Record linkage to enhance data from perinatal registries

Tromp, M.

Citation for published version (APA):
Tromp, M. (2009). Record linkage to enhance data from perinatal registries
General discussion
General discussion

The first objective of this thesis was to develop tools to combine the existing, independent source data from the various registries, maintained by Dutch perinatal caregivers. With the combined data we subsequently aimed to evaluate the performance of the perinatal care system in the Netherlands given the constraints imposed by the observational nature of the data and the variables contained in the various registries.

The first part of the thesis describes the linking techniques used to combine reliably and validly the separate registries in perinatal healthcare. It presents the collected evidence in favour of a probabilistic rather than a deterministic linkage approach, it discusses two refinements to improve the probabilistic strategy, and it presents an efficient validation method of probabilistic linkage. In the second part of the thesis epidemiological studies are presented that illustrate the added value of the integrated national perinatal datasets after they have been successfully linked. This final chapter provides a summary and critical discussion of the principal findings, presents recommendations for current linkage projects and identifies areas for future research.

Summary of principal findings

Improving existing record linkage techniques

What is the difference in performance between deterministic and probabilistic linkage under varying conditions?

Probabilistic record linkage efficiently uses distributional information and data pattern information from the candidate datasets to improve the linking decision. The strength of the probabilistic algorithm is its ability to find a fine balance between accepting sufficient disagreements to overcome data entry errors and leaving enough discriminating power to uniquely identify different persons. This meant that the performance of probabilistic record linkage was superior (fewer misclassifications) to that of a deterministic record linkage across a range of linking conditions (chapter 1).

How to optimize the performance of probabilistic record linkage strategies in situations with different types of error in linking variables?

To account for typical data entry errors, the probabilistic linking algorithm can be extended with a third possible outcome when comparing corresponding values: ‘close’ agreement (chapter 2). Addition of the concept of ‘close’ agreement for typical errors can lower the number of record pairs in the so called grey area and thereby improve the results of a linking project.

How to handle dependencies among linking variables?

Violation of the independence assumption of the probabilistic algorithm apparently could bias the outcome in a situation with few linking variables. A method was introduced to account for dependencies in linking variables without losing information and the validity of this approach was demonstrated in both empirical and simulated datasets (chapter 3).
General discussion

How to validate a record linkage procedure?
The ranking of all record pairs (by total linking weight) as a result of the probabilistic method can be used to focus validation. The validation sample selected based on total linking weight (oversampling in the grey area) and questions arising from linked data appeared to be very informative with only a minimal sample size. Validation of the linkage of the LNR register showed that the quality of most linkage results was high, but insufficient in case of linkage of readmissions of multiple births (chapter 4).

Studying perinatal care in the Netherlands using integrated data from linked registries
The linked perinatal dataset is a valuable source of data on perinatal health care and pregnancy outcomes in the Netherlands, with near to complete coverage. Its value rests on the epidemiological and health services contents per se, but also on the instrumental value for other dedicated datasets.

How high is the perinatal mortality in the Netherlands, and is there a trend?
Dutch perinatal mortality among singletons decreased from 10.5 to 9.1 per 1,000 births in the period 2000-2006 (chapter 5). This trend could not be explained by changes in known risk factors including maternal age and non-western ethnicity. The decline was present in all risk groups except in the very premature births.

Are there regional variations in perinatal mortality within the Netherlands and can these differences be explained?
The perinatal mortality in the northern rural region was higher than in the other regions and this increased risk could not be explained by the demographic risk factors age, parity and ethnicity (chapter 6). Correction for urbanisation and social economic status explained only a small part of the elevated risk. Analysis of clinical risk groups showed that the perinatal mortality risk was especially increased in term women who were transferred from midwife to gynaecologist during delivery.

