

## ORIGINAL RESEARCH ARTICLE

# Incidences of obstetric outcomes and sample size calculations: A Danish national registry study based on all deliveries from 2008 to 2015

Stinne Hoegh<sup>1</sup>  | Line Thellesen<sup>1</sup> | Karl Bang Christensen<sup>2</sup> | Thomas Bergholt<sup>1</sup> | Morten Hedegaard<sup>3</sup> | Jette Led Sorensen<sup>1</sup>

<sup>1</sup>Department of Obstetrics, Juliane Marie Center for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup>Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>Hedegaard Clinic, Copenhagen, Denmark

## Correspondence

Stinne Hoegh, Department of Obstetrics, Juliane Marie Center for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, DK-2100 Copenhagen, Denmark.

Email: stinne.hoegh@regionh.dk

## Funding information

The study was funded by TrygFonden, the Aase and Ejnar Danielsens Foundation, Oestifterne, Toemmerhandler Johannes Fog, and the Department of Obstetrics and the Juliane Marie Center at Rigshospitalet, University of Copenhagen, Denmark.

## Abstract

**Introduction:** In high-income countries the majority of pregnancies have a good outcome, and many adverse obstetric outcomes rarely occur. This makes demonstrating clinically relevant and statistically significant effects of new interventions a challenge. The objective of the study was to report incidences of important obstetric outcomes and to calculate sample sizes for tentative studies.

**Material and methods:** The study was a registry-based study. Data were retrieved from the Danish Medical Birth Registry and included all deliveries in Denmark from 2008 to 2015. The total population included 465 919 deliveries. The study population comprised intended vaginal deliveries with a single fetus in cephalic presentation at term ( $n = 381\ 567$ ). Incidences were reported for 20 outcomes considering the relevance for the patients and the severity of the outcomes. We calculated the sample sizes required in tentative obstetric studies to detect risk reductions of 25 and 50%, for tests at the 5% level, using a power of 80 and 90%. For the randomized controlled trials we calculated the sample size required for comparing two proportions with equal-sized groups. For the cohort study we calculated the sample size also required for two proportions but with unequal sized groups. Outcome measures for sample size calculation were neonatal mortality, Apgar score  $<7$  at 5 minutes and emergency cesarean section.

**Results:** The incidence of neonatal mortality, Apgar score  $<7$  at 5 minutes and emergency cesarean section was 0.05, 0.58 and 10.5%, respectively. Using neonatal mortality as the outcome in a tentative randomized controlled trial with an expected risk reduction of 50% and power of 80%, our calculation showed a sample size of 195 036 deliveries. Using Apgar score  $<7$  at 5 minutes or emergency cesarean section as the outcome, 16 254 and 818 deliveries, respectively, were required. In tentative cohort studies, the required sample sizes were larger due to the unequal proportion of exposed/non-exposed women.

**Conclusions:** Most adverse obstetric outcomes occur rarely; thus, very large sample sizes are required to achieve adequate statistical power in randomized controlled

**Abbreviations:** ECS, emergency cesarean section; RCT, randomized controlled trial.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Acta Obstetrica et Gynecologica Scandinavica* published by John Wiley & Sons Ltd on behalf of Nordic Federation of Societies of Obstetrics and Gynecology (NFOG)

trials. Multicenter studies, international collaborations or alternative study designs to randomized controlled trials could be considered.

#### KEYWORDS

emergency cesarean section, incidence, methods, obstetric outcome, obstetrics, pregnancy outcome, research design, sample size

## 1 | INTRODUCTION

In high-income countries the majority of pregnancies have a good outcome, and many adverse obstetric outcomes rarely occur. This makes demonstrating clinically relevant and statistically significant effects of new interventions a challenge.

