Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

# Validity of a hospital-based obstetric register using medical records as reference

This article was published in the following Dove Press journal:
Clinical Epidemiology
26 November 2015
Number of times this article has been viewed



Carina Sjöberg Brixval[1]
Lau Caspar Thygesen[1]
Nanna Roed Johansen[2]
Christina Rørbye[3]
Tom Weber[3]
Pernille Due[1]
Vibeke Koushede[1]

[1]National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark; [2]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [3]Copenhagen University Hospital, Department of Obstetrics and Gynecology, Hvidovre Hospital, Hvidovre, Denmark

**Background:** Data from hospital-based registers and medical records offer valuable sources of information for clinical and epidemiological research purposes. However, conducting high-quality epidemiological research requires valid and complete data sources.

**Objective:** To assess completeness and validity of a hospital-based clinical register – the Obstetric Database – using a national register and medical records as references.

**Methods:** We assessed completeness of a hospital-based clinical register – the Obstetric Database – by linking data from all women registered in the Obstetric Database as having given birth in 2013 to the National Patient Register with coverage of all births in 2013. Validity of eleven selected indicators from the Obstetric Database was assessed using medical records as a golden standard. Using a random sample of 250 medical records, we calculated proportion of agreement, sensitivity, specificity, and positive and negative predictive values for each indicator. Two assessors independently reviewed medical records and inter-rater reliability was calculated as proportion of agreement and Cohen's κ coefficient.

**Results:** We found 100% completeness of the Obstetric Database when compared to the Danish National Patient Register. Except for one delivery all 6,717 deliveries were present in both registers. Proportion of agreement between the Obstetric Database and medical records ranged from 91.1% to 99.6% for the eleven indicators. The validity measures ranged from 0.70 to 1.00 indicating high validity of the Obstetric Database. κ coefficients from the inter-rater reliability ranged from 0.71 to 1.00.

**Conclusion:** Completeness and validity of the Obstetric Database were found acceptable when using the National Patient Register and medical records as golden standards. The Obstetric Database therefore offers a valuable source for examining clinical, administrative, and research questions.

**Keywords:** obstetric register, register-based, hospital register, validity, completeness

## Introduction

In Denmark, approximately 60,000 children are born each year. During the past years, the proportion of interventions in the birth process has increased, eg, the rate of epidural analgesia has increased from 18% in 2006 to 24% in 2013.[1] Monitoring prevalence and time trends in health outcomes and medical procedures requires valid and complete data sources. All residents in Denmark are included in the Danish health registers and accurate linkages are possible due to the unique personal identification number[2] making Denmark a suitable setting for register-based research.

The advantages of register-based research is the representativeness of the study population, that risk of diagnostic process is not affected by the research question, and that data already exist which minimize time consumption and costs. A disadvantage

Correspondence: Carina Sjöberg Brixval
National Institute of Public Health, University of Southern Denmark,
Øster Farigmagsgade 5A,
1353 Copenhagen, Denmark
Tel +45 6550 7845
Email cabr@niph.dk

of using register-based data for research purposes is that data collection and data quality are not under control of the researcher and information on more specific and detailed treatment procedures of clinical interest present in hospital-based clinical registers is often not included in national registers.[3,4] Therefore, clinical registers are often necessary when conducting clinical epidemiological research.

Validation of register-based data is necessary to ensure the quality of the data. A well-conducted validation study includes sample size calculations[5] and comparison of information in the given register with information in other registers as well as primary data sources, eg, medical records.[6] The use of two assessors to extract data when validating information from registers against medical records reduces the risk of registration error. An examination of the inter-rater reliability will also give an indication of how difficult a given indicator in the medical record is to assess.

The Obstetric Database at Hvidovre Hospital has been used for internal monitoring of prevalence of obstetric interventions as well as research and high validity is therefore essential. The Obstetric Database has not been validated previously and the aim of this study was therefore to assess completeness and validity of the Obstetric Database using the Danish National Patient Register and medical records as references. Furthermore, the detailed description of the validation process in this paper may be helpful to fellow researchers or clinicians wanting to examine completeness and validity of a clinical register.