What is the influence of increasing maternal age on adverse outcomes?
The effects of increasing maternal age on fertility and pregnancy complications and outcomes of both mother and child were studied in detail by using hypothetical cohorts of women with a certain age (chapter 7). One important consequence of pregnancy planning at a higher age at is that more women will suffer from an unfulfilled child wish. The fewer older women that do reach an ongoing pregnancy have higher complication risks including twin pregnancy and non-spontaneous delivery.

What is the added value of linking cohort data with national registry data?
In chapter 8 the instrumental value of the linked datasets was shown for a large cohort study into the relationship between lifestyle during pregnancy and the health of Amsterdam children (ABCD-study). The identification of all ABCD cohort members in the national perinatal registry data through record linkage provided aggregate information on their non-responders. Selective non-response could be demonstrated after this linkage, but the associations found between risk factors and adverse perinatal outcomes were similar for respondents and non-respondents.
In light of the findings described in this thesis, some general points of discussion are put forward.

**Record linkage**

*Record linkage strategy: which method to choose*

*Probabilistic linkage: the method of choice*

The increasing availability of registries in healthcare makes it attractive to use existing data for answering research questions. If different pieces of information on the same person are present in multiple datasets, these datasets have to be linked. This situation is common in care which is provided by several caregivers (also called care ‘chain’; perinatal care is a typical example), and more generally in situations where the outcome is remote from the etiological factors (e.g. secondary cancer in adult life after successful treatment of cancer in childhood).

Probabilistic record linkage is the preferred method for combining data on the same entity from different files as it leads to the fewest number of linking errors, is flexible in adapting to changes in the characteristics of the files to be linked and provides metadata on the quality of a link.

The total linking weight is a reflection of the likelihood that two records belong to the same or to different persons. This weight guides the decision to categorise a pair of records as link or non-link. After the linking procedure, this meta-data has additional benefits. The likelihood of a (non-)link can be used to guide validation of the record linkage procedure, which was shown in chapter 4. Uncertainty about the linking status is known to be highest for record pairs with a linking weight around the threshold value and validation efforts can focus on this area. The total linking weight also presents the opportunity to take into account that in many situations either false links or false non-links have more negative consequences for the particular research question at hand. We can adjust the threshold value to obtain fewer links (by increasing the threshold) at the expense of more false non-links or lower the threshold in order to miss fewer matches but at the expense of more false links. Another possible application of the linking weight provided by the probabilistic approach is to weigh record pairs in the subsequent analysis and put more weight on record pairs linked with higher certainty (= higher linking weight). This possibility has not been applied in the studies described in this thesis.

*Complexity of the probabilistic record linkage method*

An argument against applying probabilistic record linkage techniques is its more complex structure and the fitting of statistical models. Latent class models are needed to estimate the probabilities for agreement on a linking variable among matches and among non-matches. Most statistical software packages however provide procedures to estimate these probabilities (for example SAS proc nlin). It is also possible to use commercially and freely available record linkage software that perform probabilistic techniques. The refinements of the probabilistic strategy presented in chapters 2 and 3 are not commonly available, but can easily be incorporated in existing algorithms. The probabilistic strategy calculates for each pattern of agreement and disagreement on the values of linking variables the
probability that it belongs to the same individual. The probability can be used to define a threshold above which patterns should be classified as links. The statistical background of probabilistic record linkage is sound, and the subjective part of linking is minimized. In linking the perinatal datasets a fully automated procedure was used, instead of using the linking weight to indicate the grey area for clerical review.\(^1\) Contrary to common belief subjectivity of judgement is larger within deterministic record linkage. In deterministic linkage, many ‘common sense’ decisions have to be made that are often guided by personal experience and/or belief in the quality of the available linking variables. This makes the deterministic strategy prone to error, time-consuming, and less sensitive in changes over time in the recording of data in the registries. The perinatal registers have been linked by others for the years 1995-2000 using a deterministic linking strategy with extensive manual checks.\(^3\) The results obtained from linking the perinatal registers for the year 2001 (see appendix) showed that a deterministic approach would missed several matches. The probabilistic linkage procedure for linking the Dutch perinatal registers is a reproducible and transparent method which has been described in detail in technical reports accompanying every linked annual dataset.\(^6\) The result of a probabilistic strategy is a collection of patterns with agreement and disagreement on linking variables that are accepted as link and non-link. Afterwards, this can be implemented as a deterministic strategy, which can be advocated if the underlying datasets have proven to be stable. Nevertheless, from the experience with the perinatal datasets – which are generally be considered as very stable – apparently small distributional changes may suggest changes of the linking procedure\(^7\), in particular if the number of linking variables is large.

