Comparative Effectiveness Review
Number 35

Effective Health Care Program

Terbutaline Pump for the Prevention of Preterm Birth

Executive Summary

Background

Preterm birth is defined as delivery before the completion of the 37th week of gestation, and it affects 13 percent of live births in the United States.[1] According to the 2010 National Vital Statistics report, there were 542,893 preterm births in the United States in 2006.[2] Rates of preterm birth result in a significant disease burden to the health care system. Although overall rates of neonatal mortality continue to decline, infants born too early are at risk for long-term morbidity.[3]

Tocolytics are drugs used to delay or inhibit contractions during the labor process. Several tocolytics are available to prevent preterm birth. These agents may be administered as primary therapy to control acute episodes of preterm labor or as maintenance therapy to prevent subsequent episodes. Maintenance tocolysis is usually provided for prolonged periods beyond 48 to 72 hours after arrest of acute preterm labor to inhibit the process of parturition until full term. While several studies have examined these agents for the control of acute episodes of preterm labor, the evidence to support their safety and efficacy as maintenance therapy is limited.

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at **www.effectivehealthcare.ahrq.gov/reports/final.cfm.**

The ß-agonist agent, terbutaline sulfate, has been used orally and subcutaneously as maintenance tocolytic therapy in women following acute treatment and arrest of

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care

confirmed preterm labor. As with all other contemporary tocolytics, the use of terbutaline for maintenance tocolysis is off-label. The Food and Drug Administration (FDA) has approved terbutaline for the management of acute and chronic obstructive pulmonary disease only. When administered through the subcutaneous (SQ) route, terbutaline may be administered by a pump that provides a steady continuous infusion with allowance for boluses. Compared with the oral route of administration, the SQ terbutaline pump uses lower doses (usual basal rate is 0.03–0.05 mg/hr with an intermittent bolus of 0.25 mg every 4 to 6 hours) and has less potential for tachyphylaxis.[4]

The effectiveness and safety of the SQ terbutaline pump for maintenance tocolytic therapy was examined in two systematic reviews. One review, which was based on two small randomized controlled trials (RCTs), concluded that the SQ terbutaline pump offers no advantages compared with the saline pump or oral terbutaline.[4] The second review found contradictory results among RCTs and observational studies; the RCTs found no difference between the SQ terbutaline pump and comparators, although the observational studies demonstrated positive effect estimates in favor of the pump.[5]

Despite previous systematic reviews, uncertainty surrounding the use of terbutaline and other tocolytics as maintenance therapy to prevent recurrent episodes of preterm labor still exists. No clear first-line maintenance tocolytic therapy has yet emerged. The possibility of maternal side effects and unclear evidence on perinatal outcomes contribute to the ambiguity of terbutaline's role in obstetrical practice. Moreover, in a recent cost analysis of four tocolytic agents, subcutaneous terbutaline had the highest cost.[6] The expense is due not only to the device, but also to the need for increased monitoring and management of adverse events associated with this therapy.[6]

Given the importance and associated uncertainty about the appropriateness of ongoing use of the terbutaline pump for maintenance tocolysis for clinicians, patients, and policymakers, a review about the effectiveness and safety of SQ terbutaline pump was commissioned by the Agency for Healthcare Research and Quality (AHRQ) to address six Key Questions. This evidence report will add to previous systematic reviews by performing an up-to-date search of the literature, synthesizing evidence in the context of specific populations of women, addressing confounding by level of maternal activity and level of care, and grading the strength of evidence for important outcomes to help decisionmakers develop evidence-based recommendations and policies.

Objectives

The objectives of this review were to examine the efficacy, effectiveness, and safety of the SQ terbutaline pump as prolonged maintenance tocolysis for inhibiting progression of parturition in women with arrested acute preterm labor. The SQ terbutaline pump was compared with placebo, conservative treatment, or any other active intervention in the following specific populations: women delivering at various gestational ages, classified as extremely preterm (<28 weeks of gestation), very preterm (28 weeks to 31 weeks of gestation), preterm (32 weeks to 33 weeks of gestation), and later preterm (34 weeks to 36 weeks of gestation); women with multiple gestation; women of different racial or ethnic backgrounds; women with previous preterm birth; women with history of preeclampsia; and women with recurrent preterm labor (RPTL) during the same pregnancy. Clinical endpoints, which included neonatal health outcomes and maternal/neonatal harms, were assessed in addition to several surrogate outcomes, such as birth weight and prolongation of pregnancy. The potential confounding effects of maternal activity and maternal care on the above endpoints were explored. Lastly, the pump device was evaluated by examining the incidence of pump-related outcomes, such as missed doses, dislodgment, and overdose.

These objectives were framed in the following Key Questions:

In women with arrested preterm labor, does treatment with an SQ infusion of terbutaline delivered by a pump, in comparison with placebo, conservative treatment, or other interventions:

Key Question 1: improve neonatal health outcomes, including bronchopulmonary dysplasia, neonatal death, death within initial hospitalization, significant intraventricular hemorrhage (grade III/IV), necrotizing enterocolitis, periventricular

leukomalacia, retinopathy of prematurity, seizures, sepsis, and stillbirth for the following subgroups:

a. women <28 weeks of gestation (extremely preterm)?
b. women between 28 weeks and 31 weeks of gestation (very preterm)?
c. women between 32 weeks and 33 weeks of gestation (preterm)?
d. women between 34 weeks and 36 weeks of gestation (later preterm)?
e. multiple gestation?
f. racial or ethnic subgroups?
g. women with previous preterm birth?
h. women with history of preeclampsia?
i. women with RPTL and women without RPTL?

Key Question 2: improve other surrogate outcomes, including gestational age at delivery, incidence of delivery at various gestational ages (<28 weeks, < 32 weeks, <34 weeks, <37 weeks), mean prolongation of pregnancy (days), birth weight, ratio of birth weight/gestational age at delivery, pregnancy prolongation index, need for assisted ventilation, need for oxygen per nasal cannula, and neonatal intensive care unit (NICU) admission for the following subgroups:

a. women <28 weeks of gestation (extremely preterm)?
b. women between 28 weeks and 31 weeks of gestation (very preterm)?
c. women between 32 weeks and 33 weeks of gestation (preterm)?
d. women between 34 weeks and 36 weeks of gestation (later preterm)?
e. multiple gestation?
f. racial or ethnic subgroups?
g. women with previous preterm birth?
h. women with history of preeclampsia?
i. women with RPTL and women without RPTL?

Key Question 3: increase the maternal harms of arrhythmia, heart failure, hyperglycemia, hypokalemia, maternal mortality, myocardial infarction, pulmonary edema, or refractory hypotension, or result in an increased rate of maternal discontinuation of therapy or maternal withdrawal due to adverse effects (Withdrawal-AE)?

Key Question 4: increase the neonatal terbutaline-related harms of hypoglycemia, hypocalcemia, and ileus?

