Neonatal outcomes with water birth: A systematic review and meta-analysis

Jennifer Vanderlaan, MSN MPH CNM Clinical Instructora,b,⁎, Priscilla J. Hall, PhD CNM Clinical Instructora, MaryJane Lewitt, PhD CNM Assistant Clinical Professorc

a Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Road, Atlanta, GA 30322, United States
b Nell Hodgson Woodruff School of Nursing, Emory University, United States
c Practice Chair University of New Mexico College of Nursing, United States

Introduction

Water immersion in labor and birth is a valuable strategy for pain relief, associated with a significant reduction in the use of epidural anesthesia (Cluett and Burns, 2009). Water immersion has the potential to reduce the cost of care and provide women who want physiologic births with an effective tool for non-pharmacological pain management. Though water immersion in the first stage of labor is generally considered a safe and low-cost method of pain management for women in labor, concerns persist regarding the safety of second stage immersion and birth under water. The most recent Cochrane Review of hydrotherapy for labor and birth reported a lack of randomized controlled trials examining neonatal safety of second stage immersion (Cluett and Burns, 2009). Citing a lack of evidence, some organizations discourage use of water birth (Papile et al., 2014).

While randomized controlled trials protect against some forms of bias, limitations exist in randomized controlled trials with water birth. Blinding of subjects and providers is never possible with water birth introducing a risk for performance bias. Ethical treatment of research subjects requires following protocols that remove women from water if there is concern for fetal well-being or when there is a desire for stronger pain control. This means women at higher risk of poor outcomes are likely removed from the water birth intervention group resulting in attrition bias. In the only study to report results for both women randomized to water birth and women who chose water birth, women willing to be randomized to water birth were less likely to deliver in the water than women who show a preference for water birth (Woodward and Kelly, 2004). This causes additional problems for estimating the sample size necessary for an adequately powered sample in a randomized controlled trial.

Though there are only a few randomized controlled trials on neonatal outcomes during water birth, multiple non-randomized studies and clinical audits exist. It is appropriate to include non-randomized studies in a meta-analysis when gaps exist in the evidence from randomized controlled trials, though in such cases the risk of bias and confounding in non-randomized studies must be addressed (Norris, 2011). For the non-randomized studies that report neonatal outcomes with water birth, using existing medical records data for water birth protects against some types of bias such as selective detection of poor outcomes and missing data. However, these studies may be at higher risk of bias and confounding because women who delivered in water may be different from, or have different labor trajectories than, women who did not deliver in water. This may lead to selection bias, bias in defining the intervention and control groups, or attrition bias.

Two recently published meta-analyses of neonatal outcomes synthesized experimental and observational studies, however limitations of these studies make translation to hospital practice difficult (Davies et al., 2015; Taylor et al., 2016). First, these reviews combine hospital and out of hospital deliveries, which may make generalization of findings to hospital populations difficult. Second, the reviews relied on separate synthesis of continuous and binary outcomes. Szenesizing effect sizes that use different metrics is acceptable if the studies would have otherwise been considered appropriate to combine (Borenstein et al., 2009). Use of statistical analysis that correctly combines continuous and binary outcomes allows for larger samples which facilitates identification of differences in rare outcomes necessary to determine if second stage immersion is safe. Third, the syntheses did not account for differences in water birth protocols between the studies which may contribute to the heterogeneity of reported results. Stratifying studies by specific water birth protocols can help identify practices which may contribute to any identified poor outcomes, an important step to create evidence based protocols for water birth. Finally, neither study performed cumulative meta-analysis. Cumulative meta-analysis identifies the stability of findings over time and can help determine the likelihood of future research to alter the results of synthesis.

To address these gaps in the literature, we aimed to synthesize the broadest possible collection of evidence regarding neonatal outcomes with hospital water birth to identify any association with poor neonatal outcomes, including associations with specific water birth policies. The specific question was whether water birth was associated with poor neonatal outcomes as measured by Apgar, need for resuscitation, pneumonia, neonatal infection, neonatal respiratory distress, neonatal...
hypothesis, umbilical pH, shoulder dystocia, cord avulsion, neonatal intensive care unit (NICU) admission, and neonatal mortality.

Methods

A study protocol was completed and registered with PROSPERO, registration number CRD42014015487.