**Record linkage in the future**

**Refinement of the linkage methodology**

With the growing number of medical and administrative registries, record linkage is often applied. However, publications on the validity and quality of linkage methods are less numerous. The studies in this thesis have shown that there is room for improvement of the linking methods currently in use. For that purpose it is important to communicate on the methodology and bring together researchers in the field to exchange ideas and experiences. There are some initiatives, for example the conference on ‘Exploiting Existing Data for Health Research’ held in Scotland in 2007 which will be repeated in 2009. Further refinements of the probabilistic linkage method should focus on the following aspects: value specific weights, handling of missing values and dependencies among errors. A value specific weight is based on the concept that the amount of evidence provided by agreement depends on the value of the variable, where more weight is given to agreement on a rare value.\(^1\) The application of value specific weights can help to get more information relevant for linkage out of the same data. This is not yet incorporated in the linking of the perinatal datasets. In addition, the handling of missing values in a record linkage procedure requires further attention. For linking the perinatal registers the linking weight was arbitrarily set to 0 (non-informative) in case of missing values.\(^8\) A missing value in records from both files might indicate a true unknown value and therefore provide information that two records belong to the same person (‘missing not at random’). Data entry errors in registry files complicate record linkage, but if typical errors occur than this knowledge can
be incorporated in a probabilistic strategy as was shown in chapter 2. Another area for improvement would be to analyse dependencies in the occurrence of errors in linking variables, for example there might be a higher time pressure when registering complex cases, resulting in several errors during data entry within the same period. In that case, the presence of one variable disagreeing within a match increases the probability that another variable will disagree within that match.

Requirements for combining registry data

In order to link two datasets they should have enough variables in common that are of sufficient quality and discriminative power. Some redundancy in this context is therefore desirable. Although all registries contain some identifying variables, the recording of such variables is not the goal when setting up a registry. While it seems natural to create multi-institutional, all institutes covering registries, it is preferable to start with improving current specific registries including the addition of high quality linking variables. The efforts to add information to improve the quality of record linkage are minimal (for example to register one extra variable) compared to the extreme demands to combine the data of several registries into one entity. This is difficult enough as was seen by the recurrent missing value of date of birth of the mother in the neonatal registry.

After the combination of data from different registries, observed discrepancies in values of overlapping variables have to be resolved. Decision rules can be applied based on external information on the quality of the datasets, based on focussed validation studies or the linking status can be used to resolve discrepancies. For example in linking the Dutch perinatal registries, the linking of a midwife record with a gynaecologist record means that a woman has been referred. This information is also captured in the registries itself (sometimes inconsistent between sources). The information captured in the registries was not always in agreement with the information from the linkage procedure. After external validation it was shown that the information from the record linkage procedure was correct in almost all cases (see also appendix). Another example is the linkage of two successive admissions of a child where information is recorded on the referring hospital. Validation of the internal linkage of the neonatal register of which the results are presented in chapter 4 also showed that the information obtained with record linkage was correct in almost all cases.