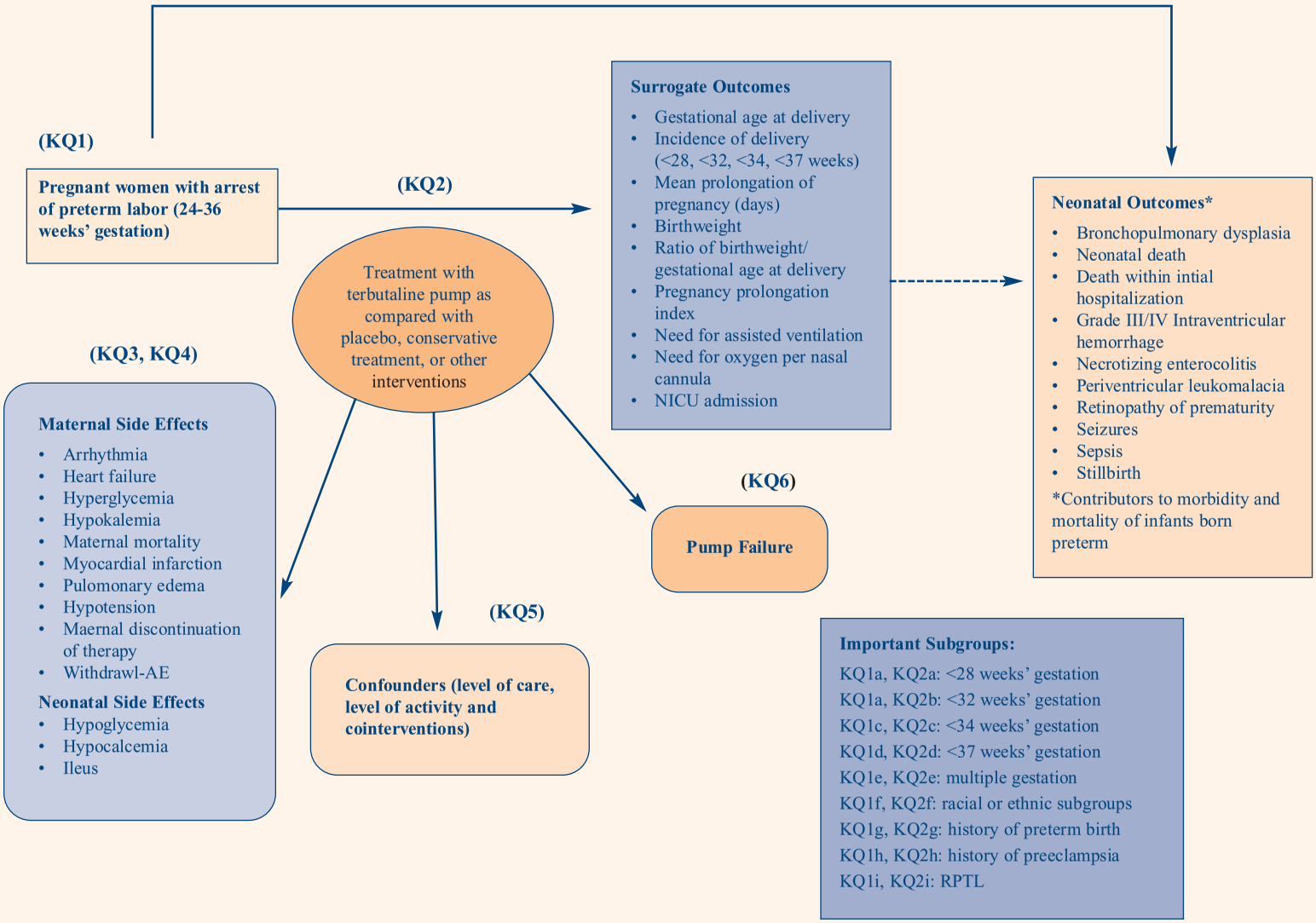
Key Question 5: Can the differences in the outcomes above be partially explained by the differences in level of care (e.g., frequency of followup, nurse visits, concomitant treatment, etc.) and level of activity (e.g., other children in the home, marital/support status, working status, bedrest, etc.) between the terbutaline pump group and the comparator group?

Key Question 6: What is the incidence of failure of the pump device used for terbutaline infusion, including missed doses, dislodgment, and overdose?

Analytic Framework

We developed an analytic framework depicting links between the intervention and related clinical and intermediate efficacy and harms outcomes and other unintended adverse effects (Figure A). In the framework below, the key questions of interest can be seen to encompass a holistic inquiry of the topic.

Figure A. Analytical framework of terbutaline pump for maintenance tocolysis



Methods

Input From Stakeholders

We formulated the population, intervention, comparator, outcome, timing, setting (PICOTS) conceptual framework and Key Questions in consultation with key informants during a topic refinement stage. The public was invited to provide comments on the Key Questions. During the review process, we followed a research protocol we developed with the clinical and methodological input of a technical expert panel. The protocol followed the Effective Health Care Program's Methods Guide for Effectiveness and Comparative Effectiveness Reviews.[7]

Data Sources and Searches

We developed a peer-reviewed search strategy and searched the following databases: MEDLINE In-Process & Other Non-Indexed Citations and MEDLINE (1950 to April 1, 2011); Embase (1980 to April 1, 2011); Cumulative Index to Nursing and Allied Health Literature (CINAHL) via EBSCOhost (1985 to December 7, 2009), the Cochrane Library via the Wiley interface (April 1, 2011) (including CENTRAL, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects – DARE, Health Technology Assessment – HTA, and the National Health Service Economic Evaluation Database – NHS EED), and the Centre for Reviews and Dissemination (CRD) databases (January 2, 2010). Appendix A provides details of the search strategies. We hand-searched the bibliographies and text of review articles, letters to editors, and commentaries and the reference lists of included studies for additional references. We also reviewed grey literature sources and information received from pharmaceutical companies (see Appendixes B and C), and sought unpublished information from Matria (now called Alere) Healthcare about their perinatal program and associated database.

In February 2011, the FDA issued new warnings against the use of terbutaline to treat preterm labor, so we also accessed a summary of the FDA postmarketing surveillance results. This decision was made post hoc.

Study Selection

Two reviewers screened abstracts and full-text reports with conflicts resolved by consensus or third-party adjudication. Studies were included if they met the following criteria: evaluated pregnant women between 24 and 36 weeks' gestation having had acute preterm labor arrested with primary tocolytic therapy; contained at least one group that was administered the SQ terbutaline pump; and assessed one of the specified outcomes listed in the key questions or described a long-term childhood outcome. Noncomparative studies (i.e., case series) were assessed only for pump-related harms outcomes, such as incidence of pump failure, missed doses, or overdose. Non-English records without an English abstract were excluded. We also excluded case reports, but in a post hoc decision sought FDA summaries of postmarketing data highlighting serious harms.

Data Extraction and Risk of Bias Assessment

One reviewer extracted data into a standardized electronic form and assessed study risk of bias and applicability. Extraction items included general study characteristics (e.g., year of publication, study design), population characteristics (e.g., inclusion/exclusion criteria, age, race, level of activity), intervention characteristics (e.g., dose, duration, details about comparators, level of care), and outcomes with their estimates. A second reviewer verified outcomes data and study risk of bias assessments. Ratings for level of activity, level of care, and assessments of applicability were verified by a clinical expert. Level of activity and level of care were rated based on composite assessments across preidentified variables.

