Archives of
Disease in Childhood

Neonatal outcomes of waterbirth; a systematic review and meta-analysis.

Journal:	*Archives of Disease in Childhood*
Manuscript ID	fetalneonatal-2015-309600.R2
Article Type:	Review
Edition:	not in use
Date Submitted by the Author:	23-Mar-2016
Complete List of Authors:	Taylor, Henry; University College London, Institute of Child Health Kleine, Ira; Luton & Dunstable NHS Foundation Trust, Bewley, Susan; King's College London, Women's Health Academic Centre Loucaides, Eva; University College London, Institute of Child Health Sutcliffe, Alastair; University College London, Institute of Child Health
Keywords:	Neonatology, Waterbirth, Neonatal Outcomes

Title: Neonatal outcomes of waterbirth; a systematic review and meta-analysis.

Authors: Henry Taylor[1], Ira Kleine[2], Susan Bewley[3], Eva Loucaides[1], Alastair Sutcliffe[1*]

Institutional affiliation:

1. Institute of Child Health, University College Hospital, London, UK,
2. Luton and Dunstable University Hospital Trust, Luton, UK
3. Women's Health Academic Centre, King's College London, London, UK

*Corresponding author: Professor Sutcliffe, 'Population, Policy & Practice', Institute of Child Health, UCL, 30 Guilford Street, London, WC1N 1EH, UK. Tel: 020 7905 2190. Email: a.sutcliffe@ucl.ac.uk

Key words: Waterbirth; Neonatology; Neonatal Outcomes.

Word count: 2498 (excluding abstract, figures, tables, acknowledgements, ethics approval, competing interests and references)

Figures: 5, **Tables:** 2, **References:** 69, **Supplementary files:** 5

Acknowledgements: We would like to thank Dr Bev Botting for assistance with statistical analysis, and the following for manuscript translation: Aghileh Djafari Marbini and Dougal Hargreaves (Persian), Tomas Cudrnak, Mary Dashfield, and Daniel Roman (Czech), Magda Kolanko (Polish).

Funding: This research received no specific grant from any funding agency, but was supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. This paper presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Protocol registration: CRD42015030119

Ethics approval: not required as secondary reanalysis of primary studies

Competing interests: SB was paid to chair the NICE intrapartum care guideline update CG190 although this did not examine waterbirth

Contributions: Conception (IK, SB, AS), design (SB, HT, EL), data acquisition (HT, IK, EL), meta-analysis (HT), writing initial draft (HT, IK, SB), editing (SB, HT, EL, AS), approving final draft (all), guarantor (AS).

ABSTRACT

Introduction: In 2015 nine percent of babies born in the UK were delivered underwater.(1) Waterbirth is increasing in popularity, despite uncertainty regarding its safety for neonates. This systematic review and meta-analysis appraises the existing evidence for neonatal outcomes following waterbirth.

Methods: A structured electronic database search was performed with no language restrictions. All comparative studies which reported neonatal outcomes following waterbirth, and that were published since 1995, were included. Quality appraisal was performed using a modified CASP scoring system. The primary outcome was neonatal mortality. Data for each neonatal outcome was tabulated and analysed. Meta-analysis was performed for comparable studies which reported sufficient data.

Results: The majority of the 29 included studies were small, with limited follow-up, and methodological flaws. They were mostly conducted in Europe and high-income countries. Reporting of data was heterogeneous. No significant difference in neonatal mortality, NICU/SCBU admission rate, Apgar scores, umbilical cord gases, or infection rates was found between babies delivered into water or on land.

Conclusion: This systematic review and meta-analysis did not identify definitive evidence that waterbirth causes harm to neonates compared to land birth. However, there is currently insufficient evidence to conclude that there are no additional risks or benefits for neonates when comparing waterbirth and conventional delivery on land.

INTRODUCTION

In 2015 nine percent of babies born in the UK were delivered underwater.(1) Waterbirth is an increasingly popular choice for women in labour, despite uncertainty regarding its safety for neonates.

Proponents argue that neonates are protected by the diving reflex of the newborn and benefit from an increased chance of uncomplicated vaginal delivery with delayed cord clamping. Concerns have been raised over possible increased risk of neonatal infection, aspiration, cord avulsion, and mortality.(2) In addition, waterbirth could influence early bacterial colonisation of the intestine, affecting the development of the gut microbiome. This mechanism is thought to be responsible for the altered infant microbiome which develops following caesarean section and has been linked to immunological disorders and obesity in childhood.(3-6)

In the USA, the American Academy of Pediatrics and the American College of Obstetricians and Gynecologists do not endorse waterbirth as a routine delivery option,(7) citing rare and serious adverse events in the newborn. In the UK, the Royal Colleges of Midwives (RCM) and Obstetricians and Gynaecologists (RCOG) advocate giving all healthy women with uncomplicated pregnancies at term the option of waterbirth.(8) However, they note that true informed choice on the benefits and risks of waterbirth is clouded by the lack of good quality safety data. This is partly because serious adverse events in low risk pregnancy are rare. To be adequately powered to detect a difference in neonatal mortality rate a study would need to have 3500 participants in each group.(9)

Waterbirth (WB, delivering a baby underwater) should be differentiated from water immersion (WI) during the first stage of labour, which has known maternal benefits including reduced duration of the first stage, and reduced need for epidural anaesthesia.(10)

The Physiology of Waterbirth

Aquatic mammals, such as whales and dolphins, give birth underwater with the newborn not breathing until it reaches the surface.(11, 12) This is facilitated by an enhanced antioxidant system and the diving reflex.(13) The diving reflex also exists in humans and provides some protection from drowning.(14, 15) Some argue that this reflex of apnoea, bradycardia and peripheral vasoconstriction protects the human neonate from aspiration during waterbirth. However, the presence of this reflex in newborns, and the 'naturalness' of relying on an emergency reflex, have been questioned.(16)

Postnatally, facial temperature (cold) receptors and laryngeal chemoreceptors trigger the trigeminal diving reflex and laryngeal chemoreflex respectively, leading to apnoea when cold water comes into contact with either the face or the larynx.(17-19) 94% of newborns demonstrate this response between 24-72 hours postnatally, and 100% do so at 2-6 months.(20) However, it is not known whether the reflex exists at birth or whether it is activated after the first breath.(16)

Even if the diving reflex does exist at birth, it may not be triggered by birth into water at body temperature. Fetal breathing movements persist in utero until the late third trimester despite warm amniotic fluid surrounding the fetus.(21) Meconium aspiration syndrome testifies that matter can be inhaled by the fetus or the immediate newborn. The presence of the diving reflex in newborns and its relevance to waterbirth has been challenged, undermining the physiological arguments commonly used to support WB. Any potential risk posed to babies born into water depends on the presence or absence of other factors that regulate the first breath.

The difficulty in elucidating whether, and why, newborns do not inhale when submerged arises from the uncertainty over the mechanism controlling the switch from fetal to extra-

uterine breathing. Hypothesised triggers to breathing in conventional birth on land include a combination of physical stimulation (such as light, temperature and handling), pain, hypercapnia, hypoxia, chronic endocrine changes, elastic recoil of thoracic tissue, and diaphragmatic contraction.(22, 23) Healthy babies delivered into warm water would not receive all of these stimuli. However, if a baby compromised by prior hypoxia and acidosis was born gasping, there would be a risk of aspiration of pool water.(24)

Inhibition of breathing in waterbirth may therefore be determined by the balance of inhibitory and stimulatory triggers. Whether or not this mechanism is sufficient to prevent morbidity in the neonate is yet to be determined.

The aim of this study was to determine the safety of waterbirth for the neonate compared to conventional vaginal delivery on land.

METHODS

A systematic review and meta-analysis of comparative studies of waterbirth versus conventional land birth, reporting neonatal outcomes, was carried out in accordance with current guidance,(25) using a pre-specified and registered protocol (CRD42015030119 registered 10/12/15).(26)

Eligibility Criteria

Inclusion criteria were peer-reviewed comparative studies reporting neonatal outcomes of waterbirth versus vaginal delivery on land. This comprised randomised controlled trials (RCTs), prospective and retrospective cohort studies (PCS, RCS), case control studies (CCS), cross sectional studies (CSS), and surveillance studies. Exclusion criteria were non-comparative studies, case series, opinions, reviews, and studies reporting neonatal outcomes following water immersion (WI) during labour without subsequent WB. No language restrictions were applied.

The primary outcome was neonatal mortality. Secondary outcomes were combined neonatal intensive care unit (NICU) or special care baby unit (SCBU) admission, resuscitation at birth, Apgar scores at one and five minutes, arterial and venous umbilical cord blood pH, post-natal infection, and knot in umbilical cord.

Search Strategy and Information Sources

Five databases were searched from 1st Jan 1995 to 8th December 2015; PubMed, EMBASE, CINAHL, British Nursing Index, and Ovid Maternity and Infant Care. The search protocol is detailed in Supplementary File A. Following removal of duplicates, titles and abstracts were screened by a single reviewer. Papers were hand searched for references. Foreign language papers were translated by a medically qualified native speaker. All papers excluded following full text review were independently read by two authors, with reference to a third senior author in case of disagreement.

Data Extraction and Quality Appraisal

Included studies were grouped by study design; data extraction was then performed using a pre-designed form. Data on study design, methodology, primary outcome, and secondary outcomes was captured when reported.

Risk of bias was considered during a quality assessment process. This used one of four appraisal tools (according to study design) modified from the CASP system (Supplementary File B).(27) Studies with score ≥11 were assigned as 'higher' quality.

Data extraction and quality appraisal were performed independently by two reviewers (HT, EL) for a random sample of 25% of included papers to check for inter-observer error. Data was tabulated for analysis (Supplementary File C).

Statistical analysis

Aggregate data was extracted from original studies. Where percentages were reported without numbers the numbers were calculated. For binary outcomes (neonatal mortality, NICU/SCBU admission, Apgar <7 or <8, infection) the risk difference and 95% confidence interval (CI) were calculated (28). Mean and median Apgar scores were compared between groups and reported as difference in average score (with 95% CI when standard deviation available). The median and range, or mean and calculated 95% range, of umbilical cord gas results were compared between groups. Results of non-parametric significance tests performed by authors of the original studies are included.

Meta-analysis was performed for comparable studies which reported sufficient data, specifically those with a low-risk maternal cohort and a matched control group (for retrospective studies) which reported binary outcomes or means with standard deviation. Risk differences (for binary outcomes) or mean differences (for numerical outcomes) were combined using inverse-variance weighing and a random effects model. Heterogeneity was measured using I^2. Results for studies with mixed or indeterminate risk cohorts were presented in the tables and figures, and subject to narrative analysis.

Two sensitivity analyses were performed using only RCTs or higher quality studies (quality score ≥11).

RESULTS

The initial search found 2470 articles, of which 47 underwent full text appraisal (Figure 1) and 28 were included. Excluded studies are listed in Supplementary File E.(29-39) One article included a description of two studies (an RCT and PCS) which were considered separately.(40) There was complete agreement between the two independent reviewers about choice of exclusion. There were only minor differences in quality appraisal (two points or less in four of seven studies) which were resolved by discussion.

Table 1 lists the design, quality score, and specific limitations and risks of bias of included studies.

The 29 included studies comprised: 5 RCTs, 9 PCS, 12 RCS, 1 CCS, 1 CSS, and 1 nationwide surveillance study. They were performed in 12 countries, but the majority in Europe and high income countries (8 in UK, 4 in Iran, 3 in Poland, 2 each in Australia, Austria, Switzerland, South Africa and Italy, and one each in New Zealand, China, Turkey and the Czech Republic).

Study sizes ranged from 20 to 14309 births (total number of included births = 39,302). All studies were set in a hospital or birth centre; most were small, single-centre, and reported on a limited number of neonatal outcomes with short follow up (see Table 1). 18 studies were limited to low risk women, 3 specified a mixed risk population, and 8 had an indeterminate risk cohort.