Randomized controlled trials (RCTs) are considered the gold standard for establishing causal inference in healthcare interventions, and are therefore frequently applied as study designs.<sup>1,2</sup> If RCTs have adequate statistical power, the expectation is that significant differences between groups will be a result of the intervention.<sup>1</sup>

To ensure the quality of scientific work, calculating and reporting a study's sample size is fundamental.<sup>2</sup> However, sample size calculations are sparsely reported in scientific papers, and many trials do not achieve the target sample size stated before starting the trial.<sup>3</sup> When a study is underpowered, there is a risk of not finding the true difference between the groups, affecting the quality of the study.<sup>4</sup> Two obstetric papers from 1997-2000 on the introduction of continuous electronic fetal monitoring in obstetric care and on the potential bias when comparing small and large maternity institutions when studying stillbirth rates, respectively, discuss the implications of study design, rare outcome measures and large sample sizes.<sup>5,6</sup> The two studies concluded that, for rare outcomes, very large sample sizes were needed to detect statistically significant differences between study groups. Even though these two studies emphasize the implications of sample size calculation in obstetric outcomes, many researchers still include rare outcomes in their study design without having sufficient power to do so.

Our study had three objectives: first, to report incidences of 16 obstetric outcomes; secondly, to calculate sample sizes for tentative studies using three selected outcomes: neonatal mortality, Apgar score <7 at 5 minutes and emergency cesarean section (ECS); and thirdly, to discuss the implications for study design in obstetrics when choosing outcome measures.

## 2 | MATERIAL AND METHODS

### 2.1 | Population and study design

The study was a registry-based study. Data were retrieved from the Danish Medical Birth Registry and included all deliveries in Denmark from 2008 to 2015. The Danish Medical Birth Registry contains information on all deliveries in Denmark, thus providing data on the mother, the child, the pregnancy and the delivery.

#### Key message

The majority of obstetric outcomes occur with a very low incidence. Our sample size calculations showed that when using rare obstetric outcomes large sample sizes were required. Multicenter studies, international collaborations or alternative study designs to randomized controlled trials could be considered.

The annual number of deliveries in Denmark is approximately 60 000, with 96-98% of deliveries in public hospitals and 2-3% at home, mostly attended by midwives from public maternity departments.<sup>7</sup> Currently there are 23 public maternity departments in Denmark, all with access to specialists in obstetrics and anesthesiology. Midwives attend all deliveries and an obstetrician is only involved in the event of complications.

In this study we operated with two populations: the total population and the study population. We used the former, which included all deliveries in Denmark in the study period, to report incidences of important obstetric outcomes.

For calculating sample sizes for tentative studies, we used a study population that included all intended vaginal deliveries with a term (gestational age  $\geq 37$  weeks) singleton in cephalic presentation. Stillbirths and homebirths were excluded. Incidences of the obstetric outcomes were also reported for the study population.

### 2.2 | Outcome measures

It is the scientific question raised that defines whether an obstetric event is an intervention, an outcome or even a population, not the event itself. In our study we have defined 16 relevant and used obstetric outcomes in the literature and are aware that these outcomes in other studies could be defined as interventions or constitute a study population.

We chose 16 obstetric outcomes for reporting incidences of the total population: preeclampsia, hemolysis, elevated liver enzymes, and low platelets (HELLP) syndrome, eclampsia, induction of labor, oxytocin augmentation, umbilical cord prolapse, shoulder dystocia, vacuum extraction, ECS, postpartum hemorrhage  $\geq 1000$  mL, manual exploration of the uterus, stillbirth, Apgar score <7 at 5 minutes, preterm delivery <37 weeks of gestation, low birthweight <2500 g and neonatal mortality. We chose these outcomes considering the

relevance for the patients and the severity of the outcomes. For the study population we report incidences of 14 obstetric outcomes.

For sample size calculation in tentative studies, we chose three outcomes from the core outcome set for key stakeholders in maternity care: neonatal mortality, Apgar score <7 at 5 minutes and ECS.<sup>8</sup> The chosen outcomes reflect different incidences: one extremely rare, one rare and one more common.

Neonatal mortality is defined as death before the age of 28 completed days after live birth.<sup>9</sup> Apgar score is used to assess the condition of the newborn at 1 and 5 minutes after birth and it is a validated predictor of neonatal survival.<sup>10</sup> The Apgar score at 5 minutes is the best predictor of neonatal survival.<sup>10</sup> Cesarean section is linked to a wide range of complications, such as uterine rupture and abnormal invasive placenta, which leads to higher risk of maternal and neonatal morbidity and mortality.<sup>11</sup>

### 2.3 | Statistical analyses

The data have been used as part of another study.<sup>12</sup> Before analysis, the dataset was checked for logical errors. We recoded missing data for maternal weight and height with unrealistic values and checked whether there was consistency between the diagnosis of the delivery and the surgical intervention or procedure coded.