# Methods

We assessed the completeness and validity of the Obstetric Database using three data sources; the National Patient Register, the Obstetric Database, and medical records. Assessment of completeness of the Obstetric Database was performed by comparing information on the presence of the unique personal identification number in the National Patient Register and the Obstetric Database. The validity of the Obstetric Database was assessed by comparing information on the presence of selected indicators in the Obstetric Database and medical records.

## Registers

The Obstetric Database is a hospital-based clinical register comprising initial obstetric and neonatal data, eg, Apgar score, obstetric interventions and outcomes from all deliveries at the three largest delivery wards (Herlev Hospital, Rigshospitalet, and Hvidovre Hospital) in the Capital Region of Denmark. At Hvidovre Hospital, the database was established in 1996. We selected Hvidovre Hospital to assess validity of the Obstetric Database as this is the largest labor ward in Denmark with more than 6,500 deliveries each year. During and after labor, midwives register the obstetric and neonatal baseline data and interventions in the database by ticking an electronic list. Postpartum, a specialist doctor or senior midwife goes through every file and adds left out information and supplies specialist diagnoses based on information from the medical records.

The Danish National Patient Register was established in 1977 and records in- and outpatient contacts from all hospitals in Denmark. The register contains information on the dates of admission and discharge, and information on diagnoses and major clinical procedures performed at hospitals. The purpose of the National Patient Register is among others to provide information for the production of statistical data and serve as the basis for the payment of hospitals via the Diagnostic Related Group system.[7]

## Completeness assessment

We assessed the completeness of the Obstetric Database by using the National Patient Register as a gold standard. Reporting to the National Patient Register is compulsory to obtain reimbursement from health authorities for patient contacts and we therefore anticipate this register to be complete. Completeness of the Obstetric Database was defined as the percentage of deliveries in the Obstetric Database that were also registered in the National Patient Register. Registrations to the National Patient Register are performed by the secretaries at the Department of Obstetrics and Gynecology at Hvidovre Hospital based on information from the Obstetric Database. For the assessment of completeness of the Obstetric Database, we selected data on all deliveries planned to be performed at Hvidovre Hospital in 2013. In the Obstetric Database, all planned deliveries at a given hospital are registered under that hospital regardless of whether or not they actually ended up being carried out there. Stillbirths were also included. Twin- and triplet births counted as one delivery. All deliveries at Hvidovre Hospital in the Obstetric Database and the National Patient Register were linked by the unique personal identification number.

## Validity of selected indicators

The validity of the Obstetric Database was assessed using medical records as gold standard. For the purpose of this study, the following eleven indicators were considered of most importance to be validated: use of oxytocin due to dystocia, epidural analgesia, vacuum extraction, emergency and elective cesarean

delivery, perineal tear – degree 3 and 4, test for gestational diabetes, scalp blood pH, arterial pH from the umbilical cord, and Apgar score after 5 minutes. We assessed the presence of the indicators in the Obstetric Database and medical records. Three of the indicators are continuous measures (scalp blood pH, arterial pH from the umbilical cord, and Apgar score after 5 minutes). We did not assess the agreement of the values of these indicators between the Obstetric Database and the medical records. Some of the indicators are obstetric interventions such as administration of oxytocin and cesarean delivery only performed among patients with complications; others are routine measurements such as arterial pH from the umbilical cord and Apgar score. Tests for gestational diabetes are performed only among pregnant women with an increased risk of diabetes.

Sample size calculations were based on the primary measure of validity; positive predictive value. We expected a valid registration of 95% between the Obstetric Database and medical records and wanted to estimate this with a confidence interval of 92%–98%. In order to fulfill this, a sample of 203 deliveries was required. To take into account that data on certain outcomes might be incomplete due to, eg, stillbirth, and further that agreement between some outcomes could potentially be lower than the expected 95%, a random sample of 250 deliveries in 2013 was selected from the Obstetric Database and the corresponding electronic medical records were retrieved. We excluded two deliveries not performed at Hvidovre Hospital due to transfers to other hospitals during labor and one delivery due to missing data on all indicators for unknown reasons leaving 247 deliveries for validation. These exclusions were necessary due to lack of information in the medical records for these women and their children. For the assessment of validity of arterial pH from the umbilical cord and Apgar score after 5 minutes, we additionally excluded three records due to stillbirths leaving 244 deliveries for validation of these indicators.