We assessed study risk of bias given the study design, by outcome, using generic items to assess confounding and various types of bias (e.g., selection, performance, detection bias, attrition bias). Selected items from the McMaster Quality Assessment Scale of Harms were also incorporated into the risk of bias assessment for harm-related outcomes.[8] Certain criteria were specific to particular study designs (e.g., allocation generation and concealment applied only to RCTs). We rated each relevant outcome in a study with an overall risk of bias rating designated as high, medium, or low. Outcomes were rated as high risk of bias if there was an apparent

and major flaw in the study that would invalidate results.

Appendix D of the full report provides the data extraction, risk of bias, and applicability forms.

Data Synthesis and Analysis

We meta-analyzed the RCTs with a random effects model, following a DerSimonian and Laird approach, when they were clinically and methodologically similar. To assess statistical heterogeneity and the magnitude of heterogeneity, we used Cochran's Q (a=0.10) and the I^2 statistic respectively. Odds ratios (ORs) were calculated for dichotomous outcomes and mean differences for continuous outcomes. All analyses were performed using Comprehensive Meta Analysis version 2.2.046 or version 2.2.055 (New Jersey, USA). We did not meta-analyze observational studies because of potential differences in confounders, nor did we combine studies of singleton and multiple pregnancies. Synthesis of evidence from observational studies was, therefore, undertaken qualitatively. Due to the small number of studies, we could not perform any meta-regression to explore statistical heterogeneity in effect estimates.

Strength of Evidence and Applicability

Based on published guidance for the Effective Health Care Program,[9] two reviewers graded the strength of evidence using the four primary domains (i.e., risk of bias, consistency, directness, and precision) for the following outcomes: incidence of delivery at various gestational ages (<28 weeks, <32 weeks, <34 weeks, <37 weeks), mean prolongation of pregnancy, bronchopulmonary dysplasia, significant intraventricular hemorrhage (grade III/IV), neonatal death, death within initial hospitalization, and maternal withdrawal due to adverse effects (Withdrawal-AE). We described population, intervention, comparison, outcome, timing, and setting characteristics to summarize the applicability of the body of evidence.

Results

Study Selection

We screened 427 citations and included 14 unique records in the review. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram below depicts the flow of records from identification to inclusion (Figure B). Most records were excluded at full-text screening (n=197) based on the reasons listed in the diagram. Appendix E provides a list of excluded studies, and Appendix F provides individual-level study data.

Study Characteristics

Table A presents general summary characteristics of the included studies. Most studies were observational and included cohorts and case series. Two studies were RCTs, and one was a nonrandomized trial. Sample sizes ranged from 9 to 1,366, but greater than 70 percent of studies included at least 200 participants (average 291 ± 395). All studies were from the United States, and participants were recruited either from single-center study sites or from a national proprietary database run by Matria Healthcare. The Matria database provides an outpatient perinatal program consisting of 24-hour nursing and pharmacy support, home uterine activity monitoring, individualized education, and provision of tocolytic therapy to women with preterm labor. Because five studies originated in the Matria database, and not all reported geographic region and/or years over which participants were recruited, the question of overlap in participants across these studies was an important concern of reviewers. Through the Scientific Resource Center (SRC), we requested this missing information from Matria (now called Alere) Healthcare but did not receive a response. Therefore, where appropriate, we report this risk of double-counting of participants.

Several studies included women with RPTL and singleton gestation. Comparator groups included placebo, no treatment, oral terbutaline, oral nifedipine, and mixed oral tocolytics. The definition of labor was unclear in 36 percent of the included studies. The remaining studies included women with persistent contractions and cervical change.

