any one individual, let alone between organisations. This is a real barrier to integration. Moreover, the absence of an identifier used in all parts of the NHS means that we are having to depend on name link processes. That makes the task of preserving confidentiality much more difficult.... (Duncan Nichol, CEO, NHS Executive 1992)

Technological improvements lead to a growth in electronic tools for linkage, and the ability to manipulate the data more easily. Smith described the three key technical issues in the development of record linkage (Smith 1984) as first, using personal identifiers to discriminate between the person to whom the record refers and all the other persons in the population, secondly, deciding whether discrepancies in identifiers are due to mistakes in reporting for a single individual or to the presence of other individuals, and thirdly, processing the large volume of data necessary for record linkage within a reasonable amount of computer processing time.

There are a number of previous record linkage studies in Britain and worldwide which are used for a wide range of health research. One of the first registers to be established in the UK was the Oxford Record Linkage Study (ORLS) set up in 1962. This linked together records for the same individual from a number of different sources, including birth, death and hospital in-patient data. By the mid-
1990s data collection covered a population of approximately 2.3 million. Its aim was for a system of linked health records which brings together selected data of biological interest for a whole population commencing with conception and ending with death, in a series of personal cumulative files, the files being organised so that they can be assembled into family groups. The term record linkage may apply specifically to the techniques of assembling the files in spite of errors and omissions in the identifying particulars, or may be used in a more general sense to apply to the organisation involved. (Acheson 1967).

There are precedents elsewhere for the type of linkage that is reported on in this Chapter. In Denmark a comprehensive national individual based register was created, known as the Danish Prevention Register. It comprised eight registers that were linked using their unique individual identification number. The register contains data from 1977 and onwards and comprised health-related registers, registers on living conditions, and a register of population statistics. The national register allowed for cross-sectional and follow-up studies and was also used for administrative purposes (Roed 1999). The congenital anomaly register was subsequently linked to these data sources and used in analyses of associations between congenital anomalies and other life events (Hansen 2000).

4.2.3 Previous experience of record linkage with ONS data sources

Since 1975, infant deaths (deaths under the age of one year) had been linked to birth records for the continuous monitoring of infant mortality by birthweight and other risk factors. These risk factors had not been collected at death registration, but were derived from the birth record through linkage. This enabled analyses showing that babies born to very young mothers, of very low birthweight, or as one of a twin or higher order birth were at greater risk of dying in infancy than other babies (ONS Series DH3 publications). The main linking variable was the NHS number, which was present on birth registrations and was derived from the National Health Service Central Register (NHSCR) on the death of a child. Over 99 per cent linkage was achieved, and half the unlinked death records related to
children born overseas who did not have a birth registration record in England and Wales with which to link. Such a high linkage rate, however, was probably due at least in part to the fact that it is a legal requirement to register both births and deaths. Also, the same close family members are often the informant at both events so give the same information. In addition, the Registrars of Births and Deaths are required to ask identical questions about some data items such as date of birth and names at both birth and death registration which maximises the ability to link using these data items.

In the early 1990s, when ONS computer systems were being redeveloped, I amended the underlying principles of the infant mortality and birth linkage. This extended the linkage to deaths of all children, including those aged over one year, born during or after 1993. When the death of a child born since 1 January 1993 was registered, their birth and death records were linked, thus extending this linkage beyond infancy. For example, a child born in 1993 who died aged seven in 2000 had their death record linked to their birth record. This allowed the known risk factors for infant mortality described above to be monitored, and to show how these risk factors extended beyond infancy. For example, these analyses showed that mortality rates for children under four years of age were highest for those with teenage mothers, for those of low birthweight, and for those children whose mother was born in the New Commonwealth (primarily Bangladesh, India, Pakistan, East Africa and the Caribbean) (Dattani 1999). This linkage of birth registration and death registration records was used in the study reported here to determine survival rates for children with congenital anomalies.

ONS birth registration data have also been linked with data sources external to the registration service. NCAS data are external to the registration service, so this previous work provides experience for the linkage study described in this chapter. One example was a study linking ONS birth records to those reported to the Confidential Enquiry into Stillbirths and Deaths in Infancy (CESDI). To maximise completeness of ascertainment, CESDI records were matched with both ONS infant mortality and stillbirth records. A 4-stage matching was performed, the
essential fields being postcode, sex, date of birth, and date of death. Ninety six per cent of records in 1998 were matched. A similar algorithm was used in a more recent feasibility study linking birth registration data with Hospital Episode Statistics data held by the Department of Health (Abrahams 2002).

This latter linkage study was more difficult as names are not recorded on the HES records provided to the Department of Health. Although not a primary linking data item, name is useful to confirm matches. However, to help preserve confidentiality many NCAS records are notified without names. Therefore this linkage with HES records provided valuable experience of matching in the absence of name.

An additional issue in linking highly confidential data is that it may not be possible for those using the data to have access to all the confidential information. At birth registration, mother’s date of birth is collected in confidence for statistical purposes only and does not appear on the child’s birth certificate. The algorithm used in the study reported here was modified from an algorithm that was developed for linking ONS births data with RICHS (Regional Interactive Child Health System) data, which had used mother’s date of birth as a primary linking data item. The study investigated the prevalence of specific infections in population groups born overseas (Ades 1999). Permission was given to use mother’s date of birth in this study to link records, provided this information was not held on any resulting databases. The RICHS system collected birth notification data from midwives for all births in the North Thames East area. The matching algorithm used NHS number or, where this was missing, sex, date of birth and postcode as the primary matching fields, supplemented by information on first and surname, hospital code and birthweight. A 99 per cent matching rate was achieved. The matching algorithm identified the parents’ countries of birth from the ONS birth record for each child on the RICHS file, and added this information to the RICHS record. The RICHS records with the additional country of birth information were anonymised before sending to the researchers.
Each of these examples contributed to the development of linking methodology, appropriate combinations of data items, issues when working with confidential data, how to deal with confidential data (such as names or mother's date of birth), and the specific problems of linking data collected through the health service to that collected at birth registration. This experience was brought together to develop the algorithm used in the study reported in this chapter.
4.3 Methods

4.3.1 Choice of methodology for record linkage

Record linkage brings together records from different data sources, often collected at different times, relating to the same person, family or locality. In this study methods of 'exact matching' are considered. These aim to link information on the same individual in two files. An alternative method often used in marketing studies is 'statistical matching', which aims to link information on individuals who are similar to each other.

The first stage of this study to evaluate record linkage was to identify appropriate 'exact matching' methodology to test the feasibility of linking NCAS records to birth registration records. Two major reviews of the literature (Gomatam 1999, Baldwin 1987) identified two principle methods for exact matching record linkage; deterministic matching and probabilistic matching. These are described in more detail below.

4.3.1.1 Deterministic matching

Deterministic methods have been used extensively in record linkage for medical purposes, but there is very little written about them in the peer-reviewed literature. Deterministic matching depends on all records in both data sets having a unique identifier, or a unique combination of data items. The earliest computer record linkage systems used the deterministic method to search for records that matched exactly on defined variables.

Deterministic methods have the advantage that they are straightforward to programme, seeking an exact match on one or more combinations of variables. The disadvantage, however, is that the high specificity of this method means that it
has lower sensitivity, and can miss true matches where one database records the
date of birth as one day different for example.

4.3.1.2 Probabilistic matching

When there is no unique identifier available on both data sources, and key data
items may be recorded differently between the sources, a best estimate has to be
made of the records that match best using the range of different data items
available on both sources. Probabilistic linkage was developed to determine the
most likely records to represent a true match and was formalised by Fellegi and

In probabilistic linkage, records are examined in pairs, one from each source. A
probability is calculated of the chance that these records refer to the same
individual. Using either previous experience of record matching in similar
applications, or from preliminary analysis using the data set of interest, a
probability is calculated of how likely the agreement would be in correctly matched
pairs. Some variables are more useful than others in dividing the population into
smaller subgroups for comparison. For example, date of birth is a useful variable
since day and month alone divide the population into 365 subgroups. Weights are
calculated for each of the variables used in the matching algorithm; positive
weights suggest agreement and negative weights disagreement. The weights are
summed and compared with threshold values that determine whether it is low
enough to definitely not be a match, high enough to classify as a true pair, or
uncertain since it lies between the two values.

The advantages of probabilistic linking is that it can identify true matching records
where there are small differences between the values of a data item recorded by
both sources, thus achieving higher sensitivity and a higher matching rate than
deterministic matching. Its main disadvantages are that by accepting probabilities
of less than unity for potential matches, it will link more false matches than
deterministic linkage and thus have a lower specificity. Probabilistic linking also
requires weights to be calculated for all the linking variables used, which can be a
large processing task.

4.3.1.3 Methodology used to link NCAS and birth registration records

The chosen methodology was deterministic matching since a unique combination
of data items is available for most births. Probabilistic matching was not selected to
link birth registration records with NCAS records in the study reported here since it
had two disadvantages. First, it would require weights to be calculated for each
date of birth in the four years covered by this study, and for each postcode in
England and Wales. Secondly, the probability algorithm would need to be applied
to the 700,000 birth registration records in each year covered by the study. Both
issues would require substantial additional development work which was beyond
the resources available for this study and hence beyond the scope of this study.

Only one study was found in the literature which compares deterministic and
probabilistic matching methods quantitatively when applied to linking files for which
the truth was known (Gomatum 2002). Medical records and subsequent
educational performance records for children born between 1989 and 1992 and
treated in Florida’s Regional Intensive Care Centres were linked. The accuracy of
linkage was assessed by calculating sensitivity, specificity and positive predictive
values (PPVs). The sensitivity determines the percentage of false-negative results,
and specificity determines the percentage of false-positive results, when a large
number of positive and negative samples are tested. The positive predictive value,
however, expresses the probability that a person with a positive result is truly
affected; negative predictive value is the probability that a person with a negative
result is not affected.

In the Gomatum study, probabilistic matching gave a higher sensitivity but lower
PPV than deterministic matching. Deterministic matching had a PPV close to unity.
but a lower sensitivity than probabilistic matching. The lower sensitivity of deterministic matching means that there will be fewer false negatives, or fewer unmatched records which in reality could have been linked, than for probabilistic matching, and its higher PPV means that there will be a higher proportion of true matches. Therefore, for the algorithm developed for the evaluation presented in this chapter, needing a high PPV, together with an understanding that from October 2002 the unique NHS number should be available at birth for each child which would provide a single unique identifier in place of the unique combinations available for this study, the chosen methodology was deterministic matching.

Other studies have tried to use a combination of both methodologies to maximise both sensitivity and PPV. A study in Scotland (Ramsay 1999) used a two stage record linkage algorithm to link neonatal morbidity records to Community Health Index records. The first stage was an exact match on date of birth and the first four characters of the postcode, together with a probabilistic match of surname. The second stage was a probabilistic computer match based on date of birth, postcode and surname, supplemented with hand matching. This algorithm matched 89.8 per cent of records. The authors tested the sensitivity of the matching using first name which had not been part of the linkage algorithm. Of the records matched, 98.9 per cent had the same first name, suggesting that a true match had been made. This study also identified that in this study one fifth of all babies had a different surname in the two records.

Names can be recorded differently on different data sources, either due to transcription errors, to reporting by different people about the same individual, or due to using different names for different situations. In both the study in Scotland (Ramsay 1999) and the study reported in this chapter, a child having different names is more likely to reflect changes of surname between birth and birth registration for children born outside marriage, rather than suggesting that the two records related to different individuals. A baby may be recorded in hospital notes with the surname of the mother, but at birth registration, where the parents are not married, the child may be registered with the father’s surname.
The authors of the study in Scotland recognised that their linking rate would have been improved if there had been a unique NHS number present to use in a deterministic procedure. Within England and Wales, the National Health Service (NHS) number is a unique identifier for health records. A new 10-digit NHS number was allocated to the entire population of Great Britain in January 1997 (Secretaries of State 1989, National Health Service and Department of Health 1990). Currently, however, this number is rarely used in the main national databases containing data on child health such as congenital anomaly registers and hospital maternity systems. The NHS number was allocated to a baby at birth registration, which could legally take place up to 42 days after live birth. Therefore the child might have left hospital before its NHS number was allocated and hence this identifier failed to be recorded on the maternity records. As a result, the NHS number was rarely reported on the NCAS SD56 notification because most NCAS notifications were derived from information collected at birth notification before the NHS number had been allocated.

If the NHS number was not given at NCAS notification, it could have been found from the NHS Central Register using the child’s date of birth, the first three letters of their forename and surname, plus other information on residence if available, to provide a unique match from several potential matches. However, the disadvantage was that some records could only be traced with intensive clerical intervention, which was expensive. Since no additional resources were available for the study reported here, this additional source to derive the child’s NHS number was not explored further in this study.
4.3.2 Data sources for record linkage

The evaluation of linkage between NCAS and birth registration records was based on all NCAS notifications for babies born between 1997 and 2000 using data taken from the NCAS database in January 2002. These data were selected because they were the most recent four complete calendar years of NCAS data.

Resources were only available to evaluate four years of data. Since this study was a feasibility study for the introduction of routine record linkage the four most recent years, 1997-2000, were selected because the results would be the most appropriate on which to base decisions on future linkage. Ideally data would have been analysed for 1993 to 2000, the most recent year’s data available. Prior to 1993 there was no name information on the birth registration records, and name was used to a limited extent in the algorithm used in the study presented here.

Complete calendar years data were included for two reasons. First, they are important for etiological analyses of the resultant data, and second, it was planned that analyses from the linked data could be included in routine ONS annual publications, which are based on data from complete calendar years.

To try to understand risk factors for congenital anomalies, a range of social and demographic characteristics are collected at NCAS notification. These include the first three initials of the child’s first and surname, their date of birth, birthweight, gestation, multiplicity (number of babies delivered from the pregnancy), and a text description of the anomalies. The mother’s date of birth, address and number of previous pregnancies are also collected. Appendix A shows the standard notification form used by NCAS (SD56).

Birth registration records are held by ONS in two separate databases; public records of birth registrations, and statistics files derived from birth registrations. The public birth registration record includes full names of the child and parents, text descriptions of the parents’ country of birth and occupation, the child’s sex, date
and place of birth, and NHS number, whereas the mother's age, birthweight, and the parents’ country of birth and occupation codes are in the statistics record. Fields required to enable the linkage and for the subsequent analyses presented here were extracted from both birth files and merged into single records for each birth between 1997 and 2000. Appendix E shows the standard form used to record the information collected at birth registration.

Permission was obtained from data custodians at ONS to access the birth records and NCAS records. In addition, permission was specifically obtained to use mother’s date of birth from the birth record as a linkage variable since this data item was collected under the Population Statistics Act 1938 and hence confidential. During the processing of the data all work was done in a secure and restricted area on the ONS file server.

Table 4.1 gives the number of records extracted from the NCAS and birth data files for the selected calendar years.

<table>
<thead>
<tr>
<th>Year</th>
<th>NCAS records</th>
<th>Birth records</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>5879</td>
<td>787333</td>
</tr>
<tr>
<td>1998</td>
<td>6043</td>
<td>673142</td>
</tr>
<tr>
<td>1999</td>
<td>7189</td>
<td>648400</td>
</tr>
<tr>
<td>2000</td>
<td>7395</td>
<td>643451</td>
</tr>
</tbody>
</table>

Table 4.1 Number of records extracted from NCAS and birth data files, 1997 - 2000.

Numbers of births include late registrations for that year and exclude births registered in that year but born in the previous year.

The final stage of the data preparation was to remove duplicate records. Births are only registered once so should not appear more than once on the file. Additional information can be entered onto the birth register subsequently, however, which could generate duplicate birth records. This problem was avoided by selecting
records using a flag on the record which indicated that this record was the most up
to date version.

A baby with anomalies could be notified to NCAS more than once as a result of
different sources reporting the same case or subsequent anomalies being
diagnosed. NCAS records where the mother's details were the same in two or
more records (same date of birth and postcode), could reflect either a duplicate
record or the baby being one of a multiple birth. The child's forename and
birthweight was used to indicate individual babies in a multiple birth. Remaining
records where the child's details and the mother's details were the same were
likely to be duplicate records. These were resolved manually; one record was
edited to include additional information from the other record, then the duplicate
record was removed from the NCAS database.

The computer systems for both NCAS and birth registration records had already
checked that the values of the different data items were within the expected range
for that variable (for example, that mother's age was between 14 and 49), before
the record had been accepted by the system. Additional crosschecks built into the
routine processing systems also identified inconsistencies between different data
items. For example, a stillbirth at 36 weeks gestation was unlikely to weigh 300g
but could weigh 3,000g. Such records would be highlighted for manual scrutiny.

There was some overlap between information collected at NCAS notification and
that collected at birth registration. This had allowed notification rates for these risk
factors to be calculated, using NCAS data as the numerator and births as the
denominator. The data items collected both by NCAS and at birth registration were
the baby's date of birth, the mother's date of birth, the parents' occupations,
multiplicity, birthweight, postcode, and, for stillbirths only, gestation. At NCAS
notification all data items except occupation and the child's names had a stated
value for over 90 per cent of records. Forty per cent of NCAS records, however, did
not record parents' occupations.
4.3.3 Algorithm for linking NCAS and birth registration records

The deterministic matching algorithm used in this study sought alternative exact matches using combinations of variables available on both birth and NCAS records. The algorithm is an enhancement of the matching and linkage algorithms described in section 4.2.3. These were based on comparisons of a small set of variables, which were collected by both data sources. For the study presented here, postcode, dates of birth (DOB) of both the baby and their mother, and sex were the main variables used. The algorithm developed for this study was a stepwise hierarchical algorithm, based on 5 passes of the data using different sets of variables.

The following abbreviations are used in the description of the algorithm:

- Bwt = birthweight
- DOB= date of birth
- MDOB=mother's date of birth
- Mage= mother's age
- Pc3=First 3 characters of postcode
- Fname=forename
- Sname=surname

- Stage 1

Baby's full DOB + Mother's full DOB + postcode + sex

Sex was only required to match if it was given on the NCAS record. If it was missing, then a successful match was established if the other variables matched completely. These were the same variables that had been used successfully for the RICHS and CESDI matching programmes described in Section 4.2.3. Duplicate matches, when one NCAS record could potentially match with more than one birth registration record, were resolved as described below.
• Stage 2

Baby’s full DOB+ Mother’s full DOB+sex+pc3

These data items were therefore the same as in Stage 1 except that the postcode criterion had been broadened to just the first three characters to allow for slight coding errors (or a house move within the same postcode sector). Duplicate matches were resolved using the same criteria as at stage 1.

• Stage 3

Surname + either (i) postcode + sex + (bwt or mage)
  or (ii)mdb + sex + (bwt or postcode)
  or (iii) dob + sex + mage + (bwt or pc3)
  or (iv) fname + either (a) dob + (bwt or mage or pc3)
    or (b) postcode + (dob or mage or bwt)
    or (c) mdob + (dob or bwt or pc3)

This was the first time the child’s names had been used as part of the algorithm, other than to resolve duplicate matches.

• Stage 4

If the child was stillborn or died within 7 days:
surname + one from (month of baby’s birth, postcode, mother’s DOB, bwt)

If the child was live born and still surviving:
surname + month of the baby’s birth, +
  either (i) postcode and one of (mage, baby’s DOB, bwt, fname),
    or (ii) baby’s DOB + either (a) the mother’s DOB
      or (b) any two from (mage, bwt, pc3, fname).
Stage 5

Bwt + either (i) mother's DOB + postcode

Or (ii) mother's DOB + baby's DOB

or (iii) baby's DOB + postcode.