Eligibility Criteria

Studies were eligible if they compared neonatal outcomes between hospital water birth and a hospital control group for full term neonates. Non-English language papers were included if the information needed for assessment was available in an English abstract or tables could be translated. Papers were excluded if they reported results from immersion during first stage only, reported maternal outcomes only, included births outside the hospital, or did not provide adequate comparison data.

Information Sources

Studies were found by searching five databases: PubMed, Embase, Web of Science, CINAHL, and WorldCat Dissertations. No restrictions were placed on language or dates of publication. Initial searches were conducted on November 13, 2014 and all database searches were updated on May 17, 2016 to collect any studies published during the process of review. In addition, reference lists of existing reviews and selected articles were reviewed, as well as a bibliography available from Waterbirth International.

Search Strategy

Keywords and index terms were used to gather as wide a range of papers as possible. The search terms combined neonatal outcomes of interest (neonatal resuscitation, neonatal mortality, respiratory distress syndrome, APGAR, neonatal intensive care, infant mortality, and pneumonia) with terms describing the water birth intervention (water immersion, hydrotherapy, water birth). The search strategy is included in Supplement 1.

Study Selection

Two researchers independently screened all abstracts to remove obviously non-eligible studies. Any article without unanimous agreement to exclude was included in the full text review. Two researchers independently reviewed each full text article. When the researchers disagreed about article inclusion, the paper was reviewed by the researchers jointly to obtain a consensus decision. All selected studies were included in the systematic review. Studies that provided adequate data for statistical synthesis were included in the meta-analysis.

Data Collection Process

Data were extracted from the reports using coding sheets designed to capture outcome data as well as information about participant and intervention characteristics which could be used to determine reasons for heterogeneity. The forms were pilot tested prior to use. Characteristics of the intervention included inclusion and exclusion criteria as well as any description of the protocol used. Characteristics of the participants included maternal age, parity, gestational age, use of oxytocin and analgesics, proportion of women with prior cesarean, matching variables, and selection of controls. Outcome data included the measure used, whether analysis was intention to treat or per protocol, presence and direction of effect, and any control variables included in analysis.

Two researchers collected data from each article independently. Where researchers differed in their interpretation of a study, researchers reviewed the study together until consensus was achieved.

Outcome Data Items

For each study, outcome data were extracted for as many categories as reported without limiting based on measurement or variable type. Outcomes were selected based on the theoretical risks and benefits of water birth for the neonate, including shoulder dystocia, cord avulsion, 1- and 5-minute APGAR, need for resuscitation, umbilical pH, neonatal hypothermia, infections, respiratory distress, neonatal intensive care, and neonatal mortality.

Risk of Bias in Individual Studies

Risk of bias was assessed using a modification of the ROBINS-I tool for all studies regardless of methodology. This assessment included seven domains: confounding, selection, intervention measurement, attrition, missing data, outcome measurement, and reporting. The ROBINS-I tool is designed to identify risk of bias individually for each outcome, however for this study, we assessed risk of bias once for each study and applied the same risk of bias to all outcomes reported.

Two reviewers made independent domain-level judgments about risk for bias for each study. Studies which did not provide enough information to assess risk of bias were determined to be at unclear risk of bias. For any domain in any study where reviewers disagreed, the reviewers discussed the study until consensus was determined.

Summary Measures and Synthesis of Results

Outcomes for this study were synthesized as log odds ratio and converted to odds ratio as the summary measure. Studies were included in the meta-analysis as long as enough information was provided to calculate a log odds ratio for synthesis. Combining different measures of effect is acceptable because all studies addressed the same question and would have been combined if they had reported the same outcome measure (Borenstein et al., 2009). Continuous measures were transformed to an odds ratio in Comprehensive Meta-Analysis v3 (CMA3) using the procedure proposed by Hasselblad and Hedges (da Costa et al., 2012).

Effect sizes were calculated in CMA3 using a fixed-effects analysis weighted with inverse variance. Fixed-effects was appropriate because it was assumed that the standard policy limiting water birth to low-risk women meant all studies would share a common effect size which can be generalized to all low-risk women (Borenstein et al., 2009). As such, we assumed large differences in effect size between studies, known as heterogeneity, would not be due to the underlying risk of water birth for low-risk women but due to differences in water birth or study protocols which could be identified through stratification. Heterogeneity was addressed by stratification on water birth protocol characteristics to identify water birth policies which result in poor neonatal outcomes. Heterogeneity of effects was assessed by the Q statistic and I2.