All studies had some risk of bias, as assessed by the modified CASP criteria (Supplementary File B). None of the RCTs were blinded (due to the nature of the intervention). Funnel plots were reviewed for all meta-analysable outcomes; no clear evidence of publication bias was noted.

Primary outcome

Neonatal mortality was reported in 10 studies (Figure 2) with a total of 27 deaths. Four studies were suitable for meta-analysis, one neonatal death was reported in these studies. Combined risk difference per thousand live births (RD_{1000}) was 0 (95% CI -10 to 10).

None of the remaining studies reported a significant difference in neonatal mortality. Two had sufficient power to detect a difference, one of which reported no deaths.(41) The other was a large nationwide surveillance study comparing 4032 WB to 10307 low risk deliveries, RD_{1000} was 0 (95% CI -10 to 20).(42)

Secondary outcomes

Data on NICU/SCBU admission was reported in 15 studies (Figure 3). Meta-analysis of eight studies found no significant difference between groups, RD_{1000} -10 (95% CI -20 to 10).

Narrative review of remaining studies identified that the majority concurred with the meta-analysis. The nationwide surveillance study reported WB infants had lower rates of NICU/SCBU admission, RD_{1000} -28 (95% CI -33 to -24).(42) One large PCS reproduced this finding, however the control group in this study comprised women of all levels of risk.(41) One RCS reported higher rates of NICU/SCBU admission following WB.(43)

Apgar scores were the most widely reported neonatal outcome (26 studies, Figure 4) but variable reporting complicates any synthesis. Seven studies reported the proportion of neonates scoring <7; there was no significant risk difference in any study at one or five minutes. Combined percentage risk difference ($RD_{\%}$) from meta-analysable studies was 0% (95% CI -1 to 1) at five minutes; at one minute data was heterogeneous (I^2 = 86%). Similarly, four studies reported the proportion of neonates with an Apgar score <8. Combined $RD_{\%}$ from three studies was 1% (95% CI -5.0 to 8.0). The remaining PCS reported a $RD_{\%}$ of -14% (95% CI -24 to -4), however risk profiles in this study were undefined.(44)

Combined data from studies reporting numerical Apgar scores identified marginally higher scores amongst WB neonates at one minute, mean difference 0.09 (95% CI 0.0 to 0.18). At five minutes data was heterogeneous (I^2=69%). Average scores were high (Supplementary File C).

Nine studies reported cord gas analysis. However, five only performed cord gases on a subset of neonates, six only reported arterial results and only three reported both arterial and venous results (Figure 5 and Supplementary File C). WB was associated with significantly higher arterial pH in two studies,(41, 45) and higher venous pH in one study.(46)

Of the eleven studies reporting on infection ten did not report any significant differences; one PCS found significantly more infections in controls (Table 2).(41)

Serious adverse events, not otherwise covered above, were also described. In one RCS three knotted umbilical cords were noted in the WB group versus none in controls.(47) In a

national surveillance study five incidents of snapped umbilical cord were recorded following waterbirth, however no reliable comparator data is available for this outcome.(42) Resuscitation was another adverse event, which was variably reported. Only two studies specifically reported on neonatal resuscitation as an outcome.(48, 49) Both reported resuscitation events in the waterbirth group, and none in the control group; however the differences were not significant.

Sensitivity analyses

Two sensitivity analyses were conducted separately; the first included 12 studies with higher quality scores (≥11), and the second 5 RCTs (Supplementary File D). Neither sensitivity analysis identified any significant findings compared to the primary analysis. In the primary analysis WB neonates had greater Apgar scores at one minute, this finding was conserved amongst high quality studies but not in RCTs.

DISCUSSION

Key findings

Most of the 29 studies addressing the comparison of neonatal outcomes following waterbirth were small, observational, and based on low-risk mothers. This perhaps reflects the ethical difficulty associated with randomisation. There was no difference in neonatal mortality following waterbirth compared to land birth. Analysis of the five measures of neonatal morbidity did not identify any consistent findings.

No meta-analysis was possible for umbilical cord gases (non-normal data), or infection rate (inconsistent definition of outcome between primary studies).

There is some evidence of higher mean Apgar scores and higher cord gas pH following WB, however this describes variation within the normal range and is of uncertain clinical significance.

Comparison with previous work

Previous systematic reviews of WB have largely concentrated on maternal outcomes.(10, 50) Only one recent systematic review specifically addressed neonatal outcomes.(51) The present study covers a longer time interval and contains a larger number of studies. We corroborate their findings that, for the majority of neonatal outcomes, there are no significant differences between waterbirth and land birth.

Strengths and limitations

This is the largest systematic review and meta-analysis considering neonatal outcomes following waterbirth performed to date. Strengths include a comprehensive search strategy, 20 year timespan, inclusion of 8 foreign language papers, wide inclusion criteria, and use of sensitivity analyses. Limitations include use of a single reviewer to perform the literature search, quality appraisal, and data capture; though a random sample of included papers were checked for consistency.

Implications for clinicians and research

Clinicians should inform women about the present, largely reassuring, data about the safety of waterbirth for their baby. There is no evidence of a difference in neonatal mortality or morbidity. However, uncertainties remain as existing evidence is not strong enough to examine the relative risk of rare and potentially devastating adverse events. Nor is there any evidence evaluating potential long-term implications of waterbirth versus land birth.

In order to assist informed decision making by pregnant women, their companions and health professionals, a large multi-centre RCT or PCS is a priority. There are undoubted maternal benefits of water immersion as well as practical and emotional difficulties in exiting the pool immediately prior to delivery.(10) Further research must consider the full safety profile of waterbirth by evaluating whether underwater delivery aids physiological fetal-to-

neonatal transition (possibly by avoiding interventions), affects the risk of rare adverse events, or causes any long-term benefits or harms, for example by influencing the developing microbiome.

REFERENCES

1. CQC. 2015 survey of women's experiences of maternity care. Care Quality Commission; 2015 [cited 2015 22 December]; Available from: http://www.cqc.org.uk/sites/default/files/20151215b_mat15_statistical_release.pdf.
2. Byard RW, Zuccollo JM. Forensic issues in cases of water birth fatalities. The American journal of forensic medicine and pathology. 2010;31(3):258-60. Epub 2010/05/04.
3. Sevelsted A, Stokholm J, Bønnelykke K, Bisgaard H. Cesarean section and chronic immune disorders. Pediatrics. 2015;135(1):e92-e8.
4. Kristensen K, Henriksen L. Cesarean section and disease associated with immune function. Journal of Allergy and Clinical Immunology.
5. Huh SY, Rifas-Shiman SL, Zera CA, Edwards JWR, Oken E, Weiss ST, et al. Delivery by caesarean section and risk of obesity in preschool age children: a prospective cohort study. Archives of Disease in Childhood. 2012.
6. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. Gut. 2014;63(4):559-66.
7. ACOG Committee Opinion no. 594: Immersion in water during labor and delivery. Obstetrics and gynecology. 2014;123(4):912-5. Epub 2014/05/03.
8. Royal College of Obstetricians and Gynaecologists, Royal College of Midwives. Immersion in water during labour and birth. London: Royal College of Obstetricians and Gynaecologists April. 2006;5.
9. Davies MW. Water births and the research required to assess the benefits versus the harms. Journal of paediatrics and child health. 2012;48(9):726-9. Epub 2010/07/06.
10. Cluett ER, Burns E. Immersion in water in labour and birth. The Cochrane database of systematic reviews. 2009(2):CD000111. Epub 2009/04/17.
11. McBride AF, Kritzler H. Observations on Pregnancy, Parturition, and Postnatal Behavior in the Bottlenose Dolphin. Journal of Mammalogy. 1951;32(3):251-66.
12. Hoelzel AR. Marine mammal biology: an evolutionary approach: John Wiley & Sons; 2009.
13. Cantu-Medellin N, Byrd B, Hohn A, Vazquez-Medina JP, Zenteno-Savin T. Differential antioxidant protection in tissues from marine mammals with distinct diving capacities. Shallow/short vs. deep/long divers. Comparative biochemistry and physiology Part A, Molecular & integrative physiology. 2011;158(4):438-43. Epub 2010/12/15.
14. Campbell LB, Gooden BA, Horowitz JD. Cardiovascular responses to partial and total immersion in man. The Journal of physiology. 1969;202(1):239-50. Epub 1969/05/01.
15. Foster GE, Sheel AW. The human diving response, its function, and its control. Scandinavian journal of medicine & science in sports. 2005;15(1):3-12. Epub 2005/02/01.
16. Walker JJ. Birth underwater: sink or swim. British journal of obstetrics and gynaecology. 1994;101(6):467-8. Epub 1994/06/01.
17. de Burgh Daly M, Elsner R, Angell-James JE. Cardiorespiratory control by carotid chemoreceptors during experimental dives in the seal. The American journal of physiology. 1977;232(5):H508-16. Epub 1977/05/01.
18. Perkett EA, Vaughan RL. Evidence for a laryngeal chemoreflex in some human preterm infants. Acta paediatrica Scandinavica. 1982;71(6):969-72. Epub 1982/11/01.
19. Goksör E, Rosengren L, Wennergren G. Bradycardic response during submersion in infant swimming. Acta Pædiatrica. 2002;91(3):307-12.
20. Pedroso FS, Riesgo RS, Gatiboni T, Rotta NT. The diving reflex in healthy infants in the first year of life. Journal of child neurology. 2012;27(2):168-71. Epub 2011/09/02.