Blinded for information in the Obstetric Database, two of the authors (NRJ and CSB) independently assessed the medical records for registration of the aforementioned indicators. In cases where Apgar score or arterial pH from the umbilical cord was not present in the mother's medical record, these indicators were assessed from the child's medical record. Assessment of performance of test for gestational diabetes was estimated from an overview of blood test results found in a separate section of the medical record. Data from each assessor was entered into separate Excel sheets. In cases of doubt as to whether an obstetric intervention had been performed, the authors consulted two skilled obstetricians (CR and TW) for clarification independently of one another. Next, the datasets from the two assessors were compared and any disagreements were solved by an obstetrician (CR) who was blind to the assessor. The final dataset was linked to data in the Obstetric Database.

## Statistical analyses

Sensitivity, specificity, positive and negative predictive values, as well as proportion of agreement were calculated for each of the eleven indicators. We calculated exact 95% binomial proportion confidence intervals. Definitions of the validity measures are given in Table 1.

Inter-rater reliability between the two assessors after consultation with the obstetricians was calculated as proportion of agreement as well as Cohen's κ coefficient for each of the indicators. We used the Landis and Koch's scale[8] to categorize strength of agreement from the κ coefficients.

All statistical analyses were performed using SAS version 9.3, SAS Institute Inc., Cary, NC, USA.

## Ethical issues

This study fulfills all Danish ethical standards and was approved by the Danish Data Protection Agency (No 2014-54-0714) and by the Department of Obstetrics and Gynecology at Hvidovre Hospital.

# Results

## Completeness

In the Obstetric Database, 6,718 deliveries were registered in 2013, whereas 6,717 deliveries were registered in the Danish National Patient Register. When linking data from the Obstetric Database and the National Patient Register,

**Table 1** Definition of measures of validity

| Obstetric Database | Medical record (gold standard) | | Total |
|---|---|---|---|
| | Present | Absent | |
| Present | a | b | a + b |
| Absent | c | d | c + d |
| Total | a + c | b + d | |

**Notes:** The sensitivity is the proportion of patients with registration of the indicator according to both medical records and the Obstetric Database (a), compared to all patients with the indicator according to medical records (a + c) = True positive (a)/(True positive [a] + false negative [c]). The specificity is the proportion of patients without registration of the indicator according to both medical records and the Obstetric Database (d), compared to all patients without the indicator according to medical records (b + d) = True negative (d)/(True negative [d] + false positive [b]). The positive predictive value is the proportion of patients with registration of the indicator according to both medical records and the Obstetric Database (a), compared to all patients with the indicator according to the Obstetric Database (a + b) = True positive (a)/(True positive [a] + false positive [b]). The negative predictive value is the proportion of patients without registration of the indicator according to both medical records and the Obstetric Database (d), compared to all patients without the indicator according to the Obstetric Database (c + d) = True negative (d)/(True negative [d] + false negative [c]).

**Table 2** Number of registrations in the Obstetric Database and in medical records and the proportion of agreement (%) for each indicator

| Indicator | In the Obstetric Database | | Not in the Obstetric Database | | Proportion of agreement, %, (95% confidence interval) |
|---|---|---|---|---|---|
| | In medical records | Not in medical records | In medical records | Not in medical records | |
| Oxytocin due to dystocia | 46 | 20 | 2 | 179 | 91.1 (86.8–94.3) |
| Epidural analgesia | 68 | 0 | 3 | 176 | 98.8 (96.5–99.8) |
| Vacuum extraction | 21 | 0 | 2 | 224 | 99.2 (97.1–99.9) |
| Emergency cesarean delivery | 35 | 0 | 1 | 211 | 99.6 (97.8–100.0) |
| Elective cesarean delivery | 23 | 1 | 1 | 222 | 99.2 (97.1–100.0) |
| Perineal tear degree 3 | 8 | 1 | 0 | 238 | 99.6 (97.8–100.0) |
| Perineal tear degree 4 | 0 | 0 | 0 | 247 | 100.0 (100.0–100.0) |
| Scalp blood pH | 69 | 1 | 0 | 177 | 99.6 (97.8–100.0) |
| Arterial pH from the umbilical cord | 221 | 3 | 3 | 17 | 97.9 (94.7–99.1) |
| Apgar score after 5 minutes | 241 | 2 | 1 | 0 | 98.8 (96.5–99.8) |
| Test for gestational diabetes | 84 | 6 | 9 | 148 | 93.9 (90.2–96.6) |

6,717 deliveries were present in both data sources. No deliveries present in the National Patient Register were missing in the Obstetric Database and only one delivery was present in the Obstetric Database but not in the National Patient Register, indicating almost exact agreement (rounded to 100%) between the two registers.

