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

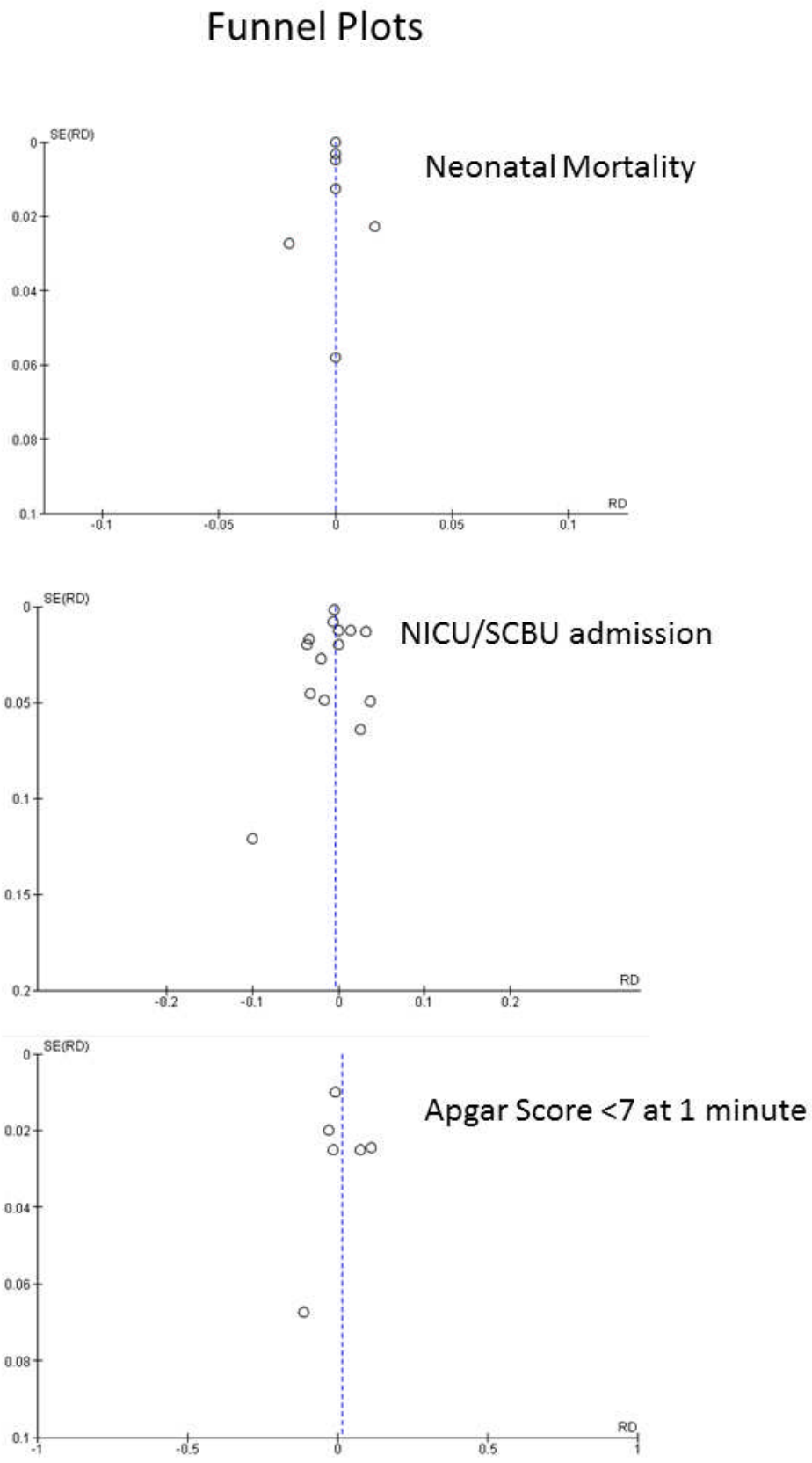
Author, Year, Country	Study Design	Modified CASP Criteria										
		1	2	3	4	5	6	7	8	9	10	Total
Nikodem, 1999.(48) South Africa	RCT	Y	Y	Y	N	Y	Y	Y	N	Y	Y	16
Woodward et al. 2004.(40) UK	RCT	Y	Y	Y	U	Y	N	Y	N	Y	Y	14
Ghasemi et al. 2013.(52) Iran	RCT	Y	Y	U	N	Y	Y	Y	N	Y	Y	14
Gayiti et al. 2015.(53) China	RCT	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.