

IQWiG-Berichte – Nr. 593

# **Evidenzbericht zur Modifizierung der Umgebungsgestaltung**

## **Rapid Report**

Auftrag: V16-01B  
Version: 1.0  
Stand: 30.01.2018

# Impressum

**Herausgeber:**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

**Thema:**

Evidenzbericht zur Modifizierung der Umgebungsgestaltung

**Auftraggeber:**

Bundesministerium für Gesundheit

**Datum des Auftrags:**

14.09.2016

**Interne Auftragsnummer:**

V16-01B

**Anschrift des Herausgebers:**

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Im Mediapark 8

50670 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

E-Mail: [berichte@iqwig.de](mailto:berichte@iqwig.de)

Internet: [www.iqwig.de](http://www.iqwig.de)

**ISSN: 1864-2500**

Dieser Bericht wurde ohne die Beteiligung externer Sachverständiger erstellt.

Für die Inhalte des Berichts ist allein das IQWiG verantwortlich.

**Mitarbeiter des IQWiG**

- Nina Peterwerth
- Susanne Unverzagt
- Elke Hausner
- Corinna Kiefer
- Claudia Mischke
- Wiebke Sieben

**Schlagwörter:** Termingerechte Geburt, Entbindungsräume, Systematische Übersicht

**Keywords:** Term Birth, Delivery Rooms, Systematic Review

# Inhaltsverzeichnis

	Seite
<b>Tabellenverzeichnis</b> .....	<b>vi</b>
<b>Abbildungsverzeichnis</b> .....	<b>vii</b>
<b>Abkürzungsverzeichnis</b> .....	<b>viii</b>
<b>Kurzfassung</b> .....	<b>ix</b>
<b>1 Hintergrund</b> .....	<b>1</b>
<b>2 Fragestellung</b> .....	<b>3</b>
<b>3 Projektverlauf</b> .....	<b>4</b>
<b>4 Methoden</b> .....	<b>5</b>
<b>4.1 Kriterien für den Einschluss von Studien zur modifizierten     Umgebungsgestaltung in die Untersuchung</b> .....	<b>5</b>
4.1.1 Population.....	5
4.1.2 Prüf- und Vergleichsintervention .....	5
4.1.3 Endpunkte.....	5
4.1.4 Studientypen .....	7
4.1.5 Studiendauer .....	7
4.1.6 Publikationszeitraum .....	7
4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss .....	8
4.1.8 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen.....	8
<b>4.2 Umfassende Informationsbeschaffung</b> .....	<b>8</b>
4.2.1 Primäre Informationsquellen .....	8
4.2.1.1 Bibliografische Datenbanken.....	8
4.2.1.2 Studienregister .....	9
4.2.1.3 Weitere Informationsquellen und Suchtechniken.....	9
4.2.1.3.1 Studiengruppen und Fachgesellschaften .....	9
4.2.1.3.2 Weitere Suchtechniken.....	9
4.2.1.4 Autorenanfragen .....	9
4.2.2 Selektion relevanter Studien.....	9
<b>4.3 Informationsextraktion und -bewertung auf Studienebene</b> .....	<b>10</b>
4.3.1 Datenextraktion .....	10
4.3.2 Bewertung des Verzerrungspotenzials der Ergebnisse.....	10
4.3.3 Bewertung des Evidenzniveaus aller eingeschlossenen Studien.....	11
<b>4.4 Informationssynthese und -analyse</b> .....	<b>12</b>

4.4.1	Gegenüberstellung der Ergebnisse der Einzelstudien .....	12
4.4.2	Metaanalysen .....	12
4.4.3	Aussagen zur Qualität der Evidenz .....	13
4.4.4	Sensitivitätsanalysen .....	15
4.4.5	Subgruppenmerkmale und andere Effektmodifikatoren .....	16
4.4.6	Evidenzprofil .....	17
<b>5</b>	<b>Ergebnisse.....</b>	<b>18</b>
<b>5.1</b>	<b>Umfassende Informationsbeschaffung.....</b>	<b>18</b>
5.1.1	Bibliografische Datenbanken .....	18
5.1.2	Studienregister .....	19
5.1.3	Weitere Informationsquellen und Suchtechniken .....	20
5.1.3.1	Studiengruppen und Fachgesellschaften.....	20
5.1.3.2	Anwendung weiterer Suchtechniken .....	20
5.1.3.3	Autorenanfragen .....	20
5.1.3.4	Zusätzliche relevante Studien bzw. Dokumente.....	22
5.1.4	Resultierender Studienpool .....	22
5.1.5	Studien ohne berichtete Ergebnisse.....	22
<b>5.2</b>	<b>Charakteristika der in die Bewertung eingeschlossenen Studien.....</b>	<b>23</b>
5.2.1	Studiendesign und Studienpopulationen .....	23
5.2.2	Übersicht der bewertungsrelevanten Endpunkte.....	29
<b>5.3</b>	<b>Einschätzung des Verzerrungspotenzials .....</b>	<b>31</b>
5.3.1	Einschätzung des Verzerrungspotenzials auf Studienebene .....	31
5.3.2	Einschätzung des Verzerrungspotenzials auf Endpunktebene .....	32
<b>5.4</b>	<b>Ergebnisse zu kritischen Endpunkten.....</b>	<b>33</b>
5.4.1	Geburtsmodus.....	33
5.4.2	Geburtserleben.....	34
5.4.3	Apgar-Score nach 5 Minuten < 6 .....	35
<b>5.5</b>	<b>Ergebnisse zu wichtigen Endpunkten .....</b>	<b>35</b>
5.5.1	Wehenmittel .....	35
<b>6</b>	<b>Evidenzprofil für die kritischen und wichtigen Endpunkte .....</b>	<b>36</b>
<b>6.1</b>	<b>Geburtsmodus .....</b>	<b>43</b>
<b>6.2</b>	<b>Geburtserleben .....</b>	<b>44</b>
<b>6.3</b>	<b>Wehenmittel.....</b>	<b>45</b>
<b>6.4</b>	<b>Gesamtbeurteilung der Qualität der Evidenz für alle Endpunkte.....</b>	<b>45</b>
<b>7</b>	<b>Diskussion.....</b>	<b>47</b>
<b>8</b>	<b>Fazit.....</b>	<b>49</b>
<b>9</b>	<b>Literatur .....</b>	<b>50</b>