21. Florido J, Padilla MC, Soto V, Camacho A, Moscoso G, Navarrete L. Photogrammetry of fetal breathing movements during the third trimester of pregnancy: observations in normal and abnormal pregnancies. Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology. 2008;32(4):515-9. Epub 2008/08/07.
22. van Vonderen JJ, Roest AA, Siew ML, Walther FJ, Hooper SB, te Pas AB. Measuring physiological changes during the transition to life after birth. Neonatology. 2014;105(3):230-42. Epub 2014/02/08.
23. van Kaam A. Transition to Extrauterine Life. In: Elzouki A, Harfi H, Nazer H, Stapleton FB, Oh W, Whitley R, editors. Textbook of Clinical Pediatrics: Springer Berlin Heidelberg; 2012. p. 115-20.
24. Johnson P. Birth under water - to breathe or not to breathe. British journal of obstetrics and gynaecology. 1996;103(3):202-8.
25. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009;6(7):e1000097.
26. Taylor H, Kleine I, Loucaides E, Bewley S, Sutcliffe A. Neonatal outcomes of birth in water: a systematic review of the health outcomes of neonates born into water compared to into air. PROSPERO International prospective register of systematic reviews2015 [cited 2015 22/12/2015]; Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015030119.
27. Critical Appraisal Skills Programme (CASP) checklists. Oxford2015 [cited 2015 21/12/15]; Available from: http://media.wix.com/ugd/dded87_40b9ff0bf53840478331915a8ed8b2fb.pdf.
28. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Statistics in medicine. 1998;17(8):873-90.
29. Burns EE, Boulton MG, Cluett E, Cornelius VR, Smith LA. Characteristics, interventions, and outcomes of women who used a birthing pool: a prospective observational study. Birth (Berkeley, Calif). 2012;39(3):192-202. Epub 2013/01/04.
30. Henderson J, Burns EE, Regalia AL, et al. Labouring women who used a birthing pool in obsteric units in Italy: prospective observational study. BMC pregnancy and childbirth. 2014;14(17).
31. Rush JPB. A randomized controlled trial of the effects of the bath in labour: University of Toronto (Canada); 1999.
32. da Silva FM, de Oliveira SM. The effect of immersion baths on the length of childbirth labor. [Portuguese] O efeito do banho de imersao na duracao do trabalho de parto. Revista da Escola de Enfermagem da U S P. 2006;40(1):57-63.
33. Geissbuehler V, Stein S, Eberhard J. Waterbirths compared with landbirths: an observational study of nine years. Journal of perinatal medicine. 2004;32(4):308-14. Epub 2004/09/07.
34. Zanetti-Dallenbach R, Lapaire O, Maertens A, et al. Water birth, more than a trendy alternative: a prospective, observational study. Archives of gynecology and obstetrics. 2006;274(6):355-65.
35. Thoni A, Zech N, Ploner F. Giving birth in the water: Experience after 1,825 water deliveries. Retrospective descriptive comparison of water birth and traditional delivery methods. [German] Gebaren im wasser: Erfahrung nach 1825 wassergeburten. Retrospektiv deskriptive vergleichsanalyse zwischen wassergeburten und traditionellen geburtsmodalitaten. Gynakologisch-geburtshilfliche Rundschau. 2007;47(2):76-80.
36. Damodaran S, Khatri P, Mahmood TA, Monaghan SC. Waterbirths in Fife: A 6-year observational study. Journal of Obstetrics and Gynaecology. 2010;30 (7):759.
37. Lim KMX, Tong PSY, Chong YS. A comparative study between waterbirths and conventional vaginal deliveries in an obstetricianled unit in Singapore. BJOG: An International Journal of Obstetrics and Gynaecology. 2015;122:283.
38. Ziolkowski R, Dabrus D, Czerniawski W, Dudek K, Darmochwal-Kolarz D, Oleszczuk J. An assessment of water births based on the author's own research. [Polish, English] Ocena przebiegu porodu w wodzie w oparciu o doswiadczenia wlasne. Ginekologia i Poloznictwo. 2009;14(4):57-65.
39. Price CA. Waterbirth and land birth: a comparative study of maternal and neonatal outcomes in one Australian birth centre. Treatise submitted in partial fulfilment of requirements for the award of master of nursing coursework (independent practitioner). Camperdown, Australia: University of Sydney, Faculty of Nursing November. 1995.
40. Woodward J, Kelly SM. A pilot study for a randomised controlled trial of waterbirth versus land birth. BJOG : an international journal of obstetrics and gynaecology. 2004;111(6):537-45. Epub 2004/06/17.
41. Geissbuhler V, Eberhard J. Experience with water births: A prospective longitudinal study of 9 years with almost 4,000 water births. [German] Erfahrung mit der unterwassergeburt: Eine prospektive longitudinale studie uber 9 jahre mit fast 4000 wassergeburten. Gynakologisch-geburtshilfliche Rundschau. 2003;43(1):12-8.

42. Gilbert RE, Tookey PA. Perinatal mortality and morbidity among babies delivered in water: surveillance study and postal survey. BMJ (Clinical research ed). 1999;319(7208):483-7.
43. Menakaya U, Albayati S, Vella E, et al. A retrospective comparison of water birth and conventional vaginal birth among women deemed to be low risk in a secondary level hospital in Australia. Women and Birth: the Journal of the Australian College of Midwives. 2013;26(2):114-8.
44. Torkamani SA, Kangani F, Janani F. The effects of delivery in water on duration of delivery and pain compared with normal delivery. Pakistan Journal of Medical Sciences. 2010;26(3):551-5.
45. Schrocksnadel H, Kunczicky V, Meier J, Brezinka C, Oberaigner W. Water birth: Experience at a university clinic and a district hospital in Austria. [German] Gebaren im wasser: Erfahrungen einer universitatsklinik und eines bezirkskrankenhauses in Osterreich. Gynakologisch-geburtshilfliche Rundschau. 2003;43(1):7-11.
46. Zanetti-Daellenbach RA, Tschudin S, Zhong XY, et al. Maternal and neonatal infections and obstetrical outcome in water birth. European Journal of Obstetrics and Gynecology and Reproductive Biology. 2007;134(1):37-43.
47. Otigbah CM, Dhanjal MK, Harmsworth G, et al. A retrospective comparison of water births and conventional vaginal deliveries. European Journal of Obstetrics and Gynecology and Reproductive Biology. 2000;91(1):15-20.
48. Nikodem C. The Effects of Water on Birth: A Randomised controlled trial: Rand Afrikaans University; 1999.
49. Ros HB. Effects of waterbirths and traditional bedbirths on outcomes for neonates. Curationis. 2009;32(2):46-52 7p.
50. Nutter E, Meyer S, Shaw-Battista J, et al. Waterbirth: an integrative analysis of peer-reviewed literature. Journal of Midwifery and Women's Health. 2014;59(3):286-319.
51. Davies R, Davis D, Pearce M, Wong N. The effect of waterbirth on neonatal mortality and morbidity: a systematic review and meta-analysis. JBI database of systematic reviews and implementation reports. 2015;13(10):180-231. Epub 2015/11/17.
52. Ghasemi M, Tara F, Ashraf H. Maternal-fetal and neonatal complications of water-birth compared with conventional delivery. [Persian]. Iranian Journal of Obstetrics, Gynecology and Infertility. 2013;16(70):9-15.
53. Gayiti MRY, Li XY, Zulifeiya AK, Huan Y, Zhao TN. Comparison of the effects of water and traditional delivery on birthing women and newborns. European review for medical and pharmacological sciences. 2015;19(9):1554-8.
54. Chaichian S, Akhlaghi A, Rousta F, Safavi M. Experience of water birth delivery in Iran. Archives of Iranian medicine. 2009;12(5):468-71. Epub 2009/09/03.
55. Mollamahmutoglu L, Moraloglu O, Ozyer S, Su FA, Karayalcin R, Hancerlioglu N, et al. The effects of immersion in water on labor, birth and newborn and comparison with epidural analgesia and conventional vaginal delivery. Suda dogumun, travay, dogum ve yenidogan uzerine etkileri ve epidural analjezi ile normal dogum ve konvansiyonel vajinal dogum ile karsilastirilmasi. Journal of the Turkish German Gynecology Association. 2012;13(1):45-9.
56. Hawkins S. Water vs conventional births: infection rates compared. Nursing times. 1995;91(11):38-40. Epub 1995/03/15.
57. Sipinski A, Poreba R, Cnota W, Poreba A. The analysis of 135 water births. [Polish] Analiza 135 porodow w wodzie. Ginekologia polska. 2000;71(4):208-12.
58. Bodner K, Bodner-Adler B, Wierrani F, Mayerhofer K, Fousek C, Niedermayr A, et al. Effects of water birth on maternal and neonatal outcomes. Wiener klinische Wochenschrift. 2002;114(10-11):391-5. Epub 2003/04/24.
59. Kolivand M, Almasi A, Heydarpour S. Comparison between the Outcomes of Water Birth and Normal Vaginal Delivery. Journal of Midwifery and Reproductive Health. 2014;2(4):220-6.
60. Pagano E, De Rota B, Ferrando A, Petrinco M, Merletti F, Gregori D. An economic evaluation of water birth: the cost-effectiveness of mother well-being. Journal of evaluation in clinical practice. 2010;16(5):916-9. Epub 2010/07/02.
61. Kowalewska M, Welfel E, Kawczynski P, Pokrzywnicka M. Clinical condition of newborns from water birth at the Perinatology Clinic, Institute Of Gynecology and Obstetrics of the Medical University in Lodz, in the years 1996-2001. [Polish] Stan kliniczny noworodkow z porodow w wodzie urodzonych w Klinice Perinatologii Instytutu Ginekologii i Poloznictwa Akademii Medycznej w Lodzi w latach 1996-2001. Ginekologia polska. 2004;75(4):267-73.
62. Pellantova S, Vebera Z, Pucek P. Under water deliveries - A five-years retrospective study. [Czech] Porody do vody - Petileta retrospektivni studie. Ceska Gynekologie. 2003;68(3):175-9.

63. Aird IA, Luckas MJM, Buckett WM, et al. Effects of intrapartum hydrotherapy on labour related parameters. Anzjog. 1997;37(2):137-42.
64. Burke E, Kilfoyle A. A comparative study: waterbirth and bedbirth. Midwives (13558404). 1995;108(1284):3-7 5p.
65. Thoni A, Mussner K, Ploner F. Water birthing: Retrospective review of 2625 water births. Contamination of birth pool water and risk of microbial cross-infection. [Italian] Partorire in acqua: Esperienza dopo 2.625 parti in acqua - Contaminazione dell'acqua nella vasca da parto e rischio di infezione con diversi microrganismi. Minerva ginecologica. 2010;62(3):203-11.
66. Garland D. Collaborative waterbirth audit -- 'supporting practice with audit': a collaborative audit between ten units with data collected during 2001. MIDIRS Midwifery Digest. 2002;12(4):508-11 4p.
67. Moneta J, Okninska A, Wielgos M, Przybos A, Chrostowska J, Marianowski L. The influence of water immersion on the course of labor. [Polish] Wplyw immersji wodnej na przebieg porodu. Ginekologia polska. 2001;72(12):1031-6.
68. Carpenter L, Weston P. Neonatal respiratory consequences from water birth. Journal of paediatrics and child health. 2012;48(5):419-23. Epub 2011/11/17.
69. Dahlen HG, Dowling H, Tracy M, et al. Maternal and perinatal outcomes amongst low risk women giving birth in water compared to six birth positions on land. A descriptive cross sectional study in a birth centre over 12 years. Midwifery. 2013;29(7):759-64.

Figure titles:

Figure 1: PRISMA flow chart of search results and paper selection

Figure 2: Difference in risk of neonatal mortality

Figure 3: Difference in risk of NICU/SCBU admission

Figure 4: Comparison of Apgar scores

Figure 5: Comparison of umbilical cord blood pH

Figure Legends:

Figure 1: A flow chart detailing search results and paper selection process

Figure 2: A forest plot showing the risk difference (RD) of neonatal mortality per thousand live births between the waterbirth (WB) and control group. RD and 95% confidence interval is plotted for each study. Studies with an asterisk (*) were not included in the meta-analysis (combined data).

Figure 3: A forest plot showing the risk difference (RD) of NICU/SCBU admission per thousand live births between the waterbirth (WB) and control group. RD and 95% confidence interval is plotted for each study. Studies with an asterisk (*) were not included in the meta-analysis (combined data).

Figure 4: Three forest plots comparing waterbirth (WB) and control groups. A: mean difference (with 95% CI) of Apgar score at one and five minutes. B: risk difference (RD) and 95% CI of having an Apgar score <7 at one and five minutes. C: RD and 95% CI of having an Apgar score <8 at five minutes. For each plot studies with an asterisk (*) were not included in the meta-analysis (combined data). A (^{T}) symbol indicates significant heterogeneity (I^2 >60%) in the combined data.

Figure 5: A side by side bar chart comparing arterial or venous cord blood pH in waterbirth (light grey bars) and control (dark grey bars) groups. Median and range (bars with capped lines) or mean and calculated 95% range (bars with uncapped lines) are plotted for each study. An asterisk (*) indicates a statistically significant difference.

