OPEN

Effects of induction of labor prior to post-term in low-risk pregnancies: a systematic review

Eva Rydahl[1,2,3] · Lena Eriksen[4,5] · Mette Juhl[1]

[1]Department of Midwifery and Therapeutic Sciences, University College Copenhagen, Copenhagen, Denmark, [2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, [3]Danish Center of Systematic Reviews: a Joanna Briggs Institute Centre of Excellence, Department of Health Science and Technology, University of Aalborg, Aalborg, Denmark, [4]The Research Unit Women's and Children's Health, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark, and [5]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

ABSTRACT

Objective: The objective of this review was to identify, assess and synthesize the best available evidence on the effects of induction prior to post-term on the mother and fetus. Maternal and fetal outcomes after routine labor induction in low-risk pregnancies at 41+0 to 41+6 gestational weeks (prior to post-term) were compared to routine labor induction at 42+0 to 42+6 gestational weeks (post-term).

Introduction: Induction of labor when a pregnancy exceeds 14 days past the estimated due date has long been used as an intervention to prevent adverse fetal and maternal outcomes. Over the last decade, clinical procedures have changed in many countries towards earlier induction. A shift towards earlier inductions may lead to 15–20% more inductions. Given the fact that induction as an intervention can cause harm to both mother and child, it is essential to ensure that the benefits of the change in clinical practice outweigh the harms.

Inclusion criteria: This review included studies with participants with expected low-risk deliveries, where both fetus and mother were considered healthy at inclusion and with no known risks besides the potential risk of the ongoing pregnancy. Included studies evaluated induction at 41+1–6 gestational weeks compared to 42+1–6 gestational weeks. Randomized control trials (n = 2), quasi-experimental trials (n = 2), and cohort studies (n = 3) were included. The primary outcomes of interest were cesarean section, instrumental vaginal delivery, low Apgar score ($\leq$ 7/5 min.), and low pH ($<$ 7.10). Secondary outcomes included additional indicators of fetal or maternal wellbeing related to prolonged pregnancy or induction.

Methods: The following information sources were searched for published and unpublished studies: PubMed, CINAHL, Embase, Scopus, Swemed+, POPLINE; Cochrane, TRIP; Current Controlled Trials; Web of Science, and, for gray literature: MedNar; Google Scholar, ProQuest Nursing & Allied Health Source, and guidelines from the Royal College of Obstetricians and Gynaecologists, and American College of Obstetricians and Gynecologists, according to the published protocol. In addition, OpenGrey and guidelines from the National Institute for Health and Care Excellence, World Health Organization, and Society of Obstetricians and Gynaecologists of Canada were sought. Included papers were assessed by all three reviewers independently using the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI). The standardized data extraction tool from JBI SUMARI was used. Data were pooled in a statistical meta-analysis model using RevMan 5, when the criteria for meta-analysis were met. Non-pooled results were presented separately.

Results: Induction at 41+0–6 gestational weeks compared to 42+0–6 gestational weeks was found to be associated with an increased risk of overall cesarean section (relative risk [RR] = 1.11, 95% confidence interval [CI] 1.09–1.14), cesarean section due to failure to progress (RR = 1.43, 95% CI 1.01–2.01), chorioamnionitis (RR = 1.13,

Correspondence: Eva Rydahl, evry@kp.dk

There is no conflict of interest in this project.

DOI: 10.11124/JBISRIR-2017-003587

95% CI 1.05–1.21), labor dystocia (RR = 1.29, 95% CI 1.22–1.37), precipitate labor (RR = 2.75, 95% CI 1.45–5.2), uterine rupture (RR = 1.97, 95% CI 1.54–2.52), pH < 7.10 (RR = 1.9, 95% CI 1.48–2.43), and a decreased risk of oligohydramnios (RR = 0.4, 95% CI 0.24–0.67) and meconium stained amniotic fluid (RR = 0.82, 95% CI 0.75–0.91). Data lacked statistical power to draw conclusions on perinatal death. No differences were seen for postpartum hemorrhage, shoulder dystocia, meconium aspiration, 5-minute Apgar score < 7, or admission to neonatal intensive care unit. A policy of awaiting spontaneous onset of labor until 42+0–6 gestational weeks showed, that approximately 70% went into spontaneous labor.

Conclusions: Induction prior to post-term was associated with few beneficial outcomes and several adverse outcomes. This draws attention to possible iatrogenic effects affecting large numbers of low-risk women in contemporary maternity care. According to the World Health Organization, expected benefits from a medical intervention must outweigh potential harms. Hence, our results do not support the widespread use of routine induction prior to post-term (41+0–6 gestational weeks).

Keywords 41 gestational weeks; 42 gestational weeks; expectant management; labor induced; perinatal mortality

JBI Database System Rev Implement Rep 2019; 17(2):170–208.

GRADE Summary of Findings

What is the effect of routine labor induction in low-risk pregnancies at 41+0 to 41+6 gestational weeks (prior to post-term) compared to routine labor induction at 42+0 to 42+6 gestational weeks (post-term) on selected maternal and perinatal outcomes?					
Bibliography: Rydahl E, Eriksen L, Juhl M. Effects of induction of labor prior to post-term in low-risk pregnancies: a systematic review. JBI Database System Rev Implement Rep 2019; 17(2):170–208.					
Outcomes	**Anticipated absolute effects* (95% CI)**		**Relative effect (95% CI)**	**№ of participants (studies)**	**Certainty of the evidence (GRADE)**
	Risk with routine labor induction at 42+0 to 42+6 gestational weeks (post-term)	**Risk with routine labor induction at 41+0 to 41+6 gestational weeks (prior to post-term)**			
Cesarean section (overall)	323 per 1000	**359 per 1000** (352 to 368)	**RR 1.11** (1.09 to 1.14)	74,860 (1 observational study)	⨁⨁◯◯ LOW
Cesarean section (overall)	149 per 1000	**165 per 1000** (146 to 187)	**RR 1.11** (0.98 to 1.26)	5109 (4 RCTs)	⨁⨁⨁◯ MODERATE [a,b]
Apgar score < 7 after 5 min.	13 per 1000	**12 per 1000** (10 to 14)	**RR 0.90** (0.79 to 1.03)	74,952 (1 observational study)	⨁⨁◯◯ LOW
Apgar score < 7 after 5 min.	11 per 1000	**10 per 1000** (6 to 18)	**RR 0.90** (0.52 to 1.56)	4671 (3 RCTs)	⨁⨁⨁◯ MODERATE [a]
pH < 7.10	44 per 1000	**84 per 1000** (65 to 107)	**RR 1.90** (1.48 to 2.43)	4017 (2 RCTs)	⨁⨁⨁◯ MODERATE [a]

Instrumental vaginal delivery	121 per 1000	**157 per 1000** (150 to 164)	**RR 1.30** (1.24 to 1.36)	49,628 (1 observational study)	⊕⊕◯◯ LOW
Instrumental vaginal delivery	223 per 1000	**223 per 1000** (201 to 248)	**RR 1.00** (0.90 to 1.11)	4509 (3 RCTs)	⊕⊕⊕◯ MODERATE [a,b]

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

RCT: randomized controlled trial; **CI:** confidence interval; **RR:** risk ratio

GRADE Working Group grades of evidence

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

a. Lack of allocation concealment or blinding

b. Allocation not truly random

Introduction

Induction of labor is defined as any intervention performed with the aim of inducing labor before a spontaneous onset of labor. Labor induction is a common obstetric procedure, but it is also described as ''one of the most drastic ways of intervening in the natural process of pregnancy and childbirth''.[1(p375)] Historically, labor induction has been performed to terminate progressing pathological conditions that could potentially harm mother or fetus, such as preeclampsia, intrauterine growth restriction or diabetes.[2] Over the last decades, however, induction rates have increased two- to four-fold in the developed world,[3-5] and the World Health Organization (WHO) estimates that 25% of women in developed countries have their labor induced.[3] There seems to be no single reason for the increase. Suggested explanations include the availability of cervical ripening agents, an increasing demand from patients and a shift in professional clinical culture towards increased use of interventions in childbirth.[5,6] Even though healthcare interventions are usually performed with the intention to improve the health of the patient, interventions are also prone to iatrogenic mechanisms, where the intervention itself can cause harm.[1,7] This may be an overlooked element in current maternity care, and there is a lack of knowledge on the benefits and harms following a more liberal use of labor induction among large groups of low-risk pregnancies.

Current induction rates do not correspond with WHO recommendations, which emphasize that labor induction should be performed only when there is a clear medical indication and the expected benefits outweigh potential harms; studies suggest that 25–50% of inductions are performed without a medical reason.[8-11] Of these, many are inductions that are performed prior to post-term. A pregnancy is defined as post-term when it reaches two completed weeks past estimated due date (EDD).[5,12] In these cases, mother and fetus are healthy when induction is initiated, and the intervention is made as a precautionary treatment to prevent potential fetal demise or fetal death.

Elective induction after EDD may be beneficial in regards to risks imposed by the ongoing pregnancy, such as the progress of adverse conditions in pregnancy (e.g. pre-eclampsia, oligohydramnios, macrosomia), shoulder dystocia, postpartum hemorrhage, asphyxia, meconium aspiration syndrome, neonatal birth injury, low umbilical artery pH, admission to neonatal intensive care unit, or intrauterine fetal demise, which can develop into fetal death.[6,13] On the other hand, labor induction has also been found to be an independent risk factor for birth complications and associated with increased fetal and maternal morbidity and possibly also mortality.[2,14,15] Iatrogenic effects (iatrogenesis or iatrogenic effect is a Greek term meaning ''brought forth by the

healer'' and defined as any consequence of medical treatment or advice to a patient)[7] of induced labor include longer labor, uterine hyperstimulation, precipitate labor, uterine ruptures, meconium stained liquid, bleeding, and fetal asphyxia.[16] In addition, induced labor can cause a cascade of interventions, such as continuous electronic fetal monitoring, amniotomy, confinement to bed due to fetal monitoring or intravenous treatment and increased requirement for analgesia (e.g. epidural), each with their own individual risk of potential subsequent complications.[16]

Between approximately the 1980s to the mid-2000s, obstetric textbooks and influential obstetric medical societies recommended close fetal surveillance and/or labor induction, when a pregnancy exceeded 14 days past EDD. In contrast to other mammals, humans may go into spontaneous labor over a wide window of time. Hence, term delivery is defined as occurring over a time span of five weeks, i.e. between 37 and 42 gestational weeks (GW) (259–293 gestational days), and a post-term delivery is defined as occurring at 42 GW or later.[5,12,17] During 2008–2010 both the UK and the US obstetric societies and the WHO changed the recommended gestational timing for induction from two weeks past EDD to one week past EDD in order to reduce adverse conditions, especially cases of intrauterine death due to fetal demise.[3,17-19] This shift occurred as part of a general trend towards increased medicalization of pregnancy and childbirth, and today inductions are widely performed as a routine medical procedure before the pregnancy exceeds 14 days past EDD.[20] However, concerns have been raised that routine induction prior to post-term puts a large number of normal pregnancies at risk of iatrogenic effects from the induction.[11,21,22] It is estimated that 15–24% of pregnant women have not yet gone into spontaneous labor or given birth at 41 weeks and zero days (41+0 GW), and that most of them are likely to do so within the following week. Hence, in settings where routine induction is not offered until after 42 GW, it has been found that only 1–6% have not gone into spontaneous labor by 42+0 GW.[23,24] This suggests that induction prior to post-term leads to a substantially increased number of inductions in low-risk pregnant women and, thus, imposes possible additional complications related to the intervention itself upon a large group of low-risk pregnancies. Additionally, this could be more costly because of increased need of care and longer hospital stays.[25,26]

Research has been carried out to explore the benefits and harms of changed guidelines from routine induction of labor two weeks past EDD to only one week past EDD. However, existing evidence suffers from a lack of consensus as to which groups to compare in studies on routine labor induction after EDD. This means that current clinical guidelines are based on studies (including meta-analyses) with varying comparison group definitions. For example, some studies compare routine induction one week past EDD with groups that are defined by unlimited time for the maximum length of pregnancy, e.g. four weeks past EDD.[27,28] This diminishes the relevance of the findings, because it is far from current obstetric practice and because the risk of severe adverse maternal and fetal outcomes is largely increased at such extended time points.[6,28] Also, failure to go into spontaneous labor long beyond term may be related to pathological conditions, e.g. severe maternal obesity and certain congenital fetal malformations,[5,29] which impose methodological problems in the analysis.

Current evidence on routine induction is based on different approaches regarding what groups to compare. There seems to be three different ways of calculating, which presumably affect the results.

The "index week method"

Initially, most studies on routine induction past EDD used the ''index week method''. This method compares induction at 41+0–6 GW (intervention group) with pregnancies with spontaneous onset of labor at 41+0–6 GW (expectant group), i.e. the expectant group is defined by spontaneously initiated deliveries in the same week as the intervention group (Figure 1).

Findings based on this method generally speak against induction at 41+0–6 GW and find induction associated with a higher risk of adverse outcomes.[30] This is not surprising, because women with spontaneous onset of labor have better cervical status, which, in addition to spontaneous contractions, is a strong predictor for an uncomplicated delivery. Hence, by comparing with spontaneous labor in the same week only, this method neglects the risk of the ongoing pregnancy. However, the method has been criticized for being ''unfair'' and said not to reflect clinical daily life,[6] because without induction,

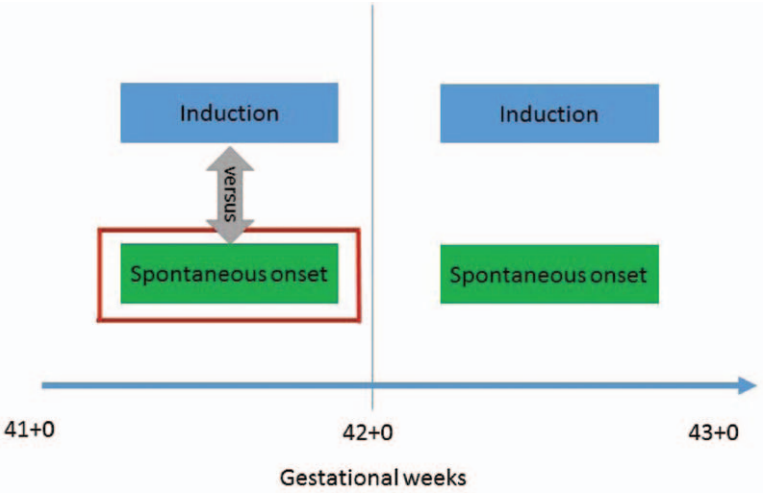


Figure 1: Index week method

women will not necessarily go into spontaneous labor during the index week; these women may have a pre-labor cesarean section (CS) or an induction later in pregnancy.

The "next week method"

As a response to the above discussion, Caughey *et al.*[6] suggested another method. The ''next week method'' compares induction at 41+0–6 GW with all deliveries at 42+0 GW or later.[30,31] This approach was claimed to be more relevant, because defining the expectant group as all women who give birth after the index week ensures the risk of ongoing pregnancy to be taken into account. When this method is used, the expectant group includes any mode of labor onset (induced or spontaneous) as well as pre-labor CS, as long as the delivery takes place after the index week (Figure 2).