<b>10 Studienlisten .....</b>	<b>56</b>
<b>10.1 Liste der eingeschlossenen Studien.....</b>	<b>56</b>
<b>10.2 Liste der gesichteten systematischen Übersichten .....</b>	<b>56</b>
<b>10.3 Liste der ausgeschlossenen Publikationen zum Thema mit         Ausschlussgründen.....</b>	<b>56</b>
<b>Anhang A – Suchstrategien .....</b>	<b>57</b>
A.1 – Bibliografische Literaturrecherche .....	57
A.2 – Suche in Studienregistern .....	60
<b>Anhang B – Verzerrungspotenzial auf Endpunktebene.....</b>	<b>62</b>
B.1 – Geburtsmodus .....	62
B.2 – Geburtserleben .....	62
B.3 – Wehenmittel.....	63
<b>Anhang C – Ergebnisse.....</b>	<b>64</b>
C.1 – Geburtsmodus .....	64
C.2 – Geburtserleben .....	66
C.3 – Ergänzende Informationen zum Apgar-Score nach 5 Minuten < 6 .....	68
C.4 – Wehenmittel.....	69

**Tabellenverzeichnis**

	<b>Seite</b>
Tabelle 1: Durch die Leitliniengruppe konsentierete Endpunkte für vergleichende Interventionsstudien und ihre Bewertung .....	6
Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss.....	8
Tabelle 3: Übersicht über Autorenanfragen .....	21
Tabelle 4: Zusätzlich identifizierte relevante Studie bzw. identifiziertes Dokument .....	22
Tabelle 5: Studienpool der Evidenzbewertung .....	22
Tabelle 6: In der Informationsbeschaffung identifizierte Studien ohne berichtete Ergebnisse .....	22
Tabelle 7: Charakteristika der eingeschlossenen Studien: Evidenztabelle .....	26
Tabelle 8: Matrix der kritischen und klinisch wichtigen Endpunkte .....	30
Tabelle 9: Bewertung des Verzerrungspotenzials auf Studienebene .....	32
Tabelle 10: Zusammenfassung der Ergebnisse zu den kritischen Endpunkten .....	33
Tabelle 11: Zusammenfassung der Ergebnisse zu den wichtigen Endpunkten .....	35
Tabelle 12: GRADE-Evidenzprofil.....	37
Tabelle 13: Bewertung des Verzerrungspotenzials auf Endpunktebene: Geburtsmodus .....	62
Tabelle 14: Bewertung des Verzerrungspotenzials auf Endpunktebene: Geburtserleben .....	62
Tabelle 15: Bewertung des Verzerrungspotenzials auf Endpunktebene: Wehenmittel intrapartal .....	63
Tabelle 16: Ergebnisse zum Geburtsmodus .....	65
Tabelle 17: Ergebnisse Geburtserleben (Teilaspekt Wirkung des Gebärraumes).....	66
Tabelle 18: Ergebnisse zum Geburtserleben (Teilaspekt Unterstützung durch das Pflegepersonal).....	66
Tabelle 19: Ergebnisse zum Geburtserleben (Teilaspekt Qualität der Pflege aus der Patientenperspektive) .....	67
Tabelle 20: Ergänzende Informationen zum Apgar-Score nach 5 Minuten < 6 (Apgar-Score nach 5 Minuten < 7) .....	68
Tabelle 21: Ergänzende Informationen für den Apgar-Score nach 5 Minuten .....	68
Tabelle 22: Ergebnisse für die Häufigkeit einer Wehenmittelgabe während der Geburt .....	69

## Abbildungsverzeichnis

### Seite

Abbildung 1: Ergebnis der bibliografischen Recherche und der Studienselektion..... 19