## Validity of indicators

The proportion of agreement between the Obstetric Database and medical records was high for most indicators (Table 2). For nine of the indicators, the proportion of agreement was 97.9% or above. Oxytocin due to dystocia (91.1%) and test for gestational diabetes (93.9%) had lower proportions of agreement.

Sensitivity for all indicators was high and ranged from 0.90 (test for gestational diabetes) to 1.00 (perineal tear degree 3, scalp blood pH, and Apgar score after 5 minutes) (Table 3). Also, specificity was high and ranged from 0.85 (arterial pH from the umbilical cord) to 1.00 (epidural analgesia, vacuum extraction, emergency and elective cesarean delivery, and perineal tear degree 3).

The predictive values were generally high. Except for oxytocin due to dystocia, the positive predictive values ranged from 0.89 (perineal tear degree 3) to 1.00 (epidural analgesia, vacuum extraction, and emergency cesarean delivery). However, the results revealed false positive registrations of the indicator oxytocin due to dystocia in the Obstetric Database resulting in a positive predictive value of 0.70.

Negative predictive values ranged from 0.85 (arterial pH from the umbilical cord) to 1.00 (emergency and elective cesarean delivery, perineal tear degree 3, and scalp blood pH).

## Inter-rater reliability

Proportion of agreement between the two assessors ranged from 94.3% (oxytocin due to dystocia) to perfect agreement

**Table 3** Sensitivity, specificity, positive, and negative predictive values (95% confidence interval) for eleven indicators in the Obstetric Database

| Indicator | Sensitivity | Specificity | Positive predictive value | Negative predictive value |
|---|---|---|---|---|
| Oxytocin due to dystocia | 0.96 (0.86–0.99) | 0.90 (0.85–0.94) | 0.70 (0.57–0.80) | 0.99 (0.96–1.00) |
| Epidural analgesia | 0.96 (0.88–0.99) | 1.00 (0.98–1.00) | 1.00 (0.95–1.00) | 0.98 (0.95–1.00) |
| Vacuum extraction | 0.91 (0.72–0.99) | 1.00 (0.98–1.00) | 1.00 (0.84–1.00) | 0.99 (0.97–1.00) |
| Emergency cesarean delivery | 0.97 (0.85–1.00) | 1.00 (0.98–1.00) | 1.00 (0.90–1.00) | 1.00 (0.97–1.00) |
| Elective cesarean delivery | 0.96 (0.79–1.00) | 1.00 (0.98–1.00) | 0.96 (0.79–1.00) | 1.00 (0.98–1.00) |
| Perineal tear degree 3 | 1.00 (0.63–1.00) | 1.00 (0.98–1.00) | 0.89 (0.52–1.00) | 1.00 (0.98–1.00) |
| Perineal tear degree 4[a] | – | – | – | – |
| Scalp blood pH | 1.00 (0.95–1.00) | 0.99 (0.97–1.00) | 0.99 (0.92–1.00) | 1.00 (0.98–1.00) |
| Arterial pH from the umbilical cord | 0.99 (0.96–1.00) | 0.85 (0.62–0.97) | 0.99 (0.96–1.00) | 0.85 (0.62–0.97) |
| Apgar score after 5 minutes[b] | 1.00 (0.98–1.00) | 0 | 0.99 (0.97–1.00) | 0 |
| Test for gestational diabetes | 0.90 (0.82–0.95) | 0.96 (0.92–0.99) | 0.93 (0.86–0.98) | 0.94 (0.89–0.97) |

**Notes:** [a]No perineal tear degree 4 was registered in either Obstetric Database or medical records. Therefore, statistics are not presented for this indicator; [b]specificity and negative predictive value equals 0 as Apgar score after 5 minutes was always registered in either Obstetric Database or the medical record or both (no true negative).