Figure B. PRISMA diagram

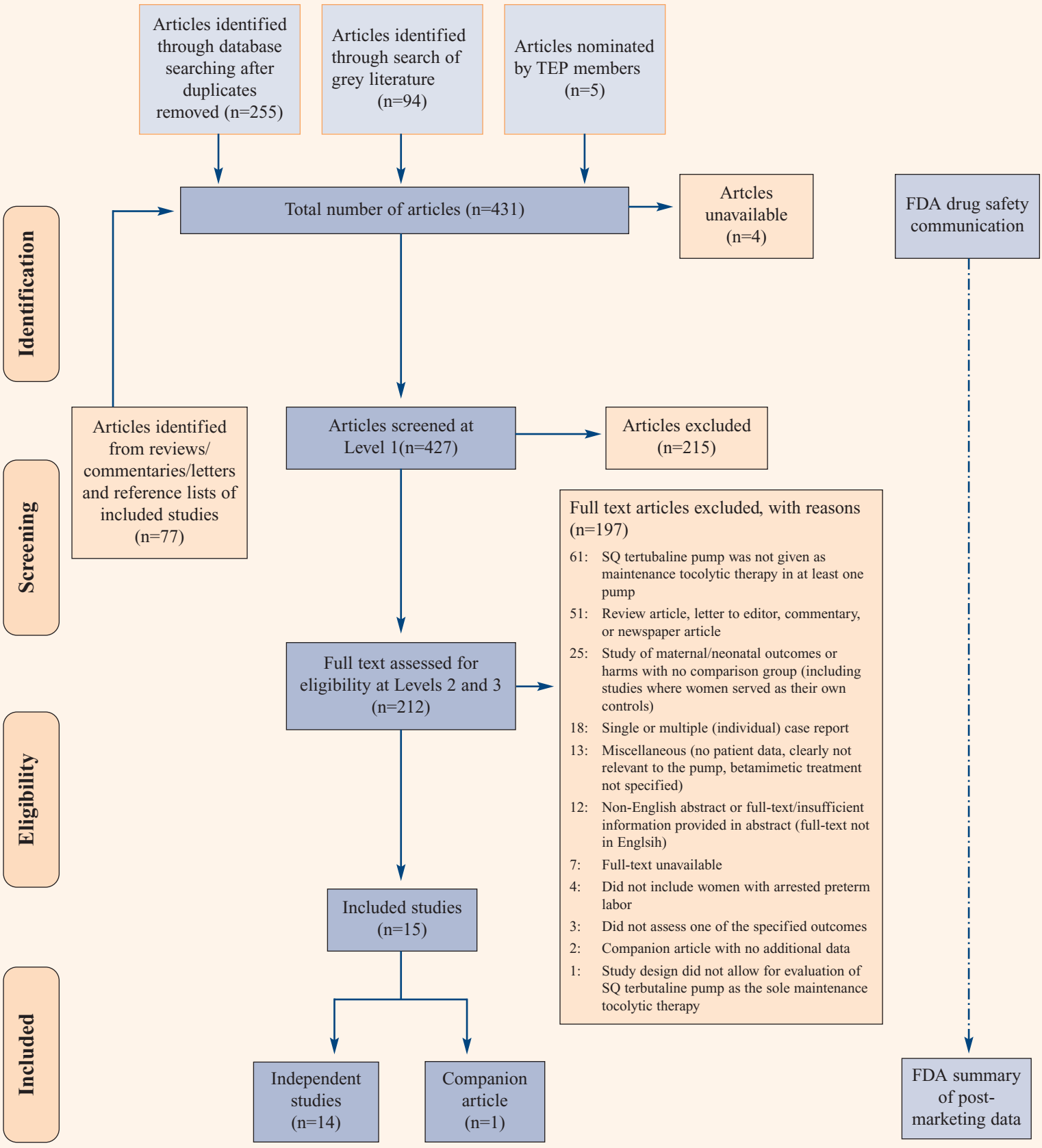


Table A. Summary characteristics of the included studies			
Characteristic		**Number of Studies**	**References**
Study Design	RCT	2	10,11
	Nonrandomized trial	1	12
	Prospective cohort	2	13,14
	Retrospective cohort	7	15-21
	Case series	2	22,23
Participant Recruitment	Single center sites	9	10-14,20-23
	Matria database	5	15-19
Funding	Industry	2	10,22
	Nonindustry	3	14,20,21
	Not reported	9	11-13,15-19,23
Comparator*	Oral nifedipine	3	15-17
	Oral terbutaline	4	11,12,20,21
	Oral tocolytics	3	14,18,1
	Placebo (saline pump)	2	10,11
	No treatment	1	13
	No comparison group	2	22,23
Primary Tocolytic Treatments	IV magnesium sulfate only	1	23
	IV magnesium sulfate and/or other agents	5	10-13,22
	Not reported	8	14-21
Gestation	Singletons only	6	10,12,13,15,17,18
	Twins only	2	16,19
	Singletons and twins	2	11,22
	Not reported	4	14,20,21,23
Definition of Labor	Not reported	5	15,17-19,21
Risk of Bias**	Low	1	10
	Medium	7	12,16,17,19,20,22,23
	High	7	11-15,18,21
By Key Question	Key Question 1	6	10,11,13,17-19
	Key Question 2	12	10-21
	Key Question 3	6	10,12,13,18,19,21
	Key Question 4	1	11
	Key Question 6	3	11,22,23

RCT=randomized controlled trial; IV=intravenous
* One study contained two comparison groups. [11]
**Risk of bias of one study differed by outcome.[12]

Risk of Bias Assessment

We rated studies as low, medium, or high risk of bias for the relevant reported outcomes. Although the randomization procedures in the two RCTs were appropriate, we rated one RCT as low risk of bias10 and the second RCT as high risk of bias because more than 90 percent of eligible participants declined to participate, the study was underpowered, and blinding was ineffective.[11]

The single nonrandomized trial was high risk of bias for the outcomes of birth weight and gestational age at delivery due to potential prognostic imbalances in groups. However, we did not anticipate that such imbalances would impact the outcome of maternal hyperglycemia, which we rated as medium risk of bias, due to insufficient information to assess several other criteria.[12]

We rated most of the cohort studies as high risk of bias because there were important group imbalances in baseline characteristics or prognostic factors.[13-15,18,21] The other cohort studies we rated as medium risk of bias; although these studies had no identifiable flaws, several criteria could not be assessed due to incomplete reporting.[16,17,19,20]

Lastly, we rated the two case series as medium risk of bias because neither study provided clear definitions for the pump-related harm outcomes, and several criteria, such as compliance, adequacy of sample size, and selective outcome reporting, were unclear.[22,23]

Neonatal Health Outcomes (Key Question 1)

Strength of evidence is insufficient for bronchopulmonary dysplasia, death within initial hospitalization, and significant intraventricular hemorrhage (grade III/IV). Based on one retrospective cohort of medium risk of bias, the strength of evidence favoring the SQ terbutaline pump compared with oral tocolytics for neonatal death in women with twin gestation and RPTL is low (Table B). This study investigated women from the Matria database and reported a statistically significant difference in neonatal death in favor of SQ terbutaline pump (OR = 0.09, 95% CI: 0.01, 0.70).19 Sparse evidence from underpowered studies addressed necrotizing enterocolitis, retinopathy of prematurity, and sepsis with inconclusive results.[11,13] No data were available for periventricular leukomalacia and seizures.

Three retrospective cohort studies from the Matria database reported stillbirths in women with RPTL and single or twin gestation.[17-19] All three studies found nonsignificant differences between the SQ terbutaline pump and oral tocolytics. However, these studies were likely underpowered to detect a difference in still birth, given the small number of events (<1%).

Table B. SQ terbutaline pump versus comparator: Strength of evidence for populations of interest

Outcome	Number of Studies	Number of Participants	Total Number of Events	Population of Interest	Comparator	Effect Estimate	Strength of Evidence
Neonatal Health Outcomes (KQ1): BPD	0	0	N/A	N/A	N/A	N/A	Insufficient
Neonatal Health Outcomes (KQ1): Neonatal death*	1[19]	706	12	Twin gestation + RPTL	Oral tocolytics	OR = 0.09 (0.01, 0.70)	Low
	1[17]	284	0	Singleton gestation + RPTL	Oral nifedipine	OR = 1.00 (0.02, 50.75)	Insufficient
Neonatal Health Outcomes (KQ1): Death within initial hospitalization	0	0	N/A	N/A	N/A	N/A	Insufficient
Neonatal Health Outcomes (KQ1): Significant IVH (Grade III/IV)†	113	60	4	Singleton gestation + RPTL	No treatment	OR = 0.30 (0.02, 5.85)	Insufficient
Other Surrogate Outcomes (KQ2): Incidence of delivery < 28 weeks	0	0	N/A	N/A	N/A	N/A	Insufficient
Other Surrogate Outcomes (KQ2): Incidence of delivery < 32 weeks	1[16]	656	192	Twin gestation + RPTL	Oral nifedipine	OR = 0.47 (0.33, 0.68)	Low
	1[19]	706	124	Twin gestation + RPTL	Oral tocolytics	OR = 0.52 (0.35, 0.76)	Low
	2[15,17]	1650	106	Singleton gestation + RPTL	Oral nifedipine	OR = 0.20-0.29 (lower CI range 0.07-0.16, upper CI range 0.52-0.61)††	Low
	1[18]	558	37	Singleton gestation + RPTL	Oral tocolytics	OR = 0.21 (0.09, 0.50)	Low
	1[13]	60	21	Singleton gestation + RPTL	No treatment	OR = 0.04 (0.00, 0.65)	Low
Other Surrogate Outcomes (KQ2): Incidence of delivery < 34 weeks‡	0	0	N/A	N/A	N/A	N/A	Insufficient
Other Surrogate Outcomes (KQ2): Incidence of delivery < 37 weeks§	2[15,17]	1650	925	Singleton gestation + RPTL	Oral nifedipine	OR = 0.72-0.75 (lower CI range 0.47-0.58, upper CI	Insufficient

Table B. SQ terbutaline pump versus comparator: Strength of evidence for populations of interest (continued)