Studies based on this method have generally not found induction at 41+0–6 GW associated with an increased risk of adverse outcomes,[28,30] and the findings have been used as an argument for routine induction at 41+0–6 GW. The method, however, has been criticized because it excludes a substantial

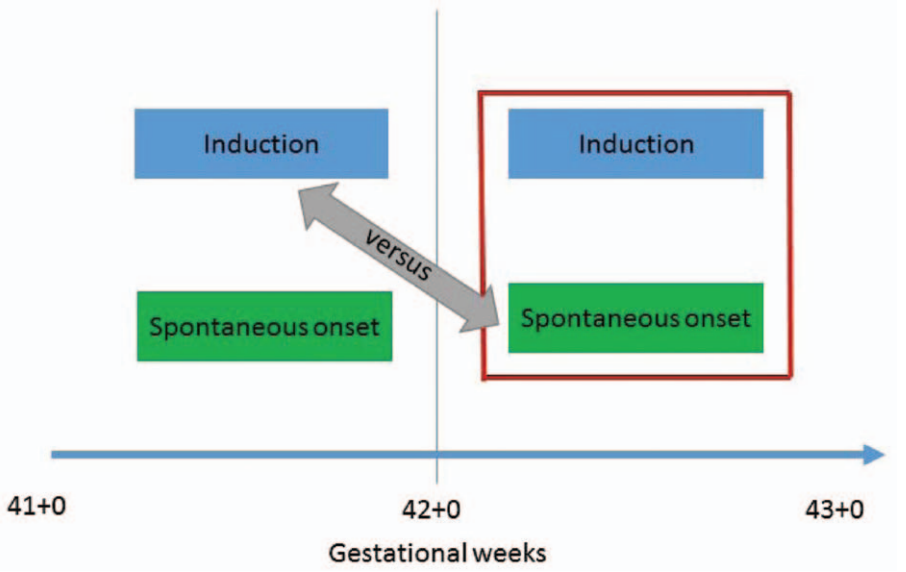


Figure 2: Next week method

number of spontaneous labor deliveries during the index week (41+0–6 GW) from the analysis, e.g. by Glantz *et al.* who have argued that spontaneous labor deliveries at 41+0–6 GW should be included in the expectant group.[21]

The "index+next week method"

The concerns illustrated above gave rise to a third approach, the ''index+next week method'', in which induction at 41+0–6 GW is compared with all non-induced deliveries during the index week and all deliveries during the following week (42+6 GW) (Figure 3). The expectant group may therefore go into spontaneous labor, have a pre-labor CS, or eventually end up with an induced delivery after the index week. This approach has been used in a few recent studies with inconclusive findings.[21,32-34]

The ''index+next week method'' ensures that the substantial number of women who go into spontaneous labor during the index week are accounted for, as well as risks related to the ongoing pregnancy. Despite these advantages, this approach is prone to an important bias source, because the pregnancy duration in the expectant group may exceed 42+6 GW. This introduces excessively increased risks for mother and child in the expectant groups, because the pregnancy can continue far beyond 42 GW. Expectant group definitions vary largely between studies that have used this approach,[6,28,35,36] which makes findings difficult to compare. Defining maximum pregnancy duration as 42+6 GW in the expectant group would reduce such bias.

The motivation for this review must be understood in the context of a generally increasing trend towards labor induction prior to post-term in low-risk pregnancies and the consequential need to clarify whether this shift in clinical practice implies negative consequences for mothers or babies. Objectives have been outlined in a previously published protocol.[37] A search of PubMed, Cochrane, Web of Science, and *JBI Database of Systematic Reviews and Implementation Reports* did not identify any similar reviews or protocols.

Review question/objective

The objective of this review was to identify, assess and synthesize the best available evidence on the effects of induction prior to post-term on the mother and fetus. Maternal and fetal outcomes after routine labor induction in low-risk pregnancies at 41+0 to 41+6 GW (prior to post-term) were compared to routine labor induction at 42+0 to 42+6 GW (post-term).

Definitions

For the purpose of this review, the following definitions will be used:

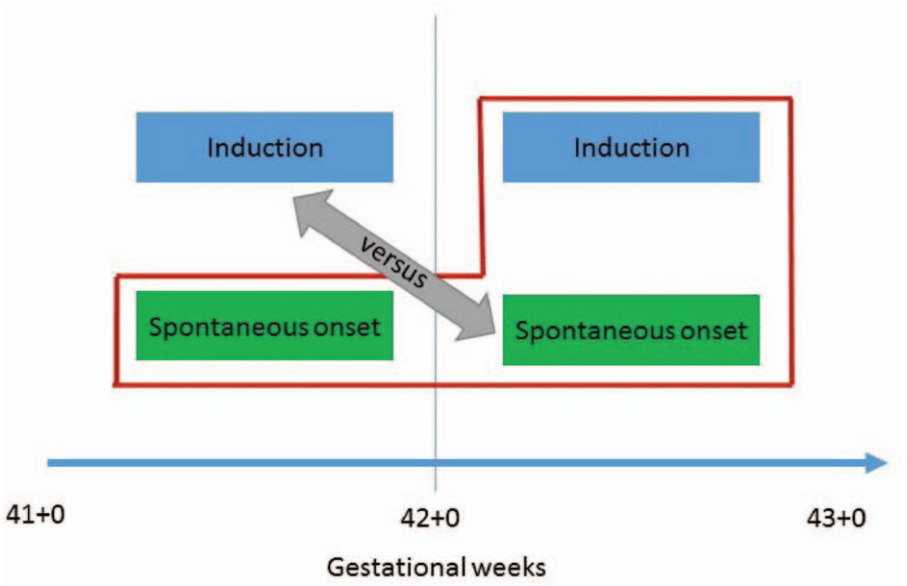


Figure 3: Index + next week method

Estimated due date (EDD): Only studies with EDD determined by ultrasound were included. The ultrasound estimation must be from the first half of the pregnancy. Estimation of due date calculated by last menstrual period differs by an average of 5 days from EDD derived from early ultrasound examination.[38] It is thus necessary to eliminate this possible source of bias by restricting to studies that estimate due date with the same method. More recent studies tend to use ultrasound estimated due date, therefore this method was chosen in the present review.

Low-risk pregnancy: A low-risk pregnancy is defined as one in which neither mother nor baby at the time of enrolment in the study are affected by conditions or circumstances that can complicate the delivery. Hence, in the present review this is defined as a singleton pregnancy with a fetus in vertex presentation and with no known risk factors or complications and a normally grown and formed baby.[2]

Elective induction: An induction performed for a nonmedical reason, e.g. a wish to schedule the delivery on a specific date or long transport to hospital. Also, some women request delivery because they are uncomfortable in the last weeks of pregnancy.[39]

Post-term: A pregnancy is post-term when it reaches two weeks past the EDD, i.e. 42+0 weeks (294 days) of gestation.[5,12]

Prior to post-term: In this study, we define "prior to post-term" as a pregnancy length of 41+0 to 41+6 weeks (287–293 days) of gestation.[12]

Inclusion criteria

Participants

Participants included pregnant women at low risk of complications at the time of enrolment. To be included, both the fetus and the mother should be healthy, with no known risks besides the potential risk of the ongoing pregnancy. There should be intended vaginal birth with no contraindications, and EDD should be determined from ultrasound examination in order to obtain reasonably comparable measures of pregnancy length estimation.

Intervention

The intervention of analysis was labor induction at 41+0–6 GW. Thus, the intervention group was composed of those who underwent labor induction at 41+0–6 GW. The expectant group was defined by all non-induced deliveries between 41+0 GW and 42+6 GW and all deliveries at 42+0–6 GW. This corresponds to the "index+next week method" described above with an additional restriction of the maximum pregnancy duration in the expectant group.

We included studies that used generally accepted induction methods: in cases of unfavorable cervix status, the induction agent could be prostaglandins (PGE1 or PGE2) or Foley catheter, while in cases of favorable cervix status, the induction method could be artificial rupture of membranes, oxytocin infusion, or artificial rupture of membranes followed by oxytocin infusion.[40]

Outcomes

The perinatal and maternal outcomes examined in this review were selected based on their known or suggested association with prolonged pregnancy and/or induced delivery.

Perinatal outcomes (primary):

- Low Apgar score (< 7 after 5 minutes)
- pH < 7.10.

Perinatal outcomes (secondary):

- Severe fetal complications (e.g. meconium aspiration, cerebral palsy, brachial plexus injury, hypoxic-ischemic encephalopathy)
- Perinatal mortality
- Hospitalization at neonatal intensive care unit
- Pathological cardiotocography (CTG)
- Shoulder dystocia
- Infection.

Maternal outcomes (primary):

- Cesarean section (CS)
- Instrumental vaginal delivery.

Maternal outcomes (secondary):

- Severe maternal complications (e.g. postpartum hemorrhage > 1000 ml, uterine rupture, intensive care, sepsis, obstetric shock)
- Maternal morbidity (e.g. chorioamnionitis)
- Epidural anesthesia
- Labor dystocia
- Abnormal contractions (e.g. uterine tachysystole, hyperstimulation, precipitate labor)
- Episiotomy or third or fourth degree lacerations
- Length of hospital stay
- Breastfeeding.

Measurement of outcomes varied between studies. Following an inclusive approach, the present

meta-analysis included a varied range of outcomes. This approach resulted in the inclusion of one additional outcome, namely pH < 7.10, which is now listed as a primary perinatal outcome together with low Apgar score, both being indicators of fetal distress. This is a deviation from the published protocol.[37] Composite measures that varied between studies (e.g. neonatal morbidity score),[33] were not considered comparable across studies and, thus, not included. Finally, data allowed for analyses of submeasures of the overall primary maternal outcome: CS (including CS due to failure of progress, CS due to fetal distress, and CS after previous vaginal birth).

Types of studies

We included randomized controlled trials (RCTs), quasi-experimental trials, and cohort studies. Cohort studies were included because large-scale studies are needed when investigating rare outcomes, such as mortality and low Apgar scores. Randomized and quasi-experimental studies were analyzed together and separate from cohort studies.

Methods

Search strategy

A three-step search strategy was used in order to identify publications from peer-reviewed journals as well as gray literature. The initial search terms were decided through discussions between authors and a research librarian to ensure identification of the maximum number of relevant articles.

An initial search limited to PubMed and CINAHL was undertaken and followed by an analysis of the text words contained in titles and abstracts and of the index terms used to describe the articles. All identified keywords and index terms were used in a second search performed across all included databases (please see below). The second search revealed keywords and terms from index text, which lead to further exploration, where terms such as a different spelling of cesarian and labor were added. Thirdly, the reference lists from each identified publication were searched for additional studies. The search was conducted Jan 2014–June 2018. Individual search strategies are presented in Appendix I.

Studies published in English, Danish, Swedish and Norwegian were considered for inclusion, and authors of primary studies were contacted in cases of missing information or to clarify unclear data.

According to rapid changes in fetal surveillance management, labor induction routines, and methods for estimating due date, studies were restricted to publications within the last two decades (1998–2018).

Databases included PubMed, CINAHL, Embase, Scopus, Swemed+, POPLINE, Cochrane, TRIP, Current Controlled Trials, and Web of Science.

Additional searches for published literature included hand searching of reference lists and bibliographies of included articles.

Search for gray literature included MedNar, Google Scholar, OpenGrey, ProQuest Nursing and Allied Health Source, and guidelines from the National Institute for Health and Care Excellence, the WHO, the Royal College of Obstetricians and Gynaecologists, the American College of Obstetricians and Gynecologists, and the Society of Obstetricians and Gynaecologists of Canada.

Assessment of methodological quality

For methodological validity prior to inclusion of studies, papers selected for retrieval were reviewed by all three authors independently using the 2014 edition of a standardized critical appraisal instrument from the JBI SUMARI.[41] Please find the exact wording of the appraisal instrument questions in the Results section (Tables 1 and 2). Any disagreements between the authors were resolved through discussion. When information was unclear, we attempted to contact the authors of the assessed studies for clarification.

We included all studies that fulfilled the inclusion criteria and that the authors considered of acceptable methodological quality, according to the critical appraisal tool from JBI SUMARI.[41]

Data extraction

Data were extracted from the included papers using the standardized JBI data extraction tool.[41] Studies selected for retrieval are presented in Appendices II-IV. Appendix II presents general characteristics on included studies. Appendix III presents details on methods for estimation of due date, comparison groups, parity, and induction agents in included studies. Appendix IV presents an overview of outcome measures in included studies. Reasons for exclusion for screened, excluded studies are shown in Appendix V.

Data synthesis

Data from the included studies were pooled in a statistical meta-analysis model using RevMan 5 (Copenhagen: The Nordic Cochrane Centre, Cochrane). To minimize errors all pooled statistics were subject to double data entry. Effect sizes expressed as risk ratios (for categorical data) and weighted mean differences (for continuous data) and corresponding 95% confidence intervals (CI) were calculated. Statistical heterogeneity was assessed in the meta-analysis using the I^2 and chi-squared statistics, and heterogeneity was considered substantial if $I^2 > 50\%$ and P value < 0.10 in the chi-square test for heterogeneity.[42] Subgroup-analysis according to type of study design were performed. Where statistical pooling was not possible the findings have been presented in narrative form, and with tables and figures where appropriate to facilitate interpretation.

Results

Study inclusion

Figure 4 shows the results of our search strategy, where 4212 records were identified in the first step. Of these 3442 were identified through database search and 770 from other sources. After abstract reading 33 studies remained for full text-reading, which further excluded 25 studies (Appendix V). We assessed eight full text papers for eligibility and excluded one due to inconsistent methods for due date estimation (i.e. many of the participants in the study had calculation by last menstrual period as the only calculation method). The remaining seven studies were selected for retrieval (Appendix II).

Seven studies were included: two randomized trials[33,34] and two quasi-experimental trials[13,43] with a total of 5119 participants,[13,33,34,43] and three cohort studies with 356,338 participants.[44-46]

Methodological quality

Assessment of risk of bias in randomized and quasi-experimental studies

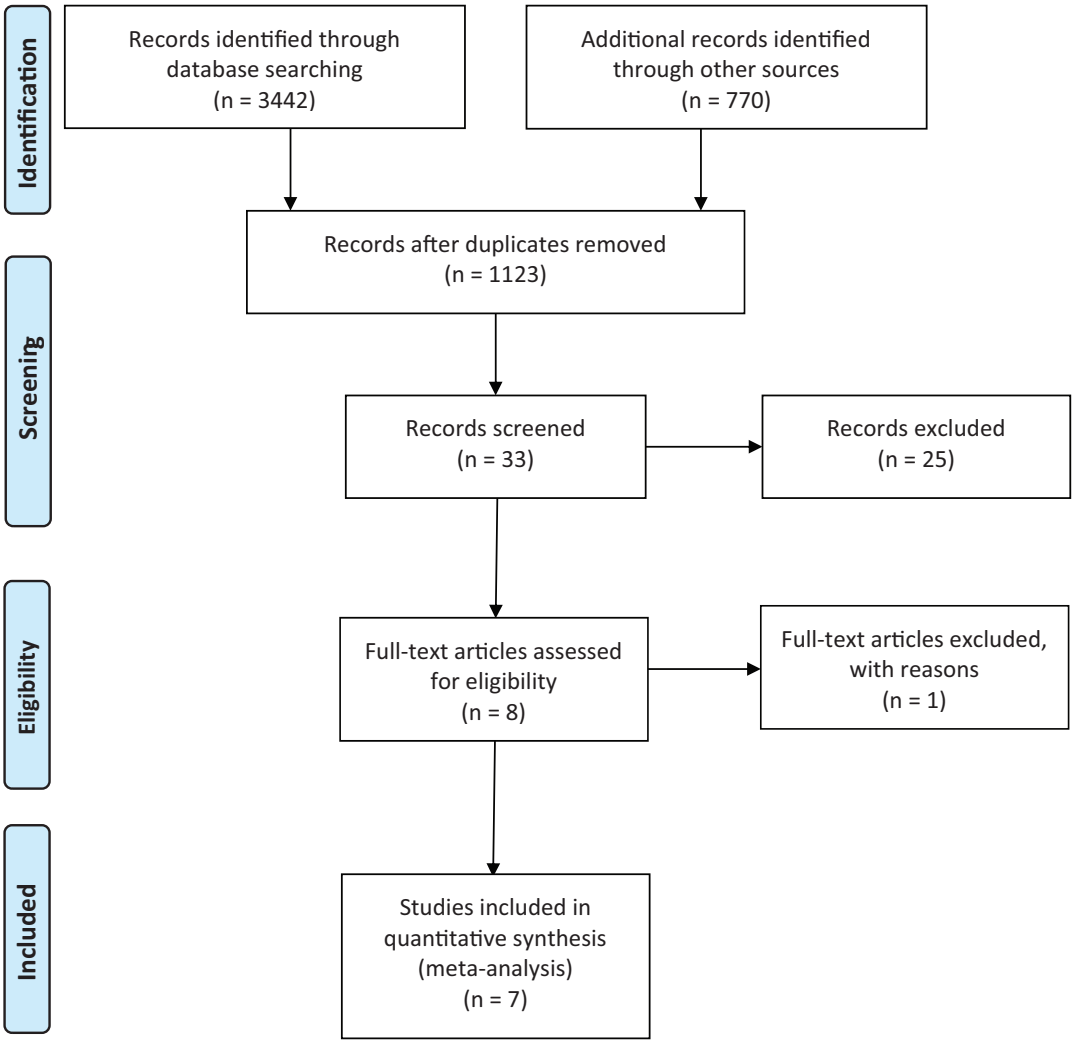
Table 1 shows the scoring of the included randomized trials and quasi-experimental studies according to the JBI SUMARI critical appraisal tools for RCTs/quasi-experimental studies.[41]