When a study provided data on more than one comparison group, a single comparison was selected for synthesis by selecting the group most like standard hospital care. For example, if water birth was compared to both birth in a hospital bed and birth on a birth stool, the hospital bed group was selected as the comparison. When results were reported differently for independent groups in a study, such as nulliparas and multiparas, the groups were combined for analysis. In the case of continuous outcomes, a random effects meta-analysis was used to combine the means. The exception to this was a study that included two arms, a randomized group and a preference group, which remained separate for analysis due to the difference in study type.

The risk of bias graph and forest plots were prepared using Review Manager 5.3 because of the feature that allowed inclusion of risk of bias
data in the figures.

**Additional Analysis**

**Generalizability of Findings**

Generalizability of the estimate was assessed by examining the stability of the results when the synthesis was restricted to studies that explicitly described a water birth protocol or control group characteristic.

**Publication bias**

Publication bias was assessed using funnel plots with observed and imputed studies. We suspected publication bias when there was evidence the published studies differed systematically from the studies which would be published in a balanced funnel plot. Including both observed and imputed studies allowed estimation of the expected change if publication bias was not present.

**Effect of potential bias**

To assess effect of potential bias the analysis was repeated after excluding any study that was at risk of confounding, selection bias, measurement bias or attrition bias because these biases had the potential to skew the estimate in favor of water birth. Sensitivity analysis could only be performed when at least two studies remained for an outcome after risk of bias exclusion. To assess the overall effect of inclusion of observational studies, the analysis was repeated stratifying the studies by observational vs experimental study type.

**Effect of wrong statistic**

After examining the studies, we determined no studies using matched cohorts used the McNemar’s test, the standard for analyzing matched data. Though McNemar’s test does use a chi-square distribution, it analyzes discordant pairs rather than individual observations. Given the age of many studies, we did not seek paired data to calculate odds ratios using McNemar’s test. The effect of using the wrong statistic was assessed by simulating the results of McNemar’s when raw counts were provided. To simulate McNemar’s, we calculated the odds ratio assuming the highest proportion of discordant outcomes possible.

**Stability of findings over time**

Cumulative meta-analysis, analysis of the change in estimate over time, was performed to assess the stability of the evidence about neonatal outcomes with water birth. Cumulative meta-analysis begins with the earliest estimate, and tracks the change in estimate as studies are added chronologically.

**Findings**

**Study Selection**

After removing duplicates, 638 studies were available for screening. In the first round of review, evaluation of titles and abstracts, 547 studies were excluded from review because it was clear to both reviewers that they were either not reporting about water birth, not reporting neonatal outcomes, not limited to hospital births, or did not include primary data. A total of 91 studies were considered appropriate for full text review.

Full text review resulted in the exclusion of 52 articles. The most common reasons for exclusion were failure to report on a control group (k=12), failure to provide primary data (k=12), study of water immersion during first stage only (k=9) or publication of results from a duplicate population (k=9). The final sample included 39 articles with 34 providing data necessary to complete meta-analysis for at least one outcome. See Fig. 1 for the study flow diagram.

**Study Characteristics**

Individual study characteristics are available in Table 1. The sample included randomized controlled trials (k=5), two non-randomized controlled trials (k=2), cohort studies (k=21) and matched cohorts (k=12), though one publication reported on two studies (Woodward & Kelly, 2004). Most utilized retrospective enrollment (k=22). About half the studies (k=15) had a sample size of 200 or less, while about a quarter (k=8) had a sample larger than 1000. The studies provided data for 28,529 births, of which 12,592 were reported to be under water (45.1%). Two studies did not provide data about sample size. Three of the studies presented updated data on a sample used in a prior study; in those cases, the prior report was included in synthesis only for the outcomes not included in the updated paper.