Initially the algorithm linked using very tight matching criteria, and removing records from both sources as unique matches were found. At stages 1 and 2 if duplicate matches were found for a given NCAS record, additional criteria (such as the child’s first name) were included to confirm potential matches and remaining duplicates were resolved manually. At later stages, all duplicates were resolved manually.

The algorithm also included a manual assessment using partial data to identify which, of a number of records (usually less than 10) suggested by the algorithm, were likely to belong to the same person. This was necessary because information on the computer records could have been subject to error, both in the information given and in transcription. Postcodes often vary between data sources, due to transcription errors or because the child may have moved house between birth (the primary source for NCAS notification) and birth registration. Dates of birth could have transpositions of day and month, or vary by one day between sources. All these complications reduce the power of a computerised deterministic matching system to detect which records relate to the same person and should be linked. Data items included in this manual evaluation were dates of births with an adjacent or transposed date, and names, in the knowledge of common abbreviations (for example – Edward being commonly abbreviated to Ted).

A child’s surname might be recorded differently at NCAS notification and at birth registration due to different collection processes and different people providing the information. At NCAS notification the name information was taken usually from the information recorded by the midwife on the birth notification form shortly after birth.
At birth registration, however, the child might be registered with a different surname, for example their father’s surname where this was different to the mother’s.

Name was only used in the latter stages of the algorithm, or to discriminate between two records manually where the matching algorithm had identified more than one possible match. Whenever surname was used in the algorithm, the computer programme used three cycles of comparisons, comparing the first three characters of the child’s surname as recorded on the NCAS record against birth registration information for the child’s surname, the father’s surname and the mother’s maiden name. Similarly, if the forename was used as a part of the matching criteria, the first three characters of the child’s forename as recorded on the NCAS record were matched against the child’s first or subsequent forenames, as given on the birth registration record.

Name was not used as a primary discriminating variable because combinations of surnames and forenames are rarely unique. Therefore, an exact match on these variables alone might relate to different individuals who share a common name. In addition, only the first three characters of the child’s first name and of their surname are collected by NCAS which made name a poor matching variable. The first three initials of each name might be insufficient to discriminate (e.g. McC) and in other cases might be incorrect (e.g. BAB for ‘baby’).

The algorithm developed for this study included the decision processes necessary to discriminate as far as is possible whether discrepancies between potential matches were due to recording errors, or because they did not relate to the same person. Data storage was no longer a major concern for ONS computer systems, so much larger arrays of data could be stored and manipulated using easily available commercial spreadsheets and databases. Therefore all the technical issues raised in section 4.2.2 were addressed in this algorithm.
4.3.4 Determination of linkage rate

The linkage rate, expressed as a percentage of the original number of NCAS records, is important if the linked records are subsequently analysed for absolute risk and to determine the risk of anomalies in different subgroups of the population.

Absolute risk rates are calculated using the linked data as numerator and total birth registration records as denominator. For example, in 2000 there were 607,304 live and stillbirths in England and Wales, and 7,284 notifications to NCAS (Table 1.1), giving an absolute birth prevalence of 119.9 cases per 10,000 births. However, if 95 per cent of NCAS records for 2000 were to link (6,920 cases), then the birth prevalence rate using linked data would be 113.9 per 10,000. This latter rate is 95 per cent of the rate based on the total number of notifications to NCAS. In both calculations the denominator is the same, being the total number of live and stillbirths, but the numerator for the linked records is reduced by the proportion of records linked. Since the denominator is unaffected by linkage rates, but the numerator is reduced to the linked proportion, any absolute risk rates will be reduced in proportion to the linked rates. This means that the rates from the linked data appear to be lower than the true rates. This is important when comparing with other data sets, for example with rates from other countries.

Provided the linkage rate does not differ between subgroups of the population, relative risks are unaffected by the linkage rate. If linkage rates differ for different subgroups, however, comparisons between subgroups will be biased. Using the example from the previous paragraph, 246 of the 7,284 notifications in 2000 had a mother aged 40 or over, compared with 15,169 of the 607,304 live and stillbirths (ONS series MB3 2001). This gave a notification rate of 162.2 per 10,000 live and stillbirths for mothers aged 40 and over, compared with 118.9 for younger mothers, giving a relative risk of 1.36. If linkage rates differ for these two groups of mothers, however, such that linkage is achieved for 90 per cent of notifications with mothers aged 40 and over (221 cases) compared with 95 per cent of younger mothers (6,686 cases), then the notification rates using linked data would be 145.7 and
112.9 respectively giving a relative risk of 1.29. Thus, if a group at higher risk (in this example, cases with older mothers) has a lower linkage rate than other groups, then the relative risk will be reduced. In contrast, if the group at higher risk has a higher linkage rate than other groups, then the relative risk will be increased.

### 4.3.5 Comparison of linked and unlinked records with the NCAS database

The second aim of the linkage study was to determine whether there were any systematic differences between linked records and those on the whole NCAS database. The percentage of records linked was calculated. Linked, unlinked and total records were compared for a selection of data items to determine any evidence of bias that would influence inferences from the linked records analysed later in this chapter.

The selected data items were birthweight, mother’s age, and whether the baby was live, live but died in the first week of life, or stillborn (perinatal mortality). As these data items were all included in the linking algorithm, it was expected that agreement for these data items would be high. Birthweight is a key data item in congenital anomaly epidemiology. Mother’s age is another key data item in analyses of congenital anomalies given the known association of anomalies such as Down syndrome with maternal age.

Comparisons were also made for data items not used in the algorithm; year of birth, region of England, and multiplicity. These data items were collected at birth registration and NCAS notification, and were important for subsequent analyses of the linked data set.

Gestation was collected by NCAS and also at stillbirth registration, but not live birth registration so was excluded from this analysis. Names and postcodes, the remaining data items collected by both sources, were not included in this analysis because of the difficulty in electronically providing meaningful comparisons of alpha numeric variables.
Chi-square tests were performed to test the null hypothesis that, for the given data item, linked records followed the same distribution as total records. This test was also used to identify subgroups in which linkage rates had differed significantly to other groups.

4.3.6 Evaluation of agreement between selected data items collected by both sources

The third aim of this study was to analyse the linked records to evaluate agreement between NCAS and birth records for data items recorded by both sources, and to compare levels of missing data between the two data sources.

The four data items selected to evaluate data collected by both sources were birthweight, mother's age, multiplicity and whether the child was live or stillborn. These data items were selected as they were affected differently by the algorithm. For mother's age the assessment of agreement is confounded by the algorithm because mother's date of birth is one of the primary linking variables in the first and subsequent stages. Birthweight and whether a child is live or stillborn is affected to a much smaller extent since it is only included in the later stages of the algorithm. Multiplicity is not used directly in the algorithm.

These data items also reflect different ways of collecting information. Birthweight is recorded by the midwife, who passes this information to the Health Authority (HA) or Trust. Therefore differences in agreement may relate to the methods of ascertainment (Figure 4.1). The HA or Trust in turn informs the registrar of birth and deaths of the birthweight, and also records this information on notification to NCAS. Thus, the midwife is the provider of the data for both sources. Differences in birthweight could only be due to a transcription error by either data source as the original notifier is the same.
Birthweight

Same notifier
Same question

Birth notification
Includes birthweight
Registrar

HA/ Trust copy birthweight
Registrar copies birthweight

NCAS notification
Birth certificate

Mother's age

Different notifier
Same question

Birth notification
Includes Mother's dob/age
Registrar

HA/ Trust copy Mother's dob/age
Registrar asks Mother's dob/age

NCAS notification
Birth certificate

Multiplicity

Different notifier
Different question

Birth notification

How many born?
Registrar

How many legally registrable?

NCAS notification
Birth certificate

Figure 4.1 Notification pathways for birthweight, mother's age and multiplicity in NCAS and at birth registration.
Mother's age is calculated from the mother's date of birth. This is collected at birth registration by the Registrar, and for NCAS notification by the midwife. There were no 'not stated' ages on the birth records as ages were imputed by ONS if these were not stated on the birth registration record. It is generally recommended that missing data are replaced using multiple imputation (modelling the uncertainty due to missing data, while using the existing data) or regression (predicting the missing value based on the other available data) or some form of hot deck imputation (Dennis1997, Little 1989). A 'hot deck' procedure retains information on the most recently processed records where all the data items are complete. Then, whenever a record is processed which has one or more missing data items, the computer selects the most recently processed record which matches the current record on all other data items. It then imputes from that record the value for the missing data item. If the mother's date of birth is not given at birth registration, an age is imputed by ONS using a hot deck procedure from the last processed record with completely stated and otherwise matching particulars. In 2000, the mother's date of birth was not stated for 0.51 per cent of all live births (3,081 births) and an age was imputed using a hot deck procedure. These cases can be identified because the mother's date of birth field is blank but there is a valid age given. Differences in mother's age between NCAS and the birth registration could therefore be transcription error, different information being given to the data sources, or due to missing age on the birth record and a value imputed (see Figure 4.1).

Multiplicity is collected at birth registration by the Registrar and for NCAS notification by the midwife. For this data item, however, a different question is asked at each source of data collection. At birth registration the informant is asked how many legally registrable births resulted from this maternity. The midwife, however, records the number of delivered babies and fetuses. If a woman delivers twins at 23 weeks of gestation, and one child is born live and the other born dead, only the live born child is legally registrable whereas both would be recorded by the midwife. Differences in multiplicity between the two sources, therefore, could be transcription error, different information being given, or a result of the different questions which are asked.
Information on whether a child is live or stillborn, or dies in the first week of life (perinatal mortality) is collected at birth registration by the Registrar and for NCAS notification by the midwife.

Birthweight was compared using both exact recorded birthweight and data grouped into 500 gram categories. All comparisons of mother's age are made based on the mother's completed years of age. In the analyses multiplicity is recorded as singleton, twin, or higher order birth, and perinatal mortality is recorded as live born, live but died in the first week, and stillborn.

A chi² test was used to test agreement between the distributions of the population according to the data item on both records. For the continuous data items, mother's age and birthweight, the two sets of values were plotted. If the values are the same or very close they will be tightly scattered about the X=Y line. Mean differences, 95 per cent confidence intervals and correlation coefficients were calculated using STATA. The assessment aimed to determine which groups have the biggest influence on the chi² statistic and how the differences are affected by the linkage algorithm.

4.3.7 Illustrative analyses enabled through linkage

The fourth aim of this study was to illustrate new analyses enabled through linkage with birth records. Two analyses are presented: First, notification rates for specific anomalies by the mother's country of birth, and secondly, survival of babies with specific anomalies up to 48 months of age.

In the first analysis, countries were included if they had 20 or more notifications to NCAS in 1997-2000. The specific anomalies included were neural tube defects, eye anomalies, cleft lip and palate, heart and circulatory system anomalies,
respiratory anomalies, limb reduction defects, abdominal defects and Down syndrome. These anomalies were selected as they were well defined, universally collected, easily detected at birth or subject to minimal clinical variation in diagnosis. They are the same categories as were analysed in Chapter 3.

Notification rates from each of the countries were compared against those for the UK using the Poisson model. The score statistic was calculated and compared against a standard normal table, based on a two sided test.

The survival analysis used the death registration information already linked to the birth record. Since 1993, death registration records had been routinely linked to their birth records, as described in section 4.2.3. Therefore it was possible to link death information to NCAS records through the linked birth record. To summarise survival to 48 months of age, Kaplan-Meier survival curves were calculated using the linked NCAS-birth-death data. Survival of these children was followed up to 31 December 2001. By this date the oldest children included in the study, born between 1 January and 31 December 1997, would have reached 47 months of age. Babies born in 2000 would have reached their first birthday in 2001 so only contributed to the population at risk of dying when aged less than 1 year 11 months (for the child born in January 2000). Similarly those born in 1999 could have contributed to the population at risk of dying up to and including age 2 years 11 months, those born in 1998 up to age 3 years 11 months.

The survival curves were tested for heterogeneity using log-rank test for censored data. Data are presented separately for anencephaly, Down syndrome, and all other anomalies combined. In addition, this section also tests the association of survival for these same three categories of anomalies, separately according to whether they are isolated anomalies or in combination with additional anomalies.
4.4  Results

4.4.1 Analysis of linkage rate

Table 4.2 shows the number of records linked at each pass of the algorithm according to year of birth. Three quarters of the records were matched using the first stage of the algorithm, which used both mother's and baby's full date of birth, postcode and baby's sex. The second phase of the algorithm increased the proportion of matched records to over 80 per cent by reducing the postcode to postcode sector (the first three characters of the postcode). The third step of the algorithm, based on dates of births and one of the names or birthweight, increased those matched to 93 per cent for 1999 and 95 per cent or higher for the remaining years.

The latter stages of the algorithm generated a number of possible matches that were resolved manually by examining other fields on the two or more potentially matching records, a labour intensive task.

The final percentage of unlinked records was less than 2 per cent for 1997 and 1998, increasing to over 3 per cent for 1999 and 2000. The final linkage rate was 97 to 98 per cent.
<table>
<thead>
<tr>
<th></th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number matched</td>
<td>Cumulative percentage linked</td>
<td>Number matched</td>
<td>Cumulative percentage linked</td>
</tr>
<tr>
<td>Original number of records</td>
<td>5879</td>
<td>---</td>
<td>6043</td>
<td>---</td>
</tr>
<tr>
<td>Duplicate records</td>
<td>6</td>
<td>0.1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Linked after stage 1</td>
<td>4449</td>
<td>75.7</td>
<td>4778</td>
<td>79.1</td>
</tr>
<tr>
<td>Linked after stage 2</td>
<td>613</td>
<td>86.1</td>
<td>526</td>
<td>87.8</td>
</tr>
<tr>
<td>Linked after stage 3</td>
<td>549</td>
<td>95.4</td>
<td>495</td>
<td>96.0</td>
</tr>
<tr>
<td>Linked after stage 4</td>
<td>130</td>
<td>97.7</td>
<td>128</td>
<td>98.1</td>
</tr>
<tr>
<td>Linked after stage 5</td>
<td>21</td>
<td>98.0</td>
<td>27</td>
<td>98.5</td>
</tr>
</tbody>
</table>

Table 4.2 Linked NCAS records by stage of algorithm, 1997-2000
4.4.2 Comparison of linked records with the NCAS database

Year of birth

Table 4.3 shows the total and unlinked records by year of birth.

<table>
<thead>
<tr>
<th>Year</th>
<th>Total records</th>
<th>Total unlinked</th>
<th>Percentage unlinked</th>
<th>Expected unlinked</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>5879</td>
<td>111</td>
<td>1.89</td>
<td>152.9</td>
</tr>
<tr>
<td>1998</td>
<td>6043</td>
<td>87</td>
<td>1.44</td>
<td>157.1</td>
</tr>
<tr>
<td>1999</td>
<td>7174</td>
<td>257</td>
<td>3.58</td>
<td>186.5</td>
</tr>
<tr>
<td>2000</td>
<td>7395</td>
<td>235</td>
<td>3.18</td>
<td>192.3</td>
</tr>
<tr>
<td>Total</td>
<td>26491</td>
<td>690</td>
<td>2.60</td>
<td></td>
</tr>
</tbody>
</table>

$\text{Chi}^2 = 78.89 \quad 3 \text{ df} \quad P<0.001$

Table 4.3 Unlinked records as a percentage of all notifications by year, 1997-2000

There was a significant stepwise increase in the proportion of unlinked records, from less than 2 per cent in 1997 and 1998, to over 3 per cent of all records in 1999 and 2000.

Region and year of birth

Table 4.4 presents unlinked data by year and area, with data for the years when the local register began data transfer to NCAS shown in bold.

Table 4.4 shows that between 1997 and 1998 there was a fall of 22 per cent in the total number of unlinked records from 111 unlinked records to 87. Wales was the only area to experience an increase in unlinked records. Wales began data transfer to NCAS in 1998.
### Table 4.4 Number and percentage of unlinked records by year for areas with a local congenital anomaly register participating in data transfer, 1997-2000

<table>
<thead>
<tr>
<th>Register</th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wales</td>
<td>13</td>
<td>18</td>
<td>23</td>
<td>17</td>
<td>71</td>
</tr>
<tr>
<td>Trent</td>
<td>25</td>
<td>17</td>
<td>81</td>
<td>79</td>
<td>202</td>
</tr>
<tr>
<td>NTW*</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>38</td>
<td>66</td>
</tr>
<tr>
<td>Mersey</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>32</td>
<td>44</td>
</tr>
<tr>
<td>Remainder</td>
<td>64</td>
<td>44</td>
<td>130</td>
<td>69</td>
<td>307</td>
</tr>
<tr>
<td>Total</td>
<td>111</td>
<td>87</td>
<td>257</td>
<td>235</td>
<td>690</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Number of unlinked records</th>
<th>Percentage of total records for the area and year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wales</td>
<td>2.7</td>
<td>2.2</td>
</tr>
<tr>
<td>Trent</td>
<td>3.0</td>
<td>2.4</td>
</tr>
<tr>
<td>NTW*</td>
<td>2.8</td>
<td>2.5</td>
</tr>
<tr>
<td>Mersey</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Remainder</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Total</td>
<td>2.0</td>
<td>1.4</td>
</tr>
</tbody>
</table>

*NTW= North Thames West

Between 1998 and 1999 the number of unlinked records trebled, from 87 to 257 records. Thirty eight per cent of this increase was for records in Trent which increased from 17 to 81 unlinked cases. Data transfer by Trent began in 1999. The number of unlinked records for Mersey increased from no records in 1998 to 11 records in 1998, one year prior to beginning data exchange, and the number of unlinked records trebled for the remainder of England. Between 1999 and 2000 the number of unlinked records fell slightly; the only increases were for Mersey and North Thames West, which both began data transfer in 2000.

### Birthweight

Table 4.5 compares the number of linked and all records by the baby's birthweight as given on the NCAS record.
Table 4.5 Numbers and percentage of linked and unlinked records by birthweight groups, 1997-2000

For most birthweight categories, over 97 per cent of all records linked. For babies without a stated birthweight, however, 93 per cent linked, and for babies weighing less than 1000 grams at birth only 89 per cent linked. The distribution of linked compared with all NCAS records did not differ significantly (p>0.1). However, an analysis of NCAS data based on linked records would underestimate the number of babies with congenital anomalies weighing less than 1000 grams by 67 (11 per cent), and those without a stated birthweight by 277 (7 per cent).

Mother's age

Table 4.6 compares the number of linked and all records by the mother's age as given on the NCAS record.
Table 4.6 Distribution of mother's age for linked and total records, 1997-2000

For most age groups, over 97 per cent of records were linked. For women aged 45 and over, however, 95 per cent of records linked. The smallest proportion of linked records, 81 per cent, is for those records where the mother's age was not stated. Since mother's age is a primary linking data item, NCAS records without a stated mother's age would be less likely to find a linking birth record. The chi\(^2\) test, comparing the observed and expected distribution of linked records by mother's age, was significant, showing that the linked records have a different distribution to the total NCAS database. If records with a not stated mother's age are excluded from the analysis, the resultant chi\(^2\) is not significant (p>0.5), showing that it is the group of mothers without a stated age which exert most influence on the chi\(^2\) statistic.