Comparability at entry (selection bias, confounding) (Q1, Q3, Q6)

Randomization (Q1) and concealment of allocation to allocator (Q3) are important factors in the evaluation of whether the intervention and control groups were comparable at baseline. In two studies,[33,34] randomization was conducted via computer or use of opaque sealed envelopes and, thus, scored as truly randomized and with allocation concealed from the allocator. In another study,[43] allocation was determined by calendar period meaning that no clinician or researcher could affect allocation to treatment group in the clinical setting. Hence, this study was also considered to have truly random allocation. The Daskalakis study allocation was considered not truly random, as allocation was determined according to the attending physician's preference, and no information was presented as to how the participants were assigned a given physician.[13] Question 6 serves to corroborate the randomization in the studies, i.e. only when baseline characteristics turn out to be similar in the two groups is confounding truly minimized, as compared to observational studies. For all four studies, we considered treatment and comparison groups to be comparable at entry. Even though the Daskalakis study only presented a few baseline characteristics (age, parity and ethnicity), the groups were considered comparable, because no statistically significant differences were seen for the examined characteristics.

Blinding (Q2, Q5)

Blinding is used in trials to enhance group comparability and minimize confounding. When the intervention under study is labor induction, concealment of treatment group to participants is not possible. Hence, all included trials were assigned a "no" in Q2, meaning that in none of the studies the participants were blinded to treatment allocation. In the two quasi-experimental studies, the studies were planned and data extracted retrospectively.[13,43] Hence, by the time of data collection, neither participants nor clinicians were aware that a comparison would be made between induction and expectant management, and, thus, were not affected by a knowledge of an ongoing trial in the assessment of outcomes. Even so, in the clinical situation the assessment and registration of outcomes for a participant is likely affected by the individual clinician's knowledge on induction vs. expectant management. This poses a risk of bias on the part of clinicians. For the included outcomes, however, we considered lack of blinding by the time of data collection of limited importance due to the character of the outcomes. Concealment of



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Figure 4: PRISMA flowchart of the search and study selection process

treatment group by the time of analysis is also difficult for labor induction. One study, however, blinded the researcher to some extent by blinding information on tachycardia and hyperstimulation in order to prevent the effects of researcher knowledge of treatment when evaluating outcomes.[34]

Loss to follow-up and treatment differences (selection bias) (Q4, Q7)

Question 4 addresses selection bias due to loss of follow-up and concerns information on outcome for participants who withdrew from the study. By design, the two quasi-experimental trials could not have any withdrawals and were therefore assigned "not applicable" in Q4. The two RCTs fulfilled this criterion. Question 7 addresses treatment differences between groups other than the intervention of interest, which may induce selection bias. In two studies we considered treatment of participants to be similar in the compared groups.[13,33] In one study it was unclear whether groups were treated identically, since deliveries took place during different calendar periods, and no information was given as to any other potential changes in procedures between the two periods.[46] The total study period, however, was no longer than three years, and the intervention

period followed the control period immediately, indicating a limited risk of important time-dependent bias. In one study, the design allowed for systematic differences in induction methods between groups[34] and, thus, we considered groups not to be properly comparable. In this study, the intervention group included deliveries induced by misoprostol medical treatment, oxytocin medical treatment, or Foley-catheter, while those in the expectant group, who had not gone into spontaneous labor at 41+0–6 GW and were induced at 42+0–6 GW, were all treated with misoprostol as induction agent.

Outcome assessment (Q8, Q9)

Questions 8 and 9 relate to differences in outcome measurement between groups and reliability in the measurements. For all four randomized trials, we considered outcomes as measured in the same reliable way, including the use of international standardized measures (e.g. CS, Apgar score). In one study, the authors had generated a composite outcome of neonatal morbidity[33] that did not rely on solid empirical basis, which may impose some degree of uncertainty about this measure.

Statistical analysis (Q10)

All studies presented intention-to-treat analyses, which is considered a relevant method in RCTs. In one study additional sensitivity analyses were performed.[33] The same study also presented analyses with imputation of missing data (e.g. on pH-values in the newborn). The three other studies did not explain how they handled missing data.

Assessment of risk of bias in cohort studies

Table 2 shows the scoring of the included cohort studies according to the JBI SUMARI critical appraisal tools for cohort studies.[41]

Generalizability (Q1)

Question 1 concerns generalizability of included cohort studies. For all three studies, we considered study populations to be representative of the background population.

Comparability at entry (selection bias, confounding) (Q2, Q3, Q4)

Question 2 concerns possible differences between groups at study entry. For two studies, we considered groups to be comparable at the time of inclusion.[44,46] In one study, it was not clear whether participants in the expectant group were at a poorer health status at 41+0–6 GW than the intervention group, because risk status was assessed at 39+0 GW.[45] Authors were approached, but we did not achieve contact. The degree of possible bias was therefore not able to be assessed. Question 3 concerns allocation into the intervention and expectant groups. For all three studies, we found that bias had been minimized with regard to allocation to the two groups. Question 4 asks about proper identification and handling of confounding factors. In two studies we regarded this to be appropriate,[44,46] whereas the third study lacked important confounding variables, such as smoking and BMI.[45]

Table 1: Assessment of methodological quality of randomized controlled and quasi-randomized trials

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Burgos *et al.* (2012)[43]	Y	N	Y	NA	N	Y	U	Y	Y	Y
Daskalakis *et al.* (2013)[13]	N	N	N	NA	N	Y	Y	Y	Y	Y
Gelisen *et al.* (2005)[34]	Y	N	Y	Y	N	Y	N	Y	Y	Y
Heimstad *et al.* (2007)[33]	Y	N	Y	Y	N	Y	Y	Y	Y	Y
Total (%)	75	0	75	50	0	100	50	100	100	100

Y, yes; N, no; NA, not applicable; U, unclear.
Q1: Was the assignment to treatment groups truly random? Q2: Were participants blinded to treatment allocation? Q3: Was allocation to treatment groups concealed from the allocator? Q4: Were the outcomes of people who withdrew described and included in the analysis? Q5: Were those assessing outcomes blind to the treatment allocation? Q6: Were the control and treatment groups comparable at entry? Q7: Were groups treated identically other than for the named interventions? Q8: Were outcomes measured in the same way for all groups? Q9: Were outcomes measured in a reliable way? Q10: Was appropriate statistical analysis used?.
Burgos *et al.* 2012: Unclear description of statistical analysis of continuous data, but it appears from the presentation of data that appropriate statistical tests have been used.
Daskalakis *et al.* 2013: Unclear description of risk status of participants, but when contacted, author confirmed that the study group did not have any medical complications.
Gelisen *et al.* 2005: Unclear description on withdrawals, but it appears from the presentation that all participants stayed in the study.
Heimstad *et al.* 2007: Unclear description of the definition of low-risk of the mother, but, even though we did not achieve contact with the authors, the randomization should ensure equal distribution of potential medical conditions.

Follow-up period (Q6)
All studies had follow-up periods long enough to evaluate the outcomes of interest.

Loss to follow-up (Q7)
As was the case for the randomized and quasi-experimental trials, due to the design of the studies and the character of the intervention under study, no participants withdrew, and, thus, all studies were assigned ''not applicable'' in Question 7.

Outcome measurement (comparability, measurement error) (Q5, Q8)
The questions 5 and 8 regard measurement of outcome. This review includes several different outcomes, and we estimated that the vast majority of these followed objective criteria and were measured in a reliable way in all three studies. One study lacked precise definitions on outcomes, such as birth injury, chorioamnionitis, and labor dystocia,[45] but we did not consider this an important limitation to the results.

Statistical analysis (Q9)
We considered all studies to have used appropriate statistical methods for analysis. One study, though, did not perform adjustment for confounders, because such information was not available.[44] Even though confounder adjustment is a desirable tool in cohort analyses, we concluded that the authors had used proper statistical tools, given the available data. It should be noted, that none of the three cohort studies presented information on the presence or handling of missing data. It is therefore not possible to evaluate possible biases related to missing information.

Overall, the included studies had fair methodological quality according to their respective study design. None of the experimental studies performed blinding of participants or clinicians to treatment allocation, but since this is not possible for induction of labor, we do not consider this a serious bias source. Generally, we considered treatment groups comparable at entry, though one quasi-randomized and one cohort study might be prone to selection bias. Further, we considered outcomes to be measured appropriately, data analyzed by appropriate methods, follow-up periods to be adequate, and for cohort studies, the study populations were reasonably representative of the background population.

Review findings

Perinatal and maternal outcomes are presented and evaluated according to induction vs. expectant management, as follows.

Perinatal outcomes

Table 3 gives an overview of all pooled risk estimates for perinatal outcomes. Individual meta-analyses are presented in Appendix VI.

Perinatal outcomes (primary)
Routine induction at 41+0–6 GW was associated with an almost doubled risk of low pH (< 7.10) compared to expectant management on pooled data from two RCTs (relative risk [RR] 1.90, 95% CI 1.48, 2.43). No association was observed for low Apgar score (< 7 after 5 minutes).

Perinatal outcomes (secondary)
The analysis showed a reduced risk in the induction group of oligohydramnios (RR_{RCT} 0.40, 95% CI 0.24, 0.67), meconium stained amniotic fluid (RR_{RCT} 0.82, 95% CI 0.75, 0.91), and shoulder

Table 2: Assessment of methodological quality of cohort studies

Citation	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
Cheng *et al.* (2012)[45]	Y	U	U	Y	Y	Y	NA	Y	Y
Jacquemyn *et al.* (2012)[44]	Y	Y	Y	U	Y	Y	NA	Y	Y
Liu *et al.* (2013)[46]	Y	Y	Y	Y	Y	Y	NA	Y	Y
Total (%)	100	67	67	67	100	100	0	100	100

Y, yes; N, no; NA, not applicable; U, unclear.
Q1: Is sample representative of patients in the population as a whole? Q2: Are the patients at a similar point in the course of their condition/illness? Q3: Has bias been minimized in relation to selection of cases and of controls? Q4: Are confounding factors identified and strategies to deal with them stated? Q5: Are outcome assessed using objective criteria? Q6: Was follow up carried out over a sufficient time period? Q7: Were the outcomes of people who withdrew described and included in the analysis? Q8: Were outcomes measured in a reliable way? Q9: Was appropriate statistical analysis used?
Cheng *et al.* 2012: Unclear description of risk status between groups. Data suggest that enrolment and risk status was only performed once at enrolment in 39 +0 GW for the expectant arm. No response from authors, when contacted for confirmation.
Jacquemyn *et al.* 2012: Unclear description of calculation of EDD. Author confirmed that ultrasound had been the standard method for determining EDD.
Liu *et al.* 2013: Unclear description of the risk status of participants. Author confirmed that the included women were low-risk at inclusion.

dystocia (RR_{RCT} 0.28, 95% CI 0.08, 1.00) compared to the expectant group. The latter, however, with substantially broad CI. The analysis on meconium stained amniotic fluid was compromised by substantial heterogeneity.

Perinatal death was not associated with either labor induction or expectant management; data showed a relative risk of 0.22 (95% CI 0.04, 1.32) in the intervention group compared to expectant management based on pooled data from four RCTs. This corresponds to an absolute risk reduction in perinatal death of 0.002 for routine induction at 41+0–6 GW compared to 42+0–6 GW. In other words, one prevented perinatal death per 501 inductions (95% CI 247 to 18,439) or two saved lives per 1000 inductions. Since perinatal deaths are rare, and results based on small numbers may cause important changes in clinical procedures, each perinatal death event should be described and choices on whether to include each of the events in final analysis explained. In this review, it was a criterion for inclusion of participants, that both fetus and mother were healthy and with no known risks at inclusion. This means that perinatal death events that were not related to either induction or expectant management were not included in the analysis. In the following, each perinatal death event in the included primary studies is evaluated: in the Heimstad study, one neonatal death was reported in the expectant group.[33] This fatal event occurred during labor after normal cardiotocogram at admission, and cause of death was explained by a true umbilical cord knot. This event was not included in the meta-analysis, because the death reason is not likely to be related to the intervention. Including the event did not change any conclusions (data not shown). In the Gelisen study, one intrauterine death was reported in the expectant group, but no explanation of the cause of death was given.[34] The authors were contacted for clarification, but no response was obtained. Despite the lack of explanation, this event was included in the meta-analysis in order to present ''worst case scenario''. Finally, the Burgos study comprised three perinatal deaths in the intervention group and seven

Table 3: Perinatal outcomes according to routine labor induction at 41+0–6 vs. 42+0–6 gestational weeks

Outcome	Study design	Number of studies	Number of participants	Risk measure	Effect size, 95% CI
Perinatal outcomes (primary)					
Apgar score < 7 after 5 min.	Cohort	1	74,952	RR	0.90 [0.79, 1.03]
	RCT	3	4671	RR	0.90 [0.52, 1.56]
pH < 7.10	RCT	2	4071	RR	1.90 [1.48, 2.43]
Perinatal outcomes (secondary)					
Admission to neonatal intensive care unit	Cohort	1	26,364	RR	0.94 [0.83, 1.07]
	RCT	3	4671	RR	1.08 [0.68, 1.71]
Macrosomia (birth weight > 4500g)	RCT	2	4163	RR	0.76 [0.56, 1.05]
Meconium aspiration syndrome	Cohort	1	48,518	RR	0.77 [0.57, 1.05]
	RCT	2	1108	RR	0.65 [0.31, 1.37]
Meconium stained amniotic fluid	RCT	3	4671	RR	0.82 [0.75, 0.91]
Oligohydramnios	RCT	1	508	RR	0.40 [0.24, 0.67]
Perinatal death	RCT	4	5099	RR	0.22 [0.04, 1.32]
Resuscitation	RCT	1	508	RR	1.50 [0.62, 3.61]
Shoulder dystocia	RCT	2	1038	RR	0.28 [0.08, 1.00]
Ventilation > 6 hours	Cohort	1	26,364	RR	1.36 [0.95, 1.93]

CI, confidence interval; MD, mean difference; RCT, randomized controlled trial; RR, relative risk.

in the expectant group.[43] Four (two in each group) out of the fatal events within the first 28 days after birth were explained by severe congenital malformations. These four events were not included in the present meta-analysis. In the intervention group, we included one pre-labor intrauterine death, and in the expectant management group, we included four prenatal deaths and one neonatal death within the first 24 hours after birth from the Burgos study. From this, follows that the Burgos study contributed with 62% of the data in the meta-analysis and with six of the eight perinatal deaths included. It is unclear whether a relatively high mean age (32.8 years) and proportion of nulliparous women (63–64%) in the study could contribute to the relatively high incidence of perinatal death in the expectant arm compared to other studies on the topic.[24,36] We contacted the authors for further details, who confirmed, that the expectant management group had antenatal control with CTG and amniotic fluid at 41+3/4 GW.