The selected outcomes are reported as incidences, both for the total population and for the study population. The incidences of neonatal mortality, Apgar score <7 at 5 minutes and ECS in the study population formed the basis for the tentative sample size calculations. We calculated the sample sizes for the comparison of two proportions, which necessitates a proportion of the outcome, and the researcher to consider the intervention effect and the desired maximum risk of statistical errors. The statistical tool is provided in Supporting Information Appendix S1.

We calculated the sample size necessary for tentative RCTs and cohort studies to be able to detect risk reductions of 25 and 50% at the 5% level with a power of 80 and 90%, respectively. For the RCTs we calculated the sample size required for comparing two proportions with equal-sized groups (i.e. 1:1 ratio), whereas for the cohort study we calculated the sample size also required for two proportions but with unequal sized groups. We calculated sample sizes for proportion of exposed women of 5, 10 and 25%.<sup>13</sup>

Incidences were computed using IBM SPSS® version 24 (IBM Corp., Armonk, NY, USA) and sample size calculations were made using SAS® software package version 9.4 (SAS Institute, Cary, NC, USA). As missing data were rare, imputation was not applied.

### 2.4 | Ethical approval

Approval was obtained from the Danish Data Protection Agency (file no.: 2012-58-0004). As this was a registry-based study, ethical approval was not required according to the Danish Research Ethics Committee Law.<sup>14</sup>

**TABLE 1** Sociodemographic characteristics of the total population and the study population in Denmark from 2008 to 2015

Characteristics	Total population <sup>a</sup> n (%) 465 919 (100)	Study population <sup>b</sup> n (%) 381 567 (100)
Singleton deliveries	456 014 (97.9)	381 567 (100)
Twin deliveries	9794 (2.1)	—
Triplet/quadruplet deliveries	111 (0.0)	—
Breech deliveries	19 244 (4.1)	—
Singleton vaginal breech deliveries	1987 (0.5)	—
Planned cesarean section	43 407 (9.3)	—
Gestational age		
<37 weeks	30 544 (6.6)	—
37 <sup>+0</sup> to 39 <sup>+6</sup> weeks	206 986 (44.4)	159 973 (41.9)
≥40 weeks	228 193 (49.0)	221 594 (58.1)
Missing data	196 (0.0)	—
Maternal age (years)		
<25	57 907 (12.4)	49 985 (13.1)
25-34	310 442 (66.6)	257 365 (67.4)
35-39	81 335 (17.5)	62 532 (16.4)
≥40	16 235 (3.5)	11 685 (3.1)
Parity		
Nulliparous	212 445 (45.6)	177 674 (46.6)
Multiparous	248 976 (53.4)	203 893 (52.5)
Missing data	4498 (1.0)	3605 (0.9)
Smoking during pregnancy		
No	402 816 (86.5)	330 631 (86.7)
Smoking cessation during pregnancy	13 676 (2.9)	11 588 (3.0)
1-20 cigarettes per day	41 061 (8.8)	33 215 (8.7)
>20 cigarettes per day	1417 (0.3)	1095 (0.3)
Missing data	6949 (1.5)	5038 (1.3)
Body mass index (kg/m <sup>2</sup> )		
<18.5	18 321 (3.9)	15 054 (3.9)
18.5-24.9	276 318 (59.3)	229 585 (60.2)
25-29.9	95 519 (20.5)	77 202 (20.2)
30-34.9	37 128 (8.0)	29 342 (7.7)
≥35	20 302 (4.4)	15 618 (4.1)
Missing data	18 331 (3.9)	14 766 (3.9)

<sup>a</sup>The total population included all deliveries with gestational age 20<sup>+0</sup> to 45<sup>+0</sup>.

<sup>b</sup>The study population included all term singleton (≥37 weeks of gestational age) with intended vaginal cephalic delivery.

### 3 | RESULTS

From 2008 to 2015, there were 465 919 deliveries in Denmark. The study population, including intended vaginal deliveries with term singletons in cephalic presentation, consisted of 381 567 deliveries. There were missing data for Apgar score <7 at 5 minutes in 1260 deliveries (0.3%). There was no missing data for the variables neonatal mortality or ECS.