All studies characterized their sample as low risk women though descriptions varied. The most common specific criteria mentioned for water birth eligibility were a minimum gestational age (k=25), singleton pregnancy (k=21), and cephalic presentation (k=21). The most common exclusion criteria were pregnancy complications (k=19), although these complications were rarely defined. The most common criteria that required women to exit the water and deliver on land were non-reassuring fetal heart tones (k=15) and presence of meconium stained fluid (k=9). The most common comparison group was defined by normal spontaneous vaginal delivery (k=15), though some prospective studies included any vaginal delivery (k=7). Most studies required the control group to meet the water birth eligibility criteria (k=20). Some studies required the control group to not use hydrotherapy (k=11) while others allowed first stage immersion (k=6). The most common matching variable was parity (k=10) followed by age (k=6) and gestation (k=3).

**Risk of bias within studies**

Overall risk of bias was moderate to low (see Fig. 2), though 23 studies had at least one high risk of bias within the four domains which could potentially move the estimate in favor of water birth; confounding, selection bias, measurement of intervention, or attrition bias. Risk of bias for studies which could only be accessed as an abstract were often unable to be completed.

Risk for confounding was high when there was no evidence the control and experimental groups were similar (k=8). Most studies addressed confounding by restricting the sample to a definition of low risk pregnancies and matching the control group. The most common evidence presented for group similarity was maternal age (k=14) followed by gestational age (k=9). While some studies identified differences in use of uterotonic (k=2) and use of analgesics (k=7), we could find no evidence to suggest either difference would confound the outcomes (Budden et al., 2014; Bugg et al., 2013; Klomp et al., 2012; Ullman et al., 2010). No studies controlled for confounding through analysis.

Risk for selection bias was high with retrospective selection of participants based on delivery outcomes because water birth required the woman not experience any of the reported exclusion criteria (k=13). This meant it was possible the water birth group had to meet some stricter criteria of normalcy to be included. This is meaningful because women with intrapartum complications, such as meconium stained fluid or abnormal fetal heart rate patterns, may have been excluded from the water birth group but not necessarily from the control group. If this occurred, the bias would result in an estimate that favored water birth.

Risk for measurement bias was high when it was determined the control group may have included women who used hydrotherapy at some point in labor (k=6). In practice, first stage hydrotherapy may not represent a bias as second stage immersion is commonly a continuation of first stage immersion. However, if women who were removed from the water due to exclusion criteria, such as meconium stained fluid,
were counted as conventional deliveries rather than excluded from analysis, the result would bias the estimate in favor of water birth.

Five studies were rated as high risk for outcome measurement bias because they were testing water birth as a new practice in a hospital. Our team believed the health care providers may be more cautious when testing a new procedure and therefore more likely to begin resuscitation or antibiotics after a water birth than after conventional hospital birth. If this occurred, the bias would result in an estimate that favored conventional birth.

Overall there was low risk of other biases. The studies at risk for attrition bias did not report the initial number of enrolled women and the attrition from each group (k=3). This prevented the team from ensuring there had not been unequal attrition. Missing data (k=1) was not a problem for most studies because the follow-up was limited to time in the hospital. Few studies were determined to be at risk for reporting bias (k=2) because all studies used simple analyses and nearly all indicated the outcomes of interest a priori.

Synthesis of results

Table 2 provides the full results of all syntheses and the risk for bias sensitivity analysis. Two syntheses could not be performed; cord avulsion and 1-minute APGAR. Cord avulsion was only reported by one study which found no difference between water births and conventional deliveries (Henderson et al., 2014).

The synthesis for 1-minute APGAR was unable to be performed because of heterogeneity, that is the estimates provided by the studies varied more than would be expected by chance alone. Per our protocol, we attempted to resolve the heterogeneity through stratification. However, we were unable to reduce the heterogeneity through stratification on any participant or protocol characteristics we had collected. The persistent heterogeneity indicated the variation in estimates is unlikely to be due to the randomness expected when a consistently defined intervention is performed. Given the lack of heterogeneity in any other outcome, it is not likely this heterogeneity is due to true difference in the populations being studied. We concluded there must be differences in the measurement of 1-minute APGAR between study sites causing heterogeneity. Therefore, we excluded 1-minute APGAR from further analysis and completed a random effects model synthesis which found no difference between water birth and conventional delivery as can be seen in Fig. 3 (OR 0.9 95% CI 0.66–1.21 Q=72.73 p < .001 I^2=79%).