Patient identifier and privacy aspects

Record linkage techniques are applied in the absence of a unique identifier to recognize records as belonging to the same person. In the Netherlands, a unique number was introduced in November 2007: the Civil Service Number (BSN - Burger Service Nummer). On the first of June 2008 the BSN was introduced in health care and from the first of June 2009 onwards all institutional organisations (care providers, municipal health services, indication organs, health insurance companies) are obliged to use the BSN for exchange of patient data and therefore have to work with a reliable BSN. It is hitherto unclear whether registries can or should record the BSN, and, equally important, if the BSN (unchanged or transformed) may be used for quality of data purposes, including record linkage. The advantage of a unique identifier, if truly assigned uniquely to one person, is that it discriminates a person from all other persons in a country/region. Data entry errors can also occur when registering the unique identifier, and this will directly lead to a missed match, because using only one variable to bring records together is in essence a full
General discussion

deterministic approach. Especially when the identifier has a non-informative numeric format, the variable will be susceptible to entry or transcription errors. Identification numbers are therefore sometimes a combination of identifiers of the patient\textsuperscript{13,14} and can also be subjected to a specific algorithm to check the validity of the number.\textsuperscript{15} The latter is true for the BSN.

Some obvious disadvantages are related to legal identifiers such as BSN. Their assignment is often dependent on some official status as resident\textsuperscript{16} or health insurance member. In case of illegal immigrants or foreign visitors receiving health care a unique identifier will not be assigned and linkage using other identifying variables will be needed.\textsuperscript{17} In the context of the perinatal registries it is also vital to assign a BSN to both live born and foetal deaths if e.g. pathology information is collected and has to be combined. Currently the BSN is assigned at the moment a newborn is registered in the municipal administration (Gemeentelijke Basis Administratie – GBA), approximately 1 to 3 days after birth. Especially in case of twin pregnancies an earlier assignment would be helpful in separating the information from both children. While some obvious disadvantages exist in the context of the perinatal registries, a unique patient identifier should be considered as a very strong linking variable due to its high discriminating power. Combination of this identifier with other linking variables will strengthen the linking procedure as other variables can compensate for possible errors in the unique identifier.

National registries are often anonymous to accommodate privacy laws. Three types of data are being distinguished: anonymous data, personal data indirectly identifiable and personal data directly identifiable (name, address information or unique administration number).\textsuperscript{18} Of importance are the concepts of unicity in relation to a person and reducibility. Anonymous data in this context means ‘not reducible to a unique person’. Data can be unique for a person, for example the BSN or a DNA profile, but non-reducible, because a link is needed between the BSN or DNA profile and name and address information to find that person (= indirectly identifiable data). Record linkage of different sources can increase the reducibility, because different pieces of information on the same person are brought together. Record linkage of two datasets should not bring the data to a higher level (easier reducible), unless explicit permission is obtained from the person involved.\textsuperscript{18} Privacy legislation determines if a linking procedure is permitted. In the Netherlands, medical registries fall under the Dutch Personal Data Protection Act (‘Wet Bescherming Persoonsgegevens’). Permission of the woman to register her data in the current perinatal registries is included in the information brochure handed out by the care provider at the first prenatal visit.

Role of Statistics Netherlands in use of existing national registers

In the Netherlands, Statistics Netherlands (Centraal Bureau voor de Statistiek – CBS) is officially responsible for collecting and processing data in order to publish statistics that can be used by policymakers and for scientific research.\textsuperscript{19} In that role, Statistics Netherlands assembles registry data as input for their statistics on a legal basis (‘Wet op het Centraal Bureau voor de Statistiek’). Large administrative databases in the Netherlands include the Municipal Register (Gemeentelijke Basis Administratie – GBA) and the causes of death register (Doodsoorzaken Registratie). The National Hospital Discharge Register (Landelijk Medische Registratie – LMR) is another national registry containing information on all hospital admissions in the Netherlands. Several linking projects have been undertaken to link specific registry data to one of these datasets.\textsuperscript{20,21}
The developed linkage procedures for the perinatal registers were applied for the linkage of the GBA and the vital statistics register with the perinatal registers at the office of the CBS. The objective of this study was to obtain 1 month (late neonatal mortality) and 1 year mortality (infant mortality) figures by gestational age and birth weight (chapter 2). In the pilot study the probabilistic method was the preferred method compared to the standard CBS method (stepwise deterministic linkage) used for record linkage. Transparency on the methods used to combine existing data sources is important to value the resulting figures. It is preferable to separate the methodology of record linkage from the application of the combined data. Experts should be involved in developing and improving the methodology of linking existing data sources and in answering relevant questions enabled by the combined data sources. For optimal use of existing data sources in the Netherlands, anonymous (non-reducible) data should be widely available for research purposes in addition to their use for national statistics.