**Abkürzungsverzeichnis**

<b>Abkürzung</b>	<b>Bedeutung</b>
Apgar-Score	Schema zur Beurteilung der postpartalen Adaptation und Vitalität des reifen Neugeborenen nach den Kriterien Atmung, Puls, Grundtonus, Aussehen und Reflexe
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BMG	Bundesministerium für Gesundheit
CONSORT	Consolidated Standards of Reporting Trials (konsolidierte Standards zum Berichten von Studien)
CTG	Kardiotokogramm
DAMI	Design Algorithm for Medical Literature on Intervention (Designklassifizierung für medizinische Interventionsstudien)
DGGG	Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V.
DGHWi	Deutsche Gesellschaft für Hebammenwissenschaft e. V.
EbM	Evidence-based Medicine (evidenzbasierte Medizin)
EP	Eröffnungsperiode
G-BA	Gemeinsamer Bundesausschuss
GRADE	Grading of Recommendations, Assessment, Development, and Evaluation (Bewertung von Empfehlungen, Einschätzung, Entwicklung und Auswertung)
IQTIG	Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention to treat
KI	Konfidenzintervall
OCEBM	Oxford Centre for Evidence-based Medicine (Oxford-Zentrum für evidenzbasierte Medizin)
PICO	Participants Intervention Comparison Outcomes (Akronym für Studienteilnehmer, Interventionsgruppe, Vergleichsgruppe, Endpunkte)
PDA	Periduralanästhesie
RCT	Randomized controlled Trial (randomisierte kontrollierte Studie)
RR	relatives Risiko
SSW	Schwangerschaftswoche

## **Kurzfassung**

Das Bundesministerium für Gesundheit (BMG) hat am 14.09.2016 das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) mit der Erstellung von Evidenzberichten zur Unterstützung der Entwicklung der S3-Leitlinie „Die vaginale Geburt am Termin“ beauftragt.

## **Fragestellung**

Von den bei der Leitlinienerstellung federführenden Fachgesellschaften, der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) und der Deutschen Gesellschaft für Hebammenwissenschaft e. V. (DGHWi), wurden zusammen mit der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) 8 Participants-Intervention-Comparison-Outcomes(PICO)-Fragen formuliert, zu denen das IQWiG Evidenzberichte im Format von Rapid Reports erstellen soll.

Die Evidenz für folgende PICO-Frage wurde in diesem Evidenzbericht zusammengefasst:

Wie wirkt sich eine modifizierte Umgebungsgestaltung bei Schwangeren am Termin unter der Geburt ab der aktiven Eröffnungsperiode (EP) auf verschiedene maternale und neonatale Endpunkte im Vergleich zu einer nicht modifizierten Umgebungsgestaltung aus?

## **Methoden**

Die Zielpopulation dieser Evidenzbewertung bildeten Schwangere um den errechneten Geburtstermin ab der vollendeten 37. bis zur vollendeten 42. Schwangerschaftswoche (37 + 0 bis 42 + 0) von der aktiven EP bis zum Ende der Nachgeburtsperiode. Es wurden randomisierte kontrollierte Studien mit einer modifizierten Umgebungsgestaltung des Kreißsaals im Vergleich zu einer Standardumgebung als Prüf- und Vergleichsintervention eingeschlossen. Für alle Studien wurde eine Zeitbeschränkung auf die Jahre ab 2000 festgelegt.

Die von der Leitliniengruppe festgelegten Endpunkte waren gemäß der Vorgehensweise nach dem GRADE-Ansatz (GRADE: Grading of Recommendations, Assessment, Development, and Evaluation) den Kategorien „kritisch für die Entscheidung“ (maternale Endpunkte: Mortalität, Geburtsmodus [Spontangeburt, vaginal-operative Geburt oder Kaiserschnitt], höhergradige Dammriss [III°, IV°], hoher mütterlicher Blutverlust  $\geq 1000$  ml, Aufnahme einer Stillbeziehung bei initialem Stillwunsch, teilweise oder vollständige Plazentaretention, Geburtserleben, subjektive Gesundheitseinschätzung; neonatale Endpunkte: Mortalität, Apgar-Score nach 5 Minuten  $< 6$ , schwere metabolische Acidose [pH  $< 7,0$ ], Basendefizit  $\geq -16$  mmol/l, Verlegung auf die Neugeborenen-Intensivstation) und „klinisch wichtig, aber für die Empfehlung nicht entscheidend“ (maternale Endpunkte: Episiotomie, Einsatz von Wehenmitteln, Amniotomie, Kristellerhilfe, Infektionen; neonatale Endpunkte: leichte, mittelgradige oder fortgeschrittene Acidose, Ikterus, Anämie) zugeordnet.

Eine systematische Literaturrecherche nach Primärstudien wurde in den Datenbanken MEDLINE, Embase, Cochrane Central Register of Controlled Trials und Cinahl durchgeführt. Parallel erfolgte eine Suche nach relevanten systematischen Übersichten in den Datenbanken MEDLINE, Embase, Cinahl, Cochrane Database of Systematic Reviews und HTA Database.

Darüber hinaus wurden folgende Informationsquellen und Suchtechniken berücksichtigt: Studienregister, die Sichtung von Referenzlisten, Anfragen an Studiengruppen und Fachgesellschaften sowie Autorenanfragen.

Die Selektion relevanter Studien erfolgte von 2 Reviewerinnen unabhängig voneinander. Diskrepanzen wurden durch Diskussion zwischen den beiden Reviewerinnen aufgelöst. Die Datenextraktion erfolgte in standardisierte Tabellen. Das Verzerrungspotenzial wurde für die vergleichenden Interventionsstudien auf Studien- und Endpunktebene bewertet und jeweils als gering, mittel oder hoch eingestuft. Es war geplant, die Einzelergebnisse vergleichbarer Studien für die kritischen Endpunkte mithilfe von Metaanalysen quantitativ zusammenzufassen.