Shoulder dystocia

Four studies reported quantitative data for shoulder dystocia (see Fig. 4). Synthesis of these studies resulted in no difference between water birth and conventional delivery (OR 0.84 95% CI 0.43–1.63). Sensitivity analysis could not be performed. Cumulative meta-analysis revealed this finding has been stable since 1999 when the outcome was first reported.

5. 5-minute APGAR

Seventeen studies reported quantitative data for 5-minute APGAR (see Fig. 5). Synthesis of these studies resulted in no difference between water birth and conventional delivery (OR 0.92 95% CI 0.75–1.11). An additional 15 studies reported no difference but did not provide data for meta-analysis (see studies 4, 6, 7, 9, 13, 14, 16–18, 21, 25, 29, 32 on Table 1). The result remained no difference between water birth and conventional delivery in the sensitivity analysis. Cumulative meta-analysis revealed this finding has been stable since 1995.
Need for Resuscitation
Six studies provided data for synthesis of need for resuscitation (see Fig. 6). Overall there was no difference in need for resuscitation between water birth and conventional delivery (OR 0.77 95% CI 0.97–1.24). The non-significant result remained during sensitivity analysis. Cumulative meta-analysis revealed this finding has been stable since first reported in 1999.

Umbilical pH
Seven studies provided data for synthesis of umbilical pH (see Fig. 7). Overall there was no difference in umbilical pH between water birth and conventional delivery (OR 0.97 95% CI 0.75–1.24), a finding supported by two additional studies reporting no difference but not including data needed for synthesis (studies 7 and 25 on Table 1). Sensitivity analysis for risk of bias maintained the finding of no difference. Cumulative meta-analysis revealed this finding has been stable since first reported in 1997.

Neonatal Hypothesis
Four studies provided data for neonatal hypothermia (see Fig. 8). Synthesis revealed an association with water birth and reduced odds of hypothermia (OR 0.56 95% CI 0.33–0.97). This appears to be due to the large sample size and not clinical differences in odds of hypothermia because the mean neonatal temperature was within normal range.
for both water birth and control groups in all studies reporting mean temperatures. In the sensitivity analysis, the estimate was no difference between water birth and conventional delivery. There were too few studies reporting neonatal hypothermia to determine stability of this finding over time.

**Infection**

Measurement of infection included laboratory confirmed cases, cases treated per symptomology regardless of laboratory confirmation, and comparisons of contamination regardless of symptomology. Studies reported a variety of cultured bacteria, colonization and infections including pneumonia, conjunctivitis, “sticky cord”, and nasal contamination. We divided the synthesis of infections into studies reporting pneumonia and studies reporting other types of infections due to the theoretical difference in risk for infection with water birth; the risk for inhalation of water increases the risk for pneumonia however the dilution of bacteria by the water may decrease risk for topical infections.

Four studies provided data for pneumonia infections (see Fig. 9). Overall there was no difference in odds of pneumonia between water birth and conventional delivery (OR 1.88 95% CI 0.36–9.86). Though one study was responsible for nearly half the weight in this analysis, the outcome of no difference was consistent when that study was removed.

Thirteen studies provided data for non-pneumonia infections (see Fig. 10). The synthesis revealed a lower odds of non-pneumonia infections with water birth compared to conventional delivery (OR 0.60 95% CI 0.37–0.97). Two additional studies reported qualitative results of no difference (see 13 and 29 in Table 1). In the sensitivity analysis, there was no difference in odds of non-pneumonia infections between water birth and conventional deliveries. Cumulative meta-analysis indicated this outcome first favored water birth in 2004 and has been stable since.

**Respiratory Distress**

Five studies provided data for synthesis of odds of respiratory distress (see Fig. 11). Overall the syntheses favored water birth (OR 0.44 95% CI 0.25–0.75). In this synthesis, one large study was responsible for nearly 50% of the weight. When this study was removed, the outcome was no difference between water birth and conventional delivery (OR 0.87 95% CI 0.41, 1.86), a finding supported by the finding of no difference in the sensitivity analysis. Cumulative meta-analysis without the large study indicated there has never been evidence of a statistically significant difference in respiratory distress between water deliveries and conventional deliveries, however with only three studies we are unable to determine the stability of this finding.