**Perinatal care data**

*Added value of combined perinatal registry data*

Healthcare is a multidisciplinary field and awareness grows that many medical and public health issues require longitudinal data. Multidisciplinary care increases the need for exchange of information between the different involved care providers, and long-term data usually involve several institutional units. Within one institution the exchange of information can be relatively easy, but it becomes more difficult in a multidisciplinary setting, especially if care is divided over primary, secondary and tertiary care. For evaluation of the care chain as a whole information on every step in the chain has to be available, taking care for proportionality: population resistance grows on linking all possible data, as is e.g. visible in electronic patient record (EPR) discussions and discussions on the creation of one all-encompassing dossier for every child.

*Characteristics of registry data*

The goal of a medical registry is to monitor and to improve the quality of care, by evaluating the provided care and to stimulate the implementation of new insights based hereon. Advantages of the use of registry data are that the data are already available, have been collected in a structured way and are often available over a large time span with broad coverage. On the other hand, registries are often limited in the number and detail of captured determinants and are susceptible to data entry errors. The registry data from the registers by midwives, gynaecologists and neonatologists have been linked into a single national perinatal register with information on pregnancy, childbirth and the postpartum period. The data captured in the registers focuses on the interventions conducted by the involved caregivers and on the referral pattern between the different disciplines.

*Completeness perinatal registers*

Participation of practices in the LVR1 registry is 97%, in the LVR2 registry 99% and in the LNR registry 68% (NICU hospitals 100%). For the years 2000 until 2006, no data were available for record linkage from the general practitioners providing obstetric care
General discussion

(estimated to be less than 4%). General practitioners started registering in a (regional) pilot study from 2003 onwards. Overall the completeness of the PRN registry for births is estimated to be around 96%.

The validation studies showed that the linkage procedure gave less than 1% error. The false non-links identified for the LVR1 – LVR2 linkage were mainly caused by LVR1 records without child information (the pregnant woman was referred to a gynaecologist before the child was born and no information was received thereafter). Records with no child information were excluded from the analysis to avoid double counting.

Linkage can be more difficult for complex cases, which was clear from the linkage of readmissions of multiple births in chapter 4. Multiple births are difficult in record linkage because they share the same information from the mother. A multiple pregnancy is an indication for referral and should always be supervised by a gynaecologist. Therefore, records from multiple births in the LVR1 register that could not be linked to records in the LVR2 register were excluded from the analysis to prevent double counting. Participation in the LNR register is only complete for around 68% and analyses on neonatal admissions based on PRN data might be biased.

In case of foetal or early neonatal mortality a gynaecologist or neonatologist will most likely be involved. The LVR2 registry is near to complete and NICU hospitals all participate in the register. Therefore completeness of mortality cases is probability higher than the overall completeness of births. When the registry data were compared with civil registration data, the quality of the outcome mortality was high. Foetal deaths were more often registered in the PRN register compared to the civil registration, especially the very preterm foetal deaths. Because mortality is an important outcome measure, the validation studies focussed on subgroups with these cases. The validation of the LVR1 – LVR2 linkage showed that within these subgroups the linking procedure was correct in all cases (no false non-links in case of mortality).

The validation of the linkage of readmissions in the LNR registry has also shown that in case of mortality the linkage was correct for singletons (chapter 4).