Maternal outcomes

Table 4 gives an overview of all pooled risk estimates for maternal outcomes.

Maternal outcomes (primary)

All relative risks for different measures of CS were above 1.00 indicating an increased risk in the intervention group compared to expectant management. Statistical significance was found for overall CS (RR_{cohort} 1.11 95% CI 1.09, 1.14) (one study, n = 74,860), CS due to failure to progress (RR_{RCT} 1.43, 95% CI 1.01, 2.03) (two studies, n = 1038), and CS after previous vaginal birth (RR_{cohort} 1.56, 95% CI 1.05, 2.34) (one study, n = 3165). The RCT-analysis on overall CS showed the same relative risk as the statistically significant cohort-analysis, but the estimate was just non-statistically significant. The analysis on CS due to failure to progress was compromised by substantial heterogeneity. Routine induction at 41+0–6 GW was associated with a 30% increased risk of instrumental vaginal delivery

Table 4: Maternal outcomes according to routine labor induction at 41+0–6 vs. 42+0–6 gestational weeks

Outcome	Study design	Number of studies	Number of participants	Risk measure	Effect size, 95% CI
Maternal outcomes (primary)					
Cesarean section (overall)	Cohort	1	74,860	RR	1.11 [1.09, 1.14]
	RCT	4	5109	RR	1.11 [0.98, 1.26]
Cesarean section (failure to progress)	RCT	2	1038	RR	1.43 [1.01, 2.03]
Cesarean section (fetal distress)	RCT	2	1038	RR	1.06 [0.75, 1.48]
Cesarean section (after previous vaginal birth)	Cohort	1	3156	RR	1.56 [1.05, 2.34
Instrumental vaginal delivery	Cohort	1	49,628	RR	1.30 [1.24, 1.36]
	RCT	3	4509	RR	1.00 [0.90, 1.11]
Vaginal delivery	RCT	3	4601	RR	0.97 [0.93, 1.02]
Maternal outcomes (secondary)					
Chorioamnionitis	Cohort	1	75,218	RR	1.13 [1.05, 1.21]
Epidural	RCT	1	438	RR	0.95 [0.76, 1.19]
Hospital stay (days)	RCT	1	438	MD	0.10 [-0.04, 0.24]
Hyperstimulation	RCT	1	600	RR	1.00 [0.29, 3.42]
Labor dystocia	Cohort	1	51,473	RR	1.29 [1.22, 1.37]
	RCT	1	508	RR	0.55 [0.20, 1.45]

Table 4. (*Continued*)

Outcome	Study design	Number of studies	Number of participants	Risk measure	Effect size, 95% CI
Manual removal of placenta	RCT	1	438	RR	1.08 [0.44, 2.66]
Maternal intensive care	Cohort	1	277,964	RR	1.59 [0.96, 2.62]
Obstetric shock	Cohort	1	277,964	RR	0.89 [0.38, 2.12]
Precipitate labor	RCT	1	508	RR	2.75 [1.45, 5.20]
Postpartum hemorrhage (blood transfusion)	Cohort	1	277,964	RR	1.04 [0.92, 1.17]
Postpartum hemorrhage > 500 ml	RCT	1	508	RR	0.91 [0.57, 1.45]
Postpartum hemorrhage > 1000ml	RCT	1	438	RR	1.43 [0.62, 3.33]
Postpartum sepsis	Cohort	1	277,964	RR	0.92 [0.71, 1.17]
Severe perineal lacerations	RCT	1	508	RR	1.20 [0.62, 2.33]
Tachysystole	RCT	1	600	RR	1.29 [0.49, 3.41]
Thromboembolism	Cohort	1	277,964	RR	1.49 [0.59, 3.77]
Uterine rupture	Cohort	1	277,964	RR	1.97 [1.54, 2.52]

CI, confidence interval; MD, mean difference; RCT, randomized controlled trial; RR, relative risk.

(RR_{cohort} 1.30, 95% CI 1.24, 1.36) (one study, n = 49,628), while pooled data from three RCTs did not show any association.

Maternal outcomes (secondary)

For secondary maternal outcomes, all analyses except vaginal delivery were based on only one of the included studies. We found routine induction at 41+0–6 GW associated with an increased risk of chorioamnionitis (RR_{cohort} 1.13, 95% CI 1.05, 1.21) (one study, n = 75,218), labor dystocia (RR_{cohort} 1.29, 95% CI 1.22, 1.37) (one study, n = 51,473), precipitate labor (RR_{RCT} 2.75, 95% CI 1.45, 5.20) (one study, n = 508), and uterine rupture (RR_{cohort} 1.97, 95% CI 1.54, 2.52) (one study, n = 277,964). The RCT-analysis on labor dystocia suffered from substantial lack of statistical power and could not support any conclusions. None of the other outcomes were associated with the intervention, except for a borderline non-significantly elevated risk of maternal intensive care (RR_{cohort} 1.59, 95% CI 0.96, 2.62) (one study, n = 277,964).

Additional findings: spontaneous onset of labor

Spontaneous onset of labor is a central element in the discussion on how to determine preferable timing of labor induction after due date. Even so, spontaneous onset of labor could not be examined in this review, because, by design, the intervention group will have close to zero spontaneous onsets of labor, and comparison with the expectant group would not make sense. Three studies, however, presented statistics on spontaneous onset of labor in the expectant group.[13,33,43] In the RCT by Heimstad, participants were randomized to immediate induction at 41+2 GW (n = 254) or to expectant management up to 42+6 GW (n = 254).[33] In the expectant group, 176 went into spontaneous labor (69%), 59 were induced due to medical reasons (most often oligohydramnios), while the remaining 19 women were still pregnant by 42+6 GW and had their labor induced. In the Daskalakis study, participants were selected for induction of labor at 41+1 GW (n = 211) or expectant management until 42+1 GW (n = 227) depending on the attending physician's preference.[13] In the expectant group, 168 went into spontaneous onset of labor (74%), 31 were induced due to medical reasons, and 28 were still pregnant, and induced, by 42+1 GW. Finally, the quasi-experimental trial by Burgos found that with a policy of inducing at 42+0 GW, 71% had spontaneous onset of labor.[43] This is in line with the pooled statistics from all three studies, where a total of 1584 out of 2227 women (71.4%), who followed expectant management,

went into spontaneous labor during the expectant time window.[13,33,43] Also, the two quasi-experimental studies[13,43] experienced a 20% increase in inductions when using a guideline of recommending induction 41+0–6 GW.

Communicating benefits and harms

Risk differences have been shown to seem greater when presented in relative terms compared to absolute numbers.[47,48] Researchers and clinicians need to be aware of this when communication scientific evidence to patients or policy makers. In order to make significant results more comprehensible Table 5 shows absolute risks (AR) and numbers needed to treat (NNT) for all outcomes with statistically significant results in Tables 3 and 4, and where estimates from observational and experimental analyses pointed in the same direction.

Absolute risk reduction (ARR) and absolute risk increase (ARI) are measures of the strength of the association, where the baseline risk in the induction and the expectant groups have been taken into account. For example, low pH will often be presented as RR = 1.90 or as a 90% increased risk after induction at 41+0–6 GW compared to expectant management. Since relative risks do not take into account the prevalence of low pH, it may be difficult for patients and stakeholders to consider the size of

Table 5: Absolute numbers for perinatal and maternal outcomes according to routine labor induction at 41+0–6 vs. 42+0–6 gestational weeks

Outcome	Study design (no.)	Absolute numbers (per 1000 inductions)	NNT (95% CI)	ARR	ARI
Perinatal outcomes					
Meconium stained amniotic fluid	RCT (n = 3)	50 fewer events	1 event prevented by 20 induced [13, 41]	4.9%	
Oligohydramnios	RCT (n = 1)	100 fewer events	1 event prevented by 10 induced [6, 20]	10.6%	
pH < 7.10	RCT (n = 2)	40 extra events	1 adverse event by 25 induced [18, 41]		4%
Maternal outcomes					
Cesarean section (overall)[†]	Cohort (n = 1)	36 extra events	1 adverse event by 28 induced [23, 33]		3.7%
Cesarean section (failure to progress)	RCT (n = 2)	33 extra events	1 adverse event by 30 induced [0, 7]		3.4%
Cesarean section (after previous vaginal birth)	Cohort (n = 1)	13 extra events	1 adverse event by 75 induced [39, 808]		1.3%
Chorioamnionitis	Cohort (n = 1)	5 extra events	1 adverse event by 195 induced [123, 460]		0.5%
Labor dystocia[‡]	Cohort (n = 1)	22 extra events	1 adverse event by 45 induced [36, 59]		2.2%
Precipitate labor	RCT (n = 1)	77 extra events	1 adverse event by 13 induced [8, 30]		8.3%
Uterine rupture	Cohort (n = 1)	0.7 extra events	1 adverse event by 1522 induced [1124, 2757]		0.07%

ARI, absolute risk increase; ARR, absolute risk reduction; CI, confidence interval; NNT, number needed to treat; RCT, randomized controlled trial.
[†]One cohort study was not accounted for in this calculation, as it only included multipara with at least one normal birth. An inclusion would have diluted the results regarding all women.
[‡]One minor inconclusive RCT with few events was ignored.

the problem. Using absolute numbers brings this aspect into the information given. In the case of low pH, the information could be that for every 1000 induced deliveries, 40 extra children would be born with a low pH compared to expectant management. Or, for every 25 induced deliveries, one extra child will be born with a low pH (with a 95% certainty between 18 and 41), and the last way to explain it is, that labor induction increases the number of children with low pH by 4%.

Discussion

The aim of this systematic review was to evaluate the effects of induction of labor prior to post-term in low-risk pregnancies. Routine induction at 41+0–6 GW has become a widespread obstetric practice in many Western countries despite a lack of systematic reviews that address both harms and benefits of the intervention. Acknowledging the fact that our analyses are based on a limited number of primary studies, this review qualifies current knowledge and depicts critical methodological aspects in existing research in the field. In the following, we present main findings, discuss the quality of included studies, and describe important methodological problems in systematic reviews on labor induction. Implications for practice from changed guidelines and derived organizational/economic consequences are also touched upon, as well as the communication of systematic review results to patients.

In summary, we found induction at 41+0–6 GW associated with a few beneficial outcomes (e.g. a reduced risk of oligohydramnios and meconium stained amniotic fluid) and several adverse outcomes (e.g. an increased risk of precipitate labor, labor dystocia, CS, uterine rupture, chorioamnionitis and low pH in the newborn) compared to expectant management. As emphasized by the WHO, expected benefits from a medical intervention must outweigh potential harms.[3]

Overall, previous systematic reviews have found induction associated with a reduced risk of CS, perinatal death, meconium stained amniotic fluid and perhaps meconium aspiration syndrome with induction prior to post-term.[6,35,36] Our results largely support previous findings on perinatal outcomes, except for perinatal death and aspiration syndrome. In regard to maternal outcomes, our findings generally do not support those of previous reviews, i.e. we found an increased risk of CS and several adverse outcomes.

Regarding perinatal outcomes, our findings are aligned with a systematic review by Caughey *et al.*[6] finding that evidence was insufficient to form conclusions for many of the key neonatal outcomes (e.g. neonatal death, fetal distress and asphyxia). As for perinatal death, our data indicated two possibly prevented fetal deaths per 1000 inductions with routine induction at 41+0–6 GW. However, conclusions could not be drawn due to lack of statistical power illustrated by a very broad CI in relative numbers (RR 0.22, 95% CI 0.04, 1.32) and, in absolute numbers, a range of 247 to 18,439 inductions to be performed to prevent one perinatal death. A Cochrane meta-analysis on induction of labor at or beyond term[35] found that a policy of labor induction was associated with fewer (all-cause) perinatal deaths (RR 0.31, 95% CI 0.12, 0.88), and Wennerholm *et al.*[36] found a decreased risk of perinatal death (RR 0.33, 95% CI 0.10, 1.09), even though it was not statistically significant. In the Wennerholm study, only pregnancies beyond term were included, which makes this review more comparable to ours. Wennerholm *et al.*[36] found meconium aspiration syndrome to be reduced (RR 0.43, 95% CI 0.23, 0.79), and Caughey *et al.* and Gulmezoglu *et al.* found similar conclusions.[6,35] Our data showed a decreased risk of meconium stained amniotic fluid, but we found no association with meconium aspiration syndrome. Further, we found an increased risk of pH < 7.10 in the newborn, which can be a serious acute condition and is associated with long-term consequences, such as cerebral palsy. On the other hand, we did not find low Apgar score associated with induction prior to post-term.

Regarding maternal outcomes, Wennerholm *et al.*[36] found a 13% decreased risk of CS after induction (RR 0.87, 95% CI 0.80–0.96), and the findings are supported by a Cochrane review[35] and Caughey *et al.*[6] In the present review, we found an 11% increased risk of CS. We believe this is due to differences in methodology, which will be discussed below. The pooled RR for CS was statistically significant in the cohort analysis, but not in the analysis including four RCTs. The largest RCT, however, by Burgos[43] which weighs 54%, found a significant increase in CS, which supports the conclusions from our cohort analysis. Cesarean section has been associated with an increased risk of excessive bleeding,

pain, bladder disorders, hysterectomy, and neonatal death, and, in subsequent pregnancies, placenta previa, fetal death and uterine rupture.[49-51] Finally, we found chorioamnionitis, labor dystocia, precipitate labor and uterine rupture associated with induction prior to post-term. Uterine rupture is a rare, but potentially life-threatening condition, and it imposes severe risks in subsequent pregnancies.[52] Unfortunately, even though several adverse birth outcomes are associated with long-term risks, we were not able to evaluate such effects, since none of the primary studies included long-term outcomes.

Overall, the methodological quality of the included studies was considered high when assessed according to the JBI SUMARI critical appraisal tools for RCTs/quasi-experimental studies and observational studies, respectively. This was expected, because one of the main purposes of this review was to restrict data to updated studies. Generalizability and length of follow-up period (both relevant only for cohort studies), measurement and assessment of outcome and statistical analysis were the items best covered in the included studies. Comparability of groups at entry was not satisfactory in two out of seven studies. For example, in one quasi-experimental trial, assignment to treatment groups was determined according to the attending physician's preference. Selection bias due to loss of follow-up was minimal in the two RCTs that this criterion applied to, whereas one of the two RCTs was prone to selection bias due to treatment differences, because induction agents differed between groups.