Tables:

Table 1: Details of included studies and quality scores

Table 2: Comparison of neonatal infection rate

Supplementary File Titles:

Supplementary File A: Search strategy

Supplementary File B: Risk of bias and quality assessment

Supplementary File C. Tabulated data on neonatal outcomes

Supplementary File D: Results of sensitivity analyses

Supplementary File E: Excluded papers

Table 1. Details of Included Studies and Quality Scores

Author, Year Country	Study Design	WB n	Control n	Waterbirth (WB) group	Control group	Follow up	Quality Score	Other comment
Nikodem, 1999.(48) South Africa	RCT	60	60	Low risk SVD All had WB in one of two state hospitals. 59 SVD, 1 ventouse	Low risk SVD in one of two state hospitals. 58 SVD, 1 CS, 1 ventouse	24 hours	16	Women in this trial were consented and randomised after onset of labour.
Woodward et al. 2004.(40) UK	RCT	40	20	Low risk women on labour ward; 10 WB, 13 WI, 16 did not use pool, 1 withdrawn. 33 SVD, 4 instrumental, 2 CS	Low risk women on labour ward; 1 WI, 1 WB. 14 SVD, 3 instrumental, 3 CS	6 weeks	14	Of 40 allocated WB only 10 delivered in water. 1 woman allocated to control group delivered in water.
Ghasemi et al. 2013.(52) Iran	RCT	83	88	Low risk women in hospital. 78 WB, 4 CS, 1 ventouse	Low risk women in hospital. 74 SVD, 14 CS	1 week	14	
Gayiti et al. 2015.(53) China	RCT	60	60	Low risk women, all had WB in hospital.	Low risk women, all had “traditional delivery method” in hospital. Included AROM and continuous fetal monitoring	Not defined	10	WB group did not have AROM or continuous monitoring.
Chaichian et al. 2009.(54) Iran	RCT	53	53	Low risk women, all had WB in hospital.	Low risk women "conventional delivery method of the hospital"	Not defined	8	Unclear from the description whether any women withdrew from the study.
Woodward et al. 2004.(40) UK	PCS	10	10	Low risk women on labour ward. 5 WB, 1 WI, 4 did not use pool. 7 SVD, 2 LCS, 1 ventouse.	Low risk women in labour ward, 10 did not use pool. 9 SVD, 1 ventouse	6 weeks	14	Of 10 in WB arm only 5 delivered in water.
Mollamahmutoglu et al. 2012.(55) Turkey	PCS	207	204	Low risk women, all had WB in hospital.	Low risk women having “conventional delivery” in hospital.	Not defined	12	Study also provides epidural group for comparison, not included in this review.
Zanetti-Dällenbach	PCS	89	146	Low risk women, all had WB in hospital.	Low risk women, all had “normal vaginal delivery”	Until discharge	12	Study also provides WI group for comparison, not included in this

et al. 2007.(46) Switzerland				in hospital.			review. Women having operative delivery excluded from study. Significant difference in ethnicity compared to WB group (more Swiss, less Mediterranean)
Ros, 2009.(49) South Africa	PCS	27	27	Low risk women, all had WB in one of two private birthing centres.	Low risk women having conventional delivery in government hospital labour ward	14 days	11	
Hawkins, 1995.(56) UK	PCS	16	16	Low risk women, all had WB in midwifery unit in hospital.	Low risk women, group comprised of women having next 'routine' delivery following a waterbirth in hospital	7 days	11	
Geissbühler et al. 2003.(41) Switzerland	PCS	3617	5901	Mixed risk cohort of women having WB in hospital.	All women having single cephalic SVD	Not defined	10	All women had free choice to have WB at this centre. WB cohort therefore included high risk and premature deliveries. However control group had significantly greater proportion of women with high risk antenatal histories, pre-eclampsia, pathological CTG and meconium stained liquor.
Torkamani et al. 2010.(44) Iran	PCS	50	50	Multiparous women with term pregnancies, uncertain risk profile. All had WB in hospital.	Multiparous women with term pregnancies having “normal delivery” in hospital.	Not defined	9	Unclear if additional inclusion or exclusion criteria were applied to the WB group.
Sipinksi et al. 2000.(57) Poland	PCS	135	135	Women having WB in hospital. Indeterminate risk profile	Consecutive ‘normal vaginal deliveries’ on labour ward	Not defined	4	The authors do not report inclusion or exclusion criteria for either group. Baseline characteristics are not reported.
Menakaya et al. 2012.(43) Australia	RCS	219	219	Mixed risk women, all had WB in hospital.	Women having SVD on land in hospital within 24 hours of WB. Matched for gestational age, parity, and risk.	Not defined	14	Low risk women and those with GBS & PROM were allowed into pool if no signs of chorioamnionitis (hence mixed risk).

Bodner et al. 2002.(58) Austria	RCS	140	140	Low risk women, all had WB in hospital.	Women having "normal SVD" in hospital, matched for parity.	Not defined	14	
Otigbah et al. 2000.(47) UK	RCS	301	301	Low risk women, all had WB in hospital.	Women having next low risk SVD on labour ward register, matched for parity and age	Not defined	14	
Kolivand et al. 2014.(59) Iran	RCS	43	62	Low risk women having WB in hospital.	Woman having normal vaginal delivery, meeting inclusion criteria for WB, matched for parity and age.	1 month	10	
Schröcksnadel et al. 2003.(45) Austria	RCS	218	218	Indeterminate risk women, all had WB in hospital.	Women matched for age, parity, gestational age.	Not defined	8	Significant difference in ethnicity and level of maternal education between WB and control group. This study also included unmatched data from a rural centre which was excluded from this SR as was non-comparative.
Pagano et al. 2010.(60) Italy	RCS	110	110	Low risk nulliparous women, all had WB in hospital.	Women having next low risk land delivery on birth register of hospital	Not defined	8	No description of matching process, unclear if all control group women also nulliparous.
Kowalewska et al. 2004.(61) Poland	RCS	42	71	All women having WB in hospital in study period. Indeterminate risk profile	Women who had the first live vaginal delivery on labour ward for each month during the study period.	Until discharge	7	No matching of control group, significant differences in baseline characteristics. Mortality and cord gas data not included in this review as no comparative data reported. Apgar not included as no time (1 vs 5 min) specified.
Pellantova et al. 2003. (62) Czech Republic	RCS	70	70	Low risk women having WB in hospital.	Women having 'conventional deliveries' without contraindications for WB.	Not defined	7	Controls were not matched. Different baseline parity between groups.
Aird et al. 1997.(63) UK	Retrospective cohortRCS	67	100	Low risk women, all had WB in hospital.	Group comprised of women having next SVD on birth register in hospital, matched for	Not defined	6	The authors do not report all neonatal outcomes separated for WI and WB groups. Only WB outcomes included in this

					parity and age			systematic review.
Burke et al. 1995.(64) UK	RCS	50	50	Low risk women 'randomly selected from pool register' of hospital.	Women having next low risk SVD selected from birth register of hospital, matched for age and parity	Until discharge	6	Women in WB group were not allowed analgesia, except Entonox, from 4 hours prior to pool use. Control group did not have this restriction.
Thoni et al. 2010.(65) Italy	RCS	2625	899	Low risk women, all had WB in hospital.	Controls unmatched, had vaginal delivery on bed or using birthing stool in hospital.	Not defined	6	Number of controls differs for different analyses, uncertain of sampling methodology, characteristics of controls, or comparability of groups.
Garland et al. 2002.(66) UK	RCS	680	680	Mixed risk women having WB in 10 different birthing centres operating alongside hospitals	Women on birth register delivering at similar time in same birthing centre, matched for parity, VBAC, age, ethnicity	Not defined	5	The 10 centres had distinct inclusion & exclusion criteria, and data collection methods. VBAC was allowed, hence cohort is 'mixed risk'.
Moneta et al. 2001.(67) Poland	RCS	109	110	All women having WB in hospital in study period. Indeterminate risk profile.	Randomly selected women 'giving birth in traditional way' on labour ward at same time as WB.	Not defined	2	Random selection of controls not described, no matching described. The WB group had a higher proportion of primiparous women.
Carpenter et al. 2012.(68) New Zealand	CCS	14	26	Neonates born at term in birth centres and hospitals within catchment area of tertiary NICU. All admitted to NICU with respiratory distress requiring pressure support following WB.	Neonates born at term in one of two local birth centres. All admitted to NICU with respiratory distress requiring pressure support following vaginal delivery on land	Until discharge	14	Neonates with encephalopathy or congenital heart disease excluded from both groups.
Dahlen et al. 2013.(69) Australia	Retrospective cross sectional	819	5220	All women having WB in birth centre alongside a hospital. Indeterminate risk profile.	All women having vaginal delivery in birth centre over same time period	Not defined	9	Outcomes recorded from midwives own handwritten notes. No data from women transferred out of birthing centre during labour.
Gilbert et al. 1999.(42) UK	Surveillance	4032	10307	All perinatal deaths and NICU/SCBU admissions within 48 hours in UK following WB.	Low risk deliveries from North West Thames region 1992-3	7 days for mortality, 48 hours for NICU	8	This surveillance study gives multiple control groups. The largest low risk group was used here for comparison.

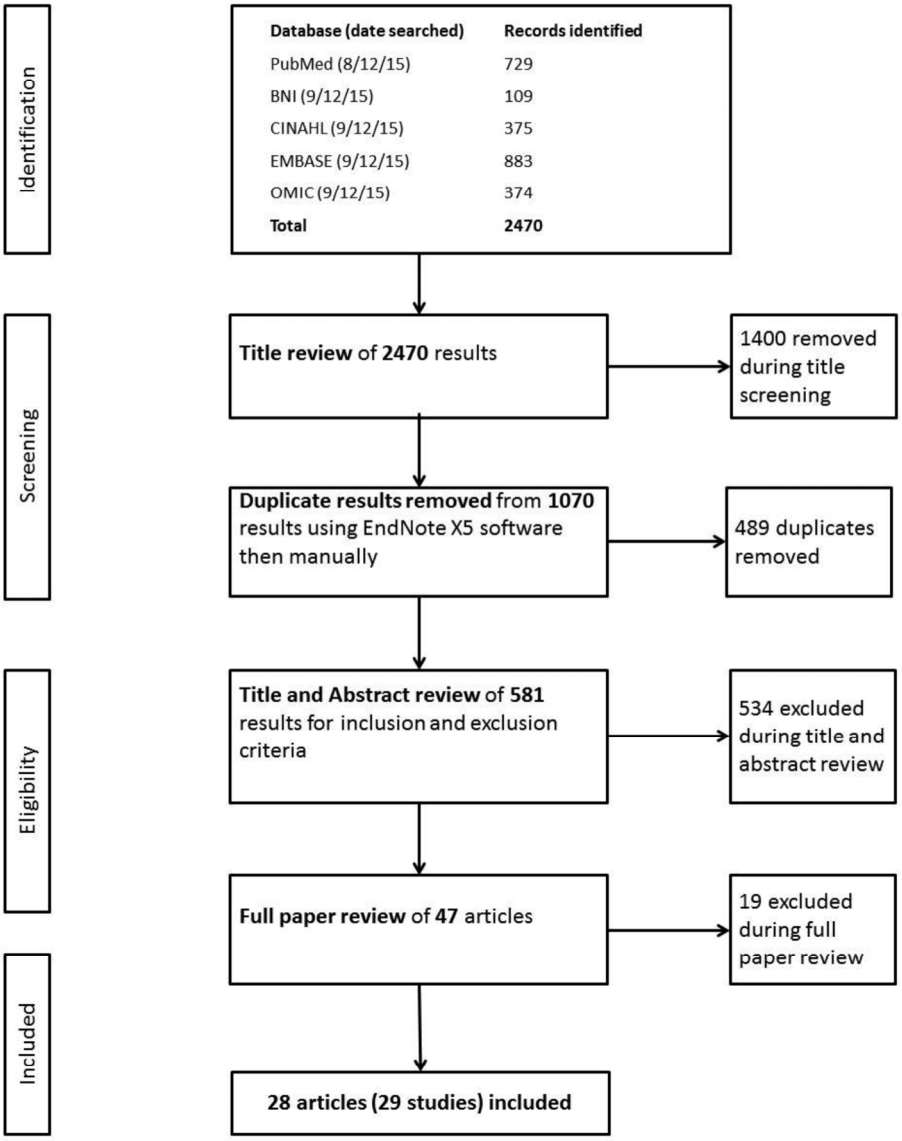
				Indeterminate risk profile.		admission		