### Multiplicity

Table 4.7 compares the number of linked and unlinked records by multiplicity as given on the NCAS record.
Table 4.7 Distribution by multiplicity for linked and total records, 1997-2000

While over 97 per cent of NCAS notifications relating to singleton and twin births linked, triplet and higher order births, and those without a stated multiplicity had lower linkage rates. Numerically these latter categories were small so had little impact on the $\chi^2$ statistic. The distribution of linked records was not significantly different to that of all NCAS notifications ($p>0.5$).

**Perinatal mortality**

Table 4.8 compares the number of linked and unlinked records by whether the child was live or stillborn as given on the NCAS record.

For live births who survived the first week, and those who died in the first week, 97 per cent of NCAS records linked with a birth record. For stillbirths and those without a stated birth status a smaller proportion linked. The resulting distribution of linked records was significantly different from the distribution of all NCAS records ($p<0.05$), mainly due to the large number of unlinked stillborn records.

A further analysis of the data showed that of the 81 unlinked NCAS stillbirth records, 35 had been born at less than 24 weeks of gestation, the legal criterion for registering a baby born dead as a stillbirth. A further 13 unlinked stillbirths did not have a stated gestation.
### Table 4.8 Distribution by perinatal mortality for linked and total records, 1997-2000

<table>
<thead>
<tr>
<th>Perinatal mortality</th>
<th>Total NCAS records</th>
<th>Linked records</th>
<th>Unlinked records</th>
<th>Percentage linked</th>
<th>Expected linked</th>
</tr>
</thead>
<tbody>
<tr>
<td>not stated</td>
<td>77</td>
<td>66</td>
<td>11</td>
<td>85.7</td>
<td>75.0</td>
</tr>
<tr>
<td>live born</td>
<td>24575</td>
<td>24006</td>
<td>569</td>
<td>97.7</td>
<td>23934.2</td>
</tr>
<tr>
<td>Early neonatal death</td>
<td>914</td>
<td>885</td>
<td>29</td>
<td>96.8</td>
<td>890.2</td>
</tr>
<tr>
<td>Stillborn</td>
<td>896</td>
<td>815</td>
<td>81</td>
<td>91.0</td>
<td>872.6</td>
</tr>
<tr>
<td>Total</td>
<td>26462</td>
<td>25772</td>
<td>690</td>
<td>97.4</td>
<td>25772</td>
</tr>
</tbody>
</table>

Chi² = 10.26  3 df  P < 0.05

### Characteristics of unlinked records

All unlinked records had a stated date of birth for the child and a valid postcode.

Table 4.9 shows the 690 unlinked NCAS records according to whether they have stated values for the other data items used in the algorithm; birthweight, mother’s age, multiplicity and perinatal mortality. Of the 690 unlinked records, 371 records (54 per cent) had a stated value for each of birthweight, mother’s age, multiplicity and perinatal mortality. However, 15 of these records had a stated mother’s age but no mother’s date of birth given showing that the mother’s age had been given directly by the person notifying NCAS and not calculated by the NCAS processing system.
### Table 4.9 Unlinked records by data items with missing values, 1997-2000

<table>
<thead>
<tr>
<th>Missing:</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>All present</td>
<td>371</td>
</tr>
<tr>
<td>Missing:</td>
<td></td>
</tr>
<tr>
<td>birthweight</td>
<td>152</td>
</tr>
<tr>
<td>mother's age</td>
<td>29</td>
</tr>
<tr>
<td>multiplicity</td>
<td>7</td>
</tr>
<tr>
<td>perinatal mortality</td>
<td>3</td>
</tr>
<tr>
<td>birthweight, mother's age</td>
<td>111</td>
</tr>
<tr>
<td>birthweight, multiplicity</td>
<td>1</td>
</tr>
<tr>
<td>birthweight, perinatal mortality</td>
<td>1</td>
</tr>
<tr>
<td>mother's age, multiplicity</td>
<td>1</td>
</tr>
<tr>
<td>mother's age, perinatal mortality</td>
<td>0</td>
</tr>
<tr>
<td>multiplicity, perinatal mortality</td>
<td>2</td>
</tr>
<tr>
<td>birthweight, mother's age, multiplicity</td>
<td>7</td>
</tr>
<tr>
<td>birthweight, mother's age, perinatal mortality</td>
<td>0</td>
</tr>
<tr>
<td>birthweight, multiplicity, perinatal mortality</td>
<td>4</td>
</tr>
<tr>
<td>mother's age, multiplicity, perinatal mortality</td>
<td>0</td>
</tr>
<tr>
<td>birthweight, mother's age, multiplicity, perinatal mortality</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>690</td>
</tr>
</tbody>
</table>

The remaining 319 records had at least one of these data items missing. A total of 277 records had birthweight missing; 152 records had just birthweight missing and a further 119 records had both mother's age and birthweight missing. A total of 149 records had mother's age missing, 120 of these records having one or more of the other data items missing too. One record had missing values for all four data items.

Of the other data items collected at NCAS notification, gestation had the highest proportion of not stated values; 276 (40 per cent) unlinked records did not have a stated gestation. Of these records, 230 also had no stated birthweight, of which 144 had neither birthweight nor mother's age stated.
4.4.3 Evaluation of agreement between selected data items collected by both sources

The third objective of this study was to evaluate agreement on linked records for data items recorded by both NCAS and birth registration. The data items analysed here are mother’s age, birthweight, multiplicity and perinatal mortality, as described in section 4.3.6. This section analyses just the NCAS records which were linked to a birth registration record.

Mother’s age
Table 4.10 compares the mother’s age at birth as given at NCAS notification, with that given at birth registration. Figure 4.2 shows these data for individual years of age. Each point plotted on Figure 4.2 may represent more than one record. For example, the point for pmage=14 and qmage=15 represents one record, compared with the point for pmage=32 and qmage=32 which represents over one thousand records.

Figure 4.2 Mother’s age as recorded by NCAS(pmage) and at birth registration(qmage), 1997-2000
<table>
<thead>
<tr>
<th>Age from birth record</th>
<th>Under 16</th>
<th>16-19</th>
<th>20-24</th>
<th>25-29</th>
<th>30-34</th>
<th>35-39</th>
<th>40-44</th>
<th>45+</th>
<th>Not stated</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 16</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>60</td>
</tr>
<tr>
<td>16-19</td>
<td>1</td>
<td>2124</td>
<td>17</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>45</td>
<td>2210</td>
</tr>
<tr>
<td>20-24</td>
<td>0</td>
<td>14</td>
<td>4700</td>
<td>33</td>
<td>20</td>
<td>6</td>
<td>7</td>
<td>1</td>
<td>128</td>
<td>4909</td>
</tr>
<tr>
<td>25-29</td>
<td>0</td>
<td>7</td>
<td>42</td>
<td>7012</td>
<td>32</td>
<td>18</td>
<td>3</td>
<td>0</td>
<td>199</td>
<td>7313</td>
</tr>
<tr>
<td>30-34</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>24</td>
<td>6752</td>
<td>52</td>
<td>3</td>
<td>1</td>
<td>169</td>
<td>7021</td>
</tr>
<tr>
<td>35-39</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>26</td>
<td>3265</td>
<td>13</td>
<td>1</td>
<td>85</td>
<td>3407</td>
</tr>
<tr>
<td>40-44</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>761</td>
<td>5</td>
<td>14</td>
<td>796</td>
</tr>
<tr>
<td>45+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>51</td>
<td>1</td>
<td>56</td>
</tr>
<tr>
<td>Total</td>
<td>59</td>
<td>2147</td>
<td>4785</td>
<td>7090</td>
<td>6845</td>
<td>3351</td>
<td>791</td>
<td>60</td>
<td>644</td>
<td>25772</td>
</tr>
</tbody>
</table>

Table 4.10 A comparison of mother's age reported on NCAS notifications with birth registration, 1997-2000
The agreement for mother’s exact age (in single years) between that given at NCAS notification and that given at birth registration was highly significant: (correlation coefficient= 0.98, p<0.001).

A separate analysis based on the difference (in complete years) between mother’s age as recorded at NCAS notification and that given at birth registration showed that for all individual ages except 45 and over, mothers’ ages agreed between the two sources for more than 96 per cent of NCAS records. Therefore, 96 per cent of linked NCAS records are represented on the X=Y diagonal on Figure 4.2. A further 1 per cent of linked NCAS records had an age difference not greater than one year.

**Birthweight**

Figure 4.3 shows NCAS and birth registration birthweight data for linked records where both records had a stated birthweight.

![Birthweight graph](image)

*Figure 4.3 birthweight as given at NCAS notification (pbwt) by that given at birth registration (qbwt), 1997-2000*
Table 4.11 Percentage distribution of birthweight as stated at NCAS notification according to that given at birth registration, linked records 1997-2000

<table>
<thead>
<tr>
<th>1997-2000 Births</th>
<th>&lt;500</th>
<th>500-999</th>
<th>1000-1499</th>
<th>1500-1999</th>
<th>2000-2499</th>
<th>2500-2999</th>
<th>3000-3499</th>
<th>3500-3999</th>
<th>4000-4499</th>
<th>4500+</th>
<th>NS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;500</td>
<td>82.0</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>0.5</td>
<td>139</td>
</tr>
<tr>
<td>500-999</td>
<td>6.5</td>
<td>94.3</td>
<td>0.4</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.9</td>
<td>2.1</td>
<td>459</td>
</tr>
<tr>
<td>1000-1499</td>
<td>0.0</td>
<td>1.1</td>
<td>95.4</td>
<td>1.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>3.0</td>
<td>2.5</td>
<td>547</td>
</tr>
<tr>
<td>1500-1999</td>
<td>0.0</td>
<td>0.0</td>
<td>1.6</td>
<td>95.8</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>4.3</td>
<td>4.1</td>
<td>928</td>
</tr>
<tr>
<td>2000-2499</td>
<td>1.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>96.7</td>
<td>1.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>8.6</td>
<td>9.0</td>
<td>2010</td>
</tr>
<tr>
<td>2500-2999</td>
<td>1.4</td>
<td>0.2</td>
<td>0.5</td>
<td>0.4</td>
<td>1.5</td>
<td>97.6</td>
<td>0.3</td>
<td>0.4</td>
<td>0.1</td>
<td>19.4</td>
<td>20.1</td>
<td>4524</td>
</tr>
<tr>
<td>3000-3499</td>
<td>1.4</td>
<td>0.7</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.5</td>
<td>98.6</td>
<td>1.5</td>
<td>0.2</td>
<td>31.2</td>
<td>31.6</td>
<td>7032</td>
</tr>
<tr>
<td>3500-3999</td>
<td>2.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.5</td>
<td>0.6</td>
<td>97.6</td>
<td>0.7</td>
<td>19.4</td>
<td>21.1</td>
<td>4748</td>
</tr>
<tr>
<td>4000-4499</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>98.3</td>
<td>20.9</td>
<td>21.1</td>
<td>1622</td>
</tr>
<tr>
<td>4500+</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>97.7</td>
<td>1.2</td>
<td>310</td>
</tr>
<tr>
<td>NS</td>
<td>4.3</td>
<td>2.4</td>
<td>0.7</td>
<td>0.3</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>1.0</td>
<td>0.4</td>
<td>3453</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>25772</td>
</tr>
</tbody>
</table>

Table 4.11 Percentage distribution of birthweight as stated at NCAS notification according to that given at birth registration, linked records 1997-2000
Table 4.11 compares the frequency of linked records by birthweight for NCAS and birth registration records.

There was a very close agreement between the birthweight as given at NCAS notification and that given at birth registration (correlation coefficient =0.98, \(p<0.001\)). For 91 per cent of birth registration records, the birthweight matched exactly with that given at NCAS notification, and a further five per cent of birth records had a birthweight within 25 grams (either side) of that on the NCAS record. The difference between the average birthweight recorded on the NCAS record and that recorded on the birth record was 2.9 grams (95 per cent confidence interval 1.07, 4.78 grams), being heavier on average in the birth record. A t-test showed that the means were not statistically significantly different from each other.

<table>
<thead>
<tr>
<th>Birthweight range</th>
<th>NCAS birthweight distribution</th>
<th>NCAS % distribution</th>
<th>Birth records birthweight distribution</th>
<th>Birth records % distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1000</td>
<td>598</td>
<td>2.3</td>
<td>682</td>
<td>2.6</td>
</tr>
<tr>
<td>1000-1499</td>
<td>547</td>
<td>2.1</td>
<td>649</td>
<td>2.5</td>
</tr>
<tr>
<td>1500-1999</td>
<td>928</td>
<td>3.6</td>
<td>1062</td>
<td>4.1</td>
</tr>
<tr>
<td>2000-2499</td>
<td>2010</td>
<td>7.8</td>
<td>2314</td>
<td>9.0</td>
</tr>
<tr>
<td>2500-2999</td>
<td>4524</td>
<td>17.6</td>
<td>5170</td>
<td>20.1</td>
</tr>
<tr>
<td>3000-3499</td>
<td>7032</td>
<td>27.3</td>
<td>8143</td>
<td>31.6</td>
</tr>
<tr>
<td>3500-3999</td>
<td>4748</td>
<td>18.4</td>
<td>5441</td>
<td>21.1</td>
</tr>
<tr>
<td>4000-4499</td>
<td>1622</td>
<td>6.3</td>
<td>1846</td>
<td>7.2</td>
</tr>
<tr>
<td>4500+</td>
<td>310</td>
<td>1.2</td>
<td>355</td>
<td>1.4</td>
</tr>
<tr>
<td>NS</td>
<td>3453</td>
<td>13.4</td>
<td>110</td>
<td>0.4</td>
</tr>
<tr>
<td>Total stated</td>
<td>22319</td>
<td>86.6</td>
<td>25662</td>
<td>99.6</td>
</tr>
<tr>
<td>Total</td>
<td>25772</td>
<td>100</td>
<td>25772</td>
<td>100</td>
</tr>
</tbody>
</table>

\(\chi^2=923.7\) on 9 d.f \(p<0.01\)

Table 4.12 Birthweight distributions of NCAS and birth records, 1997-2000

Table 4.12 compares the NCAS birthweight distribution with that of birth registration records. The two birthweight distributions are significantly different (\(p<0.01\)). Thirteen per cent of linked NCAS records did not have a stated
birthweight, however, compared with less than one per cent of birth records. Given the large difference in the proportion of missing values for birthweight, a subsequent analysis excluded not stated birthweight values before calculating the distributions by birthweight. In this second analysis the two distributions were not significantly different (p>0.5).

Multiplicity
Table 4.13 compares the stated multiplicity between the birth and the NCAS records. Multiplicity was not used as a matching variable. All birth registrations had multiplicity status recorded as this is a compulsory field at birth registration.

<table>
<thead>
<tr>
<th>As on birth record</th>
<th>As on NCAS record</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Singleton</td>
</tr>
<tr>
<td>Singleton</td>
<td>24493</td>
</tr>
<tr>
<td>Multiple</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td>24613</td>
</tr>
<tr>
<td>Percentage</td>
<td></td>
</tr>
<tr>
<td>Singleton</td>
<td>99.5</td>
</tr>
<tr>
<td>Multiple</td>
<td>0.5</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4.13 Singleton/ Multiplicity status on NCAS record compared with birth registration record, 1997-2000

Ninety four per cent of all NCAS records declared as one of a multiple birth were recorded as being part of a multiple birth at birth registration. However, 14 per cent of records registered at birth as one of a multiple birth had not been recorded as such at NCAS notification (shown in bold). Hence there was a systematic bias in the NCAS data towards recording more babies as singletons.
Vital status

Table 4.14 compares whether the child was recorded as being live or stillborn at birth registration and at NCAS notification. All birth registrations had a stated vital status as this is a compulsory field at birth registration.

<table>
<thead>
<tr>
<th>As on birth record</th>
<th>Not Stated</th>
<th>Live</th>
<th>Early neonatal death</th>
<th>Stillbirth</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live birth</td>
<td>60</td>
<td>23853</td>
<td>136</td>
<td>11</td>
<td>24060</td>
</tr>
<tr>
<td>Early neonatal death</td>
<td>4</td>
<td>140</td>
<td>733</td>
<td>15</td>
<td>892</td>
</tr>
<tr>
<td>Stillbirth</td>
<td>2</td>
<td>13</td>
<td>16</td>
<td>789</td>
<td>820</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>24006</td>
<td>885</td>
<td>815</td>
<td>25772</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>As on NCAS record</th>
<th>Live</th>
<th>Early neonatal death</th>
<th>Stillbirth</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>99.4</td>
<td>15.4</td>
<td>1.3</td>
<td>89.9</td>
</tr>
<tr>
<td>Early neonatal death</td>
<td>0.5</td>
<td>82.8</td>
<td>1.8</td>
<td>6.9</td>
</tr>
<tr>
<td>Stillbirth</td>
<td>0.1</td>
<td>1.8</td>
<td>96.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 4.14 Vital status as recorded at NCAS notification compared with that given at birth registration, 1997-2000

*Numbers of live births exclude live births dying in the first week of life (early neonatal deaths)*

There was generally good agreement between the two sources, with 99 per cent of NCAS live births being recorded as live births at birth registration, and 97 per cent of the records recorded as stillbirths at NCAS notification also recorded as such at birth registration. Only 83 per cent of the first week deaths notified to NCAS were recorded as such on the death registration linked record. However, of the 140 live births recorded as first week deaths by NCAS, a further 113 (81 per cent) were recorded through the registration records as having died after the first week of life. Infant mortality was used as a matching field in stage 4 of the algorithm.
4.4.4 Illustrative analyses enabled through linkage

4.4.4.1 Risk factors

Table 4.15 shows notification rates for specific anomalies according to the mother's country of birth.

Total notification rates were highest for mothers born in Asia and lowest for those born in Australia, New Zealand and Canada. Using the Poisson model to compare the notifications rates for mothers born in the other countries presented here with those for mothers born in the UK, gave significantly lower rates for total notification rates for the rest of Europe (excluding the UK), for Australia, Canada and New Zealand and for the rest of Africa (Africa excluding East Africa). The rest of Europe also had significantly lower rates for cleft lip and palate, cardiovascular anomalies and urogenital anomalies. Australia, Canada and New Zealand had significant lower rates for CNS and urogenital anomalies. The rest of Africa had significantly lower rates for cleft lip and palate, cardiovascular anomalies, urogenital anomalies and limb reduction defects. Babies with mothers born in the Caribbean had significantly lower rates of urogenital anomalies.