Table 1 shows the sociodemographic characteristics of the total population and the study population. In general, Danish women were most likely to deliver at term, to be 25-34 years of age, to be non-smokers, and to have a normal body mass index (i.e. 18.5-24.9 kg/m<sup>2</sup>).

Table 2 reports incidences of the 16 obstetric outcomes. Most outcomes occurred at a low incidence. The only outcomes with an incidence >10% were induction, oxytocin augmentation of labor and ECS. In the total population, the incidence of neonatal mortality, Apgar score <7 at 5 minutes and ECS was 0.4, 0.9 and 12.2%, respectively. In the study population, the incidence of neonatal mortality, Apgar score <7 at 5 minutes and ECS was 0.05% (95% confidence interval (CI) 0.04-0.06), 0.58% (95% CI 0.55-0.60) and 10.5% (95% CI; 10.4-10.6), respectively.

Table S1 reports incidences of the obstetric outcomes stratified by year.

Figure 1 and Table 3 report the sample sizes calculated for tentative RCTs and cohort studies. As shown, the incidence of the outcome measure affected the sample size. When using neonatal mortality with an incidence of 0.05% as the outcome in a tentative RCT with an expected risk reduction of 50% and power of 80%, our sample size calculation showed required sample size of 195 036 deliveries. Using Apgar score <7 at 5 minutes, with an incidence of 0.58%, as the outcome in a tentative RCT, with the same risk reduction and same power, 16 254 deliveries were required. For ECS with an incidence of 10.5%, 818 deliveries were required for a tentative RCT. Figure 1 and Table 3 also report the sample sizes required for studies with a power of 90% and for studies with a risk reduction of 25%.

Our results illustrate that an expected lower risk reduction increased the sample sizes. Using neonatal mortality as the outcome in a tentative RCT with 80% power and changing the risk reduction from 50% to 25% resulted in a fourfold increase in the required sample size to 916 518 deliveries. Using Apgar score <7 at 5 minutes or ECS as outcomes, the same fourfold increase in the required sample size was seen when the expected risk reduction changed from 50% to 25% (Figure 1, Table 3).

Our results furthermore illustrate that changing the power from 90% to 80% had a small impact on the required sample sizes. Using Apgar score <7 at 5 minutes as the outcome in a tentative RCT with 50% risk reduction and 80% power instead of 90%, the required sample size decreased from 21 758 to 16 254 deliveries. Using ECS as the outcome in an RCT with 50% risk reduction and 80% power instead of 90%, the required sample size decreased from 1092 to 818 deliveries (Figure 1, Table 3).

**TABLE 2** Incidences of obstetric outcomes in Denmark from 2008 to 2015

Outcome <sup>a</sup>	Total population <sup>b</sup> n (%) 465 919 (100)	Study population <sup>c</sup> n (%) 381 567 (100)
Pregnancy outcomes		
Preeclampsia	13 874 (3.0)	9836 (2.6)
HELLP	1177 (0.3)	472 (0.1)
Eclampsia	249 (0.05)	147 (0.04)
Induction of labor	102 499 (22.0)	93 174 (24.4)
Delivery outcomes		
Oxytocin augmentation	100 791 (21.6)	92 975 (24.4)
Umbilical cord prolapse	514 (0.1)	275 (0.1)
Shoulder dystocia	4449 (1.0)	4344 (1.1)
Vacuum extraction	32 816 (7.0)	30 943 (8.1)
Emergency cesarean section	56 619 (12.2)	40 416 (10.6)
Postpartum hemorrhage ≥1000 mL	10 627 (6.4)	8393 (6.2)
Manual exploration of the uterus	6634 (1.4)	4858 (1.3)
Neonatal outcomes		
Stillbirth	1721 (0.4)	—
Apgar score <7 at 5 minutes	4034 (0.9)	2201 (0.58)
Preterm delivery <37 weeks' gestation	30 544 (6.6)	—
Low birthweight <2500 g	21 648 (4.6)	4949 (1.3)
Neonatal mortality	1310 (0.3)	184 (0.05)