Outcome	Number of Studies	Number of Participants	Total Number of Events	Population of Interest	Comparator	Effect Estimate	Strength of Evidence
Other Surrogate Outcomes (KQ2): Incidence of delivery < 37 weeks§ (continued)						range 0.90-1.20) ††	
	1[18]	558	318	Singleton gestation L + RPT	Oral tocolytics	OR = 0.70 (0.50, 0.98)	Low
	1[13]	60	50	Singleton gestation	No treatment	OR = 0.04 (0.01, 0.23)	Low
	1[20]	64	38	Singleton/Multiple gestation + RPTL	Oral terbutaline	OR = 0.10 (0.03, 0.32)	Low
Other Surrogate Outcomes (KQ2): Mean prolongation of pregnancy (days)**	1[16]	656	N/A	Twin gestation + RPTL	Oral nifedipine	MD = 7.20 (4.10, 10.30)	Low
	2[15,17]	1650	N/A	Singleton gestation + RPTL	Oral nifedipine	MD = 6.20-7.50 (lower CI range 0.79-4.94, upper CI range 10.06-11.61)††	Insufficient
	1[18]	558	N/A	Singleton gestation + RPTL	Oral tocolytics	MD = 5.50 (2.28, 8.72)	Low
	1[13]	60	N/A	Singleton gestation + RPTL	No treatment	MD = 25.30 (16.77, 33.83)	Low
Maternal Harms (KQ3): Withdrawal-AE	0	0	N/A	N/A	N/A	N/A	Insufficient

BPD = bronchopulmonary dysplasia; IVH = intraventricular hemorrhage; MD = mean difference; N/A = not applicable; OR = odds ratio; RPTL = recurrent preterm labor; withdrawal-AE = withdrawal due to adverse effects

* One RCT also reported neonatal death.[11] No events occurred in the SQ terbutaline pump group or in the two comparator groups. We did not grade this evidence here because it did not pertain to any of the subgroups of interest.

† One RCT reported significant intraventricular hemorrhage.10 No events were observed in pump or comparator groups. We did not grade this evidence here because it did not pertain to any of the subgroups of interest.

‡ Incidence of delivery < 34 weeks was reported in one RCT, which showed a nonsignificant difference between SQ terbutaline pump and placebo (OR = 0.95, 95% CI: 0.32, 2.87). We did not grade this evidence here because it did not pertain to any of the subgroups if interest.

§ Incidence of delivery < 37 weeks was also reported in one RCT, which showed a nonsignificant difference between SQ terbutaline pump and placebo (OR = 1.57, 95% CI: 0.49, 5.02). We did not grade this evidence here because it did not pertain to any of the subgroups of interest.

** Mean prolongation of pregnancy was also reported in two RCTs, with nonsignificant effect estimates. We did not grade this evidence here because it did not apply to any of the subgroups of interest.

†† Studies were not pooled. Also, there was risk of double-counting of participants across these studies.

Other Surrogate Outcomes (Key Question 2)

Studies reported surrogate outcomes of preterm labor much more frequently than neonatal or maternal clinical endpoints. However, none of the included studies examined incidence of delivery < 28 weeks (strength of evidence is insufficient, Table B), need for oxygen per nasal cannula, or ratio of birth weight/gestational age at delivery.

Incidence of delivery at various gestational ages. Incidence of delivery < 32 weeks: The strength of evidence favoring SQ terbutaline pump compared with either oral tocolytics or no treatment is low for women with RPTL and those additionally with twin gestation (OR range = 0.04–0.52, 95% CI range: 0.00–0.35, 0.50–0.76) (Table B). The evidence originated in six, mostly Matria-based, cohort studies of medium to high risk of bias.[13,15-19]

Incidence of delivery < 34 weeks: The strength of evidence for this outcome is insufficient (Table B). One small RCT (n=52) that did not address any of the populations of interest, showed a nonsignificant difference between SQ terbutaline pump and placebo in women with singleton gestation.[10]

Incidence of delivery < 37 weeks: The strength of evidence favoring SQ terbutaline pump compared with oral tocolytics or no treatment is insufficient or low for women with RPTL (Table B). Four of five cohort studies of medium to high risk of bias, mostly from the Matria database, reported statistically significant differences in favor of SQ terbutaline pump (OR range = 0.04–0.75, 95% CI range: 0.01–0.58, 0.23–1.20).[13,15,17,18,20]

Mean gestational age at delivery. Larger cohort studies of medium to high risk of bias in women with RPTL and single or twin gestation demonstrated consistent benefit of SQ terbutaline pump compared with oral tocolytics or no treatment (RPTL and singleton gestation: difference in means range = 0.70–3.40 weeks, 95% CI range: 0.28–1.80 weeks, 0.98–5.00 weeks; RPTL and twin gestation: difference in means = 0.70 weeks, 95% CI range: 0.43–0.48 weeks, 0.92–0.97 weeks).[13,15-19] Most participants in the cohort studies came from the Matria database. RCT evidence not directly addressing the populations of interest yielded a nonsignificant effect estimate between the pump and placebo (n=52 and n=42).[10,11]

Prolongation of pregnancy. The strength of evidence favoring SQ terbutaline pump compared with oral tocolytics or no treatment is insufficient or low for women with twin gestation and/or RPTL (difference in means range 5.50–25.30, 95% CI range: 0.79–16.77, 8.72–33.83) (Table B).[13,15-18] This evidence came from five cohort studies of medium to high risk of bias, mostly from the Matria database. Two small RCTs (n=52 and n=42), which did not pertain to any of the populations of interest, showed nonsignificant differences between SQ terbutaline pump and placebo.[10,11]

In one Matria-based cohort study, more women in the SQ terbutaline pump group had pregnancy prolonged > 7 days compared with women who received oral nifedipine (OR = 7.84, 95% CI: 3.59, 17.12).[15] Other Matria-based studies reported statistically significant benefits in favor of the pump compared with oral tocolytics for prolongation > 14 days (OR range = 1.93–3.47, 95% CI range: 0.87–2.34, 2.65–5.15).[15-19]

Birth weight. Cohort studies of women with RPTL and single or twin gestation demonstrated statistically significant differences in mean birth weight in favor of SQ terbutaline pump compared with oral tocolytics or no treatment (range of mean difference in grams = 136–721, 95% CI range: 83–355, 189–1087).[13,16-19] Aside from one study, all were from the Matria database.[16-19] Two small RCTs (n=52 and n=42), which did not pertain to any of the populations of interest, reported nonsignificant differences between SQ terbutaline pump and placebo.[10,11]

Incidence of low birth weight (< 2500 g) and very low birth weight (< 1500 g) were reported in cohort studies. Most of these studies originated from the Matria database. All studies that reported low birth weight found statistically significant differences in favor of SQ terbutaline pump compared with no treatment or oral tocolytics (OR range = 0.24–0.64, 95% CI range: 0.06–0.51, 0.62–0.96).[13,15-19] Most studies also found statistically significant differences in favor of the pump for incidence of very low birth weight (OR range = 0.22-0.46, 95% CI range: 0.07–0.29, 0.60–1.06).[16-19]

Pregnancy prolongation index. Pregnancy prolongation index was reported in two cohort studies.[13,20] Both found statistically significant differences in favor of the SQ terbutaline pump

compared with either no treatment or oral terbutaline (mean difference = 0.41, 95% CI: 0.26, 0.56; and 0.14, 95% CI: 0.02–0.26).