Key: PCS = prospective cohort study, RCS = retrospective cohort study, CCS = case control study, SVD = spontaneous vaginal delivery, WB = waterbirth, WI = water immersion during first stage of labour, VBAC = vaginal birth after caesarean section, CS = caesarean section, GBS = group B streptococcus, PROM = premature rupture of membranes, NICU = neonatal intensive care unit, SCBU = special care baby unit; AROM = artificial rupture of membranes

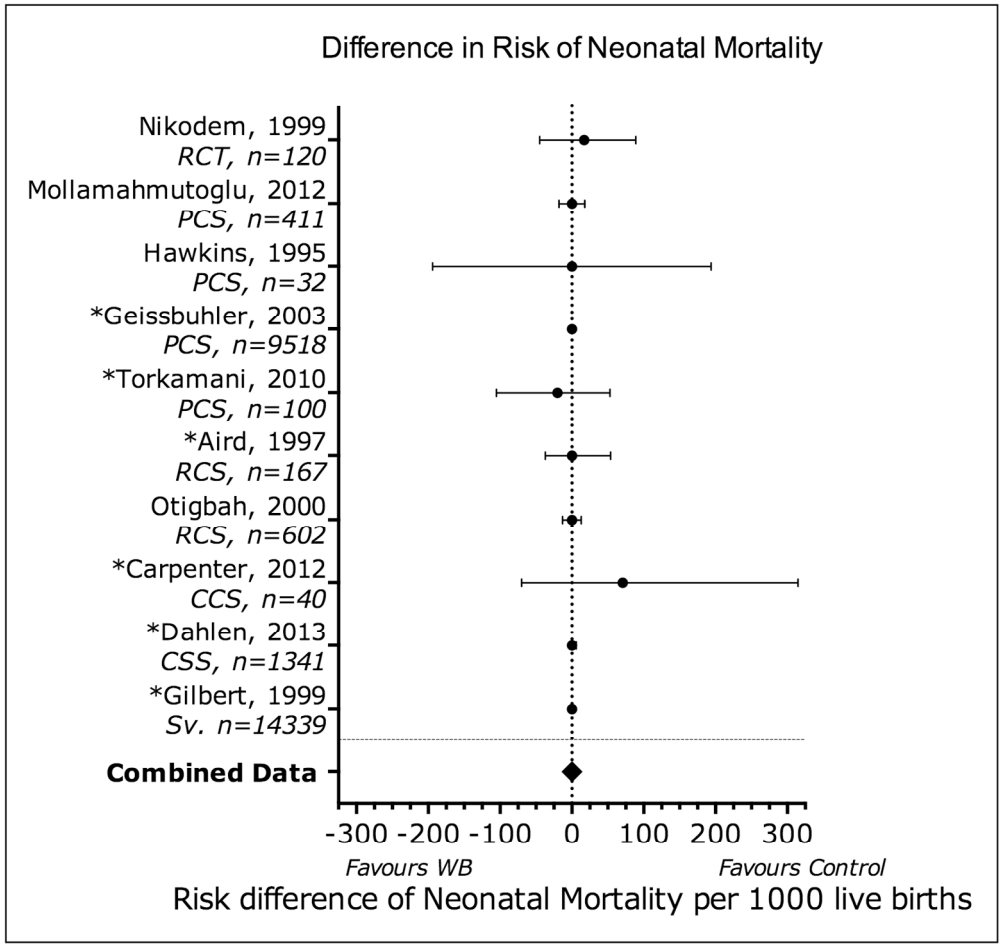
Table 2: Comparison of neonatal infection rate

Author, year	Study Design	n		Infection, n (%)		Risk difference % (95% CI)
		WB	Control	WB	Control	
Woodward et al. 2004.(40)	RCT	40	20	0	0	0% (-16, 8.8%)
Woodward et al. 2004.(40)	PCS	10	10	0	0	0% (-28, 28%)
Mollamahmutoglu et al. 2012.(55)	PCS	207	395	0	0	0% (-1, 1.8%)
Zanetti-Dällenbach et al. 2007.(46)	PCS	89	146	5[a] (5.6%)	2 (1.4%)	4.2% (-0.5, 11.2%)
Hawkins, 1995.(56)	PCS	16	16	3[b] (18.8%)	0	18.8% (-4.1, 43%)
Geissbühler et al., 2003.(41)	PCS	3617	5901	20 (0.55%)	60 (1.0%)	-0.5% (-0.8, -0.1%)
Bodner et al., 2002.(58)	RCS	140	140	0	2 (1.4%)	-1.4% (-5.1, 1.4%)
Otigbah et al. 2000.(47)	RCS	301	301	0	0	0% (-1.3, 1.3%)
Kowalewska et al. 2004.(61)	RCS	42	71	2[c] (4.8%)	5[d] (7.0%)	-2.3% (-11.4, 9.4%)
Pellantova et al. 2003. (62)	RCS	70	70	0	0	0% (-5.2, 5.2%)
Thoni et al. 2010.(65)	RCS	2625	899	26 (0.98%)	15 (1.64%)	-0.7% (-1.8, 0.1%)

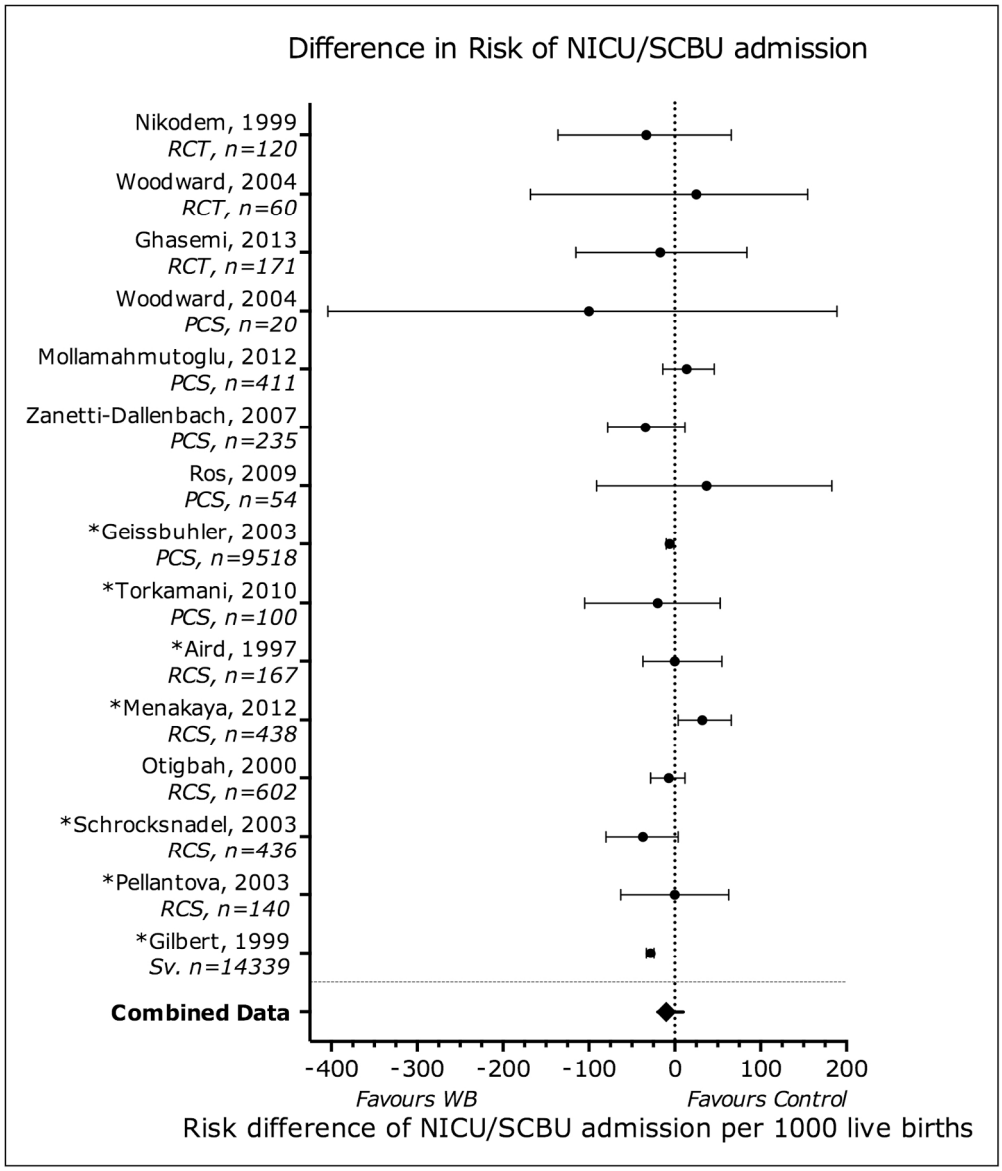
Key: Notable findings regarding specific infections, which were highlighted by study authors in discussion, included [a] five WB neonates developed conjunctivitis while no babies in the control group did; [b] one severe septicaemia following WB with pseudomonas, [c] one episode of aspiration pneumonia, and one pseudomonas skin infection following WB, [d] five intrauterine infections in the control group.



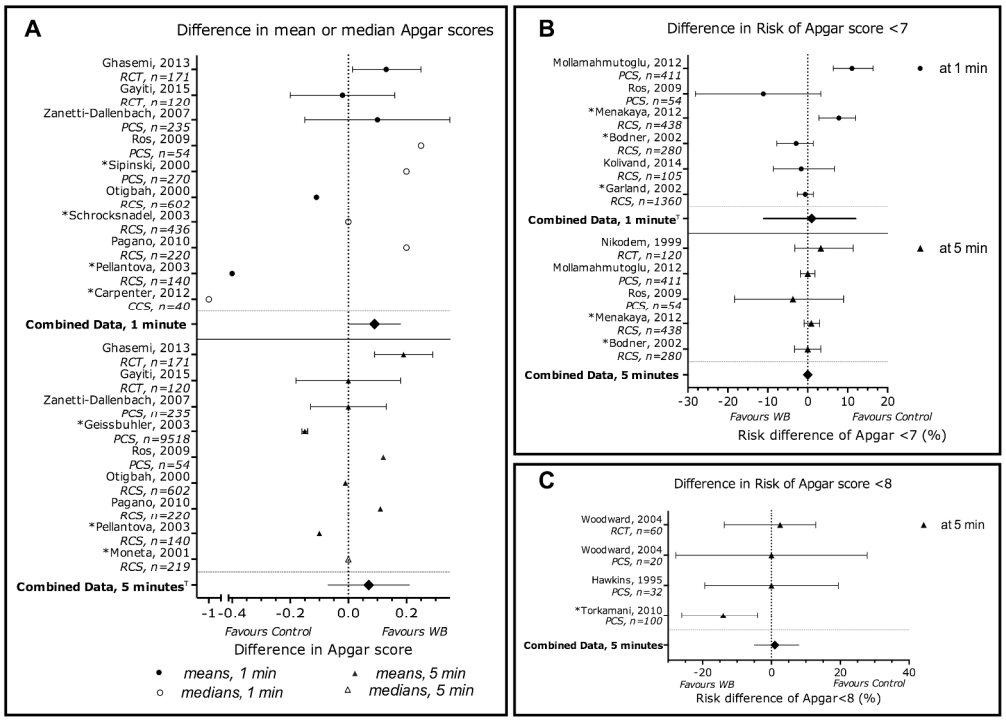
PRISMA flow chart of search results and paper selection
135x177mm (300 x 300 DPI)