The question(s) addressed in systematic reviews are said often to be (too) simple and (too) general, e.g. a title like "Elective Induction of Labor Versus Expectant Management of Pregnancy"[6] may include all varieties of women, a broad variation in gestational age, different induction agents etc. This is the reason why careful attention should be paid to inclusion criteria, when systematic reviews are interpreted. Systematic reviews form the main evidence base for clinical guidelines today, and it is crucial that policy makers and stakeholders pay careful attention to the criteria for inclusion of studies, because inclusion criteria are central for a review's external validity and relevance.[53]

Regarding gestational age criteria, even though our literature search resulted in a large amount of papers on induction of labor versus expectant management, only a few studies compared the relatively tight timeframes of 41+0–6 GW and 42+0–6 GW, which reflects the new routine versus the old standard in many Western countries. Tight timeframes for comparison of the groups were chosen in the present review, because they reflect this change in practice. In the Cochrane review, only three of 22 randomized studies met this criterion, while the remaining 19 studies included variations from 37 GW up to expectant management with no upper time limit.[35] In the present study, we only identified four RCTs or quasi-randomized trials (n = 5109) and three cohort studies (n = 356,338), that met our criteria regarding gestational age. We argue that meta-analyses that use broad criteria for gestational age in the expectant group are likely to overestimate poor fetal or maternal outcomes from expectant management. The main reason for this is that the expectant group includes excessively long pregnancies with correspondingly increased complication risk.[6,35,36] We further argue that conclusions from such studies may contribute unfairly to evidence, because this does not reflect usual practice, where expectant management beyond 42+6 GW is close to non-existent. All of the above mentioned systematic reviews by Wennerholm *et al.*,[36] Gulmezoglu *et al.*,[35] and Caughey *et al.*[6] include studies with excessively long expectant management, e.g. one of the studies weighting the most in the systematic reviews uses expectant management until 44+0 GW.[28]

Regarding temporal criteria, the cited reviews included studies dating back to the 1960s and 1970s. At this time, fetal surveillance was less developed, induction agents were different from today, and both induction and CS rates were much lower. Caughey *et al.* describes restriction of studies to include only those "published from 1966 and beyond to represent modern obstetric practice".[6(p253)] We do not agree with the authors and find that obstetric practice (e.g. induction agents, CS rates and possibilities of monitoring pregnancy and childbirth) has changed substantially since 1966. Further, population characteristics have changed over this time. One could ask if it is reasonable to base a contemporary guideline on studies that are 40–50 years old. We suggest not, and have restricted publication dates of included studies to the last 20 years. The price to pay is fewer studies but with greater relevance.

Regarding estimation of due date, when the aim is to compare policies for induction of labor according to gestational age, the estimation of due date is

paramount, and, consequently, the methods for estimation should be comparable. Previous reviews included studies with varied methods for estimating due date. Since EDD tends to differ between early ultrasound estimation and calculation by last menstrual period, this approach introduces a possible bias source to the analysis. Hence, in the present review we restricted data to studies using early ultrasound estimation.

In the present review, we allowed only for generally known methods of induction to be included (prostaglandins, balloon, oxytocin and artificial rupture of membranes). Even so, the included studies use different induction agents, and, thus, the external validity may be compromised. Furthermore, it was not possible to stratify on parity or rupture of membranes due to lack of information in the studies, and hence, both might be relevant. These factors might bias the results and we suggest the clinicians and stakeholders look thoroughly into individual studies before making local recommendations.

As described in the Methods section, the analytical approach for comparing groups is crucial to obtain a realistic scope for deciding recommendations for low-risk pregnancies past due date.[25] Two different courses of action were compared in the present review: an active induction regimen and an expectant management, and data were restricted to a maximum expectancy up to 42+6 GW, since this reflects common practice, due to the increased risk of adverse outcome beyond that point. Our data showed that when women in the expectant arm were left without induction during week 41+0–6, a large proportion (~70%) did in fact go into spontaneous labor before induction was recommended.[13,33,43] However, the cited systematic reviews include studies that failed to include this sizable group of women in their analysis.[6,35,36] When those in the expectant arm who gave birth during week 41+0–6 are excluded from analysis, the expectant arm is composed only of those who have reached 42+0 GW of pregnancy without the spontaneous onset of labor. Pregnancies that last beyond 42+0 GW are beyond the normal range of gestational time and are associated with an increased risk of complications. Therefore, comparison of outcomes between an induction arm composed of low-risk, healthy women whose labor was induced at 41+0–6 GW, and an expectant arm composed of women who have reached 42+0 GW and are therefore at heightened risk of complications is likely to bias the results in favor of induction. The risks associated with expectant management are likely to be overestimated. The problem can be circumvented if the large group of women in the expectant arm who go into spontaneous labor before 42+0 GW are included. Then a more accurate comparison can be made as to the effects of induction 41+0–6 GW versus expectant management, which, as has been noted, will in many cases result in spontaneous onset of labor before 42+0 GW. Therefore, in the present review, we draw only upon studies that include data on all participants in the expectant arm, including those who give birth between 41+0–6 and 42+0 GW. This will more correctly reflect the implications of using the two different guidelines, induction versus expectant management, and is probably the explanation as to why we found an increase in CS when performing routine induction in 41+0–6 GW, while other reviews found otherwise. We hope that issuing this topic will make stakeholders or authors of guidelines more attentive to the methodologies of the studies on which their recommendations are based.

The main limitation in this review concerns the limited number of included studies. It would, however, not be appropriate to rely on studies that do not include the gestational weeks under investigation, are up to 50 years old, use different methods for estimation of due date, or exclude the vast majority of women that go into spontaneous labor under an expectant management of pregnancies. We discussed problems with the wide variations in current systematic reviews and described how we have attempted to overcome the problem by defining strict criteria for inclusion. This decision had a major implication concerning the power of our analysis, as only seven studies were found to be eligible. To summarize, as Greenhalgh states: ''any numerical result, however precise, accurate, 'significant,' or otherwise incontrovertible, must be placed in the context of the painfully simple and often frustratingly general question which the review addressed''.[53(p.674)] Also, we were only able to separate groups by the number of weeks, even though different regimens should ideally be compared according to single days. Some guidelines recommend that the women deliver before 42+0 GW, which means that induction is not initiated until 41+3–5 GW.[18,19] In practice, this implies that a considerable number of women will have time to go into spontaneous labor before reaching the

time for recommended induction. As we compile data on a weekly basis, our findings may overestimate the effects regarding both benefits and harms if the groups in reality are only separated by a few days.

Finally, none of the studies included indicators of long-term effects on the child, such as cerebral palsy or hypoxic-ischemic encephalopathy. Therefore, we included pH < 7.10 as the best available proxy measure for possible long-term offspring outcomes. This is a deviation from the original protocol.[37]

Conclusion

The objective of this systematic review was to compare the effect of routine labor induction between one and two weeks past EDD to the practice of awaiting spontaneous onset of labor until two weeks past EDD. The review included seven primary studies. We found routine induction at 41+0–6 GW associated with a decreased risk of meconium stained amniotic fluid and oligohydramnios and with an increased risk of CS, labor dystocia, chorioamnionitis, precipitate labor, uterine rupture and low pH in the newborn. No conclusions could be drawn on perinatal death due to lack of statistical power. For perinatal outcomes, our findings largely support previous meta-analyses, except for perinatal death. For maternal outcomes (including outcomes related to the course of labor), routine induction was associated with several adverse outcomes, including an increased risk of CS, which is not in line with previous reviews. The current review used stricter inclusion criteria than most of the previous reviews, which is potentially the main reason for discrepancies from previous findings. These restrictions were applied in order to enhance the methodological quality and increase the relevance for contemporary maternity care. Our findings do not support the widespread use of induction prior to post-term, and they highlight the importance of discussing whether the threshold has been reached where risks related to the procedure of induction outweigh potential harms from the ongoing pregnancy, and whether induction of labor should be applied to large populations of low-risk women. If an intervention cannot demonstrate a positive association between higher intervention rates and an improved perinatal outcome, and evidence exists that the intervention itself may impose maternal and perinatal risks, it is normally decided to strive for the lowest intervention rates. In the case of routine induction prior to post-term, the challenge is how to prevent fetal demise leading to poor fetal outcome and in few occasions of fetal death. Nevertheless, one might ask if routine induction of labor on larger populations of healthy women and fetuses is the answer to this task, when – as demonstrated in this review – the intervention is associated with several iatrogenic effects.

Recommendations for practice

As much as the above discussion on research methodology and inclusion criteria in systematic reviews may appear as a purely academic one, it affects clinical daily life. Since clinical guidelines mainly rely on systematic reviews, and since systematic reviews are justified by their ability to weigh and sum up relevant research on a given subject, such discussions are highly relevant for practice. In the current review, we used stricter criteria for inclusion, and we found that doing so changed the conclusions from previous reviews. We suggest increased awareness of inclusion criteria in systematic reviews as they can reflect contemporary clinical practice, and the findings can be applied to guidelines and practice. We also suggest that clinicians and stakeholders look into original individual studies before making local guidelines.

The current review revealed several adverse outcomes associated with routine induction, which may not be prioritized in the information given before consent to labor induction.[54] It is of the utmost importance that women are thoroughly informed about benefits and harms before making their decision on induction by routine, as described in general requirements of health legislations. In order to present balanced and understandable information, absolute numbers may also be needed. The current review presents absolute numbers to raise awareness of the communication of scientific evidence among users and policy makers. As health professionals, we should be aware that our choice of which risk measures to present works as a tool by which we can influence people's decisions regarding routine induction or not. We suggest that future research and clinical practice include absolute numbers and absolutes risks in addition to the usual relative estimates.

According to Joanna Briggs Institute Grades of Recommendations, our findings do not support the obstetric practice for low-risk pregnancies with routine labor induction at 41+0 to 41+6 GW when

compared to expectant management (routine labor induction at 42+0 to 42+6 GW), because desirable effects do not appear to outweigh undesirable effects of such practice. There is not strong evidence supporting its use, and there does not seem to be any resource benefits from following this practice (Grade B). Values, preferences and the patient experience have not been taken into account.

Recommendations for research

The current review revealed important methodological weaknesses in existing systematic reviews on labor induction prior to post-term. One important problem is the use of studies with improper gestational timeframes in the comparison group. Another problem is the use of old studies. Obstetric practice (fetal surveillance, induction agents, clinical procedures, intervention rates etc.) as well as population characteristics have changed substantially since the 1960s–1980s. It is questionable whether findings from studies of such age can be extrapolated to a contemporary setting. A third problem is the use of studies with different methods for estimating due date and/or different induction agents within or across studies. The above aspects are all likely to bias the conclusions reported in some of the key scientific papers behind existing clinical guidelines. In the current review, we used stricter criteria for inclusion of studies in order to overcome some of the above flaws. This approach has resulted in a generally positive assessment of the methodological quality of the included studies, but the consequence has been a smaller sample size than in most previous reviews. We suggest that systematic reviews, with all their best efforts to promote evidence-based decision-making, should be studied carefully by stakeholders for external validity, since only few studies in a review may be relevant for the given clinical decision-making process.

The review process revealed a focus on few, selected outcomes in labor induction studies. Cesarean section and perinatal death have gained much focus so far, while e.g. hyperstimulation, tachysystole, and precipitate labor are understudied. This is of interest, because routine induction is performed on a healthy low-risk population, and in such cases, only a low risk of iatrogenic effects is usually accepted. We also found that several central aspects were generally not covered or inadequately accounted for in existing research. For example, it was not possible to evaluate long-term effects, the need for care, economic resources, or the experience of birth when artificially induced compared to being in a tight surveillance regimen awaiting spontaneous onset of labor. Long-term effects in the child may include cerebral palsy, brachial plexus injury, or hypoxic-ischemic encephalopathy. Even though pH < 7.10 and resuscitation may be indicators for such long-term effects, we suggest that future studies include a broad selection of both short- and long-term outcomes when evaluating the procedure of routine induction prior to post-term.

The current review revealed a 20% increase in inductions when using a guideline of recommending induction 41+0–6 GW, which affects workload in the labor ward. Inductions require extra attention compared to spontaneously initiated deliveries, and these resources may be drawn away from other laboring women. We suggest that future studies focus on economic consequences of new routines as well as health outcomes related to busy staff.

Acknowledgments

We thank Bodil K. Møller (Head of Midwifery Programme), Anna Møbjerg (Search Librarian), and Christina Ruggiero-Corliss (Visiting Student), University College Copenhagen, Copenhagen, for help and support.

References

1. Enkin MW, Keirse MJ, Neilson J, *et al*. A guide to effective care in pregnancy and childbirth. 1st ed. Oxford: OUP Oxford; 2001.
2. Dunne C, Da Silva O, Schmidt G, Natale R. Outcomes of elective labour induction and elective caesarean section in low-risk pregnancies between 37 and 41 weeks' gestation. J Obstet Gynaecol Can 2009;31(12):1124–30.
3. WHO. WHO recommendations for induction of labour.2011.
4. Caughey AB, Sundaram V, Kaimal AJ, Cheng YW, Gienger A, Little SE, *et al*. Maternal and neonatal outcomes of elective induction of labor. Evid Rep Technol Assess (Full Rep) 2009;176:1–257.
5. Mayes' midwifery: A textbook for midwives. 14th edition. Bailliere, Tindall: Elsevier; 2012.
6. Caughey AB, Sundaram V, Kaimal AJ, Gienger A, Cheng YW, McDonald KM, *et al*. Systematic review: Elective induction of labor versus expectant management of pregnancy. Ann Intern Med 2009;151(4):252–63.
7. Michell V, Rosenorn-Lanng D, Gulliver S, Currie W. Handbook of research on patient safety and quality care through health informatics. USA: Medical Records Science Reference; 2014.