Need for assisted ventilation. One cohort study from the Matria database reported a nonsignificant difference between the SQ terbutaline pump and oral tocolytics in requirement for ventilator among infants with NICU admission.[18]

NICU admission. Incidence of NICU Admission: Statistically significant differences in favor of the SQ terbutaline pump compared with oral tocolytics or no treatment were reported in cohort studies of women with RPTL and single or twin gestation (OR range 0.28–0.72, 95% CI range: 0.08–0.58, 0.63–0.97).[13,15-19] Again, most of these studies were Matria-based.[15-19] One small RCT (n=52), which did not pertain to any of the populations of interest, reported a nonsignificant difference between the SQ terbutaline pump and placebo.[10]

NICU length of stay: Statistically significant differences in favor of the SQ terbutaline pump compared with oral tocolytics or no treatment were also reported for NICU length of stay in mostly Matria-based cohort studies of women with RPTL and single or twin gestation (range of mean difference in days: -3.50 to -17.90, 95% CI range: -5.26 to -32.88, -1.74 to -3.54).[13,15,18,19] Another small RCT (n=42), which did not address any of the subgroups of interest, reported a nonsignificant difference between the SQ terbutaline pump and placebo or oral terbutaline.[11]

Maternal Harms (Key Question 3)

The strength of evidence is insufficient for Withdrawal-AE (Table B). One prospective cohort in women with singleton gestation and RPTL demonstrated highly unreliable odds favoring no treatment compared with the pump for tachycardia/nervousness (OR=25.48, 95% CI:1.23, 526.6).[13] Underpowered studies demonstrated indeterminate results for the outcomes of mortality, pulmonary edema, and therapy discontinuation (i.e., type II error cannot be excluded).[10,18,19] Two studies, a retrospective cohort and a nonrandomized trial, demonstrated nonsignificant differences between the SQ terbutaline pump and oral terbutaline in the incidence of gestational diabetes, though type II error cannot be excluded. No data were available on heart failure, myocardial infarction, refractory hypotension, and hypokalemia.

Until 2009, 16 maternal deaths and 12 cases of maternal cardiovascular events (hypertension, myocardial infarction tachycardia, arrhythmias, and pulmonary edema) in association with terbutaline tocolysis were reported to the FDA. Of these, at least three maternal deaths and three cardiovascular adverse events were clearly reported to be in association with the use of the SQ terbutaline pump.[24]

Neonatal Harms (Key Question 4)

Neonatal harms data were very sparse. Neonatal hypoglycemia was reported in only one RCT that compared the SQ terbutaline pump with placebo and oral terbutaline.[11] Differences between the SQ terbutaline pump and placebo or oral terbutaline were nonsignificant. However, given the small number of events and limited sample size (n=42), the RCT was underpowered and the results are inconclusive. No studies reported neonatal hypocalcemia or ileus.

Assessment of Confounding by Level of Activity and Level of Care (Key Question 5)

Only a small number of studies could be rated for level of activity and level of care. Therefore, we could not carry out meta-regressions to explore the effect of these variables on maternal and neonatal outcomes. Furthermore, we could not even explore the impact of level of activity on effect estimates in a qualitative manner because all studies that could be rated were designated as having "low" level of activity. No apparent trends in effect estimates according to level of care based on qualitative assessments were observed.

Incidence of Pump Failure (KQ6)

Two case series and one RCT reported outcomes related to the pump device.[11,22,23] In a case series of 51 women, one participant had dislodgment of catheter (2 percent, exact central CI: 0.5%, 10%) and there was one pump that malfunctioned (2 percent, exact central CI: 0.5%, 10%).[22] No infusion site infections or mechanical failures were observed in a case series of nine women.[23] An underpowered RCT demonstrated indeterminate results for the outcomes of local pain and local skin irritation.[11] No data were available for missed doses or overdoses.

Applicability

In Table C below, we summarize the overall applicability of the evidence base, according to the domains of population, intervention, comparison, outcomes, timing, and setting.

Table C. Overall applicability of the body of evidence

Population	The majority of evidence pertained to women with recurrent preterm labor and singleton gestation in the United States. Very little information was reported about the study populations' demographic and clinical characteristics. Nine of 14 studies (64 percent) included women judged to be in labor on account of persistent contractions and cervical change. The definition of labor was unclear in other studies. Among the studies that suggested that the pump was efficacious, 50 percent reported cervical change and contractions as part of the definition of labor while 50 percent did not report how labor was defined.
Intervention	Although there were gaps in reporting, the intervention generally did not pose any serious limitations to applicability. Very few details were reported on cointerventions that could modify the effectiveness of therapy, such as administration of corticosteroids. In several studies, participants received specialized outpatient services from Matria Healthcare.
Comparison	Comparators included oral tocolytics, no treatment, and placebo.
Outcomes	Surrogate outcomes were the most commonly reported. Data on clinical outcomes, neonatal/maternal harms, and pump-related outcomes were sparse. Long-term outcomes have not been reported at all.
Timing of Outcomes Measurement	The absence of followup beyond delivery is a major limitation because important long-term outcomes have not been evaluated.
Setting	All studies were from the United States and participant data were acquired from a national database (Matria) or from single center sites. Women from the Matria database generally received a high level of care from an outpatient perinatal program. However, the distribution of regions from which patient data were included into the national database is unknown and information about the standards followed by the individual practice sites that provided obstetrical care was not reported. Similarly, for those studies that took place at single center sites, the standards of care followed at these sites are unclear.

Discussion

In this small review of 14 studies, most data came from observational designs, and several studies analyzed data from the Matria database. Aside from two RCTs, the studies exhibited considerable clinical and methodological heterogeneity. For the gradable outcomes, the available evidence addressed only two specific populations of interest—women with RPTL or those additionally with twin gestation. The strength of evidence favoring the SQ terbutaline pump compared with oral tocolytics for neonatal death in women with twin gestation and RPTL is low (OR = 0.09, 95% CI: 0.01, 0.70). While this result is striking in the presence of insufficient findings on other neonatal health outcomes summarized below, it is apparent that it stems from the largest of studies contributing data on neonatal health outcomes with more than 700 patients. As such, it is the only outcome that appears to be adequately powered to reach statistical significance. Strength of evidence favoring terbutaline pump compared to oral tocolytics or no treatment is also low for women with twin gestation and/or RPTL for the surrogate outcomes of pregnancy prolongation. For bronchopulmonary dysplasia, significant intraventricular hemorrhage, death within initial hospitalization, and Withdrawal-AE, strength of evidence is insufficient. The evidence was inconclusive for all other neonatal health outcomes, neonatal harms, maternal harms, and pump-related outcomes.