The New Commonwealth and Asia had significantly higher than expected total notification rates. Both had significantly higher rates for CNS anomalies with the notification rate for NTDs in Asia being more than double that for mothers born in the UK. Asia also had a significantly higher rate for eye anomalies. Both country groups, however, also had significantly lower than expected rates for abdominal wall defects, and the New Commonwealth had a significantly lower rate for limb reduction defects.
<table>
<thead>
<tr>
<th>Region</th>
<th>Total</th>
<th>Central nervous system</th>
<th>Eye</th>
<th>Cleft lip and palate</th>
<th>Cardiovascular</th>
<th>Respiratory</th>
<th>Urogenital</th>
<th>limb reduction defects</th>
<th>Abdominal wall</th>
<th>Down syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>131.1</td>
<td>4.3</td>
<td>1.7</td>
<td>9.4</td>
<td>16.2</td>
<td>1.5</td>
<td>22.0</td>
<td>3.0</td>
<td>3.9</td>
<td>6.4</td>
</tr>
<tr>
<td>Rest of Europe</td>
<td>97.1**</td>
<td>4.5</td>
<td>0.9</td>
<td>6.3**</td>
<td>10.1**</td>
<td>1.1</td>
<td>16.9**</td>
<td>3.0</td>
<td>3.2</td>
<td>6.3</td>
</tr>
<tr>
<td>Australia, Canada and New Zealand</td>
<td>89.0**</td>
<td>0.0**</td>
<td>0.7</td>
<td>5.0</td>
<td>14.4</td>
<td>0.7</td>
<td>13.6**</td>
<td>2.9</td>
<td>5.0</td>
<td>4.3</td>
</tr>
<tr>
<td>New Commonwealth</td>
<td>139.2*</td>
<td>8.0*</td>
<td>2.0</td>
<td>9.4</td>
<td>16.1</td>
<td>1.7</td>
<td>20.2</td>
<td>2.0**</td>
<td>2.6**</td>
<td>6.6</td>
</tr>
<tr>
<td>Asia</td>
<td>157.0*</td>
<td>10.0*</td>
<td>2.7*</td>
<td>11.2</td>
<td>18.5</td>
<td>2.2</td>
<td>22.1</td>
<td>2.7</td>
<td>2.6**</td>
<td>6.4</td>
</tr>
<tr>
<td>East Africa</td>
<td>125.4</td>
<td>6.9</td>
<td>0.6</td>
<td>7.4</td>
<td>12.0</td>
<td>1.1</td>
<td>24.6</td>
<td>0.6</td>
<td>1.7</td>
<td>7.4</td>
</tr>
<tr>
<td>Rest of Africa</td>
<td>101.9**</td>
<td>5.4</td>
<td>1.6</td>
<td>6.0**</td>
<td>11.1**</td>
<td>0.9</td>
<td>13.9**</td>
<td>0.9**</td>
<td>2.8</td>
<td>6.6</td>
</tr>
<tr>
<td>Caribbean</td>
<td>135.3</td>
<td>2.9</td>
<td>1.0</td>
<td>6.7</td>
<td>11.4</td>
<td>1.9</td>
<td>9.5**</td>
<td>1.9</td>
<td>6.7</td>
<td>9.5</td>
</tr>
<tr>
<td>Total</td>
<td>129.3</td>
<td>4.6</td>
<td>1.7</td>
<td>9.2</td>
<td>15.8</td>
<td>1.5</td>
<td>21.5</td>
<td>2.9</td>
<td>3.7</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Table 4.15 Condition specific notification rates by mother's country of birth per 10,000 live and stillbirths, 1997-2000
Table excludes countries with fewer than 20 notifications in 1997-2000.
Rates for individual conditions based on less than 20 cases have been presented in italics.

*signifies significantly higher rates using the Poisson model

**signifies significantly lower rates using the Poisson model
4.4.4.2 Survival

Figure 4.4 shows survival to 48 months according to type of congenital anomaly for the NCAS notifications that were linked to birth records.

![Graph showing survival to age 4 years of babies born with selected conditions](image)

Figure 4.4 Survival to age 4 years of babies born with selected conditions

Just over half of all children born with central nervous system (CNS) defects survived to their 4th birthday, compared with 91 per cent of children with limb defects. Of the babies born with anencephaly almost 100 per cent were recorded as having died, half of these having been stillborn, 32 per cent dying on the first day after live birth and the remainder dying during the first week of life. In contrast, 27 per cent of babies with spina bifida died by 4 years of age, 15 per cent were stillborn, and 5 per cent died on the first day of life. Eighteen per cent of babies with eye anomalies died by 4 years of age, compared with 8 per cent of babies with cleft lip and/ or cleft palate, 13 per cent of those with Down syndrome, 25 per cent of those with cardiovascular anomalies and of those with abdominal wall defects, and 49 per cent of those with respiratory anomalies.
Data were also tested on the survival of babies with multiple anomalies. Figure 4.5 shows Kaplan Meier survival curves for babies notified with only one anomaly, and for those notified with two or more anomalies.

![Kaplan-Meier survival estimates, by otheranom](image)

**Figure 4.5** Kaplan Meier survival curves for NCAS linked records with one anomaly notified (otheranom=0) compared with those with two or more (otheranom=1), 1997-2000

At one month of age 97 per cent of babies notified to NCAS with one anomaly were still alive, compared to 92 per cent of those with two or more notified anomalies. Using a log-rank test for equality of the survivor functions, statistically more deaths occurred in the group with two or more notified anomalies than would be expected ($\chi^2 = 658.93$, $p<0.001$).

Figure 4.6 shows the linked data analysed separately for anencephaly, Down syndrome and the remainder of conditions, for survival to 4 years of age according to whether they had one, or two or more anomalies.
Figure 4.6 Kaplan Meier survival curves for NCAS linked records with anencephaly, Down syndrome and the remainder of notifications, by whether they had one anomaly notified or two or more, 1997-2000

Survival was so low for anencephaly that having one or more additional anomalies made no difference to the survival rate. For Down syndrome and the remainder of the notifications, however, having two or more notified anomalies reduced survival by a further 20 per cent.
4.5 Discussion

The algorithm linked over 96 per cent of all congenital anomaly records in 1997 to 2000. Unlinked records were overrepresented in low birthweight babies and in babies with mothers without a stated age. For linked records, data on birthweight, mother’s age and multiplicity were substantially more complete on birth registration information. These results are important for interpretation of bias in subgroups when using linked records for aetiological and outcome research. Record linkage also allowed analyses for additional risk factors and for survival of children with congenital anomalies.

4.5.1 Feasibility of record linkage

The results showed that an algorithm using deterministic linkage could be used to link NCAS and birth records. The final linkage rates, of 96.2 to 98.5 per cent linkage, met that of 98.5 per cent achieved by the Western Australia population database (Stanley 1994). The first pass of the Western Australian data required no operator intervention and about 80 per cent of all records linked during this first stage, slightly higher than the 76 to 79 per cent of records in the study reported here. One criterion for the success of the feasibility study was the final percentage of NCAS records successfully linked. There is a lack of other published research to guide the interpretation of linkage rates. Caution needs to be exercised, however, if a high proportion of records remain unlinked as this might reflect biases in the matched dataset such that any results from the new analyses would be considered unreliable.

The latter stages of the algorithm generated a number of possible matches that were resolved manually by examining other fields on the two or more potentially matching records, but increased the final linkage rate by less than one per cent. The disadvantage of this manual work was the cost as it was labour intensive. Rules developed during this manual phase, however, provided potential algorithm enhancements that could subsequently be developed to improve the proportion of automatically linked records in the future.
In principle every NCAS record should have a birth registration record with which to link unless the child was born outside England and Wales. Therefore every unlinked NCAS record must either relate to a child born outside England and Wales or there must be an error in one or more of the variables used for linkage in the NCAS or the birth registration record.

4.5.2 Comparison of linked records with the NCAS database

The analyses presented in section 4.4.2 showed differences between the linked records, the unlinked records and the NCAS database for year, region, mother’s age and perinatal mortality. Given these differences, it is possible that there could be systematic differences by other variables not analysed here. A multiple regression model could analyse the interaction between different data items, but this was outside the scope of the work presented here.

The stepwise increase between 1998 and 1999 in the proportion of records unlinked was in part due to the introduction of electronic data exchange with the four regional registers. For Trent and Mersey linkage was made more difficult because birthweight was mostly missing from their records and this was a linking variable for stages 3 and later in the algorithm. For North Thames West, the first three letters of first and family name were usually not recorded for reasons of confidentiality which again made linkage more difficult. The fact that there seems to be an association between a higher level of unlinked records for the areas engaged in electronic data exchange is counter-intuitive so requires further investigation. This additional work was outside the scope of this study.

The linked records have a significantly different distribution according to mother’s age compared with the total NCAS database. The smallest proportion of linked records, 81 per cent, was for those records where the mother’s age was not stated. This is not surprising since mother’s age is a primary linking data item. Therefore, NCAS records without a stated mother’s age would be less likely to find their corresponding birth record. As a result, the linked records underascertain records without a stated mother’s age. In the absence of any information about the distribution of these records by mother’s age, it is not
possible to estimate the effect of excluding these records from subsequent analysis.

Perinatal mortality status was also significantly different between all NCAS records and unlinked records, having been mainly influenced by the large number of unlinked NCAS stillbirth records. A further analysis of the data showed that of the 81 unlinked stillbirth records, 35 had a recorded gestation of less than 24 weeks, the legal criterion for registering a baby born dead as a stillbirth. Hence these NCAS notifications related to deliveries which are not legally registrable and therefore have no stillbirth record with which to link. A further 13 unlinked NCAS stillbirths did not have a stated gestation so some of these notifications may have also related to deliveries at less than 24 weeks gestation. However, 8 of the 896 NCAS stillbirths which did link were recorded on the NCAS notification as having a gestation of less than 24 weeks, despite their linking stillbirth record having a gestation of 24 weeks or above, suggesting that the information recorded at NCAS notification was incorrect.

Although the distribution of linked records according to birthweight was not significantly different to that of all records, any analysis of linked records would underestimate the number of babies with congenital anomalies weighing less than 1000 grams by 11 per cent, and those without a stated birthweight by 7 per cent. This is a major weakness of the linked data, since these data include a lower proportion of babies weighing less than 1000 grams, who are known to have much higher mortality rates than heavier babies. Therefore any analyses about outcomes and survival based solely on linked records would have excluded many of the babies most at risk.

Of the 690 unlinked NCAS records, 54 per cent had a stated value for each of birthweight, mother's age, multiplicity and perinatal mortality. Further work would be important to understand why these records did not link, since any solutions could help inform future linkage studies. Since the remaining records had at least one of these data items missing this would affect the ability to link these records, in particular the records that had mother's age missing.
Another weakness of the analyses presented in this chapter is that they only included a selection of the available data items to compare linked and unlinked records. It is possible that there are systematic biases in other data items not analysed here (for example, father’s age, or social class) which would need to be identified before using these data items routinely in analyses.

4.5.3 Evaluation of agreement between data items collected by both sources

There was close agreement between mother’s age, birthweight and perinatal mortality on both the NCAS and the birth registration records. This is not surprising as these are all variables used in the matching algorithm. Mother’s date of birth was one of the matching criteria in the first and second stage of the algorithm, by which stage 85 per cent of records had matched.

There was a systematic bias in the NCAS data towards recording more babies as singletons. Multiplicity was not used as a matching variable. Fourteen per cent of linked records registered at birth as one of a multiple birth had not been recorded as such at congenital anomaly notification. This is likely to be due to those completing the NCAS notification being unaware that the child was one of a multiple birth.

For other linked records, babies recorded as one of a multiple birth by NCAS were recorded as a singleton at birth notification. This could have been a result of linking the wrong record, or that the delivery was at less than 24 weeks’ gestation and one or more babies were born dead. Babies born dead at less than 24 weeks’ gestation are not legally registrable so the delivery would have been recorded at birth registration as a singleton birth. The Western Australia Register also faced the problem of two conflicting multiplicities, where a few cases with miscoded multiplicity were found each year. These were corrected during the linkage process as they were able to trace the true multiplicity through case notes (Stanley 1994).
Using the multiplicity as recorded at birth registration for linked NCAS records when calculating absolute risks would increase the number of babies recorded as being part of a multiple birth, and would mean that birth registration data were used for both numerator and denominator, hence removing numerator denominator bias. In general, anomalies are more common in multiple births, than in singletons, particularly for monozygotic twinning (Little 1988, James 1995). Analyses of NCAS data have shown this excess, but not as great as shown in other studies (Doyle 1991, Mastroiacovo 1999). Therefore, if the true multiplicity was that recorded at birth registration, then rates of anomalies among multiple births previously published by NCAS were likely to have been underestimated. More marked differences would be expected between the prevalence of congenital anomalies in singleton and multiple births if the multiplicity recorded at birth registration was used.

Vital status was only used as a linking variable in stage 4 of the algorithm, by which stage over 93 per cent of the records had already matched. Three per cent of NCAS stillbirths were recorded as live born at birth registration, with 38 per cent of these still surviving. This could be a result of either linking together records which related to different children, to coding errors, or to recording incorrect information by one source. Intuitively birth registration would be more likely to correctly record the true vital status as registration of a stillbirth or infant deaths requires the appropriate medical certificate of death completed by a doctor.

The main problem in using linked records for analysis is sensitivity because not every record was linked, and hence any analysis of prevalence using linked records would underestimate the true rates. If the data are required for absolute risk estimates then a low linkage rate is a problem. For etiological purposes, when calculating relative risks between different subgroups, however, it is not essential to have all records, provided those which were linked and included in the analysis are an unbiased sample of all NCAS records.

Completeness of data items is important as missing or omitted data often reflect selection bias. Rates of less than one per cent missing data are generally
considered trivial, one to five per cent manageable, but five to 15 per cent missing data require sophisticated methods to handle, and more than 15 per cent severely impacts on any kind of interpretation (McDermid 1999). Missing data tend to occur more in high risk situations or in some non-random fashion (McCarthy 1980, Halliday 1992). It is also important for the linked data to be as complete as possible so that analyses of subgroups maximise the numbers in cells of tables and therefore maximise the possibility of picking up any differentials between the subgroups. Therefore if it is possible to reduce the levels of missing data, bias can be reduced.

Where a data item was collected by both sources, completeness would be improved and biases reduced by using the mother’s age, birthweight, multiplicity and perinatal mortality from the birth record in analyses in place of that recorded by the NCAS record. This would reduce levels of missing data from 13 per cent for each data item to less than 1 per cent. However, there were no ‘not stated’ ages on the birth records as ages were imputed by ONS if these were not stated on the birth registration record. If these cases linked, therefore, the mother’s age on the birth record would not be based on the true date of birth. Since from the linked file only 0.5 per cent of birth records had an age imputed, this would have a smaller impact on the overall distribution of mother’s age than the 13 per cent of NCAS records without a stated mother’s age.

For multiplicity and perinatal mortality, values were missing from the NCAS record for 1.3 per cent and 0.3 per cent respectively compared with no missing values on the birth record. So in all cases the birth registration data was more complete which would enable more precise risk estimates using these data items.

The analyses presented here also showed that other groups were under-represented in the linked records, particularly babies weighing less than 1000 grams at birth. If the under-represented group is one at higher risk than other groups then any relative risk will understate the true relative risk, and
conversely the relative risk for any under-represented group at lower risk than other groups will overestimate the true relative risk.

4.5.4 Illustrative analyses enabled through linkage

Country of birth
The new findings presented showed significantly different rates for some specific congenital anomalies by mother’s country of birth. These results are important to show the possibilities for new analyses enabled through linkage. They are also important to help seek explanations for the differences between different groups of the population. These might suggest targeted public health interventions to reduce differentials.

The country of birth of the parents had traditionally been used as a proxy for ethnic origin, although as many mothers born in Britain were of a black ethnic origin it was increasingly becoming less reliable as such. These crude associations by mother’s country of birth, however, can be confounded. Possible confounders include mother’s age, parity and consanguinity. Data on mother’s age were available from the linked records, so one extension to this analysis would be to adjust the rates for maternal age, although this work was outside the scope of this study. Data on the other confounders are either not collected or are unreliable measures.

Rates for Down syndrome for mothers born in Africa excluding East Africa were similar to those for mothers born in the UK. This is comparable with findings in African births. In the past Down syndrome was thought to be rare in Africa. The first reliable analysis of Down syndrome was a retrospective study in Nigeria which concluded that the incidence of Down syndrome in African births was similar to other populations throughout the world (Adeyokunnu 1982). Previous underascertainment in African populations was thought to be due to problems in clinically diagnosing Down syndrome in African newborns since the clinical craniofacial features of African Down syndrome newborns are more similar to the craniofacial features of normal newborns than is the case for white babies (Christianson 1996).
Rates of some anomalies are higher in babies of Asian born women. One explanation has been the increased rates of consanguinity in Pakistani marriages resulting in higher rates of recessive inherited disorders (Leck 1995). Several studies have reported raised risks of anomalies in children from consanguineous families. A study of armed forces families in Pakistan found that the overall prevalence of congenital malformations in children of related parents was 4.0 per cent compared with 2.6 per cent for non-related parents (Hashmi 1997). A study in the West Midlands showed that Asian children had a significantly higher proportion of congenital heart disease that required hospital admissions (Sadiq 1995). Another population study in North-Eastern France reported that 1.2 per cent of children with congenital anomalies had consanguineous parents compared with 0.3 per cent of children without anomalies. This was statistically significant (Stoll 1999). A trend, not reaching statistical significance, was also seen in the prevalence of congenital anomalies in the Irish traveller community (Barry 1997), being higher in the offspring of first cousin marriages (6.5 per cent) than those born to parents without a blood relationship (3.8 per cent).

Survival

Generating survival data for specific anomalies from national surveillance data would enable the survival analyses presented here to be extended each year with an additional year’s data extending the age range by one year. This would meet a current gap in knowledge and help those advising parents on the prognosis for survival of babies with congenital anomalies, especially for those children with very rare conditions for which no good data currently exist nationally.

Survival curves for Down syndrome were presented in this chapter. From studies elsewhere, the life expectancy of children with Down syndrome has improved dramatically over the past 50 years, largely due to a large increase in the proportion of children surviving the first year of life. Three international studies have shown similar survival rates. In Jerusalem in 1996, 91 per cent of children with Down syndrome survived to the end of the first year; compared
with 54 per cent in 1979. In 1996 90 per cent survived to the age of 14, compared to 42 per cent in 1979 (Merrick 2000). In Western Australia, for babies with Down syndrome born between 1980 and 1996, survival to 1 year was 91 per cent, and 85 per cent of those eligible survived until the age of 10 years (Leonard 2000). Although survival in those with heart disease showed improvement over the period studied, overall this was still a strong predictor of mortality. A study in Dublin found survival rates of 88 per cent at one year and 82 per cent at 10 years (Hayes 1997). The data presented here show a similar survival rate of 89 per cent to one year of age, but survival beyond age four is not yet available.

Almost 100 per cent of the babies born with anencephaly were recorded on the linked records as having died. It is assumed that the remaining cases which appeared to still be alive at the time of analysis were errors, since anencephaly is incompatible with life. This highlights the problems of specificity for the linked records. Either the NCAS records were not true cases of anencephaly, or the NCAS records may have linked with incorrect birth records. This problem may also have occurred for other congenital anomalies but they are not as straightforward to identify.

4.5.5 Implications for the future

A high rate of record linkage was achieved by the algorithm presented here, but it did require several iterations and some manual linkage. Linkage would be considerably enhanced by the inclusion of a single unique identifier on both the NCAS record and the birth record. The algorithm developed for this study would ideally have used NHS number as the key linkage variable. However, the NHS number was not generally recorded on notifications to NCAS, although there is a box on the SD56 notification form to do so.

This problem should be redressed on 29 October 2002 when it was planned that midwives would allocate the NHS number at birth rather than, as prior to October 2002, by Registrars at birth registration (NHS Information Authority
In the future the NHS number should be included on all NCAS notifications of live births. Currently stillbirths are not given an NHS number and it was not known whether they would be given a unique identifier in the future.

In the future the NHS number would be the primary linkage field for live births, with the algorithm developed here used for stillbirths, for subsequent matching stages if the NHS number was not stated, or if an NHS number was stated but a match with a birth record could not be found. If there is absolute compliance and recording of the NHS number, NCAS notifications for live births would only need the NHS number to identify the child, other data items to enable linkage, and the details of all congenital anomalies. All other information currently required at congenital anomaly notification can be directly obtained from birth records, for example multiplicity, and parents’ occupations. Given the high levels of consistency between NCAS and birth records for mother’s age and birthweight, information will not be lost by relying on one source alone, and the quality of data on multiplicity status would be improved. It would no longer be necessary to collect the child’s name at NCAS notification which would help to preserve patient confidentiality. Stillbirths might need more identifiable information if they were not allocated a unique identifier.