**NICU Admission**

Seventeen studies provided data for synthesis of odds of NICU admission (see Fig. 12). Overall the syntheses favored water birth (OR 0.70 95% CI 0.55, 0.90), while the two additional studies reported no difference between water birth and conventional delivery (see studies 13 and 25 on Table 1). In the sensitivity analysis, there was no difference in odds of NICU admission between water birth and conventional delivery. Cumulative meta-analysis revealed this outcome moved from no difference to favor water birth in 2001 and has been stable since.

**Neonatal Death**

Four studies reported estimates for neonatal death (see Fig. 13). In this synthesis, there was no difference in odds of neonatal death between water birth and conventional delivery (OR 0.83 95% CI 0.19–3.39). Cumulative meta-analysis indicated there has never been evidence of a statistically significant difference in odds of neonatal death between water birth and conventional delivery, but four studies may be too few to determine the stability of an estimate for such a rare outcome.

**Additional Analyses**

**Generalizability**

Stratification to determine generalizability of findings across specific criteria or conditions was performed to the extent we were able given the sparsity of reporting on eligibility criteria and practice protocols. Full results of these syntheses are in Supplement 3.

The results were stable when synthesis was restricted to studies

---

**Table 2**

Results of meta-analysis of neonatal outcomes with hospital water birth compared to conventional delivery.

<table>
<thead>
<tr>
<th>Outcome (Included studies by number)</th>
<th>Total Births</th>
<th>Meta-Analysis</th>
<th>Sensitivity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Odds Ratio (95% CI)</td>
<td>Stable Since</td>
</tr>
<tr>
<td>Shoulder Dystocia (12, 22, 23, 38)</td>
<td>10,478</td>
<td>0.84 (0.43, 1.63)</td>
<td>1999</td>
</tr>
<tr>
<td>5-Minute APGAR (5, 10, 12, 15, 19, 20, 22, 24, 26–30, 35–37)</td>
<td>15,332</td>
<td>0.92 (0.754, 1.11)</td>
<td>1995</td>
</tr>
<tr>
<td>Resuscitation (15, 16, 22, 27, 38, 30)</td>
<td>3,106</td>
<td>0.77 (0.39, 1.54)</td>
<td>1999</td>
</tr>
<tr>
<td>Umbilical pH (1, 6, 22, 27, 34, 36, 37)</td>
<td>4,153</td>
<td>0.97 (0.75, 1.24)</td>
<td>1997</td>
</tr>
<tr>
<td>Neonatal Hypothermia (11, 22, 28, 36)</td>
<td>290</td>
<td>0.56 (0.33, 0.97)</td>
<td>2014</td>
</tr>
<tr>
<td>Pneumonia (12, 17, 28)</td>
<td>9,685</td>
<td>1.88 (0.36, 9.86)</td>
<td>2004</td>
</tr>
<tr>
<td>Non-pneumonia Infections (1, 7, 12, 14, 17, 23, 25–27, 33, 36, 37)</td>
<td>14,744</td>
<td>0.65 (0.47, 0.90)</td>
<td>2004</td>
</tr>
<tr>
<td>Respiratory Distress (7, 11, 25–27)</td>
<td>9,144</td>
<td>0.44 (0.25, 0.75)</td>
<td>1995</td>
</tr>
<tr>
<td>NICU Admission (3, 6, 7, 12, 15, 16, 18–20, 22, 23, 25, 26, 30–35–37)</td>
<td>39,155</td>
<td>0.70 (0.55, 0.90)</td>
<td>2003</td>
</tr>
<tr>
<td>Neonatal Death (22, 25, 31, 35)</td>
<td>13,749</td>
<td>0.83 (0.19, 3.69)</td>
<td>1999</td>
</tr>
</tbody>
</table>

All qualitatively reported outcomes were no significant difference but did not provide data necessary for inclusion in the meta-analysis.
reporting explicit water birth eligibility criteria, with all results either the same as the main analyses or, in the case of odds of neonatal hypothermia and odds of NICU admission, no difference. These analyses included requiring evidence of estimated fetal weight to be within normal limits, evidence of established progression of labor, exclusion from water birth due to meconium stained fluid, allowing entry to water with ruptured membranes, and excluding women with prior cesarean.

The only water birth practice characteristic with sufficient data to test stability of findings was the speed with which the infant was removed from the water. We found no difference in reported outcomes when studies required immediate removal of the infant from water.