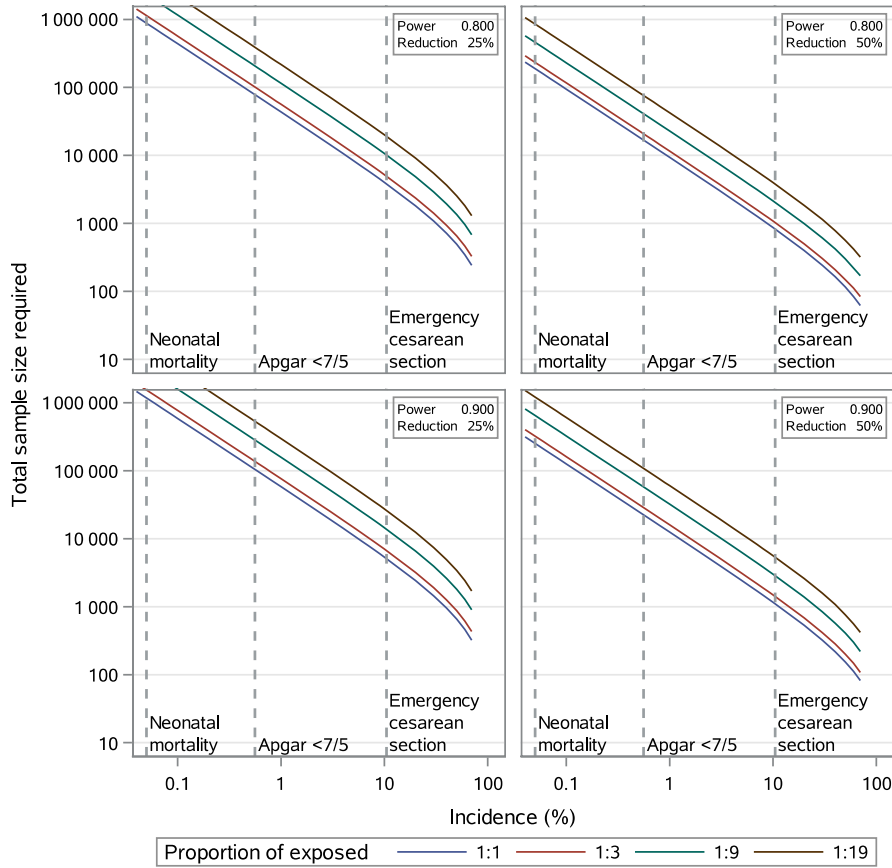
Abbreviation: HELLP, hemolysis, elevated liver enzymes, and low platelets.

<sup>a</sup>One delivery can be represented more than once.

<sup>b</sup>The total population included all deliveries in Denmark from 2008 to 2015 with gestational age 20<sup>+0</sup> to 45<sup>+0</sup> weeks. In the event of multiple fetuses in one pregnancy, an outcome among one or more of the children counts.

<sup>c</sup>The study population included all term (≥37 weeks' gestational age), singleton, intended vaginal cephalic delivery in Denmark from 2008 to 2015.

The study design furthermore affected the required sample sizes. When changing the proportion of exposed women from 50% to a smaller proportion, larger sample sizes were required. If Apgar score <7 at 5 minutes was used as the outcome in a tentative cohort study with 25% exposed instead of an RCT (50% exposed) with a power of 80% and risk reduction of 50%, the required sample size increased from 16 254 deliveries to 20 112 deliveries (Table 3). The same tendency was seen when the proportion of exposed women was even smaller. If the proportion of exposed was 10%, the required sample size was 39 680, and if the proportion of



**FIGURE 1** The total sample size required for tentative randomized controlled trials (1:1) and cohort studies (1:3, 1:9, 1:19) with neonatal mortality, Apgar score <7 at 5 minutes (<7/5) and emergency cesarean section as the outcome. Changes of a 25 and 50% reduction in outcomes are plotted against incidences, using an 80 and 90% power and a 5% significance level

**TABLE 3** Sample sizes for tentative randomized controlled trials and cohort studies<sup>a</sup>

Outcome (incidence)								
Risk reduction								
50%					25%			
Proportion of exposed					Proportion of exposed			
Power	5%	10%	25%	50%	5%	10%	25%	50%
Neonatal mortality (0.05%)								
80%	884 820	476 510	241 584	195 036	4 576 400	2 431 690	1 190 132	916 518
90%	1 259 520	673 250	334 172	261 096	6 280 940	3 327 940	1 615 172	1 226 960
Apgar score <7 at 5 minutes (0.58%)								
80%	73 680	39 680	20 112	16 254	377 260	200 460	98 112	75 786
90%	104 840	56 040	27 816	21 758	517 760	274 340	133 148	101 454
Emergency cesarean section (10.6%)								
80%	3740	2010	1016	818	18 820	10 000	4884	3764
90%	5300	2830	1400	1092	25 760	13 650	6620	5038

<sup>a</sup>Total sample sizes required for tentative classical randomized controlled trials with an allocation of 1:1 ratio (i.e. 50% exposed) and cohort studies with a proportion of exposed of 5, 10 and 25%.

exposed was 5%, the sample size was further increased to 73 680 deliveries (Table 3).