**Table 4** Prevalence of indicators in medical records by each assessor, inter-rater agreement (%) and κ coefficients (95% confidence interval) for each indicator

| Indicator | Assessor 1 | Assessor 2 | % agreement | κ coefficient |
|---|---|---|---|---|
| Oxytocin due to dystocia | 51 | 57 | 94.3 | 0.83 (0.75–0.92) |
| Epidural analgesia | 71 | 71 | 100 | 1.00 (1.00–1.00) |
| Vacuum extraction | 23 | 24 | 99.6 | 0.98 (0.93–1.00) |
| Emergency cesarean delivery | 34 | 36 | 99.2 | 0.97 (0.92–1.00) |
| Elective cesarean delivery | 26 | 24 | 99.2 | 0.96 (0.89–1.00) |
| Perineal tear degree 3 | 7 | 7 | 98.4 | 0.71 (0.43–0.98) |
| Perineal tear degree 4[a] | 0 | 1 | 99.6 | – |
| Scalp blood pH | 68 | 69 | 99.6 | 0.99 (0.97–1.00) |
| Arterial pH from the umbilical cord | 227 | 228 | 97.1 | 0.77 (0.61–0.94) |
| Apgar score after 5 minutes | 242 | 241 | 99.6 | 0.80 (0.41–1.00) |
| Test for gestational diabetes | 85 | 93 | 95.1 | 0.89 (0.84–0.95) |

**Note:** [a]It was not possible to calculate the κ coefficient for perineal tear degree 4 due to no registrations for assessor 1.

of 100% for epidural analgesia (Table 4). κ coefficients ranged from 0.71 for perineal tear degree 3 to a perfect agreement of 1.00 for epidural analgesia. Using perineal tear degree 3 as an example: although both assessors noted seven events only five of these seven events were the same, resulting in an agreement of 98.4%.

## Discussion

We examined completeness and validity of a hospital-based clinical register at the largest birth site in Denmark.

Completeness was assessed by comparing data from all women registered in the Obstetric Database as having given birth in 2013 and linking to the National Patient Register which was considered a gold standard. We found that all deliveries registered in the National Patient Register were also registered in the Obstetric Database, giving a completeness of 100%. One delivery was not registered in the National Patient Register for unknown reason. Registrations to the National Patient Register are based on information from the Obstetric Database. Reporting to the National Patient Register is compulsory to obtain reimbursement from health authorities for patient contacts and we therefore considered this register as a gold standard. This study supports that the Obstetric Database is used very actively in the clinical practice and that no deliveries therefore are missing.

We used medical records as gold standard when assessing validity of the Obstetric Database and found that sensitivity, specificity, and predictive values generally were high for the selected eleven indicators indicating high validity of the database.

For all indicators sensitivity and specificity was high (≥0.91 and ≥0.85 respectively). This implies that the Obstetric Database has high validity regarding registration from the medical record. Also, the predictive values were generally very high indicating a high probability that the registrations in the Obstetric Database are correct.

A previous systematic review of perinatal validation studies have shown that indicators related to type of delivery and perineal tear are well reported with high sensitivities and positive predictive values, whereas induction and augmentation of labor have higher degrees of underreporting.[9] The results from the present study are thus in accordance with former validity studies in the obstetric field.

Although the proportion of true positive results (the positive predictive value) in the Obstetric Database was high for almost all indicators, 20 cases of oxytocin due to dystocia were registered in the Obstetric Database but not in the medical records. In all these instances, use of oxytocin occurred as part of induction of labor according to the medical records. According to the Danish guidelines for registration of obstetric interventions, oxytocin administration should only be coded as induction if it is used as the first procedure for induction of labor. If oxytocin is administered as a secondary induction procedure it is coded as "due to dystocia".[10] In two of the aforementioned 20 cases, oxytocin was registered in the Obstetric Database as induction, while ten others were registered as induced with Misoprostol before treatment with oxytocin. The remaining eight were not registered as induced. During recent years, the registration practice has changed. Previously, indications for oxytocin administration were registered separately for induction of labor and for dystocia.[10] While the former registration practice provided an opportunity for assessing oxytocin due to dystocia and oxytocin as induction procedure separately, this is no longer possible due to the current registration practice. This implies a potential risk of misinterpretation of data if one wishes to study oxytocin due to dystocia.