Research with combined perinatal care data

As the result of the linkage procedure, the combined perinatal care data are available for the years 2000 until 2007. National perinatal registry data can be used to provide insight into the Dutch perinatal care process and its outcomes. The data from the Netherlands Perinatal Registry are used for feedback to the care providers, for research purposes and as input for policymakers. An example of feedback to the care providers is the publication of the annual national report ‘Perinatal care in the Netherlands’ by the Netherlands Perinatal Registry with detailed information on pregnancies and childbirths. In addition, the perinatal registry data are used for numerous epidemiological studies in the Netherlands. The increased use of the perinatal registry data is also visible from the increase in requests submitted to the Netherlands Perinatal Registry for use of the data (79 in 2003 to 144 in 2008). The linked data enables the analyses with respect to important indicators of perinatal care in the Netherlands, such as foetal, neonatal and perinatal mortality, but also maternal age, parity, mode of delivery and gestational age and birth weight. Moreover, with the combined perinatal registry, insight can be provided in the referral pattern between the different disciplines involved in perinatal care. An example is a study focussing on low risk pregnancies with special attention for home births. Registry data is suitable for describing
General discussion

the incidence, trend and variation of interventions and outcomes. The trend and regional variation of the important indicator perinatal mortality were presented in chapter 5 and 6. The combined perinatal registry data were also used to compare the level of foetal and early neonatal mortality between European countries, which showed a persistent unfavourable position for the Netherlands.\textsuperscript{33,36} A higher level of perinatal mortality not necessarily means that the provided care is of less quality. Explanatory factors mentioned to explain the unfavourable Dutch position are the current Dutch prenatal screening policies and subsequent termination and the restrictive policies for resuscitation of very preterm born children.\textsuperscript{29} The screening policy changed during the period 2000-2006 with the inception of the national prenatal ultrasound screening program for structural abnormalities (around 20 weeks of gestation) in the period 2004/2005 which was consolidated in 2006. The perinatal management of very preterm infants is the focus of the PRE PRE study, of which the data were recently linked with the PRN 2007 data.\textsuperscript{37}

Perinatal mortality is a rather rare outcome in western countries, preterm birth and low birth weight can be seen as mediating outcomes which are influenced by the same risk factors and are risk factors themselves for perinatal mortality. Analysing trends or variations in perinatal mortality in risk groups defined by these mediating factors can help to gain further insight especially in the role of care factors dedicated to specific groups (for example tertiary care for very preterm births). However, understanding the determinants of the trends and the variations in perinatal mortality is difficult because of the many potential explanatory factors. Registering the cause of death in case of mortality and focused audit studies could provide more insight.

Limitations of registry data

Research on other determinants like ethnicity, neighbourhood/urbanisation and on specific clinical risk groups (e.g. second stage arrest) is conducted with the data from the perinatal registry. Although the combined perinatal registry data offer a rich resource for epidemiological research, there are also some limitations to the use of registry data. Registries are limited in the number of captured determinants, especially determinants that occur earlier in time, and are less suitable for evaluation of therapy/interventions.\textsuperscript{23} Hypotheses generated through analysis of registry data should be tested in clinical trials or analysed in more detail using trial or cohort data. This is true for questions like which intervention is preferred in a specific situation or the underlying causes of higher perinatal mortality among ethnic groups. Some important determinants known to be related to perinatal outcomes are not available in the current registries, like information on maternal education, social economic status, smoking behaviour, BMI and dietary habits. Although known determinants can be included in the primary registry dataset, other types of research are needed to find unknown determinants related to perinatal outcomes.

Perinatal registry: the future

Extending perinatal care data

The availability of the national perinatal registry data also offers possibilities to link large cohort studies or clinical trials with registry data.\textsuperscript{38} An example of such a linkage was presented in chapter 8 for the ABCD cohort study to obtain information on their non-responders. National registry data can also be used to find controls to match cases in case-
General discussion

control studies (e.g. the PPP study – ‘progesterone to prevent premature birth’). Success of the linking procedure will depend on the quality and amount of information in common between the datasets.