8. Ananth CV, Wilcox AJ, Gyamfi-Bannerman C. Obstetrical interventions for term first deliveries in the US. Paediatr Perinat Epidemiol 2013;27(5):442–51.
9. Glantz JC. Labor induction rate variation in upstate new york: What is the difference? Birth 2003;30(3):168–74.
10. Le Ray C, Carayol M, Breart G, Goffinet F. PREMODA Study Group. Elective induction of labor: Failure to follow guidelines and risk of cesarean delivery. Acta Obstet Gynecol Scand 2007;86(6):657–65.
11. Ekeus C, Lindgren H. Induced labor in Sweden, 1999–2012: A population-based cohort study. Birth 2016;43(2):125–33.
12. Nguyen RH, Wilcox AJ. Terms in reproductive and perinatal epidemiology: 2. perinatal terms. J Epidemiol Community Health 2005;59(12):1019–21.
13. Daskalakis G, Zacharakis D, Simou M, Pappa P, Detorakis S, Mesogitis S, *et al*. Induction of labor versus expectant management for pregnancies beyond 41 weeks. J Matern Fetal Neonatal Med 2013;online:1–4.
14. Poignant M, Hjelmstedt A, Ekeus C. Indications for operative delivery between 1999–2010 and induction of labor and epidural analgesia on the risk of operative delivery–a population based Swedish register study. Sex Reprod Healthc 2012;3(4):129–34.
15. Grivell RM, Reilly AJ, Oakey H, Chan A, Dodd JM. Maternal and neonatal outcomes following induction of labor: A cohort study. Acta Obstet Gynecol Scand 2012;91(2):198–203.
16. Simpson KR, Thorman KE. Obstetric "conveniences": Elective induction of labor, cesarean birth on demand, and other potentially unnecessary interventions. J Perinat Neonatal Nurs 2005;19(2):134–44.
17. American College of Obstetricians, Gynecologists (ACOG). Management of postterm pregnancy.1997.
18. National Institute for Health and Clinical Excellence. Induction of labour.2008.
19. Clinical Practice Obstetrics Committee, Maternal Fetal Medicine Committee. Delaney M, *et al*. Guidelines for the management of pregnancy at 41+0 to 42+0 weeks. J Obstet Gynaecol Can 2008;30(9):800–23.
20. Wagner M. Born in the USA: How a broken maternity system must be fixed to put women and children first. USA: University of California Press; 2008.
21. Glantz JC. Term labor induction compared with expectant management. Obstet Gynecol 2010;115(1):70–6.
22. Haavaldsen C, Sarfraz A, Eskild A. Low fetal death risk in post-term pregnancy in Norway. Tidsskr Nor Laegeforen 2010;130(21):2114.
23. Vayssiere C, Haumonte JB, Chantry A, *et al*. Prolonged and post-term pregnancies: Guidelines for clinical practice from the french college of gynecologists and obstetricians (CNGOF). Eur J Obstet Gynecol Reprod Biol 2013;169(1):10–6.
24. Heimstad R, Romundstad PR, Salvesen KA. Induction of labour for post-term pregnancy and risk estimates for intrauterine and perinatal death. Acta Obstet Gynecol Scand 2008;87(2):247–9.
25. Vardo JH, Thornburg LL, Glantz JC. Maternal and neonatal morbidity among nulliparous women undergoing elective induction of labor. J Reprod Med 2011;56(1–2):25–30.
26. Kaufman KE, Bailit JL, Grobman W. Elective induction: An analysis of economic and health consequences. Am J Obstet Gynecol 2002;187(4):858–63.
27. Chanrachakul B, Herabutya Y. Postterm with favorable cervix: Is induction necessary? Eur J Obstet Gynecol Reprod Biol 2003;106(2):154–7.
28. Hannah ME, Hannah WJ, Hellmann J, Hewson S, Milner R, Willan A. Induction of labor as compared with serial antenatal monitoring in post-term pregnancy. A randomized controlled trial. the canadian multicenter post-term pregnancy trial group. N Engl J Med 1992;326(24):1587–92.
29. Arrowsmith S, Wray S, Quenby S. Maternal obesity and labour complications following induction of labour in prolonged pregnancy. BJOG 2011;118(5):578–88.
30. Caughey AB, Nicholson JM, Cheng YW, Lyell DJ, Washington AE. Induction of labor and cesarean delivery by gestational age. Am J Obstet Gynecol 2006;195(3):700–5.
31. Darney BG, Snowden JM, Cheng YW, Jacob L, Nicholson JM, Kaimal A, *et al*. Elective induction of labor at term compared with expectant management: Maternal and neonatal outcomes. Obstet Gynecol 2013;122(4):761–9.
32. Kjeldsen LL, Sindberg M, Maimburg RD. Earlier induction of labour in post term pregnancies–A historical cohort study. Midwifery 2015;31(5):526–31.
33. Heimstad R, Skogvoll E, Mattsson LA, Johansen OJ, Eik-Nes SH, Salvesen KA. Induction of labor or serial antenatal fetal monitoring in postterm pregnancy: A randomized controlled trial. Obstet Gynecol 2007;109(3):609–17.
34. Gelisen O, Caliskan E, Dilbaz S, Ozdas E, Dilbaz B, Ozdas E, *et al*. Induction of labor with three different techniques at 41 weeks of gestation or spontaneous follow-up until 42 weeks in women with definitely unfavorable cervical scores. Eur J Obstet Gynecol Reprod Biol 2005;120(2):164–9.
35. 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.
36. Wennerholm UB, Hagberg H, Brorsson B, Bergh C. Induction of labor versus expectant management for post-date pregnancy: Is there sufficient evidence for a change in clinical practice? Acta Obstet Gynecol Scand 2009;88(1):6–17.
37. Rydahl E, Eriksen LM. The effects of induction of labor prior to post-term in low risk pregnanices: A systematic review protocol. JBI Database System Rev Implement Rep 2014;12(6):36–48.
38. Hoffman CS, Messer LC, Mendola P, Savitz DA, Herring AH, Hartmann KE. Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. Paediatr Perinat Epidemiol 2008;22(6):587–96.
39. ACOG, American Congress of Obstetricians, Gynecologists. Elective delivery before 39 weeks.2013.

40. Knoche A, Selzer C, Smolley K. Methods of stimulating the onset of labor: An exploration of maternal satisfaction. J Midwifery Womens Health 2008;53(4):381–7.
41. The Joanna Briggs Institute. Joanna Briggs Institute Reviewers' Manual: 2017 edition Australia: The Joanna Briggs Institute; 2017; Available from https://reviewersmanual.joannabriggs.org/.
42. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557–60.
43. Burgos J, Rodriguez L, Otero B, Cobos P, Osuna C, Centeno Mdel M, *et al*. Induction at 41 weeks increases the risk of caesarean section in a hospital with a low rate of caesarean sections. J Matern Fetal Neonatal Med 2012;25(9):1716–8.
44. Jacquemyn Y, Michiels I, Martens G. Elective induction of labour increases caesarean section rate in low risk multiparous women. J Obstet Gynaecol 2012;32(3):257–9.
45. Cheng YW, Kaimal AJ, Snowden JM, Nicholson JM, Caughey AB. Induction of labor compared to expectant management in low-risk women and associated perinatal outcomes. Am J Obstet Gynecol 2012;207(6):502e1–8.
46. Liu S, Joseph KS, Hutcheon JA, Bartholomew S, León JA, Walker M, *et al*. Gestational age-specific severe maternal morbidity associated with labor induction. Am J Obstet Gynecol 2013;209(3):209e1–8.
47. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: Ten steps to better risk communication. J Natl Cancer Inst 2011;103(19):1436–43.
48. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. communicating statistical information. Science 2000;290(5500):2261–2.
49. The British Columbia Health Care Program. Caesarean birth task force program 2008.2008.
50. World Health Organization. WHO statement on caesarean section rates.2015.
51. Ecker J. Elective cesarean delivery on maternal request. JAMA 2013;309(18):1930–6.
52. Veena P, Habeebullah S, Chaturvedula L. A review of 93 cases of ruptured uterus over a period of 2 years in a tertiary care hospital in south india. J Obstet Gynaecol 2012;32(3):260–3.
53. Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analyses). BMJ 1997;315:672.
54. Clausen JA, Juhl M, Rydahl E. Quality assessment of patient leaflets on misoprostol-induced labour: Does written information adhere to international standards for patient involvement and informed consent? BMJ Open 2016;6(5):e011333–2016–011333.
55. Allen VM, Stewart A, O'Connell CM, Baskett TF, Vincer M, Allen AC. The influence of changing post-term induction of labour patterns on severe neonatal morbidity. J Obstet Gynaecol Can 2012;34(4):330–40.
56. Bailit JL, Grobman W, Zhao Y, Wapner RJ, Reddy UM, Warner MW, *et al*. Nonmedically indicated induction vs expectant treatment in term nulliparous women. Am J Obstet Gynecol 2015;212(1):103e1–7.
57. Bleicher I, Vitner D, Iofe A, Sagi S, Bader D, Gonen R. When should pregnancies that extended beyond term be induced?(). J Matern Fetal Neonatal Med 2017;30(2):219–23.
58. Caughey AB, Washington AE, Laros RK Jr. Neonatal complications of term pregnancy: Rates by gestational age increase in a continuous, not threshold, fashion. Am J Obstet Gynecol 2005;192(1):185–90.
59. Duff C, Sinclair M. Exploring the risks associated with induction of labour: A retrospective study using the NIMATS database. northern ireland maternity system. J Adv Nurs 2000;31(2):410–7.
60. Fok WY, Chan LY, Tsui MH, Leung TN, Lau TK, Chung TK. When to induce labor for post-term? A study of induction at 41 weeks versus 42 weeks. Eur J Obstet Gynecol Reprod Biol 2006;125(2):206–10.
61. Kiesewetter B, Lehner R. Maternal outcome monitoring: Induction of labor versus spontaneous onset of labor-a retrospective data analysis. Arch Gynecol Obstet 2012;286(1):37–41.
62. Klefstad OA, Okland I, Lindtjorn E, Rygh AB, Kaada K, Hansen ML, *et al*. A more liberal approach towards induction of labour in prolonged pregnancy does not result in an adverse labour outcome. Dan Med J 2014;61(9):A4913.
63. Kwee A, Elferink-Stinkens PM, Reuwer PJ, Bruinse HW. Trends in obstetric interventions in the dutch obstetrical care system in the period 1993–2002. Eur J Obstet Gynecol Reprod Biol 2007;132(1):70–5.
64. Nakling J, Backe B. Pregnancy risk increases from 41 weeks of gestation. Acta Obstet Gynecol Scand 2006;85(6):663–8.
65. Oros D, Bejarano MP, Cardiel MR, Oros-Espinosa D, Gonzalez de Aguero R, Fabre E. Low-risk pregnancy at 41 weeks: When should we induce labor? J Matern Fetal Neonatal Med 2012;25(6):728–31.
66. Page JM, Snowden JM, Cheng YW, Doss AE, Rosenstein MG, Caughey AB. The risk of stillbirth and infant death by each additional week of expectant management stratified by maternal age. Am J Obstet Gynecol 2013;209(4):375e1–7.
67. Pavicic H, Hamelin K, Menticoglou SM. Does routine induction of labour at 41 weeks really reduce the rate of caesarean section compared with expectant management? J Obstet Gynaecol Can 2009;31(7):621–6.
68. Raviraj P, Shamsa A, Bai J, Gyaneshwar R. An analysis of the NSW midwives data collection over an 11-year period to determine the risks to the mother and the neonate of induced delivery for non-obstetric indication at term. ISRN Obstet Gynecol 2013;2013:178415.
69. Rosenstein MG, Cheng YW, Snowden JM, Nicholson JM, Caughey AB. Risk of stillbirth and infant death stratified by gestational age. Obstet Gynecol 2012;120(1):76–82.
70. Sobande AA, Albar HM. Outcome of induced labour in pregnancies at 41 weeks gestation and over in saudi arabia. East Mediterr Health J 2003;9(3):316–23.

71. Stock SJ, Ferguson E, Duffy A, Ford I, Chalmers J, Norman JE. Outcomes of elective induction of labour compared with expectant management: Population based study. BMJ 2012;344:e2838.
72. Sue-A-Quan AK, Hannah ME, Cohen MM, Foster GA, Liston RM. Effect of labour induction on rates of stillbirth and cesarean section in post-term pregnancies. CMAJ 1999;160(8):1145–9.
73. Treger M, Hallak M, Silberstein T, Friger M, Katz M, Mazor M. Post-term pregnancy: Should induction of labor be considered before 42 weeks? J Matern Fetal Neonatal Med 2002;11(1):50–3.
74. Weiss E, Krombholz K, Eichner M. Fetal mortality at and beyond term in singleton pregnancies in baden-wuerttemberg/germany 2004–2009. Arch Gynecol Obstet 2014;289(1):79–84.
75. Yazdani M, Shakeri S, Farsipur-Naghibi M. Outcome of post-term pregnancies in southern iran. Int J Gynaecol Obstet 2006;93(2):144–5.

Appendix I: Individual search strategies

PubMed (pubmed.gov): searched on 18/06/2018

Search	Query
#1	Pregnancy [MeSH] OR pregnancy
#2	"Low risk" OR Normal OR Term OR "low-risk"
#3	Post-term OR Prolonged OR Prolonged pregnancy [MeSH] OR "Prolonged pregnancy" OR Post-date OR Past due date OR Beyond term OR "41 weeks" OR "41 gestational weeks" OR "42 weeks" OR "42 gestational weeks"
#4	Induced labor OR Induced labour OR Induction OR induce OR Cervical Ripening OR "Labor induced" OR Labor induced [MeSH] OR Elective induction OR Routine induction OR Expectant management
#5	Adverse OR adverse effects [Subheading] OR "adverse effects" OR Adverse events OR Caesarean section OR cesarean section [MeSH] OR "cesarean section" OR Cesarean OR caesarean OR caesarian.OR Mortality OR Morbidity OR complications OR "fetal complications" OR "perinatal outcome" OR "neonatal outcome" OR risk OR "stillbirth mortality " OR "Perinatal mortality" [Mesh] OR "perinatal mortality" OR "fetal death" [Mesh] OR "fetal death" OR "Stillbirth" OR "Intrauterine death"
#6	#1 AND #2 AND #3 AND #4 AND # 5
Limited to 1998 and English, Danish, Swedish, and Norwegian language (482 records)	

CINAHL (EBSCO): searched on 18/06/2018

Search	Query
#1	Pregnancy
#2	"Low risk" OR Normal OR Term OR "low-risk"
#3	Post-term OR Prolonged OR pregnancy, prolonged [MH] OR Post-date OR Past due date OR Beyond term OR "41 weeks" OR "41 gestational weeks" OR "42 weeks" OR "42 gestational weeks"
#4	Induced labor OR Induced labour OR Induction OR induce OR Cervix Dilatation and Effacement [MH] OR Labor, induced [MH] OR Elective induction OR Routine induction OR Expectant management
#5	Adverse OR Caesarean section OR cesarean section [MH] OR Mortality [MH] OR Morbidity [MH] OR complications OR pregnancy complications [MH] OR "fetal complications" OR "perinatal outcome" OR "neonatal outcome" OR Perinatal mortality [MH] OR "Stillbirth"
#6	#1 AND #2 AND #3 AND #4 AND # 5
Limited to 1998, published and English language (41 records)	

Embase (Elsevier): searched on 18/06/2018

Search	Query
#1	low risk pregnancy /exp
#2	''Low risk'' OR Normal OR Term OR ''low-risk''
#3	Post-term OR Prolonged OR Prolonged pregnancy /exp OR Post-date OR Past due date OR Beyond term OR ''41 weeks'' OR ''41 gestational weeks'' OR ''42 weeks'' OR ''42 gestational weeks''
#4	Induced labor OR Induced labour OR Induction OR induce OR uterine cervix ripening / exp OR Labor, induction /exp OR Elective induction OR Routine induction OR Expectant management
#5	Adverse OR Adverse effects OR Adverse events /exp OR Caesarean section OR cesarean section /exp OR Cesarean OR caesarean OR caesarian OR Mortality /exp OR Morbidity OR complications OR ''fetal complications'' OR ''perinatal outcome'' OR ''neonatal outcome'' OR risk OR ''stillbirth mortality '' OR perinatal mortality /exp OR fetus mortality /exp OR ''fetal death'' OR fetus death /exp OR ''Stillbirth'' OR ''Intrauterine death''
#6	#1 AND #2 AND #3 AND #4 AND # 5
Limited to 1998 and English, Danish, Swedish, and Norwegian language and human (51 records)	

Scopus (Elsevier): searched on 18/06/2018

Search	Query
#1	Pregnancy
#2	''Low risk'' OR Normal OR Term OR ''low-risk''
#3	Post-term OR Prolonged OR Prolonged pregnancy OR Post-date OR Past due date OR Beyond term OR ''41 weeks'' OR ''41 gestational weeks'' OR ''42 weeks'' OR ''42 gestational weeks''
#4	Induced labor OR Induced labour OR Induction OR induce OR Cervical Ripening OR Labor, induced OR Elective induction OR Routine induction OR Expectant management
#5	Adverse OR Adverse effects OR Adverse events OR Caesarean section OR cesarean section OR Cesarean OR caesarean OR caesarian.OR Mortality OR Morbidity OR complications OR ''fetal complications'' OR ''perinatal outcome'' OR ''neonatal outcome'' OR risk OR ''stillbirth mortality ''OR ''Perinatal mortality'' OR ''fetal death'' OR ''Stillbirth'' OR ''Intrauterine death''
#6	#1 AND #2 AND #3 AND #4 AND # 5
Limited to 1998 and English, Danish, Swedish, and Norwegian, Human within Nursing, Medicine & Health, and published in medical-, obstetrical-, gynecological- or pediatric journals (1451 records)	

Swemed+: searched on 15/03/2018

Search	Query
#1	Pregnancy
#2	Induction
#3	#1 AND #2
Limited 1998 (18 records)	

POPLINE (K4health): searched on 15/03/2018

Search	Query
#1	Pregnancy
#2	Induced labor OR Induced labour OR Induction
#3	Adverse OR Caesarean OR Cesarean OR Mortality OR Morbidity OR complications OR fetal complications OR perinatal outcome OR ''neonatal outcome'' OR risk OR stillbirth mortality OR Perinatal mortality OR fetal death OR Stillbirth OR Intrauterine death
#6	#1 AND #2 AND #3
Limited 1998 and English (90 records)	