Table S2 reports the required sample sizes for all 14 outcomes from the study population.

## 4 | DISCUSSION

We found that the majority of obstetric outcomes occurred at a very low incidence.

Our sample size calculations showed that the choice of study design, the outcome incidence and the change from 50% to 25% in the risk reduction all contributed to the required sample size of the tentative studies. Changing the power from 90% to 80% did not have large impact on the required sample size.

A strength of our study is that the reported incidences are based on a large data source of 456 014 deliveries and the sample size calculations on a data source of 381 567 deliveries. The deliveries represent the period from 2008 to 2015, making the data fairly current.

Registry-based research always involves the uncertainty associated with inaccurate reporting. Several studies show, however, that data from the Danish Medical Birth Register are valid in terms of diagnosis on most well-defined outcomes, such as preeclampsia, birthweight, oxytocin augmentation of labor, vacuum extraction and cesarean section.<sup>15,16</sup>

The study demonstrated that low incidence of the outcome affected the sample size. In Danish settings, if an RCT with 90% power was required to show a significant reduction of 25% of Apgar score <7 at 5 minutes, the study would take more than 2 years and require the inclusion of all deliveries. However, it might not be possible to include all eligible patients and some do not want to participate in the study. Thus, the time it takes to recruit patients will be longer than anticipated. This entails the researcher in the planning phase of an RCT to be realistic about recruitment and retention of participants in the study. This could be done through a feasibility study.

Our sample size calculations showed that a major contributor to the required sample size was the change from 50% to 25% in the risk reduction. Applying a risk reduction of 50% instead of 25% to the sample size calculation in a tentative RCT with a rare outcome such as Apgar score <7 at 5 minutes would still require a large-scale multicenter study. Multicenter studies have the advantage of including more participants in shorter time. However, multicenter studies are considerably more complex to run than single-site studies. Furthermore, the sample size calculations depend upon the assumption that the differences between the compared interventions in the centers are unbiased estimates of the same quantity. Based on previous studies, a reduction in risk of 50% or 25% in Apgar score <7 at 5 minutes is probably unrealistic.<sup>17,18</sup> A realistic and still clinically important reduction in Apgar score <7 at 5 minutes might be 10%, which would require an even larger sample size.

With a more common outcome such as ECS, conducting RCTs is more feasible because of the requirement of smaller sample sizes to achieve the adequate power. This might explain why ECS is often seen as an outcome in obstetric studies.<sup>17,19-21</sup> ECS may be a relevant outcome, but it is also easier to obtain power to show statistically significant results compared with a more rare outcome. Furthermore, in many studies an effect on the more common outcomes is often found and the interpretation is that a given intervention has only affected these common outcomes. The intervention, however, could potentially also have affected the rare outcomes, but the study might be underpowered to show this effect.

Meta-analyses of RCTs are a way of increasing the power of the estimated intervention effect. However, meta-analyses are, like single-site studies, prone to risk of systematic and random error.<sup>22,23</sup>

Sometimes used in studies with rare outcomes,<sup>24-26</sup> composite outcomes combine several variables, which are considered to be equivalent, into one outcome to increase the total incidence of these outcomes. Composite outcomes enable the study to be performed with a smaller sample and/or in less time. However, composite outcomes often provide an unclear reflection of the effect because the outcomes are not necessarily equivalent in terms of severity or measurements, and it is possible that the exposure increases the risk of one complication and decreases the risk of another. In the latter situation, the possible effect of the exposure may be camouflaged.<sup>27</sup>