Based on postmarketing surveillance data, the FDA has issued a new warning against the use of terbutaline in general, and as an injection in particular, as maintenance tocolysis (i.e., beyond 48–72 hours) in pregnant women.[24] Although meriting transparent disclosure in the form of a warning, evidence emerging from case reports is usually regarded as noncomparative and hypothesis generating signal rather than a hypothesis testing confirmation.[25] Furthermore, case reports are useful in identifying rare and unexpected adverse events—the rarer the adverse event, the stronger is the effect size, and the magnitude of effect size is an important criterion that increases our confidence in an estimate.[9] However, adverse events such as death, hypertension, myocardial infarction, tachycardia, arrhythmias, and pulmonary edema that were reported with the use of terbutaline are not so unexpected in any adult population—pregnant women may experience these adverse events in the absence of terbutaline therapy due to other reasons.

Observational studies of medium to high risk of bias, primarily from the Matria database, showed benefit of SQ terbutaline pump compared with oral tocolytics or no treatment for other surrogate outcomes, such as birth weight and NICU admission, for women with twin gestation and/or RPTL. In contrast, two small RCTs that did not address any of the populations of interest, reported nonsignificant differences for several surrogate outcomes.

The evidence base for this review contained several limitations. Most evidence came from observational designs of medium to high risk of bias. Several outcomes revealed nonsignificant results that could be attributed to type II error. Type II error is a statistical term that implies inability of studies to find a difference when it might truly exist because of their small sample size (false negative). Many important variables, such as race, socioeconomic status, and fetal fibronectin level were not reported. Furthermore, cointerventions, such as administration of corticosteroids, were rarely described. None of the included studies assessed long-term childhood outcomes, such as childhood development, neurobehavioral testing, long-term lung function, and long-term vision. Our review comprehensively reviewed the literature and selected reports based on well-defined inclusion and exclusion criteria. However, one potential limitation of our review process is that we excluded potentially relevant non-English publications. Also, we could not investigate the impact of publication bias. However, in completing this review, we undertook an extensive grey literature search. Further, we requested relevant scientific information from the industry and had many experts in the field participate in the review process. Despite this thorough process, the number of identified studies was very small—we had too few studies per outcome to perform statistical assessment of publication bias. We believe that all relevant data regarding the use of subcutaneous terbutaline for the prevention of preterm labor is captured in this review. Any exaggerated positive findings are more likely due to the medium to high risk of bias detected in observational studies than publication bias.

In conclusion, the available evidence suggests that pump therapy is beneficial as maintenance tocolysis. However, our confidence in the validity and reproducibility of this evidence is low. While postmarketing surveillance has detected cases of serious harms, safety of the therapy remains unclear.

Future Research

Although cohort studies have provided a glimpse of the potential for the SQ terbutaline pump to improve short-term neonatal outcomes for fetuses at risk for preterm birth, the answers to several important questions remain unanswered. Most importantly, it remains to be seen whether SQ terbutaline pump therapy alters long-term development or systemic impairment of offspring, and neonatal/maternal morbidity and mortality. The limitations of the available data must also be recognized. Most of the cohort studies were medium to high risk of bias. In addition, several of the cohort studies investigated participants from a single proprietary database (Matria), which raises concerns regarding double-counting of patients and common biases. Therefore, results showing effectiveness should be interpreted with caution, especially in light of the most recent FDA warning recommending against the use of terbutaline for maintenance tocolysis.

Information is lacking on the effectiveness and safety of SQ terbutaline pump as a maintenance tocolytic treatment in specific populations, including women who deliver at specific gestational ages, women of different racial or ethnic backgrounds, and women with previous preterm birth or preeclampsia. Future studies, whether observational or experimental in design, should focus on garnering evidence for these specific populations.

Below we provide some specific recommendations for the conduct of RCTs and observational studies to further elucidate the potential benefits and harms of SQ terbutaline pump for maintenance tocolysis.

Randomized trials. We recommend that an adequately powered randomized controlled and pragmatic clinical trial that assesses the SQ terbutaline pump as a maintenance tocolytic be conducted. A pragmatic RCT is designed to have broad applicability so that the results can guide decisions about practice.[26] Such a trial should be placebo controlled and include blinding of study participants, care providers, and study personnel. Consideration should be given to employing multiple treatment arms in order to evaluate the pump against other tocolytic agents and conservative management. Furthermore, the level of care provided to participants (i.e., nursing assessments, home uterine monitoring, education, telephone support, and restriction of activities) should be practical, feasible, and likely to be adopted in routine practice. Important cointerventions, such as administration of corticosteroids, should be reported. A full accounting of the number of women approached but not enrolled should be included to allow users to assess the impact of respondent bias. The analysis should be "intent to treat," where all participants assigned by randomization to each group are included in the primary comparisons, regardless of whether the assigned medication was received. Outcomes to be examined should go beyond those of prolongation of pregnancy and birth weight to hard clinical endpoints of neonatal morbidity, such as bronchopulmonary dysplasia, necrotizing enterocolitis, significant intraventricular hemorrhage (grade III/IV), retinopathy of prematurity, sepsis, stillbirth, and neonatal death. Lastly, there should be long-term followup to assess subsequent childhood outcomes. Pharmacodynamic and pharmacokinetic outcome measures can additionally be studied to understand inter-individual differences in effectiveness and toxicity and avoidance of ß-agonist related tachyphylaxis.

Conducting RCTs to assess the efficacy of tocolytics in general is notoriously difficult. A definitive trial in this domain must include a focus on accurate diagnosis of preterm labor (perhaps combining stringent clinical criteria with factors such as positive fetal fibronectin and shortened transvaginal cervical length). Emphasis must also be placed on securing funding and maintaining followup for an appropriate duration of time to allow assessment of long-term childhood outcomes, including neurobehavioral testing and developmental assessment.

Observational studies. Although the RCT is the ideal study design for evaluating the efficacy of interventions, it may not be feasible for a number of reasons, such as a prohibitive sample size requirements and ethical considerations. We realize that collecting

RCT evidence on clinically important outcomes may not be possible because a large number of patients will need to be recruited to detect rare events, such as maternal deaths. Therefore, we additionally propose:

- Well-designed, well powered cohort studies examining clinical outcomes. These studies should include a representative and inception cohort of all patients with arrested preterm labor. Since observational studies are susceptible to the effects of confounding, future observational studies should measure, report, and adjust for potential confounders such as fetal fibronectin, cervical length/dilation, cerclage, maternal characteristics (e.g., age, race), level of care and activity, and concomitant medications. Propensity scores based on these variables may be considered. Other considerations about power, multiple comparison groups, level of care, reporting of cointerventions, and long-term followup are the same as for RCTs.
- Record linkage studies in which mothers' prenatal and infants' NICU and childhood developmental electronic health records are linked may be a more practical research proposition for the near future with improvements in quality and accessibility of electronic patient records. NICU registries in which prenatal data of mothers are available can be a very valuable source. However, such linkage based studies may also be impacted by biases not uncommon to cohort study designs, especially confounding because of unmeasured or unrecorded variables with important prognostic implications.