Cochrane Library: searched on 18/06/2018

Search	Query
#1	Pregnancy
#2	''Low risk'' OR Normal OR Term OR ''low-risk''
#3	''Postterm'' OR ''prolonged''OR Prolonged pregnancy OR Post-date OR Past due date
#4	Beyond term OR ''41 weeks'' OR ''41 gestational weeks'' OR ''42 weeks'' OR ''42 gestational weeks''
#5	''Induced labor'' OR ''Induced labour'' OR ''Induction'' OR ''induce'' OR ''Cervical Ripening''
#6	''Labor induced'' OR Elective induction OR Routine induction OR Expectant management
#7	Adverse OR Adverse effects OR Adverse events OR Caesarean section OR cesarean section
#8	Cesarean OR caesarean OR caesarian OR Mortality OR morbidity
#9	''complications'' OR ''fetal complications'' OR ''perinatal outcome'' OR ''neonatal outcome'' OR ''risk''
#10	''stillbirth mortality'' OR ''Perinatal mortality'' OR ''fetal death'' OR ''Stillbirth'' OR ''Intrauterine death''
#11	#1 AND #2 AND (#3 OR #4) AND (#5 OR #6) AND (#7 OR #8 OR #9 OR 10)
Limited to review (1123 records)	

TRIP Medical Database: searched on 14/03/2018

Search	Query
#1	Pregnancy
#2	induction OR cervical ripening
#3	expectant OR 42.weeks OR post-term
#4	cesarean OR adverse OR mortality OR morbidity OR neonatal OR perinatal
#5	#1 AND #2 AND #3 AND #4
No limitations (24 records)	

Current Controlled Trials: searched on 14/03/2018

Search	Query
#1	Pregnancy
#2	induction
#6	#1 AND #2
No limitations (26 records)	

Web of Science: searched on 14/03/2018

Search	Query
#1	Pregnancy
#2	"Low risk" OR Normal OR Term OR "low-risk"
#3	Post-term OR Prolonged OR Prolonged pregnancy OR Post-date OR Past due date OR Beyond term OR "41 weeks" OR "41 gestational weeks" OR "42 weeks" OR "42 gestational weeks"
#4	Induced labor OR Induced labour OR Induction OR induce OR Cervical Ripening OR Labor, induced OR Elective induction OR Routine induction OR Expectant management
#5	Adverse OR Adverse effects OR Adverse events OR Caesarean section OR cesarean section OR Cesarean OR caesarean OR caesarian OR Mortality OR Morbidity OR complications OR "fetal complications" OR "perinatal outcome" OR "neonatal outcome" OR risk OR "stillbirth mortality " OR "Perinatal mortality OR "fetal death" OR "Stillbirth" OR "Intrauterine death"
#6	#1 AND #2 AND #3 AND #4 AND # 5
Limited to 1998 and English (136 records)	

MedNar: searched on 15/03/2018

Search	Query
#1	Induction of labor
Limited 1998 (203 records)	

Google Scholar: searched on 15/03/2018

Search	Query
#1	Induction of labor AND (prolonged OR "41 gestational weeks" OR cesarean OR morbidity OR mortality)
Limited 1998 and "in the title" (392 records)	

OpenGrey: searched on 15/03/2018

Search	Query
#1	Pregnancy
#2	Induction
#3	#1 AND #2
Limited to English (7 records)	

ProQuest Nursing and Allied Health Source: searched on 20/03/2018

Search	Query
#1	Pregnancy
#2	"Low risk" OR Normal OR Term OR "low-risk"
#3	Post-term OR Prolonged OR Prolonged pregnancy OR Post-date OR Past due date OR Beyond term OR "41 weeks" OR "41 gestational weeks" OR "42 weeks" OR "42 gestational weeks"
#4	Induced labor OR Induced labour OR Induction OR induce OR Cervical Ripening OR Labor, induced OR Elective induction OR Routine induction OR Expectant management
#5	Adverse OR Adverse effects OR Adverse events OR Caesarean section OR cesarean section OR Cesarean OR caesarean OR caesarian OR Mortality OR Morbidity OR complications OR "fetal complications" OR "perinatal outcome" OR "neonatal outcome" OR risk OR "stillbirth mortality " OR "Perinatal mortality" OR "fetal death" OR "Stillbirth" OR "Intrauterine death"
#6	#1 AND #2 AND #3 AND #4 AND # 5
Limited 1998 and "abstracts" (9 records)	

National Institute for Health and Care Excellence, World Health Organization, Royal College of Obstetricians and Gynaecologists, American College of Obstetricians and Gynecologists, and Society of Obstetricians and Gynaecologists of Canada. Searched on 15/03/2018

Search	Query
#1	induction of labor
#2	Post-term
#3	#1 AND #2
Unlimited (30 records) (2 records) (15 records) (97 records) (15 records)	

Appendix II: General characteristics of included studies

Burgos 2012[43]	
Methods	Study design: Quasi-randomized Duration of study: 2008–2010
Participants	Setting: Cruces University Hospital (Spain) Inclusion criteria: Singleton, cephalic with normal fetus. Exclusion: hypertensive disorders; IUGR; diabetes Participants enrolled: 3563 women; 1746 women until July 2009 (referred as 42 week) and 1817 from July 09 (referred as 41.week)
Interventions	Labor was induced at week 42+0 until July 2009, then protocol was changed to induce labor 41+0–6 GW
Outcomes	Cesarean section; instrumental delivery; inductions; umbilical cord blood pH<7.10; NICU and perinatal mortality
Notes	The trial monitors the impact of changing a guideline. Pregnant women at a given time will thus be treated according to a corresponding guideline (quasi-randomization).The quasi-randomization reveals equal baseline for the groups except gestational age.

Cheng 2012[45]	
Methods	Study design: Retrospective cohort Duration of study: 2005
Participants	Setting: USA Inclusion criteria: Low risk, nulliparous, singleton, cephalic, live birth, Exclusion criteria: chronic hypertension, gestational or pregestational diabetes mellitus, placenta previa. The indution group excluded gestational hypertension, preeclampsia/eclampsia, and oligohydramnios/polyhydramnios, but the conditions were allowed to develop in the expected management arm. Participants: 70,097 low-risk nulliparous women.
Interventions	Induction: Induction in 41+0–6 GW Expectant: Spontaneous onset of labor in or after 41+0 GW or induction 42+0 GW. Only data to the end of 42+0–6 GW are included.
Outcomes	Cesarean section, instrumental vaginal delivery, birthweight >4000 g., labor dystocia, chorioamnionitis, 5 min. Apgar<7, meconium aspiration, ventilator use>6 hours, NICU.
Notes	The original study covers all births from 39 GW and onwards. Data can be biased towards a more healthy induced population as the expectant in 41 GW according to the design. Pregnancies are only low risk at enrollment 39.GW, but the expectant management group are then prone to developing complications as a consequence of the ongoing pregnancies. This means however, using data from 41 GW can possibly imply, that the induced are selected low-risk whereas the expectant are non-selected and not low risk. Authors have been contacted due to this question, but did not reply.

Daskalakis 2013[13]	
Methods	Study design: Quasi-randomized Duration of study: 2009–2011
Participants	Setting: Alexandra Hospital (Greece) Inclusion criteria: Singleton, cephalic with normal fetus, maternal age >17 years Exclusion criteria: Low lying placenta; former cesarean; known hypersensitivity to prostaglandin; medical problems, diabetes or preeclampsia Participants enrolled: 438 women; 211 in induction group (41 week) and 227 in expectant group (42 week)
Interventions	Induction: Induced 41+1 GW Expectant: Awaiting spontaneous onset and induction 42+1 GW. Fetal surveillance with non-stress test and ultra scan every 2.day
Outcomes	Pain relief; instrumental deliveries; shoulder dystocia (more than McRoberts); cesarean section; postpartum hemorrhage >1000 ml; manual removal of placenta; Apgar score, NICU.
Notes	Low-risk are not evident at inclusion criteria, but author confirm, that only healthy pregnancies continues to 41.gestational week. The quasi-randomization is due to two treatment regimen at the ward on practitioners' preference. It's not transparent how women choose practitioner, but groups are alike at baseline.

Gelisen 2004[34]	
Methods	Study design: Randomized trial Duration of study: 2009–2011
Participants	Setting: Tertiary hospital (Turkey) Inclusion criteria: Singleton, cephalic with live fetus; intact membranes; Bishop score <5; reactive non-stress test; amniotic fluid index >= 5 cm. Exclusion criteria: Contractions at enrolment; estimated fetal weight> 4500 gr; known hypersensitivity to prostaglandins; previous cesarean section; BMI>=30; parity>5; low lying placenta; diabetes. Participants randomized: 600 women; 300 in each group
Interventions	Induction: induction 41+1 GW Expectant: Awaiting spontaneous onset until induction until 42+0 GW with non-stress test twice weekly and one time a biophysical profile.
Outcomes	Tachysystole (>=6 contractions/10 min.); uterine hyperstimulation; cesarean section; length of hospital stay; meconium stained amniotic fluid; macrosomia; pH; NICU; cesarean for fetal distress.
Notes	

Heimstad 2007[33]	
Methods	Study design: Randomized trial Duration of study: 2002–2004
Participants	Setting: Trondheim University Hospital (Norway) Inclusion criteria: singleton cephalic presentation and no prelabor rupture; speaking fluent Norwegian; healthy fetus. Exclusion criteria: none described Participants randomized: 508 women; 254 in each group
Interventions	Induction: induction 41+2 GW Expectant: Awaiting spontaneous onset until induction until 42+6 GW with antenatal fetal monitoring every 3 days
Outcomes	Meconium, birth weight, Apgar scores, umbilical cord pH, resuscitation; NICU; third and fourth- degree perineal laceration; Maternal hemorrhage (> 500 ml); precipitate labor (total length of labor < 3 hours);
Notes	Low-risk status cannot be confirmed, but it is normal procedure to induce labor before 41+3 GW if medical conditions exist. Author did not respond when asked about the issue.

Jacquemyn 2012[44]	
Methods	Study design: Retrospective cohort Duration of study: 2006–2007
Participants	Setting: Flandern (Belgium) Inclusion criteria: low-risk and the least possible risk for caesarean section; cephalic, birthweight between 2500–4500g, singletons, normotensive and without diabetes. Groups were based on spontaneous or induced labor. The outcome measures were vaginal or (secondary) caesarean section. Exclusion criteria: Multiple pregnancies, diabetes or hypertension Participants included: 3.156 women, 1428 are induced 41+0–6 GW and 1728 go into spontaneous onset or are induced 42+0–6 GW
Interventions	Induced: Induced 41+0–6 GW Expectant: Awaiting spontaneous onset until induction 42+6 GW
Outcomes	Vaginal or cesarean section
Notes	Fetal monitoring not important, as there are no fetal outcomes measured

Liu 2012[46]	
Methods	Study design: Retrospective cohort Duration of study: 2003–2010
Participants	Setting: Canadian Institute of Health registry, Canada Inclusion criteria: Singleton, cephalic, live fetus Exclusion criteria: previous cesarean section, IUGR, no medical/obstetric diagnoses as e.g. multi parity >5, preeclampsia, diabetes, antepartum hemorrhage, hypertensive disease, poly- oligohydramnios, abruption of placenta, anemia, heart disease, Herpes, HIV, pulmonary disease, macrosomia (>4000gr). Participants included: 277.964,women; 126.863 are induced 41+0–6 GW and 151.101 go into spontaneous onset or are induced 42+0–6 GW
Interventions	Induction: induction 41+2 GW by oxytocin or prostaglandin.
	Expectant: Awaiting spontaneous onset until induction until 42+6 GW
Outcomes	Focus of outcome is on severe maternal morbidity: Postpartum hemorrhage with blood transfusion, postpartum sepsis, uterine rupture during labor, intensive care unit admission, venous tromboembolism, obstetric shock.
Notes	

IUGR, Intrauterine growth retardation/restriction; NICU, neonatal intensive care unit.

Appendix III: Details on comparison groups in included studies

Randomized or quasi-randomized studies											
Reference, year	Location	Study	Estimated date of delivey	Time for enrolment	Intervention	Control Expectancy until	Parity	Women included	Induction agents	PROM	Comments
Burgos *et al.* 2012	Spain	Quasi-randomized	Ultrasaound adjusted LMP	41+0	41+0–6	42+0	All	Singleton / cephalic	PGE2 or oxytocin	Included	Authors contacted and responded due to missing information on fetal surveillance and fetal deaths.
Daskalaskis *et al.* 2013	Greece	Quasi-randomized	LMP and confirmed with 1.trim UL	41+1	41+1	42+1	Nulli- and multipara, stratified	Singleton, cephalic, >17 yrs, reactive non stress	PGE2 or ARM	Not PROM at enrolment	Authors contacted and responded regarding the low-risk status of women. Low risk confirmed by authors.
Gelisen *et al.* 2005	Turkey	Randomized trial	1 trimester ultrasound	41+1	41+1	42+0	All	Singleton / cephalic /live fetus	PGE1, Foley or Oxytocin	Not PROM at enrolment	Authors contacted due to missing information on cause of intrauterine death, but did not respond
Heimstad *et al.* 2007	Norway	Randomized trial	UL scan	41+2	41+3	42+6	All	Cephalic, % PROM, reassuring test at inclusion	PGE1, PGE2, ARM, Foley or oxytocin	Not PROM at enrolment	

Cohort studies											
Reference, year	Location	Study	Estimated date of delivey	Time for enrolment	Intervention	Control expectancy until	Parity	women Included	Induction agents	PROM*	Comments
Cheng *et al.* 2012	USA	Retrospective cohort	consistency between LMP and obstetric / clinical estimation	Each gestational week	41+0–6	42+0–6	Nulliparous	Low risk, expectant are possibly not low-risk at enrolment	No information	No information	Authors contacted but did not respond. Due to the design it is not possible to confirm that the expectant management group was at low-risk in 41+0 (at "enrolment"). This could bias results in favor of induction.
Jaquemyn *et al.* 2012	Belgium	Retrospective cohort	Ultrasound adjusted LMP	Each gestational week	41+0–6	42+0–6	Multi-parous	Low risk, cephalic >2499 and <4501; previous normal birth	No information	No information	Authors contacted due to lack of knowledge about UL-scan. Author confirm that EDD is estimated from ultrasound.
Liu *et al.* 2013	Canada	Retrospective cohort	LMP and confirmed with UL	Each gestational week	41+ 0–6	42+0–6	All	Low risk	Prostaglandin and oxytocin	No information	Authors contacted due to missing information about EDD-calculations and confirmed LMP + ultrasound.

PROM, Patient Reported Outcome Measure; LMP, last menstrual period; US, ultrasound; EDD, estimated date of delivery.

Appendix IV: Outcome measures in included studies

Author	Gelisen (2004)	Heimstad (2007)	Burgos (2012)	Daskalakis (2013)	Cheng (2012)	Jacquemyn (2012)	Liu (2013)
Design							
Randomized	x	x					
Quasi-randomized			x	x			
Retrospective cohort					x	x	x
Participants							
Participants all	600	508	3563	438	75,218	3156	277,964
Participants 41	300	254	1817	211	28,470	1428	126,863
Participants 42	300	254	1746	227	46,748	1728	151,101
Outcome measures							
Admission to NICU	x	x	x		x		
Cesarean section	x	x	x	x	x	x	
Cesarean section for failure to progress	x			x			
Cesarean section for fetal distress	x			x			
Chorioamnionitis					x		
Epidural analgesia				x			
5- min. Apgar < 7	x	x	x		x		
Hospital stay	x						
Hyperstimulation	x						
Induction of labor	x		x				x
Labor dystocia		x			x		
Macrosomic >4500 gm		x	x				
Manual removal of the placenta				x			
Maternal intensive care							x
Meconium aspiration syndrome	x	x			x		
Meconium stained amniotic fluid	x	x	x				
Obstetric shock							x
Oligohydramnios		x					
Opreative vaginal delivery		x	x	x	x		
Perinatal death	x	x	x	x			
pH < 7.10		x	x				
Post partum haemorrhage		x		x			x
Post partum sepsis							x
Precipitate labor		x					
Resuscitation		x					
Severe perineal lacerations		x					
Shoulder dystocia	x			x			
Spontaneous onset of labor			x	x		x	x
Tachysystoli	x						
Thromboembolism							x
Uterine rupture							x
Vaginal delivery	x		x	x			
Ventilator use >6 h					x		

NICU, neonatal intensive care unit.