Pregnancy, childbirth and the postpartum period form a well defined episode of care. There is a growing awareness that longitudinal data are required for many medical and public health issues. Therefore, the period preceding and following the perinatal period is also of interest. Data on preconception care, screening and applied assisted reproduction techniques (Landelijke Infertiliteits Registratie – LIR) are also of interest for outcomes in perinatal care and could be linked to the perinatal registries. Following the perinatal period, information from child health centres, vital statistic, follow-up information of NICU children and the electronic child record (‘elektronisch kinddossier’ – EKD) could be used to determine long-term outcomes of perinatal care.

Moreover, a pregnancy is an event that can occur more than once for a woman/couple. Ideally, the information on all pregnancies of a woman is available for analyses. Complications occurring during the first pregnancy of a woman can influence the outcome of a subsequent pregnancy. Longitudinal linkage of the perinatal registries has not been conducted so far. The data captured on the mother in the current registries is limited (date of birth, postal code and gravidity) which will hamper successful record linkage.

The new perinatal register

The Foundation of the Netherlands Perinatal Registry has initiated the creation of a new perinatal register where all involved caregivers register information in the same register. The process started with the definition of the new dataset by the PRN Taskforce that was agreed upon by all involved disciplines. The specifications of the new dataset are being implemented in the software used by care providers to register their data and send it to the national registry since 2008. The revised dataset also incorporates determinants not available in the current registries such as maternal education, smoking behaviour, alcohol intake, folic acid intake and country of origin of both mother and father and motivations for important interventions like caesarean sections. The definition of one perinatal dataset for all involved disciplines does not solve the problem of having to combine the data of one woman/child registered by the different caregivers afterwards. Registering more identifiable information on the woman and her child(ren), including a national identification number, could help to make the linkage process easier but linkage will remain essential.

Another important development in this light is the announced introduction of a national electronic patient record (EPR) in the Netherlands. The EPR will be introduced to exchange medical data between care providers. Exchange of information by the national EPR will be based on the BSN number. In the future the new perinatal registry should lead to the electronic exchange of information between the involved caregivers. New problems that may arise with this structure include the timing of data exchange and the responsibility for registered items and for mutation of already entered data. Two non-trivial distinctions exist between registry data for public health research and clinical data needed in the direct care process. For the latter, identification of the patient is essential while for research purposes data should only be unique and preferably non-reducible. Also, data should be direct available for clinical purposes, while research can work with data after a time lag. In addition to exchange of information between involved caregivers a subsequent processing step should be included to prepare the data for analysis after all information is captured.
General discussion

The question if perinatal care has indeed improved by higher quality perinatal care data still remains to be answered. Further research should focus on how the information is used in daily clinical practice and which new research directions arise by using the data. Focused studies should be conducted to measure the effect of a specific intervention based on increased knowledge obtained from the registry data.

Concluding remarks

Medical record linkage improves the use of existing data as new research directions can be covered by the combination of existing data sources. The registry data of the different disciplines involved in perinatal care have been successfully combined. Probabilistic record linkage is the preferred method to combine datasets because it leads to the fewest linking errors, is most flexible and provides meta-data on the quality of a link. The validity of outcomes generated with linked data is dependent on the quality of the linking procedure, which should always be confirmed in a validation study. The combined perinatal registry data enables the evaluation of the quality of perinatal care in the Netherlands. Evaluation of care based on registry data is essential in everyday clinical practice, especially with the developments in healthcare towards transparency of the quality of care by performance indicators. The combined perinatal registry is a valuable source for epidemiologic research aiming to improve perinatal care. Detailed questions about the exact contribution of a specific factor will often require additional data from primary studies.
References

31 Visser GHA, Steegers EAP. Beter baren Nieuwe keuzen nodig in de zorg voor zwangeren. Medisch Contact 2008;65(3):96-100.