The findings remained stable when synthesis was restricted for specific control group characteristics with results either the same as the main analysis or no difference. These analyses included studies with control groups that used either analgesia or uterotonic at a higher rate than the water birth group and restricting the study to women without prior cesarean. Because water birth is generally considered to be a pain management intervention, restriction of the synthesis to studies in which the control group had a higher proportion of analgesia use than the water birth group may be more similar comparison to what a hospital may expect when implementing a water birth program. In that analysis, there were no differences between water birth and conventional delivery in any outcome reported in two or more studies.

Publication Bias

Inspection of the funnel plot with observed and imputed studies for each outcome revealed adjusted point estimates equal to or more in favor of water birth for each outcome, suggesting no publication bias in favor of water birth. Funnel plots have been provided in Supplement 4.

Effect of Potential Bias

Results of sensitivity analysis for risk of bias is included in the reporting of each outcome and on Table 2. A sensitivity analysis stratifying synthesis by controlled trials and observational studies revealed consistent findings of no difference between water birth and conventional delivery for most outcomes, though synthesis of observational studies was in favor of water birth for NICU admission, non-pneumonia infections, and respiratory distress (see Supplement 5). The consistency of the two analyses suggest that, though biases may move estimates for some outcomes to favor water birth, these biases are not obscuring evidence of poor outcomes with water birth.

Wrong Statistic

The results of the sensitivity analysis estimating the correct statistics for matched data were no difference for most outcomes, and in favor or water birth for neonatal hypothermia and non-pneumonia infections (see Supplement 6). This finding suggests use of the correct test statistic is unlikely to have altered the results of the synthesis.

Stability over Time

Results of cumulative meta-analysis are included in the reporting of each outcome. The cumulative meta-analysis revealed there has never been a time when evidence indicated an association between water birth and any poor neonatal outcome. Based on the number of studies synthesized, the stability of the estimates over time suggests future research is unlikely to change the synthesized estimate for 5-minute APGAR, need for resuscitation, umbilical pH, non-pneumonia infections, and NICU admission.
Discussion

This study used statistical techniques that allowed us to create the largest synthesis of neonatal outcomes with waterbirth to date. This review of 39 studies reporting estimates for 12 different neonatal outcomes did not find evidence of an increased odds of poor neonatal outcome with water birth compared to conventional delivery. The findings from this study agree with the two meta-analyses published while this study was being prepared (Davies et al., 2015; Taylor et al., 2016). This study adds to the evidence by limiting the analysis to hospital deliveries, which may be why only one outcome, 1-minute APGAR, had heterogeneity. This study also adds to the evidence by using statistical techniques that allowed a wider range of studies to be combined and by testing the stability of these findings when assumptions about the studies changed.

Based on the cumulative meta-analysis, these findings for several outcomes are unlikely to be changed with additional research. Heterogeneity exists when the studies have enough variation in the intervention being tested that we can no longer assume we are measuring one thing. We expected heterogeneity given the differences in reported study characteristics and the assumed differences in standard practice based on the wide range of year and location for the studies. The only synthesis found to have heterogeneity was 1-minute APGAR. The consistency of the estimates suggests there was little difference in safety between the studies with the most and least restrictive water birth eligibility criteria. Though we could not specifically test for differences based on water birth eligibility criteria and protocols, the stability of the findings when analysis was restricted to those criteria specifically defined in the included studies supports the safety of all protocols included in this study. Future research should examine existing water birth protocols to identify areas of variation and to test for associations between protocol characteristics and neonatal outcomes.

These results show water birth associated with lower odds of some outcomes, but the lower odds did not persist when studies at risk for bias in favor of water birth were removed from the synthesis. It is physiologically unlikely that delivery under water reduces odds of NICU admission or respiratory distress. Instead, these findings likely provide evidence that practitioners correctly identify the cases where a fetus may be compromised, allowing removal of the woman from the water before delivery. The lower odds of non-pneumonia infections with water birth raise the question of the effects of water birth on colonization of the neonate for the normal microbiome and pathologic bacteria. Though sensitivity analysis for risk of bias was performed, our definitions of confounding and selection bias were intended to address questions of overall neonatal wellbeing and may not account for sources of potential bias in the study of topical colonization of the neonate. This association should be further investigated to understand associations between colonization of the neonate and delivery under water.