Appendix V: Excluded studies

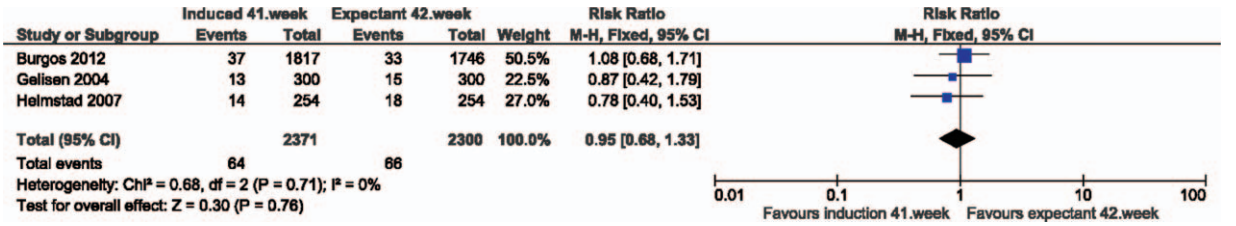
Author	Reason for exclusion
Allen 2012[55]	No data on the ongoing pregnancies in 41+ GW, but only data on induced 41 vs 42 GW.
Ananth 2013[8]	Does not compare 41 vs. 42 GW. Also, the low-risk criteria is not met.
Bailit 2015[56]	Does not compare 41 vs 42 GW.
Bleicher 2017[57]	Sample include high risk pregnancies.
Caughey 2005[58]	No explanation of EDD. Summarizes adverse events by GW without any sub-analysis over time. 1976-babies are summarized together with 2001-babies with no reflection on surveillance improvements over 25 years. Timespan back to 1976 is irrelevant according to inclusion.
Caughey 2006[30]	Includes high-risk pregnancies, no explanation of EDD, and expected management after 41 GW has no limits
Duff 2000[59]	Groups not comparable to our study as expectant management goes beyond 42+6 GW.
Fok 2005[60]	Includes possible risk-pregnancies. Authors were contacted and responded about the uncertainty of EDD: Dating scan was not available to all women, and only about 60% would had a scan before 20 GW. Excluded after scoring due to criteria of secure EDD.
Glantz 2010[21]	No explanation for the EDD, and induced are a high risk population. As the authors did not respond to questions regarding if the pregnancy was still high-risk in 41+42 gestational weeks, the study was excluded according to the protocol.
Grivell 2011[15]	Data not available at specific gestations, and the upper limit is 41+ GW and beyond.
Kiesewetter 2012[61]	Terminates at 41+3 GW and thus not relevant.
Klefstad 2014[62]	Initial population at enrolment includes high-risk pregnancies. 61% induced at 41+3 GW according to high-risk.
Kwee 2006[63]	Comparator 42+ GW is without any gestational limitations.
Nakling 2006[64]	Not low-risk. Comparator 42+ GW is without any gestational limitations.
Oros 2012[65]	The comparator is not 42+ and not relevant to the study.
Page 2012[66]	Fetal death at different gestational ages stratified by maternal age. Inductions not recorded.
Pavicic[67]	Women are not low-risk. No explanation of EDD
Raviraj 2013[68]	Does not differentiate between GW after 41+ GW. No explanation of EDD
Rosenstein 2012[69]	Investigates mortality according to GW, but does not differentiate between induced and spontaneous labor.
Sobande 2003[70]	It is not possible to see the number of pregnancies reaching 41 GW but going into spontaneous labor. No information on the ongoing pregnancies.
Stock 2012[71]	Expectant group includes anyone beyond 41 GW with no upper limits. No explanations for EDD.
Sue-A-Quan 1999[72]	Not relevant due to the groups: 40 GW serves as control vs. all pregnancies at or beyond 41 GW.
Treger 2002[73]	It is possible to segregate according to GW, but study does not differentiate between induced and spontaneous labor.
Weiss 2014[74]	Investigates intrauterine death according to GW, but study does not differentiate between induced and spontaneous labor.
Yazdani 2006[75]	Very sparse information level. Not information of baseline and possible confounders.

EDD, estimated date of delivery; GW, gestational weeks.

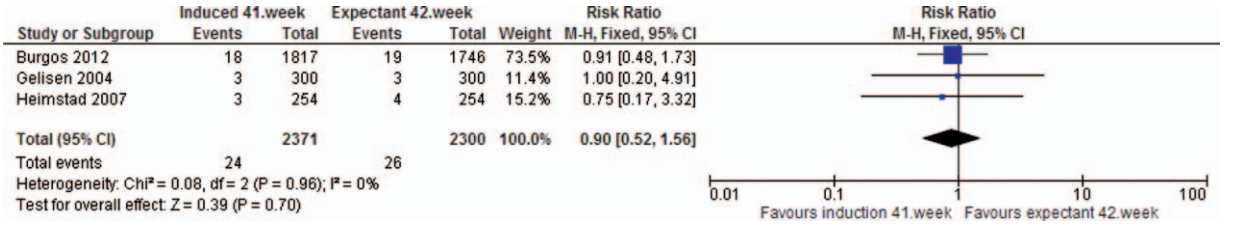
Appendix VI: Meta-analyses

Perinatal outcomes

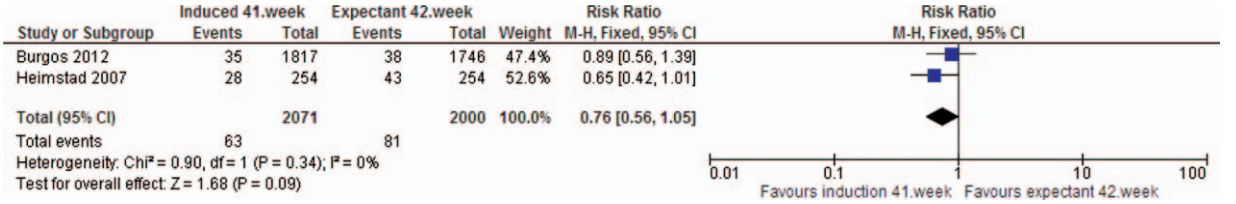
Analysis 1. Admission to neonatal intensive care unit. Quasi- or randomized studies



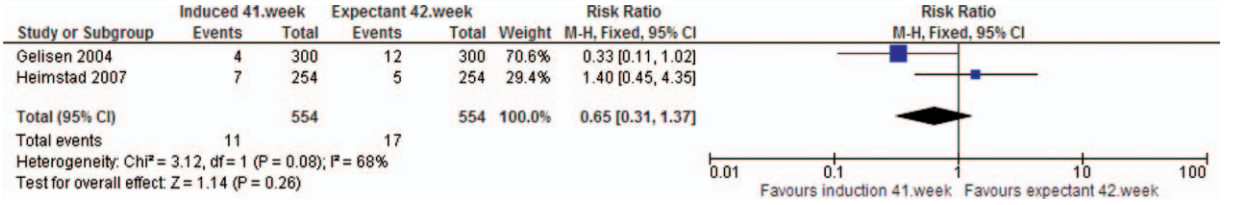
Analysis 2. Apgar score (less than 7 after 5 minutes). Quasi- or randomized studies



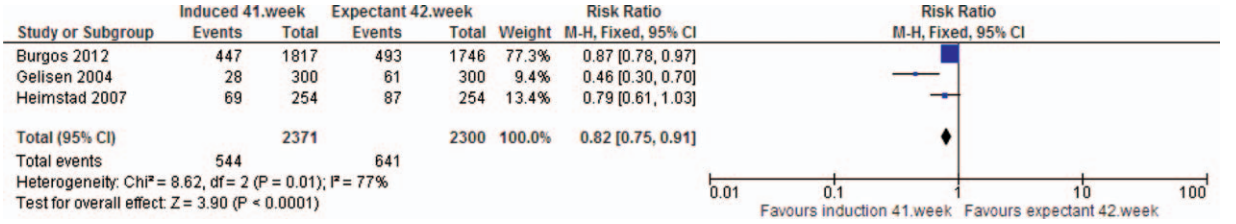
Analysis 3. Macrosomia >4500 g. Quasi- or randomized studies



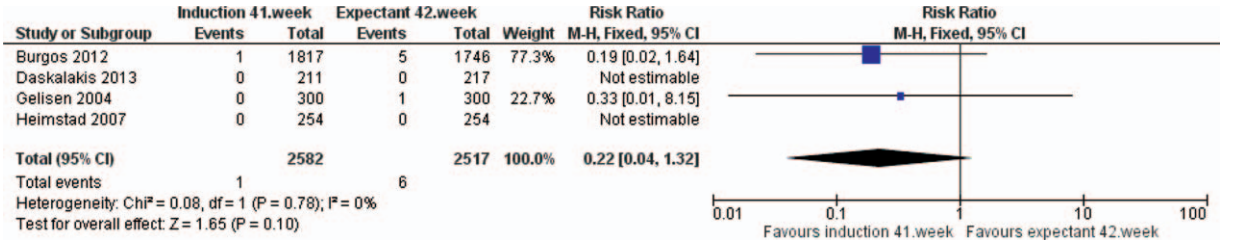
Analysis 4. Meconium aspiration syndrome. Quasi- or randomized studies



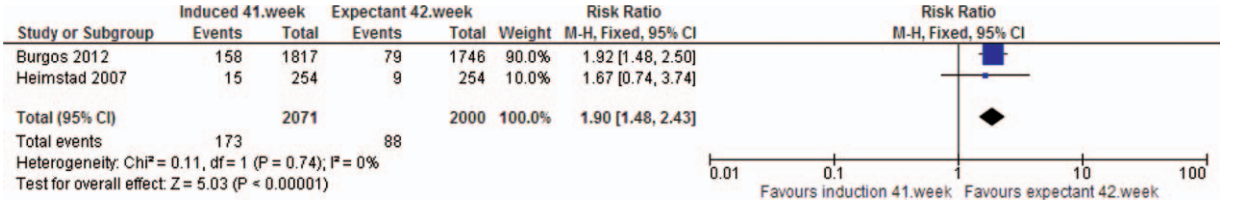
Analysis 5. Meconium stained amniotic fluid. Quasi- or randomized studies



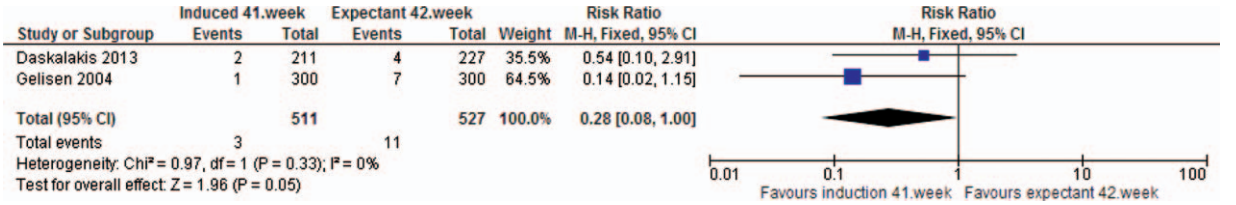
Analysis 6. Perinatal death. Quasi- or randomized studies



Analysis 7. pH <7.10. Quasi- or randomized studies

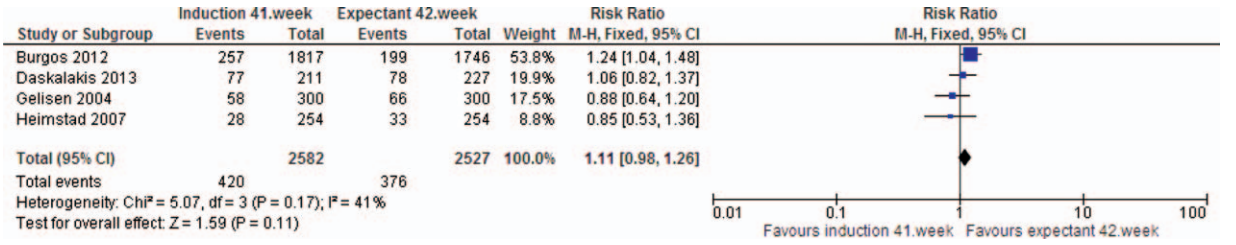


Analysis 8. Shoulder dystocia. Quasi- or randomized studies

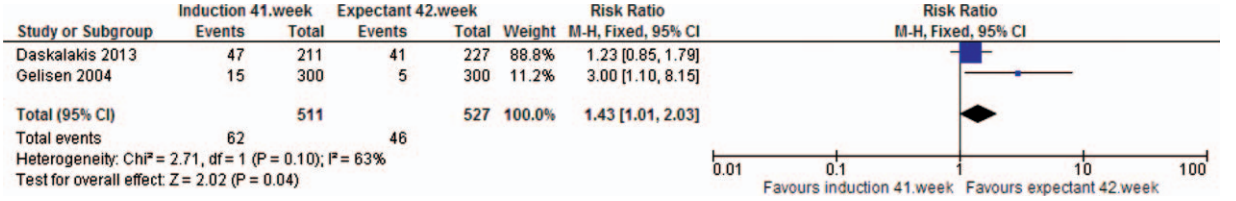


Maternal outcomes

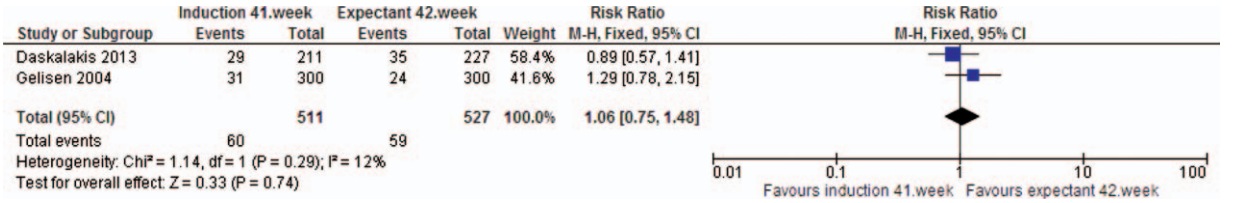
Analysis 9. Cesarean Section. Quasi- or randomized studies



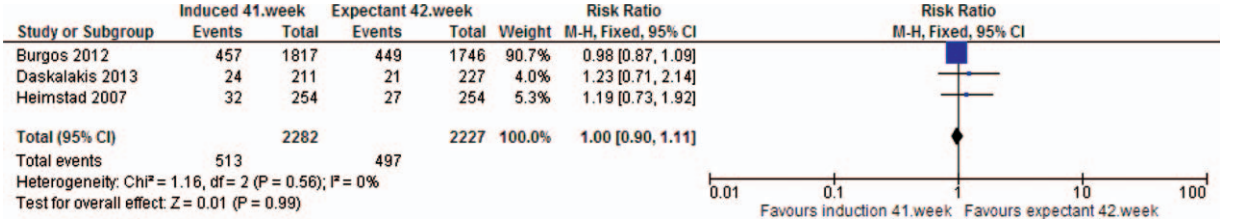
Analysis 10. Cesarean Section for failure to progress. Quasi- or randomized studies



Analysis 11. Cesarean Section for fetal distress. Quasi- or randomized studies



Analysis 12. Instrumental vaginal delivery. Quasi- or randomized studies



Analysis 13. Vaginal delivery. Quasi- or randomized studies