Occupation data are collected by both NCAS and at birth registration but were not analysed in this study. Occupation is important both for identifying the teratogenic effects of hazardous occupations and to derive social class to measure health inequalities. Occupation information for both parents is collected at birth registration for all babies (although not for the father if the child was born outside marriage and registered by the mother alone). Although occupation information for both parents is requested at NCAS notification, it is not recorded for over 40 per cent of NCAS records. This suggests that the occupation recorded at birth registration could replace that held by NCAS records.

Linkage of NCAS records to their birth records, and using existing links between birth and death information, has been shown here to provide new analyses of survival. With appropriate ethics and confidentiality safeguards, this linkage
would enable NCAS to inform local congenital anomaly registers of the death of children on their registers. This would help maintain the quality of the data held by local congenital anomaly registers.

An additional research question, although outside the scope of this study, was whether it would be possible to improve ascertainment of congenital anomalies through record linkage. Linkage of a child's death record, through the birth record to any NCAS record would enable an evaluation of children whose death certificate mentioned a congenital anomaly compared with whether the child had been notified as having one or more congenital anomalies.

For further linkage possibilities with other child health data, it would be also useful to have on the birth record the NHS numbers of both parents. This would enable analyses of generational effects, and also subsequent links with siblings, to measure recurrence risks within the same family of outcomes such as congenital anomalies and sudden infant death syndrome.
4.6 Conclusions

Record linkage of NCAS and birth records is feasible. Over 95 per cent linkage was achieved, although it was clear that the use of the NHS number would improve the linkage. Linkage should be part of routine analysis of national surveillance data to improve our understanding of risk factors for congenital anomalies, to improve the quality of congenital anomaly data, and also to help inform other record linkage projects in the future.

Linkage also has the advantage that it will enable NCAS to obtain from other databases additional information not previously collected by NCAS, without the need to go back to the notifiers to seek this information. In addition, however, several data items collected at notification to NCAS are also collected at birth registration. Therefore, for those variables where this feasibility study showed that the information collected at birth registration was of equal or better quality than the same data items collected at NCAS notification, it would be possible to remove that data item from the information collected at NCAS notification and derive it from linkage with birth records. In principle, in the future NCAS notification could be replaced by a fast notification of a child’s NHS number and details of the condition notified, leaving all currently collected demographic information to be derived from linkage with the birth record. This would reduce the burden on those notifying.

The philosophy of linking different data sources to maximise the information which can be gleaned from them is in line with current Government thinking. The Modernising Government publication ‘Adding it Up’ (Performance and Innovation Unit. 2000) promotes the linkage of records from different sources. Given the current demand for audit and understanding of rates of disease (Audit Commission 2003), record linkage can provide timely answers and form the basis of good epidemiological studies.
CHAPTER 5

Looking to the future
5.1 Summary

The purpose of this thesis has been to evaluate the contribution NCAS can make to the changing demands for congenital anomaly data, and the implications for its future role in disease monitoring systems nationally and internationally. The completeness of ascertainment and accuracy of NCAS data has been examined and different methods to improve ascertainment and the quality of the data evaluated so that NCAS can better meet its new purposes, whilst retaining the benefit of being low cost.

These demands require higher levels of ascertainment than previously possible by NCAS alone. Ascertainment by local registers and data transfer to NCAS offers a feasible and effective solution. There is also now scope to link with a much wider range of data from other record systems to improve understanding of the aetiology and outcome of congenital anomalies.

The purpose of this chapter is to discuss how these improvements can be incorporated into the system, other external influences on NCAS, and to present my vision for NCAS in the next decade building on the studies presented here.
5.2 Summary of thesis

5.2.1 Chapter 1 History and description of NCAS

Chapter 1 provided the background for the thesis. It described the original purpose of the National Congenital Anomaly System (NCAS) as a rapid system for identifying localised and national increases in notifications of major external anomalies. It also described the surveillance technique used to identify these increases and its outputs. It identified and described the major limitations to NCAS within the context of its evolution and external developments. Over three decades there have been evolving clinical, policy, research, ethical and legal demands on the system. These were summarised, together with an overview of the operational changes made to NCAS to help meet some of these new requirements.

5.2.2 Chapter 2 Estimation of ascertainment by NCAS

Chapter 2 reported on five studies undertaken by NCAS in collaboration with other researchers, which aimed to measure the extent of underascertainment of selected conditions notified to NCAS. This was possible through comparisons with data collected by other registers and research groups in response to policy or public health questions when NCAS data had been judged to be inadequate. It was not possible with the resources available to evaluate ascertainment for every condition included in NCAS, but analyses showed that ascertainment varied for different conditions depending on the ease of identification at birth, and the need for specialised clinical diagnosis or laboratory testing.

5.2.3 Chapter 3 Evaluation of NCAS receiving data from local registers

The purpose of Chapter 3 was to determine the effect on notification rates of ascertainment by four local registers and data transfer to NCAS, rather than direct reporting from Health Authorities (HAs). In this Chapter I discussed whether data transfer could be extended to other local and condition-specific
registers and whether it would be sustainable. I also discussed the impact of monitoring a step change in notification rates on surveillance alarms and feedback to local areas.

Collaboration with the local register in Wales and the three local registers in England demonstrated immediate benefits by resulting in statistically significant improvements in ascertainment. This approach for NCAS to obtain data, where possible, from local and condition specific registers is to be advocated. If electronic data transfer was extended to cover more of England, NCAS data would become more complete and therefore more sensitive to real changes in prevalence. National data would also be closer to true prevalence rates so better able to be used to inform Government policy, assist in geographical comparisons, and assist in medical research.

5.2.4 Chapter 4 Evaluation of linking NCAS and birth registration records

Chapter 4 evaluated the feasibility of linking NCAS notifications with birth registration records. This evaluation aimed to maximise the completeness of existing data, to evaluate the quality of data held by NCAS, to expand the information available for each notification by comparing data items that were collected by both data sources, and to obtain additional information about each notified baby. Appropriate linkage methodology was selected and tested.

The results of this study showed that linkage with birth registration records was feasible and with some manual intervention, over 95 per cent linkage was achievable. The completeness of some data items was measured and using values of the data item collected at birth registration was shown to provide more complete recording of these data items. New analyses were made possible through linkage. Analyses of congenital anomaly rates by mother’s country of birth, and survival of babies with selected anomalies were presented to illustrate the use of linked data in the evaluation of risk factors for congenital anomalies and for mortality rates. The results of this study are important to improve our understanding of risk factors for congenital anomalies, to improve the quality of NCAS data, and also to help inform other record linkage projects in the future.
5.3 The way forward

5.3.1 Scope for NCAS in the future

Studies in earlier chapters demonstrated the need for more complete national data on congenital anomalies with more information on aetiological factors and outcomes for rapid surveillance, clinical audit, research and policy planning. Both ascertainment and data quality could be improved by implementing local ascertainment by local congenital anomaly registers and data transfer to NCAS wherever these registers exist, and by an expansion of record linkage.

For the future, there will be a need for congenital anomaly data to meet new policy initiatives and public health, clinical or research demands. If NCAS is to fulfil these roles, it is essential that NCAS is flexible and easily able to change. Surveillance remains an important purpose of NCAS, and this function is improved through increased ascertainment making the analyses more sensitive to real changes in prevalence. The system needs to evolve to both be able to continue to provide effective surveillance and to be available for a wide range of other purposes including cluster analysis, and monitoring the effects of antenatal screening policies. It needs to react quickly to local and national concerns about increases (real or perceived) in prevalence rates of congenital anomalies. Many of the alarms in recent years have been generated from media coverage of suspected toxins (Bove 1995, Dolk 1998, Gilbert 1993, Paduano 1993), which were exposed in the media after a perceived lack of local concern from clinicians. In the past these increases were largely investigated using expensive retrospective separate studies (Busby 1998). With more complete data, NCAS would be readily available to investigate such concerns.

In addition, however, it is important that these uses do not conflict with the evolving standards on ethics and confidentiality. Recently, there has been a growing public expectation that personal privacy will be respected whilst collecting data to investigate public concerns. These issues on ethics and confidentiality are discussed in Section 5.5.3.
5.3.2 Data transfer by local registers

In Chapter 3 I explored moving from a single source to multi-source ascertainment through data transfer by local registers and showed that it improved ascertainment by NCAS. These improvements should continue as data transfer is extended to other local registers. NCAS records will become more useful as a research tool, and increasingly be used in place of separate ad hoc studies when a hypothesis needs to be evaluated. It therefore becomes even more important to ensure that these records remain of as high quality as possible.

Since the study described in Chapter 3 took place, the four local registers have continued to provide data in place of the local HAs. In addition, Wessex and Oxford congenital anomaly registers began data transfer on 1 January 2002. The Northern Region Congenital Anomaly Register has agreed to begin data transfer in January 2003. A new register being established in the South West of England also plans to provide data directly to NCAS. At that stage congenital anomalies in approximately half of all births in England and all births in Wales will be notified directly to NCAS from a local register. Future analyses will show whether similar improvements in ascertainment prove possible. Therefore the expansion of data transfer to other local registers is feasible and only limited by the lack of national coverage of England by local congenital anomaly registers. This lack of national coverage means that ascertainment varies geographically depending on the data provider, which limits the ability of NCAS to assist in a local cluster investigation if the area is not covered by a local register.

If local congenital anomaly registers were to cover all of England and Wales, one potential model would be for NCAS to act as a national co-ordinating centre for congenital anomaly notifications. This model is very similar to that used by cancer registries where ONS acts as the National Cancer Intelligence Centre, receiving data from all local cancer registries. The difference, however, is that all the local cancer registries are similarly funded and together provide national coverage. Universal coverage by local congenital anomaly registers would maximise ascertainment to NCAS, and ensure that data were available.
nationally to investigate any hypothesis, irrespective of condition or geographical location.

Advantages
The main advantage of data transfer by local registers is the improvement in ascertainment by NCAS. This could be extended to include terminations with the data transferred to NCAS. Local registers all collect information on congenital anomalies detected antenatally and on legal terminations for fetal handicap. The lack of abortion data in NCAS was identified as a problem for NCAS in Chapter 1. These data could be transferred to NCAS to supplement existing data on congenital anomalies in live and stillbirths.

Local registers also have the advantage of being able to confirm cases using links between their different notifiers and avoid duplicate notifications. For example, ultrasonographers may identify an anomaly which is not seen at birth. Therefore, the anomaly would not be notified to NCAS hence improving its specificity.

Another advantage of data transfer by local registers is that it minimises the impact of changes in health service boundaries on NCAS' surveillance and outputs. Throughout the history of NCAS there have been changes in health service boundaries. Some, such as in April 1974, were the result of a major reorganisation, but others take place each year, as two or more smaller areas combine to form a single health area. Each year, therefore NCAS has needed to recalculate new baseline levels for CUSUM analysis of the new areas, identify staff within the new area that would be prepared to accept responsibility for notification to NCAS, and amend computer programmes that produce regular outputs for local areas. Since the local registers receive notifications from a wide variety of clinicians and health service providers, these would be largely unaffected by mergers of areas, and notification to NCAS would continue from the same local register. NCAS would agree with the local register the most appropriate areas for surveillance using available geographies, and these could continue independently of health service boundaries.
Disadvantages
There are some disadvantages of data transfer by local registers, however. The main concern is its sustainability. Local registers have been shown to improve ascertainment by NCAS, but only 50 per cent of England is covered by a local congenital anomaly register. There have previously been other local congenital anomaly registers in England, such as those in Liverpool and in South East Thames, but these closed when their funding ceased. It is important for national coverage within England and Wales, that a mechanism exists to restore HA reporting to NCAS if the local register closes. Having lost the contact with the HA notifiers, however, this is likely to be a difficult task.

Data transfer by local registers also can lead to a drift in case definition, because in place of receiving a text description of anomalies, local registers tend to provide their own processed data. NCAS therefore loses its comparability of condition data which had always been centrally coded by its own trained coders.

Any major change to the system affects the ability to produce time trend data. Whenever another local register begins data transfer to NCAS this has been shown to have a significant effect on local notification rates. This in turn will also have a smaller effect on the national notification rate, which affects the ability to use NCAS data for time trend analyses. These need to therefore be stratified according to local register areas to try to isolate real changes from changes due to improved ascertainment on beginning data transfer.

Other options
Transfer of data by local registers is not the only option to improve ascertainment by NCAS. Other options include greater use of other sources of child health data. Recently there have been bigger changes than ever seen before in electronic handling of data. In the future a uniform system of computerised health data might simplify the identification and bringing together centrally of congenital anomaly data.
5.3.3  Linkage of NCAS with other data sources

Record linkage with NCAS and birth records was shown to be feasible in Chapter 4, and I demonstrated how the quality of data items can be improved and how additional data items can be derived without the need to burden those supplying the data with additional work. In addition, however, other useful information might emerge in the future which would help to test new hypotheses about the possible aetiology of congenital anomalies. Linkage of that information with NCAS records could enable these analyses. It is not necessary for NCAS to collect data items for which there is not a current research or policy need. For future research, it is possible to either link with other sources that do hold this information (for example, hospital records), or to go back and collect from clinicians additional information specific for that study. Linkage can be used to trace the natural history of the condition. It can also confirm case definition, that they are true cases, for example by linking NCAS notifications to results for a specific test required to diagnose the given condition.

Ideally, my ultimate vision is to link information from a wide range of different sources on child health to birth records for all children. This will not only allow us to derive additional risk factors for congenital anomalies and monitor outcomes, but also to describe pathways of child health, illness and subsequent survival for a range of different conditions for all children, and to improve ascertainment and data quality. This approach has been implemented successfully elsewhere. For example, in Denmark nine registers including the congenital anomaly register were linked using their unique individual identification number to create a comprehensive national individual based register, known as the Danish Prevention Register (Roed 1999), and used in analyses of associations between congenital anomalies and outcomes (Hansen 2000).

One possible source of data on child health to link with NCAS records is the Hospital Episode Statistics (HES). A link with HES records would make it possible to identify whether children have hospital admissions for congenital
anomalies. These would enable confirmation of cases already notified to NCAS, and also detection of new cases not previously notified. The feasibility of linking birth registration records with HES records was tested (Abrahams 2002). The method used to link records was based on that used in Chapter 4 to link NCAS records to birth registration records. Again, 95 per cent of HES records were matched with their corresponding birth registration records.

Further analysis of the stillbirth and death registration records already linked to birth records could provide additional notifications of congenital anomalies. For example, all babies with anencephaly would be expected to die at or very soon after birth so each NCAS record should have a corresponding stillbirth or neonatal death registration record. If such a record cannot be found, it is possible that the NCAS record is not a true case, so it should be returned to the original notifier for confirmation. In addition, all anomalies notified in stillbirth and death registration records should have been notified to NCAS, so these sources would provide additional confirmation or new notifications. Identifying additional notifications from these sources was outside the scope of the studies reported here, but is to be advocated.

Technological advances, however, are moving fast, so it would be also essential to constantly review the best way for these data to be collated nationally. For example, with current computing processing power, linkage with other data sources is now considerably easier and faster than with previous generations of computers.

Having shown that linkage between NCAS and birth registration is possible, and given the pre-existing links between birth and death records for children, Figure 5.1 shows the existing links between birth registration data and death registrations, stillbirths and NCAS records, together with the potential links with childhood cancer, HES and local child health data.
Advantages

The principal advantage of this approach to linkage is that it enables linkage for all children, not just those with congenital anomalies. For example, childhood cancer information could be linked to information about parental occupation collected at birth registration, and immunisation data recorded by the local child health system, to test hypotheses on risk factors for childhood cancers.
The advantages of population-based record linkage include the avoidance of selection bias which can occur in epidemiological case-control and cohort studies, and the avoidance of recall bias, as some data are collected before the outcome or in ignorance of the outcome. Such data can address many different hypotheses and are therefore much more cost effective than conducting several cohort studies. Other advantages of such linked files include large sample sizes leading to generalisability of the results, and rapidity of analyses as the data are already assembled. Also such studies are not intrusive on the population studied.

In addition, the feasibility study reported in Chapter 4 showed that for some data items the information collected at birth registration was of equal or better quality than the same data items collected at NCAS notification. Therefore, unless that data item is required for the linkage algorithm, it would be possible to remove that data item from the information collected at NCAS notification and derive it directly from linkage with birth records. Similarly, linkage with other data sources also has the advantage that it will enable NCAS to obtain additional information not previously collected by NCAS, without the need to go back to the notifiers to seek this information.

Disadvantages
Specificity is a problem when linking data from several sources, some of which may not have been collected for health purposes. Ascertainment may be biased by notification of anomalies which are not subsequently confirmed as anomalies. Administrative sources (such as birth registration data) are often used for health research, when health outcomes were not the reason for collecting the data. It is unlikely that an administrative data source will have sufficient detail for a specific health purpose if that purpose had not been identified in advance of data collection.
5.4 Incentives for the future scope of NCAS

This thesis has made a number of proposals to enable NCAS to be more reactive to policy, clinical, public health and research needs. These proposals, however, are not evolving in a vacuum; they meet initiatives within the wider Government context which have been designed to improve data flows and the audit of health service practices. The Government has declared children's health to be important. The only way to identify whether new initiatives are successful in improving children's health is to be able to monitor the outcomes, for which good quality data is essential.

5.4.1 The new National Service Framework (NSF)

The National Service Framework (NSF) was announced by Alan Milburn on 28 February 2001

‘... underpinning this national crusade to improve children’s health there will be a new Service Framework setting in place new national standards for children’s services.’

This NSF aims to take the key values from the NHS Plan of modernisation, breaking down professional boundaries and partnership between agencies, and apply them to services for children. It aims to cover the services, including maternity and child and adolescent mental health, which help children to grow into healthy adults.

5.4.2 NHS Information Strategy

The NHS Executive’s information Strategy ‘Information for Health’ (1998) indicated that information strategies would be developed to support each NSF. To this end the Department of Health is to develop a comprehensive Information Strategy for Child Health to underpin the development of the NSF for children’s services. The Information Strategy will be developed at the same time as the NSF, commencing in the autumn of 2001 and completed by the time of the NSF publication in 2003. One principle of this development of the strategy is to set the foundations for the development of an information culture to enable
further improvements in the quality of data capture and information provision. The strategy aims to address in the wider child health field, many of the same problems identified here with NCAS data. There are concerns about data variability, the lack of routine national data for certain conditions, variable data quality, problems from manually collecting data, and inconsistency in coding. As the NSF will include support for children with disabilities and special needs, there will be a priority for good quality congenital anomaly data.

5.4.3 The NHS numbers for babies project

The NHS number is at the heart of electronic health records proposed in the Government's 'information for health' strategy. This single unique identifier will enable a range of disparate information about an individual to be brought together into a single electronic health record. Although each individual has been allocated an NHS number since the inception of the NHS in 1948, the NHS number is rarely used in health records. In particular, up until 29 October 2002 babies were not allocated an NHS number until their civil birth registration. This could take place at any time until six weeks of age.

The NHS numbers for babies project identified solutions to this problem. It was planned that from October 2002 NHS numbers would be allocated electronically at birth. The midwife notifying the birth would request the NHS number from a Central Issue System. This would be available for use immediately within the hospital if the birth took place there. The child's details would also be automatically transferred to the local child health department.

