
Paediatric and Perinatal Epidemiology

doi: 10.1111/ppe.12146

Commentary

Validation of Registries: A Neglected, but Indispensable Investment

Lone Krebs,[a] Jens Langhoff-Roos[b]

[a]*Department of Obstetrics and Gynecology, Holbaek Hospital, University of Copenhagen, Holbaek*

[b]*Department of Obstetrics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark*

The national registries of the Nordic countries have a high international reputation. The medical birth registries are considered a 'goldmine for clinical research' for registry-based epidemiological studies especially when they are combined with analyses of biological material from the population.[1] A specific example is when data from registries form the medical background for biochemical analyses in large cohorts such as the Danish National Birth Cohort (DNBC) and Norwegian Mother and Child Cohort Study (MoBA) birth cohorts of 100 000 pregnant women each that include meticulous, prospectively collected data from questionnaires and blood samples.[2,3]

In this issue of *Paediatric and Perinatal Epidemiology*, Klungsøyr and colleagues[4] have produced a noteworthy work – a validation of the diagnosis of preeclampsia in the Medical Birth Registry of Norway (MBRN) for use in MoBa. The validation showed that less than half of women with the diagnosis of preeclampsia based on data from antenatal records or discharge diagnosis were recorded with preeclampsia in the MBRN. However, the diagnosis was confirmed for nearly all women recorded with preeclampsia in the registry. This study shows that the MBRN can be relied upon for the identification of cases with preeclampsia for further analysis, but is less useful for monitoring the incidence of the condition. Validation of these data for use in MoBa means that blood analyses can be carried out in blood samples collected during pregnancy from women with a validated diagnosis of preeclampsia and matched to validated controls.

Validation is an integral component of scientific research, but validation of registry-based data is rarely published with epidemiological studies. When we publish experimental or clinical studies that include biological material such as blood, we often describe the collection procedure, storage, and kits used for biochemical analyses. In epidemiology, these descriptions are not available or, when they are, validated variables are not always relevant for the study or relate to a different time period. Validation studies of registries are expensive to implement and not easy to publish, as editors know that the appeal of these studies remains limited. Furthermore, researchers may have a conflict of interest because they are reluctant to question the validity of their own data.

Yet routine validation of data from registries is necessary to ensure the quality of epidemiological studies. Primary validation should be done at the time of reporting to the registries, by ascertaining that values fall within an acceptable range and that data inconsistencies are corrected. These explicit criteria for routine validation as well as the proportion of missing values for each variable in the register should be freely accessible. Secondary validation should be performed on an ad-hoc basis by matching of variables within the registry and by linkage of data to other registries, as well as with primary sources, usually medical records. This type of validation – such as reported in this issue of the journal – is very resource demanding and should be carried out for a specific purpose, in selected materials and for limited periods. In particular, validation studies should be carried out when new variables are introduced as this can be associated with missing and erroneous values.

In general, validation studies shows that clinically significant conditions and interventions such as placenta previa and caesarean section are more valid than less significant and more common events such as cystitis in pregnancy. Also, interventions are more valid than diagnosis[5] and a clear and clinically relevant definition improves validity. Unfortunately, definitions of clinical conditions change over time. For instance, pregnancy oedema was originally included

Correspondence:
Lone Krebs, Department of Obstetrics and Gynaecology, Holbaek Hospital, Smedelundsgade 60, Holbaek, DK-2700, Denmark.
E-mail: lone.krebs@dadlnet.dk

as a diagnostic criterion for preeclampsia. However, since oedema also occurs in normal pregnancy, the definition was changed to a combination of hypertension and presence of proteinuria as well as less well-defined severe symptoms or biochemical changes without hypertension and proteinuria.[6] Even with quite specific cut-off values for blood pressure and proteinuria, the definition of preeclampsia is difficult to apply in clinical practice: blood pressure and proteinuria may change from time to time and a diagnosis of preeclampsia will depend on the number and intervals of measurement. Thus, despite the fact that diagnostic criteria for preeclampsia are explicit and reasonably well defined, they may be difficult to apply in the clinical setting.