RCTs are considered the gold standard for establishing causality between exposure and outcome in healthcare interventions.<sup>1,2</sup> RCTs are usually expensive and time-consuming, and not all research questions can be answered by interventional studies due to ethical considerations. A possible alternative study design to RCTs is observational studies, such as cohort or case-control studies. These designs facilitates studies using rare outcomes as they allow for inclusion of large populations due to no intervention and no treatment, just as the exposure does not have to be administered to the participants.<sup>28</sup> Thus, even though our sample size calculations revealed larger sample sizes for cohort studies of unequal groups, also when studying rare outcomes such as Apgar score <7 at 5 minutes, obtaining the required sample size in a cohort study would be easier than in an RCT. This design would even allow studying rare outcomes of low exposure and also even with a relatively low reduction in risk. There is a limitation when studying extremely rare outcomes because the required sample sizes will be extremely large, making it impossible to recruit participants within a reasonable time frame. Cohort studies, however, have the limitations that they are prone to bias, the data may be inaccurate and misclassified, and deducing causal conclusions is not possible, weakening the study's internal validity. The International Network of Obstetric Survey Systems (INOSS), a multi-country collaboration, facilitates studies of rare and severe outcomes in pregnancy and childbirth through international cooperation.<sup>29</sup> International collaborative work on registry-based data might be a good approach to obtain sufficient sample size when studying rare outcomes. Sometimes RCTs are not the most feasible study design, and many historical cases exist in which treatments with a convincing change are based on observational studies.<sup>30</sup>

The attention given to relevant outcome measures, as in the CROWN initiative<sup>31</sup> is very important to obtain high quality evidence, but it is also essential to consider whether an RCT with rare outcomes is actually feasible, or if an alternative study design must be chosen. To our knowledge, the medical literature on sample size calculation and power regarding obstetric outcomes and choice of study design is sparse. Our results support the findings from two other studies investigating this issue. Mongelli et al demonstrated that when introducing electronic fetal monitoring, it was much easier to detect an increase in the incidences of cesarean section than a reduction in morbidity because of the different sample sizes and time needed to detect a significant change in the two outcomes.<sup>5</sup>

Moster et al demonstrated that large sample sizes were needed when comparisons of safety between different sizes of delivery units were made for low-risk pregnancies, including stillbirth as the outcome measure.<sup>6</sup> Our findings furthermore provide insights into sample sizes in relation to study design in both rare and more common obstetric outcomes.

## 5 | CONCLUSION

Based on Danish national data from an 8-year period, we found that several obstetric outcomes occur rarely. Consequently, very large sample sizes are required to achieve adequate statistical power in tentative RCTs. This necessity entails a risk of studies being underpowered or only showing an effect on common outcomes when an effect on rare outcomes might also exist.

Focusing on international multicenter collaboration and prioritizing a feasible study design can provide high quality evidence when investigating rare outcomes.

## ACKNOWLEDGMENT

We would like to express our deepest gratitude to senior advisor Steen Rasmussen, MSc (econ), MPH, Rigshospitalet, for managing the national registry data.