This initiative should help the implementation of my proposed routine record linkage of NCAS data to birth records, as described in Chapter 4. NCAS records should have NHS number recorded, which will enable a simplified linkage of NCAS records, with the associated improvements in the quality of information held for each child with notified congenital anomalies.
However, the implementation of the new NHS number has created another problem for NCAS. Preceding the introduction of the NHS number at birth which relied on electronic allocation of the new NHS number, each area needed an electronic maternity system. Therefore, all areas that still relied on a paper system needed to implement a new maternity system, designed by the NHS Information Authority, by the end of July 2002. This system includes a question on whether the child had a congenital anomaly, but omitted the previous text box to describe any anomalies. As a result, local notifiers who relied on birth notification forms as their primary source of information for congenital anomalies would need to contact clinicians or midwives at the place of delivery to confirm details. This additional work may affect data quality if those notifying do not have the resources for this additional level of work.

5.4.4 E-government agenda

In addition, the e-government agenda aims to ensure that, by 2005, all transactions with government will be carried out electronically wherever possible and that records will be held and maintained in electronic form and will be accessible by a range of different methods. Again, this initiative supports my proposals for ascertainment by local registers and electronic data transfer to NCAS described Chapter 3.

The policy of the current Government is that information should be collected just once from the public, and then passed throughout Government to all agencies that need to receive such information. The philosophy of linking different data sources to maximise the information that can be gleaned from them is in line with current Government thinking. The Modernising Government publication ‘Adding it Up’ (Performance and Innovation Unit 2000) promotes the linkage of records from different sources. Given the current demand for audit and understanding of rates of disease, record linkage can provide timely answers and form the basis of good epidemiological studies.
5.5 Barriers to future expansion of data transfer to NCAS and record linkage

Three barriers potentially could hinder the future scope of NCAS. The first is funding, the second is the untested work identified in this thesis but which was outside the scope of the studies presented here, and the third is the impact of external initiatives such as the legislation concerning ethics and informed consent.

5.5.1 Lack of resources

The improvements identified in this thesis cannot happen without appropriate funding. Currently local registers are funded in different and often precarious ways. BINOCAR has a role to be an advocate for local registers and complete coverage of England.

5.5.2 Untested work outside scope of this thesis

Some ideas presented throughout this thesis have not been tested because they were not resourced or were outside the scope of the studies presented here. For example, data exchange with condition specific registers or the link to childhood cancer register. These would need to be tested fully to ensure that the proposals could achieve the benefits suggested here.

5.5.3 Ethics and confidentiality

There are competing interests of public health and privacy needs. Whilst this thesis has been concerned with the collection of good quality data on children with congenital anomalies for a wide range of public health purposes, in recent years there has been a growing concern about collecting such information without informed consent. Data provided to NCAS do not have the explicit consent of the parents, and the records currently include identifiable information; the first 3 initials of the child’s first and family name, their date of birth and their postcode to enable linkage with the child’s birth records.
The Government has made it clear that informed consent is the fundamental principle governing the use of identifiable patient information by any part of the NHS or research community. The NHS plan aims to develop a patient centred service where information is shared between all those involved in delivering or developing care. This gives an opportunity to make the best possible use of patient information. However, legally to preserve patient confidentiality, researchers need to use either identifiable information with informed consent or anonymised data.

Whilst these intentions are clearly desirable and in the patients’ best interests, evidence from other congenital anomaly registers has shown that requiring informed consent can severely reduce ascertainment. The Congenital anomaly register in the Northern Netherlands has evaluated the impact of the requirement for informed consent. Congenital anomaly registration began in the Northern Netherlands in 1981 (Cornel 1992). Dutch legislation has had an impact on notification since its inception. In 1981 a regulation to a Dutch Act described the rights of people whose details were recorded on a register. The Bill of Medical Treatment became law in the Netherlands in 1990. Informed consent from the parents is therefore necessary whenever a physician sends information on a child to their registry. The consent may be written or verbal. The Register wrote to all parents to verify that the parents are aware that their child has been registered. In 1981-85 only 12 per cent of cases had written informed consent from a parent. By 1990 the registry had written consent for 57 per cent of known cases. During the first decade the register was operating there were no complaints on privacy issues by parents, clinicians or midwives, yet the consequences of requiring informed consent were that almost one half of all cases were lost to the register.

Currently less than 10 couples per year refuse to give their consent (de Walle 2002), but in addition it is estimated that 20 per cent of parents do not respond to the request for consent made by their physician or midwife.
Although these requirements for informed consent are founded on ethical principles, there are now also important legal considerations in the UK. Patients provide information about themselves (or in the case of NCAS, about their children) in confidence and common law provides no justification for the use of these data in identifiable form other than with informed consent. The NHS must comply with the Data Protection Act 1998, which came into force on 1 March 2000, which requires that patients are fully informed about any treatment and the storage subsequently of tissue and written information. The Human Rights Act 1998 subjects any invasion of the private life of an individual to a test of necessity. The General Medical Council (GMC) Guidance in August 2000 that doctors should not pass any information about their patients without the patient's informed consent reflected the growing and evolving understanding and interpretation of the legal position. Hence, when midwives and other clinicians record information about congenital anomalies in a child's medical records, consent should be sought from the parents for this information to be passed to a congenital anomaly register.

The problem for all those working in the NHS is that most work cannot move immediately to a position where all patient information is either on an informed consent basis or anonymised. Much valuable work is currently undertaken using identifiable information obtained without consent. It is recognised that this work in the field of research, clinical trials and registries such as NCAS must continue. There are also situations where informed consent cannot practically be obtained. For example there are a number of medical research studies, and disease registers (such as NCAS), which include tens or hundreds of thousand study members where locating them all would be impracticable. Some people, such as those with mental disabilities, would not be capable of giving informed consent but their information is very valuable to the NHS. Excluding those who refuse consent is a manageable problem as the numbers are usually small, but a bigger problem is the unknown number of people who do not respond to request for consent as they can bias the data collection so as to invalidate any research findings.
As an interim measure, Section 60 of the Health and Social Care Act 2001 provides a power to ensure that in England and Wales patient identifiable information needed to support essential NHS activity can be used without the consent of the patient. This power can only be used to support medical purposes that are in the interests of the patient or the wider public, where consent is not a practical alternative and where anonymisation will not suffice. It is largely intended as a transition arrangement, reviewed annually.

NCAS and all the congenital anomaly registers decided to apply for Section 60 support to continue to collect and process congenital anomaly data without informed consent. It was decided that the best way to achieve this would be to make a collective application in the name of BINOCAR. After several years of discussion about the need for BINOCAR to develop from an informal organisation, in 2000 a formalised BINOCAR constitution was developed to provide affiliation for registers run on agreed guidelines. In order to achieve formal BINOCAR accreditation, each register signed a statement of affiliation, which involved agreeing to a set of data handling guidelines, signed by one representative of the register (usually the Register leader).

Under this more formalised constitution, NCAS and all the congenital anomaly registers affiliated to BINOCAR subsequently applied for support under Section 60 of the Health and Social Care Act 2001 for legal cover to continue to collect identifiable personal data without informed consent. However, the registers still wanted to be open to the public about all their activities, so posters were drafted which could be adapted for local use and displayed prominently in antenatal clinics. Also similar leaflets were prepared which could be included in the information pack given to pregnant women on booking for delivery. Both posters and leaflets explained that women have a right to ask that both their details and those of their children are not to be included in the register. At the same time, various projects are being developed to investigate the feasibility of obtaining parental consent for their baby's details to be held by the local registers and by NCAS.
In July 2002 BINOCAR registers including NCAS received support under Section 60 of the Health and Social Care Act 2001 to continue to collect congenital anomaly data without consent. This legal cover, however, required the registers to continue to develop information about their registers and to undertake pilot studies to test the feasibility of obtaining informed consent. Section 60 support is renewable annually. In the future it is envisaged that registers will either seek informed consent or will need specific legislation to continue to collect this information without informed consent. Our task is to identify the optimum way to seek consent which minimises attrition.
5.6 Conclusions

The Public Health agenda and Government policy has a strong emphasis on child health. This is therefore an ideal time to reshape NCAS, in particular to address the primary problems of ascertainment and data quality. In this thesis I have shown how these major weaknesses can be improved by working with existing registers, and linking with existing data sources. Given the increased participation of local registers in data provision to NCAS, by 2003 NCAS will be receiving data from all the local registers in England except the West Midlands Congenital Anomaly Register. Further improvements in ascertainment would require complete national coverage by local registers and/or linkage between as many sources as possible of child health data.

Funding will be required to address these issues, but with investment major improvements can be achieved. The priorities I see for NCAS are first to develop the model of becoming a national congenital anomaly co-ordinating centre through data transfer by local and condition specific registers and through data linkage with other child health sources which collect anomaly information. Second, BINOCAR should work to increase national coverage of local registers, particularly with a view to seeking comparable funding in a similar way to cancer registries. Thirdly, for those HAs without a local register, NCAS should seek greater involvement of local clinicians in NCAS as this has been shown to increase ascertainment. Fourthly, BINOCAR should continue to test the feasibility of gaining informed consent without a significant reduction in the number of notifications. If this is unsuccessful, then BINOCAR should lobby for its own primary legislation. This is a challenging work programme, but most of the proposals have been tested at least in part and shown to be feasible, and with the commitment of those currently working locally and nationally, this should be achievable in the next decade.
REFERENCES


Acheson ED. Medical record linkage. London: Oxford University Press, 1967


Botting B. Improving the completeness of Down syndrome notification. Health Statistics Quarterly, 2000a; 6: 14-17


Castilla EE. Clusters of ophthalmia. No further clues from global investigation BMJ 1994 Jan 15;308(6922):206

Castilla EE, Dutra M da G. Limb reduction defects and coastal areas. Letter. Lancet, 1994a; 343: 1034


Congenital Anomaly Register & Information Service for Wales (CARIS) Annual Report 2000


Cornel MC, de Walle HEK, ten Kate LP. Ten years of experience with registration of congenital anomalies in the northern Netherlands by EUROCAT Tijdschrift Sociale Gezondheidszorg 1992; 70: 637-644

Cornel MC, Erickson JD. Comparison of national policies on periconceptional use of folic acid to prevent spina bifida and anencephaly (SBA). Teratology. 1997 Feb;55(2):134-7


Darroch JN. The two-sample capture-recapture census when tagging and sampling are stratified. Biometrika. 1961; 48: 241-260


Gregg NM. Congenital cataract following German Measles In the Mother. Transactions of the Ophthalmological Society of Australia. 1941, 3,35-46


Gutteridge W, Collin C. Letter BMJ 1994; 308; 531


Hobbs CA, Hopkins SE, Simmons CJ. Sources of variability in birth defects prevalence rates. Teratology 2001; 64 Suppl 1: S8-S13


225

Laporte RE. Assessing the human condition: capture-recapture techniques. BMJ 1994; 308: 5-6


Leck IM, Millar ELM. Incidence of malformations since the introduction of thalidomide. BMJ July 7, 1962 16

Leck I, Smithells RW. The Ascertainment of Malformations. Lancet 1963 i 101-103


Lie RT, Vollset SE, Botting B, Skjaerven R. Statistical methods for surveillance of congenital malformations: when do the data indicate a true shift in the risk that an infant is affected by some type of malformation? International Journal of Risk and Safety in Medicine, 1991; 2: 289-300


Little R, Ruban DB. The analysis of social science data with missing values. Sociological Methods and Research, 1989; 18(2): 292-326


Merseyside and Cheshire Congenital Anomaly Survey Annual Report 2000

Minton SD, Seegmiller RE. An improved system for reporting congenital malformations. JAMA 1986 256. No 21:2976-2979


Nichol D, CEO, NHS Executive 1992


North Thames (West) Congenital Malformation Register Report, 2000, Perinatal Public Health Unit, Kennedy-Galton Centre, Northwick Park Hospital


Paduano M. Mystery of babies with no eyes. Observer 17 January 1993:3

Page ES. Continuous inspection schemes. Biometrika 1954; 41: 100-144


Roderick PJ, Connelly JB. The problems of monitoring tuberculosis in an inner-city health district: integrated information is required. Public Health 1992 May;106(3):193-201


Slater BS, Watson GI, McDonald. Seasonal Variation in Congenital Abnormalities. British Journal of Preventative and Social Medicine. 1964; 18 Vol 1


Stanley FJ, Croft ML, Gibbons J, Read AW. A population database for maternal and child health research in Western Australia using record linkage. Paediatric and Perinatal Epidemiology 1994; 8: 433-447


Weatherall JAC. An assessment of the efficiency of notification of congenital malformations. The Medical Officer,1969;121:65-68


Woodward RH, Goldsmith PL. Cumulative Sum Techniques. ICI Monograph No 3. Oliver and Boyd 1964


Appendix A  NCAS notification form – SD56
CONGENITAL ANOMALIES

Identification number: .............................................
(The form of this number is not important, but should be decided locally to enable a particular case to be
identified subsequently if required.)

District Health Authority of mother's usual residence: .......................... DHA

District Health Authority in which baby was born: ............................... DHA
Surname of child (first three characters only): .................................
Forename of child (first three characters only): .................................
Place of birth of child (please tick one box only): ...............................
Date of birth of child: .................................
Sex of child (please tick one box only): .................................
Whether live or still birth (please tick one box only): ..........................
Whether single or multiple birth (please enter number): ........................
Date of L.M.P.: .................................
If L.M.P. date not known, state estimated gestation: ............................
Birthweight: .................................

Home address of mother: .................................
(Please include postcode if known)
Parents' occupation (just before or early in mother's pregnancy): .............................
Mother: .................................
Father: .................................
Date of birth of mother: .................................
If date of birth not known, state age: .................................
Number and outcome of previous pregnancies resulting in: .......................... Live births: ................................. Stillbirths: ................................. Others*: .................................

Status of informant (please tick one box only): .................................
Doctor ................................. Midwife ................................. Other .................................
If 'Other', please specify: .................................

Congenital anomalies reported: .................................
(A detailed written description of each congenital anomaly is required to enable ONS to code the
abnormalities to the 4 digit ICD classification. It is important that no defects are omitted).

* 'Others' is defined as pregnancies that ended in other than a registrable live or stillbirth.
# Appendix B  NCAS monitoring groups of anomalies affecting the same anatomical system

<table>
<thead>
<tr>
<th>Central Nervous System A</th>
<th>ICD 10 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0A Anencephalus</td>
<td>Q00.0-Q00.2</td>
</tr>
<tr>
<td>0B Spina bifida</td>
<td>Q05.0-Q05.9</td>
</tr>
<tr>
<td>0C Congenital hydrocephalus</td>
<td>Q03.0-Q03.9</td>
</tr>
<tr>
<td>0E Encephalocele</td>
<td>Q01.0-Q01.9</td>
</tr>
<tr>
<td>0F Other</td>
<td>G04.9, G12.0, G12.9, G40.9, G60.0, G62.9, G70.9, G71.1, G71.2, G80.9, G83.2, G93.0, G93.1, Q02, Q04, Q06, Q07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eye and Ear B</th>
<th>Q11.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A Cystic eyeball</td>
<td>Q12.0-Q12.9</td>
</tr>
<tr>
<td>1B Congenital lens anomalies</td>
<td>Q11.3</td>
</tr>
<tr>
<td>1C Other &amp; unspecified eye anomalies</td>
<td>Q10, Q11.3, Q13-Q15, H18.5, H 50.0, H50.8, H54.0, H54.4, H55</td>
</tr>
<tr>
<td>1D Ear, all</td>
<td>Q16, Q17</td>
</tr>
<tr>
<td>1E Other anophthalmos</td>
<td>Q11.1-Q11.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alimentary System C</th>
<th>Q35</th>
</tr>
</thead>
<tbody>
<tr>
<td>2A Cleft of palate only</td>
<td>Q35</td>
</tr>
<tr>
<td>2B Cleft of lip only</td>
<td>Q36</td>
</tr>
<tr>
<td>2C Cleft palate with cleft lip</td>
<td>Q37</td>
</tr>
<tr>
<td>2D Tracheo-oesophageal fistula/stenosis</td>
<td>Q39.0-Q39.3</td>
</tr>
<tr>
<td>2E Other or unspecified anomalies of alimentary system</td>
<td>Q42.0-Q42.9</td>
</tr>
<tr>
<td>2F</td>
<td>K74.0, K80.2, Q39.4-Q39.9, Q40.0, Q40.2-Q40.9, Q41, Q43-Q45, R14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cardiovascular System D</th>
<th>Q21.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>3A Tetralogy of Fallot</td>
<td>Q21.0</td>
</tr>
<tr>
<td>3B Ventricular septal defect</td>
<td>Q21.1-Q21.2, Q21.4-Q21.9</td>
</tr>
<tr>
<td>3C Other septal defects</td>
<td>Q25.0</td>
</tr>
<tr>
<td>3D Patent Ductus Arteriosus</td>
<td>Q27.0</td>
</tr>
<tr>
<td>3E Anomalies of the umbilical artery</td>
<td>Q45.6, Q47.1, Q49.1, M30.3, Q20, Q22-Q24, Q25.1-Q25.9, Q26</td>
</tr>
<tr>
<td>3F Other congenital cardiac or great vessel anomalies</td>
<td>Q27.1-Q27.9, Q28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Respiratory System E</th>
<th>Q30-Q34</th>
</tr>
</thead>
<tbody>
<tr>
<td>4A Congenital anomalies of the respiratory system</td>
<td>Q54, Q64.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Urogenital System F</th>
<th>Q55</th>
</tr>
</thead>
<tbody>
<tr>
<td>5A Hypospadias/epispadias</td>
<td>Q54, Q64.0</td>
</tr>
<tr>
<td>5B Other anomalies of the male genitalia</td>
<td>N47, Q53, Q55</td>
</tr>
<tr>
<td>5C Anomalies of the female genitalia</td>
<td>N89.8, Q50-Q52</td>
</tr>
<tr>
<td>5D Bladder exstrophy</td>
<td>Q64.1</td>
</tr>
<tr>
<td>5E Renal agenesis</td>
<td>Q60</td>
</tr>
<tr>
<td>5F Other or unspecified defects of urogenital system</td>
<td>N25.8, Q63, Q64.2-Q64.9</td>
</tr>
<tr>
<td>5G Indeterminate sex</td>
<td>Q61</td>
</tr>
<tr>
<td>5H Cystic kidney disease</td>
<td>Q61</td>
</tr>
<tr>
<td>5J Congenital obstructive defects of</td>
<td>N13.9, Q62</td>
</tr>
</tbody>
</table>
### Limbs

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>ICD-10 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>6A</td>
<td>Polydactyly/syndactyly</td>
<td>G69, Q70</td>
</tr>
<tr>
<td>6B</td>
<td>Limb reductions</td>
<td>Q71-Q73</td>
</tr>
<tr>
<td>6C</td>
<td>Deformities of feet</td>
<td>Q66</td>
</tr>
<tr>
<td>6D</td>
<td>Dislocation of hip</td>
<td>Q65.0-Q65.6</td>
</tr>
<tr>
<td>6E</td>
<td>Other limb or limb girdles</td>
<td>M21.2, Q65.8, Q65.9, Q68.1-Q68.5, Q74</td>
</tr>
</tbody>
</table>