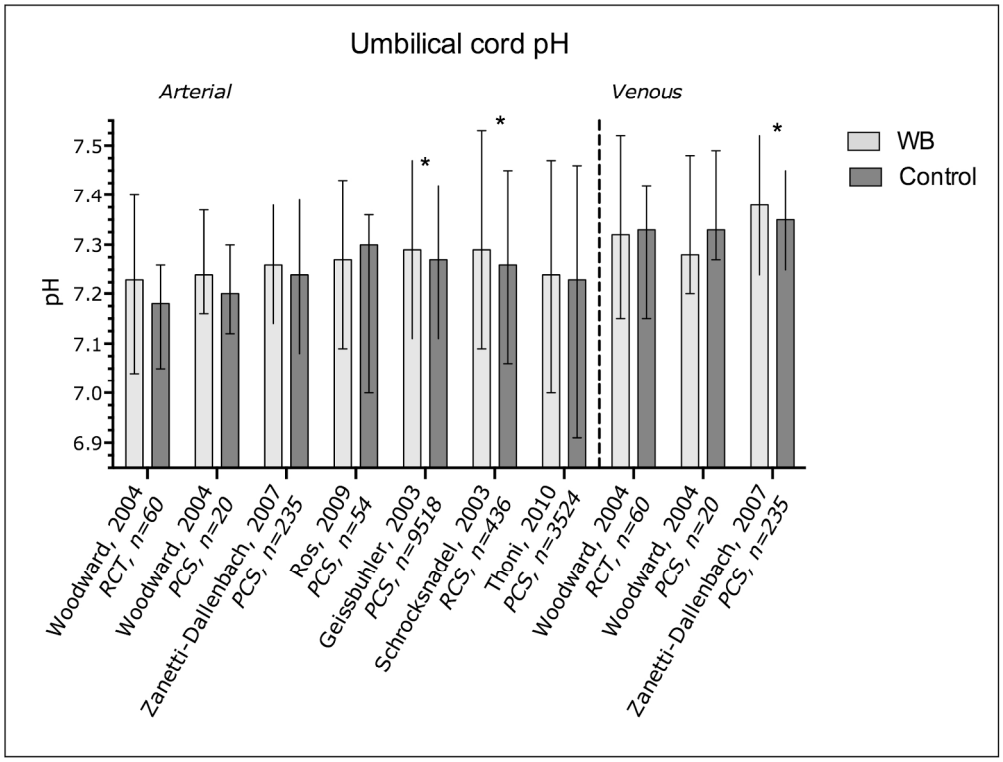
A forest plot showing the risk difference (RD) of neonatal mortality per thousand live births between the waterbirth (WB) and control group. RD and 95% confidence interval is plotted for each study. Studies with an asterisk (*) were not included in the meta-analysis (combined data).
147x139mm (300 x 300 DPI)



A forest plot showing the risk difference (RD) of NICU/SCBU admission per thousand live births between the waterbirth (WB) and control group. RD and 95% confidence interval is plotted for each study. Studies with an asterisk (*) were not included in the meta-analysis (combined data).
162x189mm (300 x 300 DPI)



Three forest plots comparing waterbirth (WB) and control groups. A: mean difference (with 95% CI) of Apgar score at one and five minutes. B: risk difference (RD) and 95% CI of having an Apgar score <7 at one and five minutes. C: RD and 95% CI of having an Apgar score <8 at five minutes. For each plot studies with an asterisk (*) were not included in the meta-analysis (combined data). A (T) symbol indicates significant heterogeneity (I2 >60%) in the combined data.
271x194mm (300 x 300 DPI)



A side by side bar chart comparing arterial or venous cord blood pH in waterbirth (light grey bars) and control (dark grey bars) groups. Median and range (bars with capped lines) or mean and calculated 95% range (bars with uncapped lines) are plotted for each study. An asterisk (*) indicates a statistically significant difference.
178x135mm (300 x 300 DPI)

1. Water
2. Underwater
3. Pool
4. Bath
5. Tub
6. Hydrotherapy
7. Baths
8. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7
9. Birth
10. Births
11. Delivery
12. Deliveries
13. Delivered
14. Labour
15. Labor
16. Intrapartum
17. Birthing
18. 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17
19. Waterbirth
20. Waterbirths
21. 19 OR 20
22. (8 AND 18) OR 21

Supplementary File B: Risk of bias assessment and Quality scoring

Modified Critical Appraisal Skills Programme (CASP) criteria

Randomised Controlled Trials (RCT):

1. **Does the trial have relevant extractable data? Yes/No**
   - Answer yes if all of the following are met:
     - Population = mothers at term (can be any risk level) and neonates not diagnosed as stillborn prior to onset of labour
     - Intervention = waterbirth (delivery of baby underwater)
     - Comparator = vaginal delivery into air
     - Outcome = One or more neonatal outcome reported

   Note if answer = No, study is excluded from systematic review as per inclusion and exclusion criteria

2. **Do the authors state that this is a randomised controlled trial, or that the allocation of participants was randomised? Yes/No**

Note if answer = No, study is not an RCT - determine study design and use appropriate tool

3. **Were participants appropriately allocated to intervention and control groups; was this process truly random, and was it valid? Yes/No/Can't tell**

Trials may stratify randomisation by age, parity or other acceptable variables. Differences in baseline characteristics between groups should be noted.

4. **Was the person performing data collection blinded? Yes /No/Can't tell**

5. **Were all participants accounted for at trial conclusion? Yes/No/Can't tell**

6. **Did study design avoid significant contamination between groups? Yes/No**

Note if per-protocol analysis was performed = No. If >25% of women in waterbirth group delivered on land score = No.

7. **Were follow up and data collection performed in the same way for intervention and control groups? Yes/No/Can't tell**

8. **Was a power calculation performed, and if so was the sample size sufficient to detect a difference in neonatal outcomes? Yes/No**

Note a study appropriately powered to detect a difference in a maternal outcome does not satisfy this criterion.

9. **Are results presented adequately? Yes/Partial/No**

Note that:

- *'No' indicates only written reporting of outcomes, such as 'no significant difference' without any numerical data.*
- *'Partial' indicates reporting of numerical data, such as a median APGAR, but without standard deviation, statistical analysis, or confidence intervals where they would have been of value.*
- *'Yes' indicates numerical data with standard deviations, statistical analysis and confidence intervals where appropriate.*

10. **Does this study report a range of neonatal outcomes? Yes/No**

Note: Yes = 3 or more neonatal outcomes. No =1 or 2

Scoring: Yes = 2, Partial = 1, Can't tell =0, No = 0

Supplementary File B: Risk of bias assessment and Quality scoring

Prospective Cohort Studies (PCS)

1. **Does the study have relevant extractable data? Yes/No**
   - Answer yes if all of the following are met:
     - Population = mothers at term (can be any risk level) and neonates not diagnosed as stillborn prior to onset of labour
     - Risk factor = waterbirth (delivery of baby underwater)
     - Outcome = One or more neonatal outcome reported

Note if answer = No, study is excluded from systematic review as per inclusion and exclusion criteria

2. **Were the same exclusion criteria for the waterbirth group applied to land birth group? Yes/No/Can't tell**

Note: women in control group should have same risk profile as waterbirth group

3. **Did women in land birth group receive equivalent intrapartum care? Yes/No/Can't tell**

4. **Were baseline characteristics of the two groups the same? If not was this accounted for in statistical analysis? Yes/No/Can't tell**

Note important confounding factors include maternal age, parity, maternal risk

5. **Was the person performing data collection blinded, or was data taken from a reliable source, e.g. the medical record? Yes/No/Can't tell**

6. **Were all participants accounted for at study conclusion? Yes/No/Can't tell**

7. **Were follow up and data collection performed in the same way for both groups? Yes/No/Can't tell**

8. **Are results presented adequately? Yes/Partial/No**

Note that:

- *'No' indicates only written reporting of outcomes, such as 'no significant difference' without any numerical data.*
- *'Partial' indicates reporting of numerical data, such as a median APGAR, but without standard deviation, statistical analysis, or confidence intervals where they would have been of value.*
- *'Yes' indicates numerical data with standard deviations, statistical analysis and confidence intervals where appropriate.*

9. **Does this study report a range of neonatal outcomes? Yes/No**

Note: Yes = 3 or more neonatal outcomes. No =1 or 2

Scoring: Yes = 2, Partial = 1, Can't tell =0, No = 0, Mixed = 0, Low = 0

Supplementary File B: Risk of bias assessment and Quality scoring

Retrospective Cohort Studies (RCS) and Case Control Studies (CCS)

1. **Does the study have relevant extractable data? Yes/No**
   - Answer yes if all of the following are met:
     - Population = mothers at term (can be any risk level) and neonates not diagnosed as stillborn prior to onset of labour
     - Risk factor = waterbirth (delivery of baby underwater)
     - Outcome = One or more neonatal outcome reported

Note if answer = No, study is excluded from systematic review as per inclusion and exclusion criteria

2. **Were the same exclusion criteria for the waterbirth group applied to control group? Yes/No/Can't tell**

Note: women in control group should have same risk profile as waterbirth group

3. **Did women in control group receive equivalent intrapartum care? Yes/No/Can't tell**

4. **Were baseline characteristics of the two groups the same? If not was this accounted for in statistical analysis? Yes/No/Can't tell**

Note important confounding factors include maternal age, parity, maternal risk

5. **Was data collected from a reliable source, e.g. the medical record? Yes/No/Can't tell**

6. **Were follow up and data collection performed in the same way for both groups? Yes/No/Can't tell**

7. **Are results presented adequately? Yes/Partial/No**

Note that:

- *'No' indicates only written reporting of outcomes, such as 'no significant difference' without any numerical data.*
- *'Partial' indicates reporting of numerical data, such as a median APGAR, but without standard deviation, statistical analysis, or confidence intervals where they would have been of value.*
- *'Yes' indicates numerical data with standard deviations, statistical analysis and confidence intervals where appropriate.*

8. **Does this study report a range of neonatal outcomes? Yes/No**

Note: Yes = 3 or more neonatal outcomes. No =1 or 2

Scoring: Yes = 2, Partial = 1, Can't tell =0, No = 0, Mixed = 0, Low = 0

Supplementary File B: Risk of bias assessment and Quality scoring

Cross Sectional Study or Surveillance Study

1. **Does the study have relevant extractable data?** **Yes/No**
   - Answer yes if all of the following are met:
     - Population = mothers at term (can be any risk level) and neonates not diagnosed as stillborn prior to onset of labour
     - Risk factor = waterbirth (delivery of baby underwater)
     - Outcome = One or more neonatal outcome reported

Note if answer = No, study is excluded from systematic review as per inclusion and exclusion criteria

2. **Are women in waterbirth group comparable to general population of women having waterbirth?** **Yes/No/Can't tell**

3. **Is the waterbirth group comparable to the control group?** **Yes/No/Can't tell**

Note: women in control group should have same risk profile as waterbirth group

4. **Did women in control group receive equivalent intrapartum care?** **Yes/No/Can't tell**

5. **Was data collected from a reliable source, e.g. the medical record?** **Yes/No/Can't tell**

6. **Was data collection performed in the same way for both groups?** **Yes/No/Can't tell**

7. **Are results presented adequately?** **Yes/Partial/No**

Note that:

- *'No' indicates only written reporting of outcomes, such as 'no significant difference' without any numerical data.*
- *'Partial' indicates reporting of numerical data, such as a median APGAR, but without standard deviation, statistical analysis, or confidence intervals where they would have been of value.*
- *'Yes' indicates numerical data with standard deviations, statistical analysis and confidence intervals where appropriate.*