This study was limited by the availability of published studies and the data available in the articles. Our search was conducted using English keywords, which limited our ability to identify studies without English abstracts. Additionally, we were unable to locate four studies.

### Table 1

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Water Birth</th>
<th>Conventional</th>
<th>Odds Ratio IV, Fixed, 95% CI</th>
<th>Risk of Bias</th>
<th>Bias of bias model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Henderson, 2014</td>
<td>4</td>
<td>5</td>
<td>986</td>
<td>29.4%</td>
<td>0.43 [0.12, 1.53]</td>
</tr>
<tr>
<td>Kovland, 2014</td>
<td>2</td>
<td>2</td>
<td>61</td>
<td>14.0%</td>
<td>0.97 [0.15, 6.05]</td>
</tr>
<tr>
<td>Nelson, 1999</td>
<td>4</td>
<td>2</td>
<td>60</td>
<td>15.7%</td>
<td>0.60 [0.26, 1.43]</td>
</tr>
<tr>
<td>Regl, 1999</td>
<td>1</td>
<td>1</td>
<td>100</td>
<td>9.1%</td>
<td>0.49 [0.04, 5.55]</td>
</tr>
<tr>
<td>Ross, 2009</td>
<td>5</td>
<td>2</td>
<td>27</td>
<td>26.3%</td>
<td>0.80 [0.25, 2.70]</td>
</tr>
<tr>
<td>Scott, 2013</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>6.9%</td>
<td>1.00 [0.06, 16.9]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>1810</td>
<td>1296</td>
<td>100.0%</td>
<td>6.77 [1.39, 1.54]</td>
<td></td>
</tr>
</tbody>
</table>

### Fig. 5

Forest plot of meta-analysis for 5-minute APGAR comparing hospital water birth to conventional delivery.

### Fig. 6

Forest plot of meta-analysis for need for resuscitation comparing hospital water birth to conventional delivery.
identified through our search (see Supplement 2 for excluded studies). This study was delimited to include only hospital births to provide evidence for use by hospitals. Results of this synthesis should not be assumed to represent differences between water birth and conventional birth in out of hospital settings.

This study was the most comprehensive attempt to synthesize the evidence about neonatal outcomes with water birth to date. In addition to finding no increased odds of poor neonatal outcomes with water birth compared to conventional delivery practice, the analysis revealed this finding was stable for many outcomes and is unlikely to change with future research. As such, water immersion during second stage may be considered a safe non-pharmacologic method for managing labor pain. Water immersion during second stage should be considered among the tools hospitals use to promote physiologic birth and made available to all low risk women who wish to use this method for pain management (American College of Nurse-Midwives, 2014; Committee on Obstetric Practice, 2017). Though this study was able to provide some evidence of the stability of findings for specific hospital water birth policies, future research should begin to focus on the best criteria for eligibility and exclusion from second stage hydrotherapy to provide clear clinical practice guidance that allows the widest access to water birth while maintaining optimal patient safety and equivalent neonatal outcomes. Such research can expand evidence for hospital policy makers by including measurement of quality outcomes such as iatrogenic neonatal sepsis and exclusive breastfeeding (American College of Nurse-Midwives, 2014; Roberts and Milton, 2015).

Acknowledgement

This project was funded by a grant from the Alpha Epsilon Chapter of Sigma Theta Tau International.
Fig. 10. Forest plot of meta-analysis for non-pneumonia infections comparing hospital water birth to conventional delivery.

Fig. 11. Forest plot of meta-analysis for respiratory distress comparing hospital water birth to conventional delivery.

Fig. 12. Forest plot of meta-analysis for neonatal intensive care unit admission comparing hospital water birth to conventional delivery.
Conflict of Interest
None declared.

Ethical Approval
Not applicable.

Funding Sources
Research grant received March, 2015 from Alpha Epsilon Chapter of Sigma Theta Tau International.

Registration
Systematic review and meta-analysis protocol was registered with PROSPERO: CRD42014015487.

Appendix A. Supporting information
Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.mijd.2017.12.023.

References
Nikodem, V., 1999. Immersion in Water During Birth: A Randomized Controlled Trial...