Für jeden berichteten kritischen und wichtigen Endpunkt wurde anschließend eine Bewertung der Qualität der Evidenz nach der GRADE-Methodik in 4 Abstufungen (hoch, moderat, niedrig oder sehr niedrig) bezüglich des Vertrauens in die Effektschätzungen vorgenommen. Anschließend erfolgte eine Gesamtbewertung der Qualität der Evidenz.

### ***Ergebnisse der Informationsbeschaffung***

Die Informationsbeschaffung identifizierte 2 randomisierte kontrollierte Studien (2 Dokumente) als relevant für die Fragestellung der vorliegenden Bewertung.

Es wurden 2 laufende und 2 abgeschlossene Studien ohne berichtete Ergebnisse identifiziert. Die letzte Suche fand am 29.09.2017 statt. Die laufenden Studien umfassen eine große Teilnehmerinnenanzahl und werden in Deutschland beziehungsweise Dänemark durchgeführt. Sie weisen eine große potenzielle Relevanz für die untersuchte Fragestellung auf.

### ***Ergebnisse der in die Bewertung eingeschlossenen Studien***

In die eingeschlossenen Studien wurden insgesamt 122 Erst- und Mehrgebärende eingeschlossen, von denen 55 einem Gebärraum mit einer modifizierten Umgebungsgestaltung und 57 einem Gebärraum entsprechend den üblichen Standards zugewiesen wurden. Für 11 randomisierte Teilnehmerinnen wurden keine Informationen zur Gruppenzuteilung oder zu Ergebnissen berichtet.

In 1 Studie lag eine umfangreiche modifizierte Umgebungsgestaltung vor, in der das Kreißbett entfernt und durch eine Doppelmatratze mit großen Kissen in einer Ecke des Raumes ersetzt wurde. Die modifizierte Umgebungsgestaltung umfasste außerdem Möbel zur Unterstützung einer aufrechten Position, eine Projektion von beruhigenden Naturbildern,

gedimmtes Licht und eine große Musikauswahl mit einer Variabilität des Tempos. Zudem wurde ein Hinweisschild mit der Bitte, vor dem Betreten des Raumes anzuklopfen, angebracht und medizinische Geräte wurden aus dem Raum entfernt oder verdeckt. Eine elektronische Herzfrequenzmessung erfolgte ausschließlich, wenn diese als medizinisch notwendig erachtet wurde, stattdessen erfolgte eine Auskultation der fetalen Herzfrequenz. Der Standard-Gebärraum enthielt ein Gebärbett, das zentral im Raum positioniert war, kein Hinweisschild, eine Infusionsstange, angeschaltete Deckenbeleuchtung, klinikübliche Umgebungsgeräusche und die Herzfrequenz des Fetus wurde als Routinekontrolle durchgeführt, in der Regel mittels kontinuierlicher elektronischer Herzfrequenzmessung.

In der 2. Studie wurde eine geringe Modifizierung in Form von Naturbildern, die über einen gesonderten Fernsehbildschirm im Gebärraum platziert wurden, vorgenommen. Im Standard-Gebärraum war kein zusätzlicher Fernsehbildschirm mit Naturbildern vorhanden.

Es konnten Ergebnisse für insgesamt 2 kritische und 1 wichtigen Endpunkt ermittelt werden. Die Qualität der Evidenz wurde für die 2 kritischen Endpunkte Geburtsmodus und Geburtserleben als sehr niedrig und für den wichtigen Endpunkt Wehenmittel als niedrig bewertet. Es lagen keine Ergebnisse aus randomisierten kontrollierten Studien zu den kritischen Endpunkten maternale und fetale Mortalität, höhergradige Dammriss (III° / IV°), hoher mütterlicher Blutverlust  $\geq 1000$  ml, Aufnahme einer Stillbeziehung bei initialem Stillwunsch, teilweise / vollständige Plazentaretention, subjektive Gesundheitseinschätzung, schwere metabolische Acidose, Basendefizit  $\geq -16$  mmol/l und Verlegung auf die Neugeborenen-Intensivstation vor. Zum kritischen Endpunkt Apgar-Score nach 5 Minuten  $< 6$  lagen nur ergänzende Informationen zum Apgar-Score nach 5 Minuten  $< 7$  vor. Zudem lagen keine Ergebnisse aus randomisierten kontrollierten Studien zu den weiteren klinisch wichtigen, aber für die Empfehlung nicht entscheidenden Endpunkten vor.

Bei einer insgesamt sehr niedrigen Qualität der Evidenz bleibt unklar, ob sich die modifizierte Umgebungsgestaltung positiv oder negativ auf den Geburtsmodus auswirkt. Jedoch führte die modifizierte Umgebungsgestaltung bei einer niedrigen Qualität der Evidenz zu einem geringeren Einsatz von Wehenmitteln (RR: 0,59; 95 %-KI [0,36; 0,97]). Das Geburtserleben wurde mithilfe nicht validierter Fragebogen erhoben, die in 1 Studie nur Teilaspekte des Geburtserlebens, die die Wirkung der verglichenen Gebärräume und die Unterstützung durch das Pflegepersonal betrafen, umfassten. In der 2. Studie wurden Informationen zur Qualität der Betreuung aus der Patientenperspektive erhoben und nur Teilfragen eines bereits vorhandenen Fragebogens verwendet. Die Informationen werden ergänzend dargestellt. Bei insgesamt sehr niedriger Qualität der Evidenz konnte für keine der Fragen ein bedeutsamer Unterschied bezüglich des Geburtserlebens beobachtet werden.