8. **Does this study report a range of neonatal outcomes?** **Yes/No**

Note: Yes = 3 or more neonatal outcomes. No =1 or 2

Scoring: Yes = 2, Partial = 1, Can't tell =0, No = 0

Supplementary File B: Risk of bias assessment and Quality scoring

Quality Scores of Included Papers

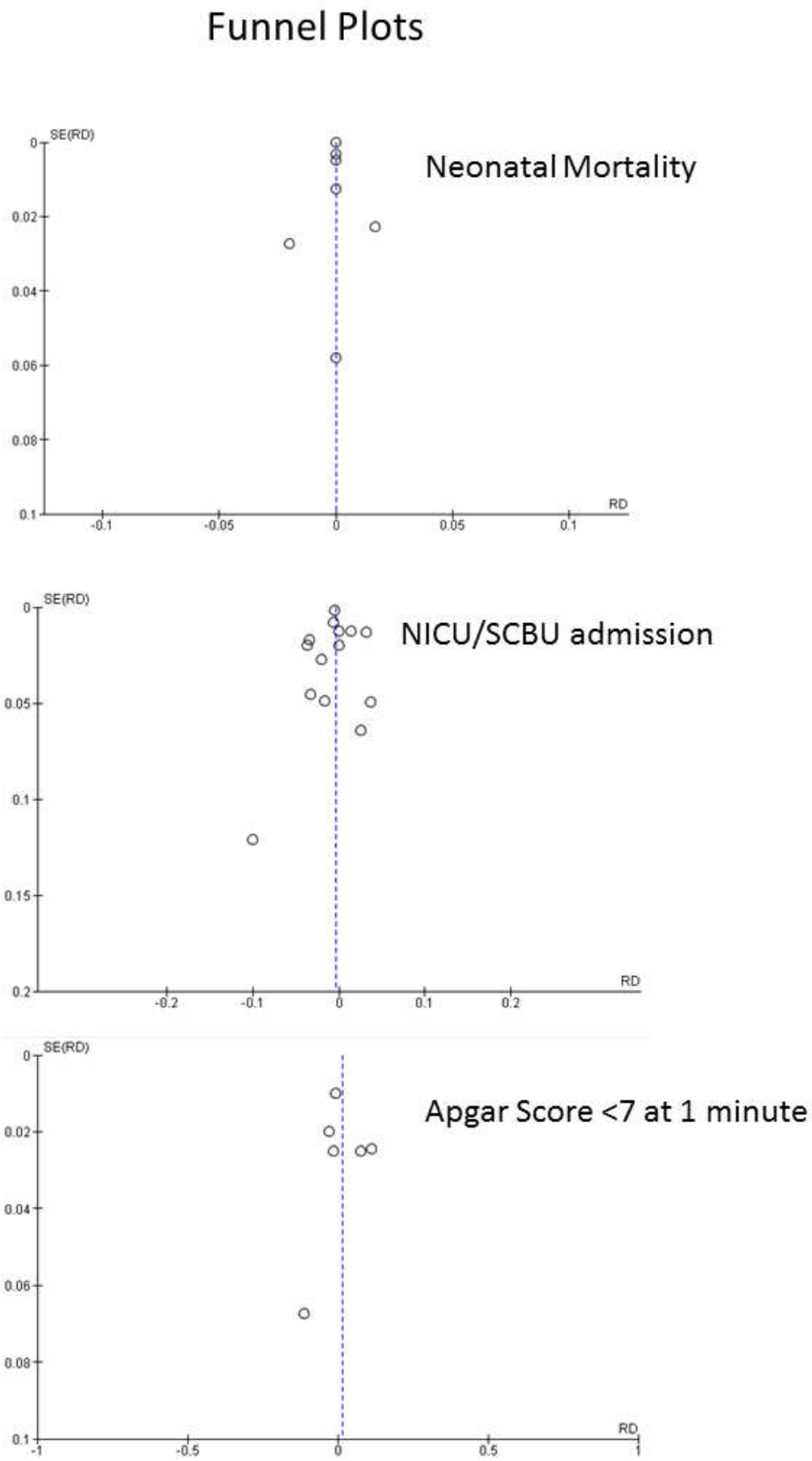
Author, Year, Country	Study Design	Modified CASP Criteria										
		1	2	3	4	5	6	7	8	9	10	Total
Nikodem, 1999.(48) South Africa	RCT	Y	Y	Y	N	Y	Y	Y	N	Y	Y	16
Woodward et al. 2004.(40) UK	RCT	Y	Y	Y	U	Y	N	Y	N	Y	Y	14
Ghasemi et al. 2013.(52) Iran	RCT	Y	Y	U	N	Y	Y	Y	N	Y	Y	14
Gayiti et al. 2015.(53) China	RCT	Y	Y	U	U	Y	Y	U	N	Y	N	10
Chaichian et al. 2009.(54) Iran	RCT	Y	Y	U	U	U	Y	Y	N	N	N	8
Woodward et al. 2004.(40) UK	PCS	Y	Y	U	Y	U	Y	Y	Y	Y		14
Mollamahmutoglu et al. 2012.(55) Turkey	PCS	Y	Y	U	N	N	Y	Y	Y	Y		12
Zanetti-Dällenbachet al. 2007.(46) Switzerland	PCS	Y	Y	N	N	N	Y	Y	Y	Y		12
Ros, 2009.(49) South Africa	PCS	Y	Y	U	U	N	Y	Y	P	Y		11
Hawkins, 1995.(56) UK	PCS	Y	Y	U	U	N	Y	Y	P	Y		11
Geissbühler et al. 2003.(41) Switzerland	PCS	Y	N	U	N	U	Y	Y	Y	Y		10
Torkamani et al. 2010.(44) Iran	PCS	Y	U	U	N	N	Y	Y	P	Y		9
Sipinksi et al. 2000.(57) Poland	PCS	Y	U	U	U	U	U	U	Y	N		4
Menakaya et al. 2012.(43) Australia	RCS	Y	Y	Y	Y	Y	Y	Y	N			14
Bodner et al. 2002.(58) Austria	RCS	Y	Y	Y	Y	N	Y	Y	Y			14
Otigbah et al. 2000.(47) UK	RCS	Y	Y	U	Y	Y	Y	Y	Y			14
Schröcksnadel et al. 2003.(45) Austria	RCS	Y	U	N	U	Y	Y	Y	N			10
Kolivand et al. 2014.(59) Iran	RCS	Y	Y	U	Y	Y	Y	P	N			10
Pagano et al. 2010.(60) Italy	RCS	Y	U	U	Y	U	Y	Y	N			8
Aird et al. 1997.(63) UK	RCS	Y	N	Y	Y	U	U	N	N			6

Supplementary File B: Risk of bias assessment and Quality scoring

Kowalewska et al. 2004.(61) Poland	RCS	Y	U	U	N	Y	Y	P	N			7
Pellantova et al. 2003. (62) Czech Republic	RCS	Y	Y	U	N	U	U	P	Y			7
Burke et al. 1995.(64) UK	RCS	Y	N / U	U	U	Y	Y	N	N			6
Thoni et al. 2010.(65) Italy	RCS	Y	U	U	U	U	U	Y	Y			6
Garland et al. 2002.(66) UK	RCS	Y	N	U	Y	U	U	P	N			5
Moneta et al. 2001.(67) Poland	RCS	Y	U	U	N	U	U	N	N			2
Carpenter et al. 2012.(68) New Zealand	Case control	Y	U	Y	Y	Y	Y	Y	Y			14
Dahlen et al. 2013.(69) Australia	Cross sectional	Y	Y	U	Y	N	Y	P	N			9
Gilbert et al. 1999.(42) UK	Surveillance	Y	Y	U	U	N	N	Y	Y			8

Supplementary File B: Risk of bias assessment and Quality scoring

Funnel Plots of selected outcomes



Supplementary File C. Tabulated data on neonatal outcomes

Outcomes of Waterbirth: Neonatal Mortality

Study	Study design	n		Mortality, n (%)		Risk Difference /1000 (95% CI)
		WB	Control	WB	Control	
Nikodem, 1999.(48)	RCT	60	60	1 (1.7%)	0	17 (-45, 89)
Mollamahmutoglu et al. 2012.(55)	PCS	207	204	0	0	0 (-18, 18)
Hawkins, 1995.(56)	PCS	16	16	0	0	0 (-194, 194)
Geissbühler et al. 2003.(41)*	PCS	3617	5901	0	0	0 (-1, 1)
Torkamani et al. 2010.(44)*	PCS	50	50	0	1 (2.0%)	-20 (-105, 53)
Otigbah et al. 2000.(47)	RCS	301	301	0	0	0 (-13, 13)
Aird et al. 1997.(63)*	RCS	67	100	0	0	0 (-37, 54)
Carpenter et al. 2012.(68)*	CCS	14	26	1 (7.1%)	0	71 (-70, 315)
Dahlen et al. 2013.(69)*	Cross Sectional	819	5220	1 (0.12%)	4 (0.077%)	0 (-1, 6)
Gilbert et al. 1999.(42)*	Surveillance	4032	10307	5 (0.12%)	14 (0.14%)	0 (-1, 2)
Combined Data		584	581	1	0	0 (-1, 1) Heterogeneity: Tau^2= 0, I^2= 0%

Key: *studies not included in meta-analysis (combined data)

Outcomes of Waterbirth: SCBU/ NICU Admission

Author, Year	Study Design	n		SCBU/NICU, n (%)		Risk Difference /1000 (95% CI)
		WB	Control	WB	Control	
Nikodem, 1999.(48)	RCT	60	60	3 (5.0%)	5 (8.3%)	-33 (-136, 66)
Woodward et al. 2004.(40)	RCT	40	20	3 (7.5%)	1 (5.0%)	25 (-168, 155)
Ghasemi et al. 2013.(52)	RCT	83	88	9 (10.8%)	11 (12.5%)	-17 (-115, 84)
Woodward et al. 2004.(40)	PCS	10	10	0	1 (10%)	-100 (-404, 189)
Mollamahmutoglu et al. 2012.(55)	PCS	207	204	5 (2.5%)	2 (1%)	14 (-14, 46)
Zanetti-Dällenbach et al. 2007.(46)	PCS	89	146	0	5 (3.4%)	-34 (-78, 12)
Ros, 2009.(49)	PCS	27	27	1 (3.7%)	0	37 (-91, 183)
Geissbühler et al. 2003.(41)*	PCS	3617	5901	25 (0.69%)	74 (1.3%)	-6 (-10, 1)
Torkamani et al. 2010.(44)*	PCS	50	50	0	1 (2%)	-20 (-105, 53)
Menakaya et al. 2012.(43)*	RCS	219	219	8 (3.7%)	1 (0.45%)	32 (4 , 66)
Otigbah et al. 2000.(47)	RCS	301	301	2 (0.66%)	4 (1.3%)	-7 (-28, 12)
Schröcksnadel et al. 2003.(45)*	RCS	218	218	6 (2.7%)	14 (5.3%)	-37 (-80, 4)
Pellantova et al. 2003. (62)*	RCS	70	70	1	1	0 (-63, 63)
Aird et al. 1997.(63)*	RCS	67	100	0	0	0 (-37, 54)
Gilbert et al. 1999.(42)	Surveillance	4032	10307	34 (0.84%)	380 (3.7%)	-28 (-33,-24)
Combined Data		817	856	23	29	-1 (-2, 1) Heterogeneity: Tau^2=0, I^2=3%