The absence of validation constrains research. For instance, serious, but rare complications of pregnancy and delivery, such as uterine rupture, are under-researched in part due to validation problems.[7] Clinicians are not used to reporting rare conditions, and a lack of strict definitions and diagnostic criteria may explain why the quality of these data is poor. Validation of the recording of uterine rupture in the Danish birth registry by a review of medical records revealed massive over-reporting as the diagnosis covered a mixture of cases with true and suspected uterine ruptures. Furthermore, at least 16% of complete uterine ruptures were not reported to the registry.[7] Studies of serious rare obstetric complications such as uterine rupture or placenta percreta and their possible association to the increasing rate of caesarean deliveries are extremely important; because of their rarity, these can only be performed using large databases. The Nordic International Network of Obstetric Survey Systems (INOSS) group is one model for countering the lack of data on serious rare obstetric complications. The group performed a 2-year validation study of peripartum hysterectomy, excessive postpartum haemorrhage, uterine rupture, and placenta percreta,[8] by using direct reporting from clinicians to a database in each country and data from the birth registries to identify and request information on overlooked cases. Initiatives like this contribute to the validity of data for the project underway, and may improve future reporting of these severe and rare complications in pregnancy. This project also reinforces collaboration between obstetricians and the birth registries which we hope will improve the general quality of reported data.

Large registry-based studies may be easier to publish than smaller studies because larger sample sizes are more persuasive to most readers. However, it is important to remember that the impact of systematic errors is the same regardless of the population size. More and better use of routinely collected databases should be encouraged, but the scientific work of validating our research materials cannot be sidestepped.

About the authors

Lone Krebs is a senior consultant at the Department of Obstetrics and Gynaecology, Holbaek Hospital and associate professor at University of Copenhagen, Denmark. Jens Langhoff-Roos is a senior consultant and associate professor at the Department of Obstetrics, Rigshospitalet, University of Copenhagen, Denmark. Both authors participate in a collaboration between Nordic Obstetricians and the data managers and epidemiologists at the Nordic Medical Birth Registries.

References

1 Langhoff-Roos J, Krebs L, Klungsøyr K, Bjarnadóttir RI, Källén K, Tapper AM, *et al*. The Nordic medical birth registers – a potential goldmine for clinical research. *Acta Obstetricia et Gynecologica Scandinavica* 2013; 93:132–137.

2 Olsen J, Melbye M, Olsen SF, Sørensen TI, Aaby P, Andersen AM, *et al*. The Danish National Birth Cohort – its background, structure and aim. *Scandinavian Journal of Public Health* 2001; 29:300–307.

3 Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C, *et al*. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology* 2006; 35:1146–1150.

4 Klungsøyr K, Harmon Q, Skard L, Simonsen I, Austvoll E, Alsaker E, *et al*. Validity of preeclampsia registration in the Medical Birth Registry of Norway for women participating in the Norwegian Mother and Child Cohort Study, 1999–2010. *Paediatric and Perinatal Epidemiology* 2014; 28:362–371.

5 Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *International Journal of Epidemiology* 1996; 25:435–442.

6 Steegers EAP, Dadelszen P, von Duvekot JJ, Pijnenborg R. Preeclampsia. *Lancet* 2010; 376:631–644.

7 Thisted DLA, Mortensen LH, Hvidman L, Rasmussen SC, Larsen T, Krebs L. Use of ICD-10 codes to monitor uterine rupture: validation of a national birth registry. *European Journal of Obstetrics, Gynecology, and Reproductive Biology* 2014; 173:23–28.

8 Knight M. The International Network of Obstetric Survey Systems (INOSS): benefits of multi-country studies of severe and uncommon maternal morbidities. *Acta Obstetricia et Gynecologica Scandinavica* 2014; 93:127–131.