### ***Fazit***

Auf der Basis von 2 randomisierten kontrollierten Studien mit einer insgesamt geringen Teilnehmerinnenanzahl bleibt der Einfluss einer modifizierten Umgebungsgestaltung des Kreißsaals auf den Endpunkt Geburtsmodus und den Endpunkt Geburtserleben, zu welchem

nur ergänzende Informationen dargestellt werden konnten, unklar. Das Geburtserleben wurde mit nicht validierten Fragebogen erhoben. Für den klinisch wichtigen, aber für die Empfehlung nicht entscheidenden Endpunkt Wehenmittel konnte in 1 Studie eine Auswirkung in Form einer signifikant geringeren Gabe von Wehenmitteln unter der Geburt festgestellt werden. Ob dies mit einem Vor- oder Nachteil für die Mutter und das Neugeborene verbunden ist, bleibt aufgrund der fehlenden Informationen zu weiteren kritischen Endpunkten unklar. Zusammenfassend wurde die Qualität der Evidenz und somit das Gesamtvertrauen in die Effektschätzungen für die vorliegende Fragestellung als sehr niedrig bewertet. Diese geringe Qualität der Evidenz ist vorrangig durch schwerwiegende Studienlimitationen wie fehlenden Angaben zur Randomisierung, einem intransparenten Teilnehmerinnenfluss, einer selektiven Berichterstattung und einer fehlenden Differenzierung zwischen primären und sekundären Endpunkten sowie einer fehlenden Genauigkeit der Effektschätzungen bedingt.

Insgesamt ist zu beachten, dass nur für 3 der vordefinierten Endpunkte Evidenz aus 2 randomisierten kontrollierten Studien vorlag. Aktuell werden 2 potenziell relevante randomisierte kontrollierte Studien mit einer großen Anzahl von Teilnehmerinnen durchgeführt, die Informationen zu den kritischen und wichtigen Endpunkten erheben. Die Ergebnisse dieser Studien lassen eine Verbesserung der Evidenzlage und eine Beantwortung der Fragestellung auf der Basis von hoher Evidenz bei hoher Übertragbarkeit erwarten.

## 1 Hintergrund

Das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) wurde vom Bundesministerium für Gesundheit (BMG) beauftragt, die Deutsche Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) und die Deutsche Gesellschaft für Hebammenwissenschaft e. V. (DGHWi) bei der Entwicklung einer interdisziplinären S3-Leitlinie „Die vaginale Geburt am Termin“ (Registernummer der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften [AWMF] 015-083) [1] zu unterstützen.

Ziel dieser Leitlinie ist die Evaluierung evidenzbasierter Interventionen bei der vaginalen Geburt am Termin, die Geburtshelferinnen und Geburtshelfern mehr Sicherheit geben soll. Die meisten Schwangerschaften führen zu einer vaginalen Geburt ohne notwendige intensive geburtshilfliche Interventionen. Bei auftretenden Risiken ist eine intensivmedizinische Betreuung von Mutter und Kind essenziell. Daraus ergibt sich die Notwendigkeit einer risikoadaptierten Versorgungsplanung und -gestaltung.

In Deutschland lag die Mortalität für alle Neugeborene gemessen an der Frühsterblichkeit (bis zum 7. Lebenstag verstorben) 2015 bei 186 je 100 000 Lebendgeborene und für die Spätmortalität (8. bis 28. Lebenstag verstorben) bei 42 je 100 000 Lebendgeborene [2]. Die Müttersterblichkeit lag bei rund 3 Frauen je 100 000 Lebendgeborene [3]. Als Risikofaktoren für Komplikationen rund um die Schwangerschaft und Geburt gelten ein höheres Alter der Mütter bei der 1. Geburt, Rauchen während der Schwangerschaft, eine geringe soziale Unterstützung, ein Migrationshintergrund, Adipositas und Mehrlingsschwangerschaften [4].

Der Anteil der Kaiserschnittentbindungen an allen Geburten in deutschen Krankenhäusern ist in den letzten Jahrzehnten von 17,3 % im Jahr 1994 auf 31,2 % im Jahr 2016 [5] angestiegen und zeigt eine hohe innerdeutsche Variabilität [4]. Bei 23,1 % aller Kaiserschnitte im Geburtsverlauf wird als Hauptindikation ein pathologisches Kardiotokogramm (CTG) oder eine protrahierte Geburt als Einzeldiagnose genannt. Außerdem ist ein pathologisches CTG bei 41,2 % beziehungsweise eine protrahierte Geburt bei 21,1 % der durchgeführten sekundären Kaiserschnitte eine der dokumentierten Indikationen [5]. Hinzu kommen eine aufgrund forensischer Motive zunehmend veränderte Risikobewertung in der Schwangerenbetreuung und Geburtshilfe, veränderte Klinikorganisationen und eine abnehmende Erfahrung des geburtshilflichen Fachpersonals in der Betreuung komplizierter Geburtsverläufe [6].