### Other Musculoskeletal H

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>ICD-10 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7A</td>
<td>Other anomalies of the diaphragm</td>
<td>Q79.1</td>
</tr>
<tr>
<td>7B</td>
<td>Anomalies of the face, skull or neck</td>
<td>Q67-Q68.0, Q75, R22.0</td>
</tr>
<tr>
<td>7C</td>
<td>Other musculoskeletal anomalies of the thorax and neck</td>
<td>Q76.8, Q76.9</td>
</tr>
<tr>
<td>7D</td>
<td>Osteodystrophy or chondrodystrophy</td>
<td>Q77, Q78</td>
</tr>
<tr>
<td>7E</td>
<td>Other or unspecified anomalies of the musculoskeletal system</td>
<td>Q68.8, Q76.0-Q76.7, Q79.5, Q79.9</td>
</tr>
<tr>
<td>7F</td>
<td>Anomalies of the abdominal wall (hernias)</td>
<td>K40-K46</td>
</tr>
<tr>
<td>7G</td>
<td>Exomphalos</td>
<td>Q79.2</td>
</tr>
<tr>
<td>7H</td>
<td>Anomalies of the lips, tongue and pharynx</td>
<td>Q18.4-Q18.7, Q38</td>
</tr>
<tr>
<td>7J</td>
<td>Congenital hiatus hernia</td>
<td>Q40.1</td>
</tr>
<tr>
<td>7K</td>
<td>Congenital diaphragmatic hernia</td>
<td>Q79.0</td>
</tr>
<tr>
<td>7L</td>
<td>Gastrochisis</td>
<td>Q79.3</td>
</tr>
<tr>
<td>7M</td>
<td>Prune belly syndrome</td>
<td>Q79.4</td>
</tr>
<tr>
<td>7N</td>
<td>Branchial cleft, auricular sinus</td>
<td>Q18.0-Q18.2</td>
</tr>
<tr>
<td>7P</td>
<td>Other anomalies of face &amp; neck</td>
<td>K07, Q18.3, Q18.8, Q18.9, R22.0</td>
</tr>
</tbody>
</table>

### Skin and Integument II

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>ICD-10 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>8C</td>
<td>Anomalies of the skin or integument</td>
<td>D18.0, D22.3, D22.6, D22.7, D22.9, D23.6, D23.7, L53.9, L81.3, Q80-Q84</td>
</tr>
</tbody>
</table>

### Other Anomalies H

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>ICD-10 Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>9A</td>
<td>Congenital neoplasms (other than benign skin)</td>
<td>C49.2, C69.2, C71.9, C74.9, D13.9, D14.3, D15.1, D16.6, D17.2, D17.9, D36.1, D37.0, D41.0, D43.2, D47.1, D48.0, D48.7, D48.9</td>
</tr>
<tr>
<td>9B</td>
<td>Endocrine and metabolic disorders</td>
<td>D66, D67, D68.0, D69.4, D81.9, D82.1, E03.0, E03.1, E07, E23-E25, E27-E30, E32, E34, E70-E80, E83-E85, E88, E90</td>
</tr>
<tr>
<td>9C</td>
<td>Trisomy 21 - Down syndrome</td>
<td>Q90</td>
</tr>
<tr>
<td>9D</td>
<td>Other chromosomal anomalies</td>
<td>Q91-Q99</td>
</tr>
<tr>
<td>9E</td>
<td>Other and unspecified congenital anomalies</td>
<td>B27.0, D18.1, D55.0, D56.3, D57.1, D57.3, D58.0, F79, I42.4, K21.9, L05.9, P02.6, P10-P15, P20-P29 (excl P29.4), P36, P50, P61, P70-P78, P80-P83, P90-P93, P95, P96, Q85-Q89, R16.0, R16.2, R18, R19.0, R22.9</td>
</tr>
<tr>
<td>9H</td>
<td>Congenital infections</td>
<td>A50.0, P29.4, P35, P37, P52.5</td>
</tr>
</tbody>
</table>
Appendix C  EUROCAT list of minor conditions that should not be notified
unless they occur in combination with other major anomalies

List of conditions for exclusion:

Reports of cases with the following anomalies are not to be transmitted to ONS
unless occurring in combination with other anomalies:

Name of condition

Spina bifida occulta uncomplicated
Stenosis or stricture of lacrimal duct
Minor or unspecified anomaly of auricle
Minor or unspecified anomaly of nose
Minor or unspecified deformity of face
Minor anomaly of nipple, accessory or ectopic nipple
Congenital umbilical hernia, inguinal or para umbilical
Undescended testicle and unspecified ectopic testis
Congenital hydrocele or hydrocele of testis
Phimosis
Hypospadias when the meatus lies before the coronary sulcus
Abnormal palmar crease
Skin tag with surface less than 4 cm²: skin tag, naevus, angioma, haemangioma, glomus tumour, lymphangioma, birthmark
Clicking hip
Clubfoot of postural original
Minor or unspecified anomalies of toe such as hallux valgus, hallux varus or "orteuil en marteau"
Functional or unspecified cardiac murmur
Absence or hypoplasia of umbilical artery, single umbilical artery

With acknowledgement to EUROCAT for permission to reproduce their
  exclusion list.
Appendix D  NCAS form for notifying limb defects.
Diagrams reproduced with the permission of EUROCAT (an EEC Concerted Action Project)
Appendix E  The standard form used to record the information given at birth registration
### BIRTH

**GRO Reference No.**

**SD No.**

**Register No.**

**Entry No.**

**Date of registration**

### CONFIDENTIAL PARTICULARS

The particulars below, required under the Population (Statistics) Acts, will not be entered in the register. This confidential information will be used only for the preparation and supply of statistical information by the Registrar General.

1. **Where the father's name is entered in register:**
   - Father's date of birth
   - Month
   - Year

2. **In all cases:**
   - Mother's date of birth
   - Month
   - Year

3. **Where the child is born within marriage:**
   - Date of marriage
   - Month
   - Year

4. **Has the mother been married more than once?**
   - Yes
   - No
   - Month
   - Year

5. **Mother's previous children (excluding registered) by her present husband:**
   - Name
   - Date of birth
   - Month
   - Year

6. **Name and surname:**
   - Informant
   - Qualification

7. **Total number of live and stillbirths at this maternity:**
   - Live births
   - Still births

**POSTCODE**

**Date of registration**

**Signature of registrar**

**BIRTH**

**Register No.**

**Entry No.**

**District & S.D. Nos.**

**Administrative area**

**Date and place of birth**

**CHILD**

2. **Name and surname**

3. **Sex**

4. **Name and surname**

5. **Place of birth**

6. **Occupation**

7. **Name and surname**

8. **Father's occupation**

9. **Surname at marriage**

10. **INFORMANT**

11. **Name and surname**

12. **Qualification**

13. **Usual address**

14. **I certify that the particulars entered above are true to the best of my knowledge and belief**

15. **Date of registration**

16. **Signature of registrar**

**FORM 309 (Rev)**
Appendix F
Botting B. The impact of more complete data from Wales on the National Congenital Anomaly System
Health Statistics Quarterly 05 2000
The impact of more complete data from Wales on the National Congenital Anomaly System

Background

The National Congenital Anomaly System began in 1964 after the thalidomide epidemic. Its primary purpose was surveillance, to quickly detect any similar increases in notifications. At that time, people were aware of the importance of monitoring anomalies, so notification was never made statutory – instead, notification was voluntary and provided by the local health area. Data are currently provided by NHS trusts on a standard ONS notification form (known as form SD56). Research estimates of the completeness of notification to the national system vary by type of anomaly, with those which are easily visible at birth being more completely notified than internal and chromosomal defects.

A review of the National System in 1993 recommended that 'where good congenital malformation registers exist outside OPCS (now ONS), information should be exchanged with these to improve the completeness and validity of both local and national data'. A number of local registers exist, but there had never been any formal contact between these registers and ONS. The local registers have the advantage that those working in it have local knowledge, and are therefore more able to obtain data. They are also more visible to local clinicians. As a result, the local registers hold more complete data. Nevertheless, the national system still provides the best estimate of national prevalence and allows comparisons across Health Authorities.

In 1998, ONS, in collaboration with Dr David Stone from the Glasgow Register of Congenital Anomalies, began the creation and co-ordination of BINOCAR (the British Isles Network of Congenital Anomaly Registers). Two new registers had recently obtained funding, and they were keen to work with ONS – indeed a condition of the Wales register (CARIS) funded by the Welsh Office was that they were to take over the notification to ONS of congenital anomalies for all of Wales. The It has long been known that notification to the National Congenital Anomaly System is incomplete. A review of the system recommended that ONS should explore data exchange with local regional registers. In 1998 ONS began data exchange with a local register (known as CARIS) which covers children in Wales and collects data from several different sources. This paper reports on the impact on monitoring a step change in ascertainment using multiple source notification from the Wales register. Following data exchange with the CARIS register there appears to be a real increase in the completeness of notification between 1997 and 1998. The number of respiratory anomalies notified in 1998 was four times higher than in 1997. The number of notifications of several other anomalies doubled between these two years. These levels of notification from Wales are close to the levels expected based on previous research from ad hoc surveys.
CARIS register notifies ONS of all congenital anomalies in live or still births known to them from any source. Although CARIS formally commenced in January 1998 it also notifies ONS of additional cases of congenital anomalies not reported from previous years. This has included a substantial number of cases for South Wales collected by an earlier register based in South Wales.

This analysis is based on notifications to ONS from the CARIS register in Wales. It compares these with previous notification levels in Wales and with those in England. It discusses the impact of improvements in completeness on local surveillance and on national statistics.

RESULTS

Figure 1 shows the number of notifications to ONS from Wales and their source, between 1992 and 1998. Table 1 shows these statistics, together with the number of notifications of selected congenital anomalies from Wales, over the same period. The anomalies included are those which are regularly analysed by ONS. For several of these anomalies, previous analyses have provided estimates of the completeness of notification to ONS.

It can be seen that the number of notifications from Wales rose to 642 in 1998, 66 per cent higher than in 1997. Previously the figures had been stable since 1992. Thus, in the absence of any suggestion that there was a real increase in prevalence, there appears to be an increase in the completeness of notification between 1997 and 1998. Figure 1 shows that in 1996 and 1997 the number of notifications for Wales were between 84 and 88 notifications per 10,000 live and still births. For England alone, however, the notification rate fell between 1997 and 1998 from 87 per 10,000 to 82 per 10,000.

Table 2 compares the notification rates for Wales with the rates for England and Wales as a whole and with those in England alone. In 1998 rates for Wales were higher than in England for all conditions except neural tube defects and hypospadias and epispadias. Rates for Wales were more than seven times as high as England for heart and circulatory anomalies, respiratory anomalies and eye anomalies (although the latter is based on very small numbers).

DISCUSSION

In the early and mid 1990s there were a number of administrative changes in Welsh Health Authorities and local providers. This may begin to explain why notification rates in Wales were lower than those for England. In the absence of any reason why true prevalence should be lower in Wales, this suggests that in this period notification was previously less complete for Wales than for England. Further evidence for this is given by the large increase in notifications seen in 1997 data.
### Table 3
Selected congenital anomalies in England and Wales 1996-98

<table>
<thead>
<tr>
<th>Condition</th>
<th>England and Wales</th>
<th>England only</th>
<th>Wales only</th>
</tr>
</thead>
<tbody>
<tr>
<td>All congenital anomalies</td>
<td>85.4</td>
<td>88.2</td>
<td>87.8</td>
</tr>
<tr>
<td>Central nervous system</td>
<td>3.8</td>
<td>3.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Neural tube defects</td>
<td>1.3</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Eye anomalies</td>
<td>1.2</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>Chloroacrodacty and epispadias</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Heart and circulatory anomalies</td>
<td>0.9</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Respiratory anomalies</td>
<td>7.9</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>Hypoplasia and epispadias</td>
<td>9.2</td>
<td>9.9</td>
<td>9.5</td>
</tr>
<tr>
<td>Deformities of feet</td>
<td>6.7</td>
<td>6.6</td>
<td>5.8</td>
</tr>
<tr>
<td>Syndactyly</td>
<td>4.0</td>
<td>3.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Limb reductions</td>
<td>3.0</td>
<td>2.1</td>
<td>3.0</td>
</tr>
<tr>
<td>Abdominal defects</td>
<td>2.2</td>
<td>2.2</td>
<td>2.3</td>
</tr>
<tr>
<td>Down's syndrome</td>
<td>5.1</td>
<td>5.9</td>
<td>5.8</td>
</tr>
</tbody>
</table>

As CARIS reported additional cases not previously reported. Data for 1998 showed a subsequent increase when CARIS formally began and started providing more complete notification. The particularly large increases in notification rates for heart and circulatory anomalies support this hypothesis, as we know these to be amongst the least well notified conditions to the national system.²

We plan to test this hypothesis further using 1999 data from another register. A second new register, the Trent Congenital Anomaly Register (CAR) began to notify ONS in 1998 on behalf of all health areas in Trent. Initially they only sent to ONS the information they had received from Health areas on the ONS notification forms. In 1999, however, they began complete electronic data exchange, based on multiple source ascerttainment. We expect to find a similar increase in notification from this register, particularly in those previously less-well notified conditions. The problem, in analysing these data, will be to disentangle any real increases from those due to improved notification.

The ONS Surveillance routine compares current notification levels for a given condition in each local area against expected values based on previous experience in that area. Therefore, any changes in reporting practice, such as more complete notification from a regional register, will only affect the surveillance outputs from that area. All other areas will be unaffected.

Given the large increase in notification rates from Wales, alarms reflecting higher than expected prevalence occurred regularly for 1998 data. Indeed, the surveillance routine for 1998 generated 96 alarms compared with just one in 1997. These are assumed to be a result of the increase in completeness of notification. It is therefore important that the surveillance programmes are quickly adjusted to these new raised levels of notification. Then any subsequent increase can be identified as a potential true increase in prevalence. Surveillance baselines change each year, so baselines in 1999 have already been partly adjusted to take on board these increases. It is our intention, however, to revise all the surveillance values for Wales at the end of 1999, based on just 1998 and 1999 data, so that the values can truly reflect the current reporting practices.

The improvements in notifications from Wales have had a small effect on the total for England and Wales as a whole. In 1996, 4 per cent of all notifications related to births in Wales. In 1998 the corresponding percentage was 11 per cent. This increase in rates in Wales in combination with the rates from English areas which have been falling slightly, has led to largely static rates for England and Wales. The biggest proportionate increases in rates for England and Wales between 1997 and 1998 were for those conditions for which there were very large increases in notification from Wales. In particular, heart and circulatory anomalies showed an 11 per cent increase in notification rate, limb reduction defects a 43 per cent increase and central nervous system defects a 22 per cent increase.

### Conclusion
Trent has now started multiple source electronic notification to ONS. We therefore expect to see similar increases in their notification rates for 1999 onwards. Other regional registers have indicated an interest in exchanging data electronically with ONS, so a similar increase in notification rates can be expected. We plan that the data exchange will be a two-way flow, with the national data benefiting from more complete notification, and local registers receiving in return information on any deaths of children in their register.

Whilst we are in this transition stage of improving the completeness of notification, we anticipate an impact on the local surveillance outputs and on national statistics. It is important that we are able to quantify the expected increases due to improvements in notification for given anomalies, so as to be able to follow trends in the underlying number of anomalies. Also local registers are funded in different and sometimes precarious ways. If a local register loses its funding, there must be a mechanism to restore local notification from health areas to the National System.

ONS’ collaboration with Wales has demonstrated immediate benefits in improving the completeness of data. As electronic data exchange extends to cover more of England and Wales, national data will become more complete, and therefore more sensitive to real changes in prevalence. This will improve the quality of the surveillance alarms. National data will be closer to true prevalence rates so better able to be used to inform Government policy, assist in geographical comparisons, and assist in medical research.

### References
Appendix G

Linking congenital anomaly and birth records

Bev Botting and Carole Abrahams, Office for National Statistics

Improving our understanding of the role played by different risk factors for congenital anomalies requires more and better information than is currently available from the notification form. The purpose of this study was to investigate the feasibility of linking congenital anomaly records to corresponding birth records and methods of doing this. If successful, this linkage would enable ONS to collect less demographic information at congenital anomaly notification as this information could be obtained from the birth record, and therefore reduce the burden on data suppliers. It would allow analyses of survival in babies with congenital anomalies using existing links to child death records, help us to measure the quality of the data, and could inform regional congenital anomaly registers of death and other health outcomes in children on their registers. The study was successful. It was demonstrated that linkage was possible for 97 per cent of records. We measured completeness of some data items and identified new variables, which allowed analyses not previously possible.

Introduction

ONS has collected notifications of congenital anomalies through the National Congenital Anomaly System (NCAS) since 1964. To try to understand some risk factors for different anomalies, a range of social and demographic characteristics are collected at notification. Much of this information is also collected at the child’s birth registration. This has allowed rates to be calculated, using births as the denominators. Record linkage between the NCAS and the birth registration records would allow us to derive some information directly from the birth record, and eliminate the need to collect it again at congenital anomaly notification. This would reduce the burden on the data suppliers, which in turn could improve the completeness of notified congenital anomalies. Also, with appropriate ethics approval and awareness of confidentiality issues, local congenital anomaly registers could be informed of the death or other health outcomes (if these are also linked to the birth records) for children on their registers.

Background

To understand better the risk factors for congenital anomalies, and to make the best use of routine data sources, we wanted to link congenital anomaly records with corresponding birth records. Ultimately the vision is to link information from a wide range of different sources on child health to birth records. This will allow us to describe pathways of child health, illness and subsequent survival for a range of different conditions. Other possibly linkable sources of data on child health include child cancer and information from Hospital Episode Statistics. Currently the only child health linkage that is routinely undertaken by ONS is for deaths of children.
ONS has long-standing experience in linkage, particularly involving links with the birth records. Since 1975, infant deaths have been successfully linked to birth records, and resulting analyses have been published annually. The main linking variable identifier is the NHS number, which is derived from the National Health Service Central Register (NHSCR) on the death of a child. NHSCR uses the name, date of birth, and sex to find the NHS number, enabling the birth record to be identified. This linkage has been extended to link all death records of children born since 1993. For example, a child who died aged 7 in 2000, having been born in 1993, would have their death record linked to their birth record. Consequently it has been possible to monitor known risk factors for infant mortality and, more recently, show how these risk factors extend beyond infancy.

Technological improvements mean that tools for linkage are becoming more accessible and the data are more easily manipulated. Congenital anomalies were selected as the next source of data for linkage as the data are held within ONS and are regularly used alongside linked infant mortality analyses.

The congenital anomaly system would be enhanced by linkage, enabling, for example, less information to be collected at notification, whilst providing improved analysis since both numerator and denominator would be derived from the same source. Since the demographic information collected at birth registration is normally from the parents themselves, it is more complete and of better quality than that recorded, usually by midwives, at congenital anomaly notification. For example, 40 per cent of congenital anomaly records do not include the mother’s or father’s occupation, whereas the mother’s occupation is collected at birth registration for all babies, and the father’s occupation is collected for all registrations where the father’s occupation is known.

Congenital anomalies are a leading cause of childhood mortality. The death certificates of children who have died, however, often do not mention the presence of an existing anomaly. This means that it is impossible to measure the survival of different cohorts of children with a given anomaly. Since birth records already contain child mortality links, linkage to congenital anomaly records would ensure that rates and causes of death for children born with particular anomalies could be obtained. Once more cases are available, survival curves could be constructed for successive cohorts. Similarly, when a child has a congenital anomaly mentioned on their death certificate, it would be possible to ascertain whether the child had been previously notified as having a birth defect. For those anomalies which are life threatening, this could help measure the completeness of notification.