Outcomes of Waterbirth: Apgar Scores

Studies reporting numerical Apgar scores

Author, year	Study Design	n		Apgar 1 min			Apgar 5 min		
		WB	Control	WB	Control	Difference	WB	Control	Difference
Ghasemi et al. 2013.(52)	RCT	83	88	8.94 (±0.23)	8.81 (±0.49)	0.13 (0.014, 0.25)	9.21 (±0.44)	9.02 (±0.14)	0.19 (0.09, 0.29)
Gayiti et al., 2015.(53)	RCT	60	60	9.26 (±0.51)	9.28 (±0.47)	-0.02 (-0.2, 0.16)	9.34 (±0.49)	9.34 (±0.52)	0 (-0.18, 0.18)
Zanetti-Dällenbach et al. 2007.(46)	PCS	89	146	8.7(±0.8)	8.6 (±1.0)	0.1 (-0.15, 0.35)	9.8 (±0.5)	9.8 (±0.5)	0 (-0.13, 0.13)
Ros, 2009.(49)*	PCS	27	27	8.4 (7-9)	8.15 (3-9)	0.25	8.93 (8-9)	8.81 (6-9)	0.12
Geissbühler et al., 2003.(41)*	PCS	3617	5901				9.83 (±0.43)	9.98 (±0.27)	-0.15 (-0.16, -0.14)
Sipinksi et al. 2000.(57)*	PCS	135	135	9.8 (7-10)	9.6 (6-10)	0.2			
Otigbah et al. 2000.(47)*	RCS	301	301	8.4	8.51	-0.11	9.57	9.58	-0.01
Schröcksnadel et al. 2003.(45)*	RCS	218	218				10 (6-10)	10 (5-10)	0
Pagano et al. 2010.(60)*	RCS	110	110	9.48	9.28	0.2	9.95	9.84	0.11
Pellantova et al. 2003. (62)*	RCS	70	70	8.5	8.9	-0.4	9.6	9.7	-0.1
Moneta et al. 2001.(67)*	RCS	109	110				10	10	0
Carpenter et al. 2012.(68)*	CCS	14	26	7	8	-1.0			
Combined Data		232	294	Combined difference: 0.09 (0, 0.18) Heterogeneity: Tau^2 =0.0, I^2=0%			Combined difference: 0.07 (-0.07, 0.21) Heterogeneity: Tau^2 =0.01, I^2=69%		

Data presented as mean (± standard deviation), median (range), or difference (95% confidence interval). Key: *study not included in meta-analysis

Studies reporting number in neonates with Apgar score <7

Author, year	Study Design	N		Apgar 1 min <7			Apgar 5 min <7		
		WB	Control	WB	Control	$RD_{\%}$	WB	Control	$RD_{\%}$
Nikodem, 1999.(48)	RCT	60	60				2 (3.3%)	0	3.3% (-3.2, 11.4%)
Mollamahmutoglu et al. 2012.(55)	PCS	207	204	26 (12.6%)	3 (1.5%)	11.1% (6.4, 16.4%)	0	0	0% (-1.8, 1.8%)
Ros, 2009.(49)	PCS	27	27	0	3 (11.1%)	-11.1% (-28.1, 3.3%)	0	1 (3.7%)	-3.7% (-18.3, 9.1%)
Menakaya et al. 2012.(43)*	RCS	219	219	25 (11.4%)	8 (3.7%)	7.8% (2.8, 13%)	2 (1%)	0	0.9% (-0.9, 3%)
Bodner et al., 2002.(58)*	RCS	140	140	2 (1.4%)	6 (4.3%)	-2.9% (-7.7, 1.4%)	1 (0.71%)	1 (0.71%)	0% (-3.3, 3.3%)
Kolivand et al. 2014.(59)	RCS	43	62	0	1 (1.6%)	-1.6% (-8.6, 6.7%)			
Garland et al. 2002.(66)*	RCS	680	680	22 (3.2%)	26 (3.8%)	-0.6% (-2.6, 1.4%			
Combined Data				Events: 26 Total: 277	Events: 7 Total: 293	$RD_{\%}$: 1% (-11, 12%)	Events: 2 Total: 294	Events: 1 Total: 291	0% (-1, 1%)
				Heterogeneity: Tau^2=0.01 , I^2= 89%			Heterogeneity: Tau^2= 0.0, I^2= 0%		

Data presented as n (%), Risk difference percentage ($RD_{\%}$) (95% confidence interval). Key: *study not included in meta-analysis

Studies reporting number in neonates with Apgar score <8

Author, year	Study Design	N		Apgar 1 min <8			Apgar 5 min <8		
		WB	Control	WB	Control	$RD_{\%}$	WB	Control	$RD_{\%}$
Woodward et al. 2004.(40)	RCT	40	20				1 (2.5%)	0	2.5% (-13.7, 12.9%)
Woodward et al. 2004.(40)	PCS	10	10				0	0	0% (-27.8, 27.8%)
Hawkins, 1995.(56)	PCS	16	16				0	0	0% (-19.4, 19.4%)
Torkamani et al. 2010.(44)*	PCS	50	50				0	7 (14.6%)	-14% (-26.2, -4%)
Combined Data							Events: 1 Total: 66	Events: 0 Total: 46	$RD_{\%}$: 1% (-5, 8%)
							Heterogeneity: Tau^2= 0.0, I^2= 0%		

Data presented as n (%), Risk difference percentage ($RD_{\%}$) (95% confidence interval). Key: *study not included in meta-analysis

Studies with descriptive report of Apgar scores

Author, year	Study Design	n		Descriptive reports of Apgar scores
		WB	Control	
Chaichian et al., 2009.(54)*	RCT	53	53	"No difference" in 1 min or 5 min Apgar
Burke et al. 1995.(64)*	RCS	50	50	"Mean Apgar scores were the same for both groups"
Dahlen et al., 2013.(69)*	CSS	819	5220	"No significant difference" in 5 min Apgar

Key: *study not included in meta-analysis

Outcomes of Waterbirth: Umbilical Cord Gases

Author, year	Design	n		Arterial pH			Venous pH		
		WB	Control	WB	Control	Analysis	WB	Control	Analysis
Woodward et al. 2004.(40)	RCT	40	20	Median: 7.23 Range: 7.04-7.40 (n=35)	Median: 7.18 Range: 7.05-7.26 (n=13)	NS	Median: 7.32 Range: 7.15-7.52 (n=36)	Median: 7.33 7.15-7.42) (n=16)	NS
Woodward et al. 2004.(40)	PCS	10	10	Median: 7.24 Range: 7.16-7.37 (n=7)	Median: 7.2 Range: 7.12-7.3 (n=7)		Median: 7.28 Range: 7.2-7.48 (n=5)	Median: 7.33 Range: 7.27-7.49 (n=10)	
Zanetti-Dällenbach et al. 2007.(46)	PCS	89	146	Mean: 7.26 (±0.06) 95% range: 7.14-7.38	Mean: 7.24 (±0.08) 95% range: 7.08-7.39	MD: 0.02 (0.0, 0.04)	Mean: 7.38 (±0.07) 95% range: 7.24-7.52	Mean: 7.35 (±0.05) 95% range: 7.25-7.45	MD: 0.03 (0.015, 0.045)
Ros, 2009.(49)	PCS	27	27	Median: 7.27 Range: 7.09-7.43 (n=21)	Median: 7.30 Range: 7.0-7.36 (n=18)				
Geissbühler et al., 2003.(41)	PCS	3617	5901	Mean: 7.29 (±0.09) 95% range: 7.11-7.47	Mean: 7.27 (±0.08) 95% range: 7.11-7.42	MD: 0.02 (0.017, 0.023)			
Schröcksnadel et al. 2003.(45)	RCS	218	218	Median: 7.29 Range: 7.09–7.53	Median: 7.26 Range: 7.06-7.45	p = 0.001			
Thoni et al. 2010.(65)	RCS	2625	899	Median: 7.24 Range: 7.0-7.47 (n=1826)	Median: 7.23 Range: 6.91-7.46 (n=1334)	NS			
				Arterial pH <7.1		$RD_{\%}$			
Bodner et al., 2002.(58)	RCS	140	140	3 (2.1%)	4 (2.9%)	-0.7% (-5.2, 3.6%)			
				Arterial pH <7.2					
Nikodem, 1999.(48)	RCT	60	60	12 (n=57)	14 (n=59)	-2.7% (-17.6, 12.5%)			

Data presented as mean (± standard deviation), or n (%). When cord gas analysis is not performed on the whole group then n is given beneath the data (n=).
Key: MD = mean difference (95% confidence interval); $RD_{\%}$= risk difference percentage (95% confidence interval); NS = not significant; p = p value. NS and p values are taken from non-parametric significance tests performed by original study authors.

Supplementary File D. Sensitivity Analyses

Outcome	Primary analysis			Sensitivity Analyses					
				High Quality Score Only			RCT only		
	n studies	n births	Outcome (95% CI)	n studies	n births	Outcome (95% CI)	n studies	n births	Outcome (95% CI)
Neonatal Mortality, RD_{1000}	4	1165	0 (-10, 10) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 0\%$	4	1165	0 (-10, 10) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 0\%$	1	120	20 (-30, 60) Heterogeneity: Not applicable
NICU-SCBU admission, RD_{1000}	8	1673	10 (-20, 10) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 3\%$	9	2111	0 (-20, 20) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 40\%$	3	351	-10 (-70, 40) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 0\%$
Apgar score at 1 min, mean difference	3	526	-0.09 (0, 0.18) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 0\%$	2	406	0.12(0.02, 0.23) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 0\%$	2	291	0.07 (-0.07, 0.21) Heterogeneity: $Tau^2 = 0.1$, $I^2 = 49\%$
Apgar score at 5 min, mean difference	3	526	0.07 (-0.07, 0.21) Heterogeneity: $Tau^2 = 0.01$, $I^2 = 69\%$	2	406	0.1 (-0.09, 0.29) Heterogeneity: $Tau^2 = 0.1$, $I^2 = 80\%$	2	291	0.11 (-0.07, 0.29) Heterogeneity: $Tau^2 = 0.1$, $I^2 = 69\%$
Apgar score <7 at 1 min, RD %	3	570	1 (-11, 12) Heterogeneity: $Tau^2 = 0.01$, $I^2 = 89\%$	4	1183	1 (-6, 11) Heterogeneity: $Tau^2 = 0.01$, $I^2 = 89\%$	0		
Apgar score <7 at 5 min, RD %	3	585	0 (-1, 1) Heterogeneity: $Tau^2 = 0.00$, $I^2 = 0\%$	5	1303	0 (0, 1) Heterogeneity: $Tau^2 = 0.00$, $I^2 = 0\%$	1	120	3 (-2, 9) Heterogeneity: Not applicable
Apgar score <8 at 5 min, RD %	3	112	1 (-5, 8) Heterogeneity: $Tau^2 = 0.0$, $I^2 = 0\%$	3	112	1 (-5, 8) Heterogeneity: $Tau^2 = 0.00$, $I^2 = 0\%$	1	60	3 (-6, 11) Heterogeneity: Not applicable

Key: RD_{1000} = risk difference per 1000 live births, RD % = risk difference percentage, CI = confidence interval

Supplementary File E. Excluded papers

Author, Year	Reason for Exclusion
Burns et al. 2012 (29)	Not a comparative study.
Henderson et al. 2014 (30)	Data is not separated for WI and WB.
Rush 1999 (31)	Data is not separated for WI and WB.
Da Silva et al. 2006 (32)	Data is not separated for WI and WB.
Geissbühler et al. 2004 (33)	There are multiple publications by the same author describing the same cohort. In addition to the excluded paper cited here, three earlier papers were also excluded. All papers describing this cohort were reviewed. The 2003 paper contained all relevant extractable data for the largest number of births and was selected for inclusion.(41)
Zanetti-Dällenbach et al. 2006 (34)	There are multiple publications by the same author describing the same cohort. In addition to the excluded paper cited here, two other papers were also excluded. All papers describing this cohort were reviewed. One of the 2007 papers contained all relevant extractable data and was selected for inclusion.(46)
Thoni et al, 2007 (35)	There are multiple publications by the same author describing the same cohort. In addition to the excluded paper cited here, three other papers were also excluded. All papers describing this cohort were reviewed. The most recent paper (2010) described all relevant extractable data for the largest number of births and was selected for inclusion.(65)
Damodaran et al. 2010 (36)	Conference abstract with no extractable qualitative data.
Lim et al. 2015 (37)	Conference abstract with no extractable qualitative data.
Ziolkowski et al. 2009 (38)	Paper is not available in any UK reference library and is unavailable online. Unable to contact author.
Price 1995 (39)	PhD thesis not available in any UK reference library and unavailable online. Unable to contact author.