Die interdisziplinäre S3-Leitlinie „Die vaginale Geburt am Termin“ soll die Evidenzbasierung der Begleitung der natürlichen (vaginalen) Geburt durch Hebammen und Entbindungspfleger, Medizinerinnen und Mediziner und weitere Gesundheitsberufe verbessern. Das S3-Leitlinienniveau fordert eine systematische Recherche, Auswahl und Bewertung der vorliegenden wissenschaftlichen Evidenz zu den relevanten klinischen Fragestellungen und die Beschreibung des methodischen Vorgehens, sodass in der Leitlinie schließlich nachvollziehbar dargelegt werden kann, auf welcher Evidenz die gegebenen Empfehlungen basieren [7].

Ziel des geplanten Evidenzberichts ist es, für eine durch die Fachgesellschaften und die AWMF vorgegebene Schlüsselfrage eine Evidenzgrundlage zusammenzustellen und diese hinsichtlich ihrer Qualität zu bewerten. Diese Frage umfasst eine modifizierte Umgebungsgestaltung bei Schwangeren am Termin unter der Geburt ab der aktiven Eröffnungsphase (EP). Der Evidenzbericht soll von der Leitliniengruppe als Grundlage für die Entwicklung von evidenzbasierten Empfehlungen für die S3-Leitlinie herangezogen werden.

Dieser Rapid Report des IQWiG nutzt die Methoden der AWMF und der GRADE-Arbeitsgruppe (GRADE: Grading of Recommendations, Assessment, Development, and Evaluation).

## 2 Fragestellung

Von den bei der Leitlinienerstellung federführenden Fachgesellschaften, der Deutschen Gesellschaft für Gynäkologie und Geburtshilfe e. V. (DGGG) und der Deutschen Gesellschaft für Hebammenwissenschaft e. V. (DGHWi), wurden zusammen mit der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) 8 Participants-Intervention-Comparison-Outcomes(PICO)-Fragen formuliert, zu denen das IQWiG Evidenzberichte im Format von Rapid Reports erstellen soll.

Die Evidenz für folgende PICO-Frage wurde in diesem Evidenzbericht zusammengefasst:

Wie wirkt sich eine modifizierte Umgebungsgestaltung bei Schwangeren am Termin unter der Geburt ab der aktiven Eröffnungsperiode (EP) auf verschiedene maternale und neonatale Endpunkte im Vergleich zu einer nicht modifizierten Umgebungsgestaltung aus?

### **3 Projektverlauf**

Das BMG hat am 14.09.2016 das IQWiG mit der Unterstützung der Entwicklung einer interdisziplinären S3-Leitlinie „Die vaginale Geburt am Termin“ beauftragt.

Auf Basis der Projektskizze wurde ein Rapid Report zur Evidenz zu den Auswirkungen einer modifizierten Umgebungsgestaltung bei Schwangeren am Termin unter der Geburt ab der aktiven EP erstellt. Dieser Bericht wurde an das BMG und die Leitliniengruppe der AWMF übermittelt und nach Abschluss aller Teilprojekte gemeinsam auf der Website des IQWiG veröffentlicht.

## 4 Methoden

### Evidenzbewertung anhand von vergleichenden Interventionsstudien

Die Evidenz zur Fragestellung lässt sich anhand von prospektiv geplanten vergleichenden Interventionsstudien unter Beachtung maternaler und neonataler Endpunkte bewerten. Idealerweise sollten Personen oder Krankenhäuser einer Gruppe randomisiert zugeteilt worden sein.

#### 4.1 Kriterien für den Einschluss von Studien zur modifizierten Umgebungsgestaltung in die Untersuchung

##### 4.1.1 Population

Eingeschlossen wurden Studien, die Gebärende um den errechneten Geburtstermin ab der vollendeten 37. Schwangerschaftswoche (SSW) (37 + 0 SSW) bis zur vollendeten 42. SSW (42 + 0 SSW) von der aktiven EP bis zum Ende der Nachgeburtsperiode untersuchen. In der aktiven EP (Aktivphase, engl. „first-stage“, „active phase of labour“, „phase one“) treten regelmäßige und schmerzhafte Wehen auf [8,9]. Der Beginn dieser Phase wird in der Literatur häufig anhand einer Muttermundweite von mindestens 3 cm beschrieben, jedoch reicht die Spanne von 3 bis 7 cm. Eine einheitliche Angabe gibt es nicht [10]. Die Nachgeburtsperiode umfasst die Zeit von der Geburt des Neugeborenen bis 2 Stunden nach der Plazentageburt [8].

##### 4.1.2 Prüf- und Vergleichsintervention

In die Bewertung wurden randomisierte kontrollierte Studien (RCTs) eingeschlossen, die eine modifizierte Umgebungsgestaltung als zu prüfende Intervention mit einer nicht modifizierten Gestaltung des Kreißsaales als Vergleichsintervention untersuchten. Die modifizierten Gestaltungsmöglichkeiten konnten ein Kreißbett außerhalb des Gebärraums, ein im Gebärraum beiseitegeschobenes Kreißbett, spezielle Sitzmöbel, die Möglichkeit der Kontrolle über die Luftzufuhr, Raumtemperatur und Beleuchtung, eine integrierte Möglichkeit zur Zubereitung von Nahrung oder Getränken, spezielle Musik- oder DVD-Angebote und Bildmaterial an den Wänden zur Entspannung umfassen. Die Vergleichsintervention sollte dem heutigen Standard in deutschen Kliniken entsprechen, bei dem eine häusliche Einrichtung widergespiegelt werden soll [11,12], um die klinische Atmosphäre zu reduzieren.