Methods

Overview

Ideally, the key matching field would be the NHS number since it is unique to each child. It was not possible to use this identifier for this study, however, since currently it is not generally included in congenital anomaly notifications. From May 2002, however, it is planned that the NHS number will be allocated at birth by midwives rather than, as at present, by Registrars at birth registration. This will only apply to live births, as stillbirths are not given an NHS number.

If an NHS number is not given it can currently be found using the NHSCR. However, the disadvantage is that some records can only be traced with intensive clerical intervention, which is expensive. Therefore, this was not included in the current linkage programme. Instead, an algorithm, which matched on other key fields, was tested.

Matching algorithms

The algorithms were developed using existing knowledge from other linkage and matching exercises recently undertaken. An example of recent matching work is for the Confidential Enquiry into Stillbirths and Deaths in Infancy (CESDI). It is important that the CESDI records have complete coverage. ONS works with CESDI to compare their records with death registrations to identify missing cases. CESDI records are matched with the Infant Mortality file and the Stillbirths file as appropriate. A 4-stage matching is performed, using postcode, sex, date of birth (DOB), and date of death (DOD). A 96 per cent matching rate was achieved for 1998 data.

Another matching exercise links births data with Regional Interactive Child Health System (RICHS) records. The RICHS system collects birth notification data from midwives for all births in the North Thames East area. This comparison identifies the parents’ countries of birth for each child on the RICHS file. The records are then anonymised before sending to the researchers investigating levels of specific infections in groups of the population born overseas. NHS number is often present in RICHS records, but the algorithm requires that sex and either DOB (within one day) or postcode must also match. Where NHS number is not given, the key fields are postcode, sex and DOB. Occurrences of more than one possible match are narrowed down by surname and hospital code. The process is complicated by multiple births, so first names and birthweight need to be included in these cases. An extremely high rate of matching is achieved (99 per cent).

All these matching and linkage algorithms are based on comparisons of a small set of variables, collected by both data sources. The key link variable is NHS number, since it provides a unique match. In the absence of NHS number, however, postcode, DOB and sex are the key variables used. Name is rarely used, and only then as either a confirming variable or where the matching routine has identified more than one possible match. These various algorithms were used as the basis for the algorithm developed for this study.

Stage 1 Link using baby’s DOB, postcode and sex. These are the same variables, which have been used successfully for the RICHS and CESDI matching programmes. Duplicate matches were eliminated by requiring that (surname or mother’s age) and (forename or birthweight) matched. The first criterion ensures that the mother is the same, the second identifies individual babies in a multiple birth. Surname was matched against child’s surname, father’s surname or mother’s maiden name, forename was matched against child’s first or second forenames.

Stage 2 Link using baby’s DOB, first three characters of postcode, sex and mother’s DOB. Duplicate matches eliminated as at stage 1.

Stage 3 Match by baby’s DOB, mother’s DOB, plus ONE of the following: surname, forename or birthweight.

Stage 4 Match by first three characters of postcode plus birthweight, plus either forename or surname.

Stage 5 Match by surname plus various combinations of forename, baby’s DOB, mother’s DOB, postcode, sex and birthweight.
DATA USED

The analysis consisted of all congenital anomaly records for 1997 and 1998.

Births records are held in two separate databases, births registration and births statistics. These files were combined to provide the fields required for linking purposes. For example, names, place of birth, and NHS number are in the registration record whereas mother’s age, birthweight, and parents’ occupation codes are in the statistics record.

A birth extract file was prepared. This consisted of specified fields from Birth Registration and Birth Statistics databases, merged into single records for each birth in 1997 and 1998. Permissions were obtained from data custodians to access the birth records. In addition, permission was specifically obtained to use mother’s date of birth as a linkage variable since this data item is collected under the Population Statistics Act. All work was done in a secure and restricted environment.

There is some overlap between the fields in the congenital anomaly system and the births system since some items are collected at both birth registration and congenital anomaly notification, for example, postcode and birthweight. The information, however, is collected from different people - the parents usually register the birth, whereas the midwife completes the birth notification form from which congenital anomaly information is often derived. Linkage provides a way to measure which source holds more complete data.

RESULTS

Matching processes

Three quarters of all the records were matched using the first stage of the algorithm, which uses date of birth, postcode and sex. The second phase of the algorithm, which included mother’s date of birth, but reduced the postcode to postcode sector, raised the proportion of matched records to over 80 per cent. The third step of the algorithm found some additional cases and generated a few duplicates and possible matches that could only be resolved manually. This raised the proportion matched to 98 per cent for 1998 and 97 per cent for 1997.

Quality checks by comparing variables collected from both sources

Table 1 shows the mother’s age as given at birth registration and the difference from that given on the congenital anomaly notification. Overall, 96 per cent of mother’s ages agreed between the two sources, and a further 1 per cent had an age difference not greater than one year. There was an equal distribution of the remainder above and below the age given on the congenital anomaly record. Therefore, mother’s age is generally consistently reported on both sources, and there is no systematic bias towards a higher or lower age.

### Table 1
Mother’s age in congenital anomaly record subtracted from mother’s age in birth record

<table>
<thead>
<tr>
<th>Mother’s age in birth record</th>
<th>Difference</th>
<th>&gt;10 less</th>
<th>-6 to -10</th>
<th>-3 to -5</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3 to 5</th>
<th>6 to 10</th>
<th>&gt;10 more</th>
<th>Missing</th>
<th>Total</th>
<th>% Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;16</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>904</td>
<td>5</td>
<td>1</td>
<td>13</td>
<td>2,054</td>
<td>18</td>
<td>1</td>
<td>6</td>
<td>32</td>
<td>942</td>
</tr>
<tr>
<td>16-19</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>16</td>
<td>2,800</td>
<td>21</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>47</td>
<td>4</td>
<td>11</td>
<td>32</td>
<td>3,015</td>
</tr>
<tr>
<td>20-24</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>23</td>
<td>2</td>
<td>1,435</td>
<td>13</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>47</td>
<td>44</td>
<td>3</td>
<td>52</td>
<td>3,319</td>
</tr>
<tr>
<td>25-29</td>
<td>3</td>
<td>10</td>
<td>6</td>
<td>7</td>
<td>23</td>
<td>1,177</td>
<td>21</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>52</td>
<td>3</td>
<td>11</td>
<td>47</td>
<td>3,015</td>
</tr>
<tr>
<td>30-34</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>23</td>
<td>2</td>
<td>1,435</td>
<td>15</td>
<td>2</td>
<td>6</td>
<td>7</td>
<td>47</td>
<td>44</td>
<td>3</td>
<td>52</td>
<td>3,319</td>
</tr>
<tr>
<td>35-39</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>16</td>
<td>1,177</td>
<td>21</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>52</td>
<td>3</td>
<td>11</td>
<td>47</td>
<td>3,015</td>
</tr>
<tr>
<td>40 and over</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>60</td>
<td>2</td>
<td>1,177</td>
<td>21</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>52</td>
<td>3</td>
<td>11</td>
<td>47</td>
<td>3,015</td>
</tr>
</tbody>
</table>

Total | 10 | 29 | 24 | 20 | 80 | 10,837 | 80 | 17 | 23 | 33 | 13 | 175 | 11,341 | 95.6   |

### Table 2
Difference in birthweight between birth records and congenital anomaly records

<table>
<thead>
<tr>
<th>Difference between birthweights (grams)</th>
<th>Birthweight (ranges in grams)</th>
<th>&lt;1500</th>
<th>1500-1999</th>
<th>2000-2499</th>
<th>2500-2999</th>
<th>3000-3499</th>
<th>3500 and over</th>
<th>Missing</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;500 less</td>
<td>10</td>
<td>3</td>
<td>9</td>
<td>15</td>
<td>15</td>
<td>1</td>
<td>33</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>-500 to -100</td>
<td>6</td>
<td>2</td>
<td>22</td>
<td>19</td>
<td>27</td>
<td>12</td>
<td>4</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>-100 to -50</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>43</td>
<td></td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>-50 to -25</td>
<td>6</td>
<td>5</td>
<td>10</td>
<td>9</td>
<td>19</td>
<td>17</td>
<td>66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-25 to -1</td>
<td>17</td>
<td>14</td>
<td>45</td>
<td>77</td>
<td>115</td>
<td>139</td>
<td>407</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weights same</td>
<td>408</td>
<td>383</td>
<td>877</td>
<td>2,048</td>
<td>3,207</td>
<td>2,937</td>
<td>9,860</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 to 25</td>
<td>15</td>
<td>10</td>
<td>17</td>
<td>54</td>
<td>82</td>
<td>91</td>
<td>269</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 to 50</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>13</td>
<td>18</td>
<td>7</td>
<td>46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51 to 100</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>7</td>
<td>13</td>
<td>38</td>
<td></td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>101 to 500</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>18</td>
<td>16</td>
<td>20</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;500 more</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>15</td>
<td>35</td>
<td>67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>25</td>
<td>13</td>
<td>30</td>
<td>59</td>
<td>78</td>
<td>79</td>
<td>10</td>
<td>294</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>506</td>
<td>446</td>
<td>1,026</td>
<td>2,342</td>
<td>3,611</td>
<td>3,363</td>
<td>47</td>
<td>11,341</td>
<td></td>
</tr>
</tbody>
</table>

Weights same as % of the total | 80.6 | 85.9 | 85.5 | 87.4 | 88.8 | 87.3 | 86.9 |
Table 2 shows the differences in stated birthweight between the birth and congenital anomaly records. Eighty seven per cent matched exactly and a further 6 per cent were within 25 grams either side. Six per cent of records were heavier in the birth records and 4 per cent were heavier in the congenital anomaly record. Hence, birthweight is generally reported consistently to both sources.

Table 3 compares the stated multiplicity between the two records. Ninety six per cent of all congenital anomaly records declared as one of a multiple birth were found to be part of a multiple birth at birth registration. However, 14 per cent of records registered at birth as one of a multiple birth had not been recorded as such at congenital anomaly notification.

Table 4 shows the vital status of the child at both birth registration and congenital anomaly notification. There was generally good agreement between the two sources, with 96 per cent of the records recorded as stillbirths at birth registration also recorded as such at congenital anomaly notification.

### New data enabled through linkage

Through linking the two records, it is now possible to analyse congenital anomalies by outcome (survival/death). Table 5 shows linked records by the grouped type of anomaly and outcome. Less than half of all children born with neural tube defects in 1997 and 1998 were surviving at the time of analysis, compared with over 90 per cent of those children born with limb defects. In general a slightly higher proportion of those born in 1998 were surviving at the time of analysis, compared to those born in 1997. This would be expected, however, as those born in 1997 were older so had been at risk of dying for longer. For urogenital and limb anomalies, however, slightly higher proportions of children born in 1997 with the given condition were surviving at the time of analysis compared with those born in 1998. Figure 1 presents the percentage distribution of outcome by condition for 1997 and 1998 combined.

### Table 4  Comparison of live/still birth indicators

<table>
<thead>
<tr>
<th>Birth record</th>
<th>Congenital anomaly record</th>
</tr>
</thead>
<tbody>
<tr>
<td>Live birth</td>
<td>Died within 7 days</td>
</tr>
<tr>
<td>Live birth</td>
<td>10,646</td>
</tr>
<tr>
<td>Stillbirth</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>10,652</td>
</tr>
</tbody>
</table>

### Table 5  Survival by Malformation group

<table>
<thead>
<tr>
<th>Malformation groups</th>
<th>Total</th>
<th>Stillbirth</th>
<th>Infant death</th>
<th>Surviving</th>
<th>% surviving</th>
<th>Total</th>
<th>Stillbirth</th>
<th>Infant death</th>
<th>Surviving</th>
<th>% surviving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central nervous system</td>
<td>308</td>
<td>36</td>
<td>58</td>
<td>114</td>
<td>56</td>
<td>259</td>
<td>56</td>
<td>58</td>
<td>145</td>
<td>56</td>
</tr>
<tr>
<td>Eye and ear</td>
<td>157</td>
<td>1</td>
<td>18</td>
<td>138</td>
<td>88</td>
<td>253</td>
<td>14</td>
<td>19</td>
<td>220</td>
<td>87</td>
</tr>
<tr>
<td>Alimentary system</td>
<td>805</td>
<td>12</td>
<td>52</td>
<td>741</td>
<td>92</td>
<td>849</td>
<td>21</td>
<td>62</td>
<td>765</td>
<td>90</td>
</tr>
<tr>
<td>Cardiovascular system</td>
<td>575</td>
<td>16</td>
<td>100</td>
<td>459</td>
<td>60</td>
<td>673</td>
<td>18</td>
<td>168</td>
<td>537</td>
<td>80</td>
</tr>
<tr>
<td>Respiratory system</td>
<td>54</td>
<td>5</td>
<td>12</td>
<td>37</td>
<td>69</td>
<td>73</td>
<td>14</td>
<td>25</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>Urogenital system</td>
<td>1,073</td>
<td>17</td>
<td>39</td>
<td>1,017</td>
<td>95</td>
<td>1,088</td>
<td>20</td>
<td>32</td>
<td>1,016</td>
<td>93</td>
</tr>
<tr>
<td>Limbs</td>
<td>1,628</td>
<td>17</td>
<td>48</td>
<td>1,563</td>
<td>96</td>
<td>1,587</td>
<td>24</td>
<td>54</td>
<td>1,509</td>
<td>95</td>
</tr>
<tr>
<td>Other musculoskeletal</td>
<td>525</td>
<td>23</td>
<td>81</td>
<td>421</td>
<td>80</td>
<td>579</td>
<td>32</td>
<td>80</td>
<td>467</td>
<td>81</td>
</tr>
<tr>
<td>Skin and integument</td>
<td>422</td>
<td>1</td>
<td>5</td>
<td>116</td>
<td>95</td>
<td>201</td>
<td>2</td>
<td>10</td>
<td>189</td>
<td>94</td>
</tr>
<tr>
<td>Other anomalies</td>
<td>837</td>
<td>52</td>
<td>100</td>
<td>685</td>
<td>82</td>
<td>1,144</td>
<td>78</td>
<td>144</td>
<td>922</td>
<td>81</td>
</tr>
</tbody>
</table>
**Discussion**

The algorithm was successful in linking 97 per cent of all congenital anomaly records. It also provided us with new information about death and survival for children with congenital anomalies.

Linkage would be simplified by the inclusion of the NHS number on the congenital anomaly record. NHS number can be supplied by notifiers where available and this will be added to the congenital anomaly record. At that stage, it will no longer be necessary to collect the child's name. Name can be an unreliable variable since forename and surname may change between birth notification and birth registration. Removing the name from the notification process would also help preserve patient confidentiality when transferring data.

Once the NHS number is allocated at birth, congenital anomaly returns for live births will only need NHS number (to identify the child and enable linkage) and the details of birth defect(s). All other information currently required at congenital anomaly notification can be directly obtained from birth records, e.g. birth multiplicity, parents' occupations. Given the high levels of consistency between the two records of mother's age and birthweight, information will not be lost by relying on one source alone. Stillbirths may need more identifiable information if they are not allocated a unique identifier.

Occupation is important both for identifying the teratogenic effects of hazardous occupations and to derive a social class to measure health inequalities. Linkage would also improve the quality of occupation data. Occupation is poorly collected at congenital anomaly notification, and although only coded for 10 per cent of births, is available as text for all birth records and so could be coded for congenital anomaly analyses.

In addition, the quality of data on multiplicity was improved by linking to the birth record. Some anomalies, such as some neural tube defects, are known to have a higher prevalence in multiple births. Other analyses of NCAS data have shown this excess. Linkage shows, however, that previous analyses of the prevalence of anomalies among multiple births are likely to have been underestimated. Therefore more marked differences would be expected between the prevalence of congenital anomalies in singleton and multiple births. This would be in closer agreement with analyses from congenital anomaly registers in other countries.

The new findings on outcomes of children with congenital anomalies will allow us to generate survival data for different cohorts of children for specific anomalies. This will meet a current gap in knowledge and help those advising parents on risks and possible survival of babies born with different conditions. Provided ethical approval is given, it will also allow data exchange with local registers to inform them of the death and other health outcome for the children on their registers. This will help epidemiological analyses. Taken together, the reduction of the burden on notifiers plus feedback on outcomes, could improve the level of congenital anomaly notification.

It will also be possible to investigate cases where a child's death registration records the death as being due to a congenital anomaly, but where there is no corresponding record in the NCAS.

There are also linkage possibilities with other child health data and with information about the parents. This would enable a link with siblings, to look at generational effects, and to measure recurrence risks within the same family of outcomes such as congenital anomalies and sudden infant death syndrome.

**Conclusions**

It has been demonstrated that linkage of congenital anomaly and birth records is feasible, and that with some manual intervention, 97 per cent linkage was achievable. The algorithm used was successful, although it was clear that the NHS number would simplify the linkage process. This success is important to improve our understanding of risk factors for congenital anomalies, to improve the quality of congenital anomaly data, and also to use the methods developed for the linkage to help inform other record linkage projects in the future.

**Key points**

- The study was successful in linking 97 per cent of all congenital anomaly records to their birth records. It also provided new information about death and survival for children with congenital anomalies.
- Record linkage between the NCAS and the birth registration records allows some information to be derived directly from the birth record, and eliminates the need to collect it again at congenital anomaly notification. This would reduce the burden on the data suppliers, which in turn could improve the completeness of notified congenital anomalies.
- Overall, 96 per cent of mother's ages agreed between the two sources, and a further 1 per cent had an age difference not greater than one year. Therefore, mother's age is generally consistently reported on both sources, and there is no systematic bias towards a higher or lower age.
- Ninety six per cent of all congenital anomaly records declared as one of a multiple birth were found to be part of a multiple birth at birth registration. However, 14 per cent of records registered at birth as one of a multiple birth had not been recorded as such at congenital anomaly notification.
- Less than half of all children born with neural tube defects in 1997 and 1998 were surviving at the time of analysis, compared with over 90 per cent of those children born with limb defects.
- The new findings on outcomes of children with congenital anomalies allows survival data to be generated for different cohorts of children for specific anomalies. This will meet a current gap in knowledge and help those advising parents on risks and possible survival of babies born with different conditions.
- This success is important to improve our understanding of risk factors for congenital anomalies, to improve the quality of congenital anomaly data, and also to use the methods developed for the linkage to help inform other record linkage projects in the future.

**References**

Introduction

ONS has collected notifications of congenital anomalies through the National Congenital Anomaly System (NCAS) since 1964. To try to understand some risk factors for different anomalies, a range of social and demographic characteristics are collected at notification. Much of this information is also collected at the child’s birth registration. This has allowed rates to be calculated, using births as the denominators. Record linkage between the NCAS and the birth registration records would allow us to derive some information directly from the birth record, and eliminate the need to collect it again at congenital anomaly notification. This would reduce the burden on the data suppliers, which in turn could improve the completeness of notified congenital anomalies. Also, with appropriate ethics approval and awareness of confidentiality issues, local congenital anomaly registers could be informed of the death or other health outcomes (if these are also linked to the birth records) for children on their registers.

Background

To understand better the risk factors for congenital anomalies, and to make the best use of routine data sources, we wanted to link congenital anomaly records with corresponding birth records. Ultimately the vision is to link information from a wide range of different sources on child health to birth records. This will allow us to describe pathways of child health, illness and subsequent survival for a range of different conditions. Other possibly linkable sources of data on child health include child cancer and information from Hospital Episode Statistics. Currently the only child health linkage that is routinely undertaken by ONS is for deaths of children.