## CONFLICT OF INTEREST

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

## ORCID

Stinne Hoegh  <https://orcid.org/0000-0002-8712-1088>

## REFERENCES

- OCEBM Levels of Evidence Working Group. Levels of Evidence: Oxford Centre for Evidence-based Medicine. <https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>.
- Schulz K, Altman D, Moher D, et al. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
- Nijjar SK, Khan KS. Threats to reliability risk erroneous conclusions: a survey of prospective registration and sample sizes of randomised trials in women's health. *BJOG*. 2017;124:1057-1061.
- Whitley E, Ball J. Statistics review 4: sample size calculations. *Crit Care*. 2002;6:335-341.
- Mongelli M, Chung TK, Chang AM. Obstetric intervention and benefit in conditions of very low prevalence. *Br J Obstet Gynaecol*. 1997;104:771-774.
- Moster D, Markestad T, Lie RT. Assessing quality of obstetric care for low-risk deliveries; methodological problems in the use of population based mortality data. *Acta Obstet Gynecol Scand*. 2000;79:478-484.
- National Health Data Authority. Medical Birth Registry. <https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/graviditet-foedsler-og-boern/foedselsregisteret> (accessed 11 April 2018).
- Devane D, Begley CM, Clarke M, et al. Evaluating maternity care: a core set of outcome measures. *Birth*. 2007;34:164-172.
- World Health Organization. *The WHO Application of ICD-10 to Deaths During Perinatal Period: ICD-PM*. Geneva: World Health Organization; 2016.
- Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med*. 2001;344:467-471.
- Silver RM, Landon MB, Rouse DJ, et al. Maternal morbidity associated with multiple repeat cesarean deliveries. *Obstet Gynecol*. 2006;107:1226-1232.
- Thellesen L, Bergholt T, Sorensen JL, et al. The impact of a national cardiotocography education program on neonatal and maternal outcomes: a historical cohort study. *Acta Obstet Gynecol Scand*. 2019;98:1258-1267.
- Altman D. *Practical Statistics for Medical Research*, 1st edn. London: Chapman & Hall; 1999.
- The National Committee on Health Research Ethics. Act on Research Ethics Review of Health Research Projects: Part 4, section 14-2. <http://www.nvk.dk/english/act-on-research> (2013, accessed 18 June 2018).
- Langhoff-Roos J, Krebs L, Klungsoyr K, et al. The Nordic medical birth registers—a potential goldmine for clinical research. *Acta Obstet Gynecol Scand*. 2014;93:132-137.
- Klemmensen AK, Olsen SF, Osterdal ML, et al. Validity of pre-eclampsia-related diagnoses recorded in a national hospital registry and in a postpartum interview of the women. *Am J Epidemiol*. 2007;166:117-124.
- Devane D, Lalor JG, Daly S, et al. Cardiotocography versus intermittent auscultation of fetal heart on admission to labour ward for assessment of fetal wellbeing. *Cochrane Database Syst Rev*. 2017;(1):CD005122.
- Thellesen L. *A National Cardiotocography Education Programme: Development, Validation and Impact on Interpretation Skills and Birth Hypoxia*. Copenhagen: University of Copenhagen; 2017.
- Bugg GJ, Siddiqui F, Thornton JG. Oxytocin versus no treatment or delayed treatment for slow progress in the first stage of spontaneous labour. *Cochrane Database Syst Rev*. 2013;(6):CD007123.
- Symon A, Winter C, Inkster M, et al. Outcomes for births booked under an independent midwife and births in NHS maternity units: matched comparison study. *BMJ*. 2009;338:b2060.
- Gulmezoglu AM, Crowther CA, Middleton P, Heatley E. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev*. 2012;(6):CD004945.
- Huusom LD, Secher NJ, Pryds O, et al. Antenatal magnesium sulphate may prevent cerebral palsy in preterm infants—but are we convinced? Evaluation of an apparently conclusive meta-analysis with trial sequential analysis. *BJOG*. 2011;118:1-5.
- Brok J, Thorlund K, Gluud C, et al. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol*. 2008;61:763-769.
- Cluver C, Novikova N, Koopmans CM, et al. Planned early delivery versus expectant management for hypertensive disorders from 34 weeks gestation to term. *Cochrane Database Syst Rev*. 2017;(1):CD009273.
- Altman D, Carroli G, Duley L, et al. Do women with pre-eclampsia, and their babies, benefit from magnesium sulphate? The Magpie Trial: a randomised placebo-controlled trial. *Lancet*. 2002;359:1877-1890.
- Grobman WA, Rice MM, Reddy UM, et al. Labor induction versus expectant management in low-risk nulliparous women. *N Engl J Med*. 2018;379:513-523.

27. Freemantle N, Calvert M, Wood J, et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA*. 2003;289:2554-2559.
28. Messerlian C, Basso O. Cohort studies in the context of obstetric and gynecologic research: a methodologic overview. *Acta Obstet Gynecol Scand*. 2018;97:371-379.
29. Knight M, INOSS. The International Network of Obstetric Survey Systems (INOSS): benefits of multi-country studies of severe and uncommon maternal morbidities. *Acta Obstet Gynecol Scand* 2014; 93:127-131.
30. Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007;334:349-351.
31. Khan K. The CROWN Initiative: journal editors invite researchers to develop core outcomes in women's health. *BJOG*. 2014;121:1181-1182.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Hoegh S, Thellesen L, Christensen KB, Bergholt T, Hedegaard M, Sorensen JL. Incidences of obstetric outcomes and sample size calculations: A Danish national registry study based on all deliveries from 2008 to 2015. *Acta Obstet Gynecol Scand*. 2020;99:34–41. <https://doi.org/10.1111/aogs.13700>