Validity was also lower for test for gestational diabetes. Six women were registered with the test in the Obstetric Database but not in the medical record. If the person entering data into the Obstetric Database did not check whether the test was actually performed, they may have falsely registered tests based on recorded indication alone. The nine tests for gestational diabetes registered in the medical record but not in the Obstetric Database have most likely been overlooked by the person entering data into the Obstetric Database. This could be due to complex registration systems, ie, that information has to be found in separate sections of the medical records.

The κ coefficients for the inter-rater reliability was above 0.70 for all the indicators, and is therefore considered either "substantial" or "almost perfect" according to the Landis and Koch categorization.[8] In cases of disagreements between the two assessors the decision was made by a skilled obstetrician. We therefore consider the reliability of the review of the medical records to be adequate and the medical records to be valid as gold standard. The κ coefficients were lower for the indicators that also had low predictive values (oxytocin due to dystocia, perineal tear degree 3, and arterial pH from the umbilical cord) indicating that these indicators were generally more difficult to assess.

Strengths of this study include the use of the National Patient Register with national coverage as well as medical records as gold standards. Further, the extensive review of medical records was performed by two independent assessors and approved by two independent clinical experts. The high agreement between the two assessor's registrations confirms that use of the medical records as gold standards was appropriate. The random sample selected among women giving birth at Hvidovre Hospital makes these results generalizable to all deliveries at Hvidovre Hospital in this period. The results might not be generalizable to other clinical databases at other hospitals as registration practices might vary between hospitals. However, the registration guidelines for the obstetric coding apply throughout the entire country which suggests that the results may be generalizable to other clinical databases.

We assessed whether the indicators were present in the medical records and the Obstetric Database. The accuracy of the values of scalp blood pH, arterial pH from the umbilical cord, and Apgar score was not assessed. Therefore, further validation of the accuracy of these indicators will be desirable before using them for research or administrative purposes.

Both the issue of using the Danish National Patient Register as gold standard and the reduced validity of a few of the indicators stress the importance of careful consideration and evaluation of the completeness and validity of the different components of registers.

In conclusion, completeness and validity of the selected indicators in the Obstetric Database are high. With data being valid and the database complete, the Obstetric Database offers a valuable source for monitoring prevalence of obstetric interventions and outcomes as well as obstetrical research studies. However, when monitoring use of oxytocin due to dystocia, care should be taken as the code for this obstetric intervention might also cover oxytocin used as part of induction of labor.

The detailed description of the validation process may be helpful to fellow researchers or clinicians wanting to examine completeness and validity of a clinical register.

## Acknowledgments

This study was conducted as part of a PhD study funded by the Danish Cancer Society. The Danish Cancer Society does not have any influence in study design, in the collection, analysis, and interpretation of data, in the writing of the manuscript, or in the decision to submit this manuscript.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Statens Serum Institut. Sygehusfødte og komplikationer [Deliveries at hospitals and complications] [database on the Internet]; 2014. Available from: http://www.esundhed.dk/sundhedsregistre/MFR/Sider/MFR03.aspx. Accessed January 22, 2015. Danish.
2. Thygesen LC, Daasnes C, Thaulow I, Bronnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. *Scand J Public Health*. 2011;39(7 Suppl):12–16.
3. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996;25(2):435–442.
4. Thygesen LC, Ersboll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. *Eur J Epidemiol*. 2014;29(8):551–558.
5. Liao JJ. Sample size calculation for an agreement study. *Pharm Stat*. 2010;9(2):125–132.
6. Krebs L, Langhoff-Roos J. Validation of registries: a neglected, but indispensable investment. *Paediatr Perinat Epidemiol*. 2014;28(5):351–352.
7. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(7 Suppl):30–33.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
9. Lain SJ, Hadfield RM, Raynes-Greenow CH, et al. Quality of data in perinatal population health databases: a systematic review. *Med Care*. 2012;50(4):e7–e20.
10. Dansk Selskab for Obstetrik og Gynækologi. Anbefalede obstetriske koder [Recommended coding for obstetrics] [report on the Internet]; 2013. Available from: http://static.squarespace.com/static/5467abcce4b056d72594db79/546e7741e4b0d969a4f6c8e0/546e7745e4b0d969a4f6ca68/1385584564000/Obstetriske-diagnoser.pdf?format=original. Accessed January 26, 2015. Danish.