##### 4.1.3 Endpunkte

Die Bewertung der Evidenzgrundlage für die Prüfinderventionen erfolgte anhand von Endpunkten, die von der Leitliniengruppe nach den GRADE-Vorgaben konsentiert und hinsichtlich ihrer Bedeutung für die Gebärenden und Neugeborenen den Kategorien „kritisch für die Entscheidung“ und „klinisch wichtig, aber für die Empfehlung nicht entscheidend“ zugeordnet wurden. Endpunkte, die aus der Perspektive der Gebärenden unbedingt in der Entscheidungsfindung für oder gegen eine Intervention berücksichtigt werden müssen, wurden der Kategorie „kritisch“ zugeordnet. Weitere Endpunkte, die ebenfalls berücksichtigt werden sollten, aber nicht unbedingt in die Entscheidungsfindung einbezogen werden müssen, wurden als „klinisch wichtig“ kategorisiert [13,14] (siehe Tabelle 1). Als Endpunkt ergänzt

wurde entsprechend den IQWiG-Methoden die Mortalität der Mutter oder des Neugeborenen. Diese Ergänzung wurde mit der Leitliniengruppe konsentiert. Es war geplant, dass alle Endpunkte mit validierten Messinstrumenten erfasst sein müssen.

Tabelle 1: Durch die Leitliniengruppe konsentierte Endpunkte für vergleichende Interventionsstudien und ihre Bewertung

<b>Rangfolge der Endpunkte</b>	<b>Endpunkte</b>
Kritisch für die Entscheidung (Bewertung mit 7–9 Punkten durch die Leitliniengruppe)	<p><b>Maternale Endpunkte:</b></p> <ul style="list-style-type: none"> <li>▪ Mortalität</li> <li>▪ Spontangeburt</li> <li>▪ vaginal-operative Geburt</li> <li>▪ Kaiserschnitt</li> <li>▪ höhergradige Dammriss (III°, IV°)</li> <li>▪ hoher mütterlicher Blutverlust (<math>\geq 1000</math> ml)</li> <li>▪ Aufnahme Stillbeziehung bei initialem Stillwunsch</li> <li>▪ teilweise / vollständige Plazentaretenion</li> <li>▪ Geburtserleben</li> <li>▪ subjektive Gesundheitseinschätzung</li> </ul> <p><b>Neonatale Endpunkte:</b></p> <ul style="list-style-type: none"> <li>▪ Mortalität (Früh- und Spätsterblichkeit)</li> <li>▪ Apgar-Score nach 5 Minuten <math>&lt; 6</math></li> <li>▪ schwere metabolische Acidose (<math>\text{pH} &lt; 7,0</math>)</li> <li>▪ Basendefizit <math>\geq -16</math> mmol/l</li> <li>▪ Verlegung auf die Neugeborenen-Intensivstation</li> </ul>
Klinisch wichtig, aber für die Empfehlung nicht entscheidend (Bewertung mit 4–6 Punkten durch die Leitliniengruppe)	<p><b>Maternale Endpunkte:</b></p> <ul style="list-style-type: none"> <li>▪ Episiotomie</li> <li>▪ Wehenmittel</li> <li>▪ Amniotomie</li> <li>▪ Kristellerhilfe</li> <li>▪ Infektionen</li> </ul> <p><b>Neonatale Endpunkte:</b></p> <ul style="list-style-type: none"> <li>▪ leichte, mittelgradige, fortgeschrittene Acidose<sup>a</sup></li> <li>▪ Ikterus</li> <li>▪ Anämie</li> </ul>

a: Eine Acidose wird wie folgt kategorisiert:  $\text{pH } 7,19-7,15$ : leichte Acidose,  $\text{pH } 7,14-7,10$ : mittelgradige Acidose,  $\text{pH } 7,09-7,0$ : fortgeschrittene Acidose,  $\text{pH} < 7,0$ : schwere Acidose [15,16].

Apgar-Score: Schema zur Beurteilung der postpartalen Adaptation und Vitalität des reifen Neugeborenen nach den Kriterien Atmung, Puls, Grundtonus, Aussehen und Reflexe

#### 4.1.4 Studientypen

Es war geplant, prospektiv geplante vergleichende Studien (RCTs, nicht randomisierte kontrollierte Studien, Kohortenstudien und Fall-Kontroll-Studien) mit möglichst hohem Evidenzniveau (Evidence-based-Medicine(EbM)-Kriterien nach Oxford [17]) einzuschließen, wobei die Designklassifizierung der Studien auf den Design-Algorithm-for-Medical-Literature-on-Intervention(DAMI)-Kriterien beruhen sollte [18].

Randomisierte kontrollierte Studien (RCTs) sind, sofern sie methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt wurden, mit der geringsten Ergebnisunsicherheit behaftet. Sie liefern daher die zuverlässigsten Ergebnisse für die Bewertung der Evidenz einer medizinischen Intervention.

Bei auf RCTs basierender nicht hinreichender Datenlage war geplant, auch nicht randomisierte prospektiv geplante vergleichende Interventionsstudien mit zeitlich paralleler Vergleichsgruppe und adäquater Confounderkontrolle zur Evidenzbewertung heranzuziehen. Nicht randomisierte Studien ohne Confounderkontrolle oder mit einem Evidenzniveau geringer als 3 nach den EbM-Kriterien nach Oxford [17] sollten nicht berücksichtigt werden.