Glossary

Preterm birth: Delivery before completion of the 37th week of gestation.

Tocolytic: An agent that inhibits labor by slowing or halting uterine contractions.

Strength of evidence: The strength of evidence grading reflects a global assessment of the evidence base. Strength of evidence may be designated as insufficient, low, moderate or high based on the domains of study risk of bias, consistency, directness, and precision.

Applicability: The relevance of the evidence base to an external population.

Bias: A systematic error, arising from participant selection or outcome measurement that produces an erroneous effect estimate.

References

1. Behrman R, Butler A. Preterm Birth: Causes, Consequences, and Prevention. Washington, DC: National Academies Press; 2007.
2. Centers for Disease Control and Prevention. Birthweight and Gestation. Available at: www.cdc.gov/nchs/fastats/birthwt.htm. Accessed May 12, 2010.
3. Mathews TJ, MacDorman MF. Infant mortality statistics from the 2006 period linked birth/infant death data set. Natl Vital Stat Rep 2010 Apr 30;58(17):1-31. PMID: 20815136.
4. Nanda K, Cook LA, Gallo MF, et al. Terbutaline pump maintenance therapy after threatened preterm labor for preventing preterm birth. Cochrane Database Syst Rev 2002 Dec;(4):
5. HAYES, Inc. Continuous subcutaneous terbutaline infusion for treatment of preterm labor (Structured abstract). Lansdale, Pa: Hayes, Inc 2006;
6. Hayes E, Moroz L, Pizzi L, et al. A cost decision analysis of 4 tocolytic drugs. Am J Obstet Gynecol 2007 Oct;197(4):383-6. PMID: 17904969.
7. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Draft]. AHRQ Publication No. 10(11)-EHC063-EF. Available at: www.effectivehealthcare.ahrq.gov/ehc/products/60/318/MethodsGuide_Prepublication_Draft_20110824.pdf. Rockville, MD: 2011.
8. Santaguida, P and Raina, P. McMaster Quality Assessment Scale of Harms (McHarm) for primary studies: Manual for use of the McHarm. Available at: http://hiru.mcmaster.ca/epc/mcharm.pdf. Accessed December 22, 2010.
9. Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. J Clin Epidemiol 2010 May;63(5):513-23. PMID: 19595577.

10. Guinn DA, Goepfert AR, Owen J, et al. Terbutaline pump maintenance therapy for prevention of preterm delivery: a double-blind trial.[see comment]. Am J Obstet Gynecol 1998 Oct;179(4):874-8. PMID: 9790362.

11. Wenstrom KD, Weiner CP, Merrill D, et al. A placebo-controlled randomized trial of the terbutaline pump for prevention of preterm delivery. Am J Perinatol 1997 Feb;14(2):87-91. PMID: 9259905.

12. Lindenbaum C, Ludmir J, Teplick FB, et al. Maternal glucose intolerance and the subcutaneous terbutaline pump. Am J Obstet Gynecol 1992 Mar;166(3):925-8. PMID: 1550166.

13. Morrison JC, Chauhan SP, Carroll CS, Sr., et al. Continuous subcutaneous terbutaline administration prolongs pregnancy after recurrent preterm labor. Am J Obstet Gynecol 2003 Jun;188(6):1460-5. PMID: 12824979.

14. Morrison JC, Allbert JR, Floyd RC, et al. Interval to delivery in high-risk patients: Do tocolytic agents really work? Int J Gynaecol Obstet 1992;38(3):189-93. PMID: 1992210309.

15. Flick AA, de la Torre L, Roca LE, et al. An examination of the clinical benefits and cost-effectiveness of tocolytic replacement following recurrent preterm labor. Am J Perinatol 2010 Jan;27(1):53-9. PMID: 19823965.

16. de la Torre L, Istwan NB, Desch C, et al. Management of recurrent preterm labor in twin gestations with nifedipine tocolysis. Am J Perinatol 2008 Oct;25(9):555-60. PMID: 18773381.

17. Fleming A, Bonebrake R, Istwan N, et al. Pregnancy and economic outcomes in patients treated for recurrent preterm labor. J Perinatol 2004 Apr;24(4):223-7. PMID: 14999214.

18. Lam F, Istwan NB, Jacques D, et al. Managing perinatal outcomes: the clinical benefit and cost-effectiveness of pharmacologic treatment of recurrent preterm labor. Managed Care 2003 Jul;12(7):39-46. PMID: 12891954.

19. Lam F, Bergauer NK, Jacques D, et al. Clinical and cost-effectiveness of continuous subcutaneous terbutaline versus oral tocolytics for treatment of recurrent preterm labor in twin gestations. J Perinatol 2001 Oct;21(7):444-50. PMID: 11894512.

20. Allbert JR, Johnson C, Roberts WE, et al. Tocolysis for recurrent preterm labor using a continuous subcutaneous infusion pump. J Reprod Med 1994 Aug;39(8):614-8. PMID: 7996525.

21. Regenstein AC, Belluomini J, Katz M. Terbutaline tocolysis and glucose intolerance. Obstet Gynecol 1993 May;81(5 (Pt 1):739-41. PMID: 8469464.

22. Adkins RT, Van Hooydonk JE, Bressman PL, et al. Prevention of preterm birth: early detection and aggressive treatment with terbutaline.[see comment]. South Med J 1993 Feb;86(2):157-64. PMID: 8240475.

23. Lam F, Gill P, Smith M, et al. Use of the subcutaneous terbutaline pump for long-term tocolysis. Obstet Gynecol 1988 Nov;72(5):810-3. PMID: 3173932.

24. U.S. Food and Drug Administration. FDA Drug Safety Communication: New warnings against use of terbutaline to treat preterm labor. Available at: www.fda.gov/Drugs/DrugSafety/ucm243539.htm. Accessed April 7, 2011.

25. Chou R, Aronson N, Atkins D, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. J Clin Epidemiol 2010 May;63(5):502-12. PMID: PM:18823754.

26. Zwarenstein M, Treweek S, Gagnier JJ, et al. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ 2008;337:a2390. PMID: 19001484.

Full Report

This executive summary is part of the following document: Gaudet L, Singh K, Weeks L, Skidmore B, Tsouros S, Tsertsvadze A, Daniel R, Doucette S, Walker M, Ansari MT. Terbutaline Pump for the Prevention of Preterm Birth. Comparative Effectiveness Review No. 35. (Prepared by the University of Ottawa Evidence-based Practice Center under Contract No. HHSA290-2007-10059I.) AHRQ Publication No. 11-EHC068-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

For More Copies

For more copies of Terbutaline Pump for the Prevention of Preterm Birth: Executive Summary No. 35 (AHRQ Pub. No 11-EHC068-1), please call the AHRQ Clearinghouse at 1–800–358–9295 or email ahrqpubs@ahrq.gov.

AHRQ Pub. No. 11-EHC068-1
September 2011