Für alle unter Abschnitt 4.1.2 genannten Interventionen und unter Abschnitt 4.1.3 genannten Endpunkte aus Interventionsstudien ist eine Evaluation im Rahmen von RCTs möglich und praktisch durchführbar. Die Randomisierung kann sowohl über die Gebärenden als auch über eine Clusterrandomisierung der Praxen oder Krankenhäuser erfolgen.

#### 4.1.5 Studiendauer

Eingeschlossen wurden Studien mit einer Mindestdauer, die den Zeitraum vom Beginn der aktiven EP bis zum Ende der Nachgeburtsperiode umfasste. Aufgrund von Unsicherheiten in den Definitionen der Geburtsphasen war geplant, geeignet erscheinende Studien im Zweifel einzuschließen und den Zeitpunkt des Einschlusses der Teilnehmerinnen wie in der Publikation berichtet in die Evidenztabelle zu extrahieren.

#### 4.1.6 Publikationszeitraum

Es wurden die unter Abschnitt beschriebenen Studientypen mit einem Publikationsdatum ab 2000 eingeschlossen. Für die Einschränkung des Publikationszeitraumes wurde die Publikation von Rose und Schmied-Knittel aus dem Jahr 2011 [19] zugrunde gelegt, da diese die Umbruchphase von der programmierten zur selbstbestimmten Geburt der 1980er-Jahre beschreibt. Diese führte bereits in den 1980er-Jahren zu einer Änderung der Umgebungsgestaltung, unter anderem um eine häusliche Einrichtung widerzuspiegeln und die klinische Atmosphäre zu reduzieren [11]. Die damaligen Änderungen sind heutzutage Standard [12,15] und daher nicht der Fokus der Prüfintervention. Diese Einschränkung soll zur Identifikation relevanter Literatur führen, die den heutigen geburtshilflichen Standard berücksichtigt und daher für die Erstellung der Leitlinienempfehlungen Relevanz besitzt.

#### 4.1.7 Tabellarische Darstellung der Kriterien für den Studieneinschluss

In Tabelle 2 sind die Kriterien aufgelistet, die Studien erfüllen mussten, um für die Fragestellung in die Bewertung eingeschlossen zu werden.

Tabelle 2: Übersicht über die Kriterien für den Studieneinschluss

<b>Einschlusskriterien</b>	
E1	Population: alle Schwangeren am Termin unter der Geburt ab aktiver EP wie in Abschnitt 4.1.1 formuliert
E2	Prüfintervention: modifizierte Umgebungsgestaltung wie in Abschnitt 4.1.2 definiert
E3	Vergleichsintervention: keine modifizierte Umgebungsgestaltung wie in Abschnitt 4.1.2 definiert
E4	Erhebung kritischer oder wichtiger Endpunkte wie in Abschnitt 4.1.3 formuliert
E5	Randomisierte kontrollierte Studien, ggf. nicht randomisierte prospektiv geplante vergleichende Studien mit möglichst hohem Evidenzniveau wie in Abschnitt 4.1.4 formuliert
E6	Vollpublikation <sup>a</sup>
E7	Publikation ab 2000 wie in Abschnitt 4.1.6 definiert
<p>a: Als Vollpublikation gilt in diesem Zusammenhang auch ein Studienbericht, der den Kriterien des CONSORT- [20], TREND- [21] oder STROBE-Statements [22] genügt und eine Bewertung der Studie ermöglicht, sofern die in diesen Dokumenten enthaltenen Informationen zu Studienmethodik und -ergebnissen nicht vertraulich sind.</p> <p>CONSORT: Consolidated Standards of Reporting Trials; EP: Eröffnungsperiode; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology; TREND: Transparent Reporting of Evaluations with Nonrandomized Designs</p>	

#### 4.1.8 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen

Für die Einschlusskriterien E1 (Population), E2 (Interventionsgruppe) und E3 (Vergleichsgruppe) reichte es aus, wenn bei mindestens 80 % der eingeschlossenen Gebärenden dieses Kriterium erfüllt war. Es war geplant, für solche Studien auf entsprechende Subgruppenanalysen zurückzugreifen, aber solche Studien, bei denen die Einschlusskriterien bei weniger als 80 % erfüllt waren, nur dann einzuschließen, wenn entsprechende Subgruppenanalysen vorlagen.

## 4.2 Umfassende Informationsbeschaffung

### 4.2.1 Primäre Informationsquellen

#### 4.2.1.1 Bibliografische Datenbanken

Die systematische Recherche nach relevanten Studien beziehungsweise Dokumenten wurde in folgenden bibliografischen Datenbanken durchgeführt:

- Suche nach Primärstudien in den Datenbanken MEDLINE, Embase, Cochrane Central Register of Controlled Trials und Cinahl
- Suche nach relevanten systematischen Übersichten in den Datenbanken MEDLINE, Embase und Cinahl parallel zur Suche nach relevanter Primärliteratur sowie Suche in den Datenbanken Cochrane Database of Systematic Reviews und HTA Database

#### **4.2.1.2 Studienregister**

Die folgenden Studienregister wurden durchsucht:

- U.S. National Institutes of Health, ClinicalTrials.gov
- World Health Organization, International Clinical Trials Registry Platform Search Portal

#### **4.2.1.3 Weitere Informationsquellen und Suchtechniken**

Mit dem Ziel, weitere veröffentlichte und unveröffentlichte Studien beziehungsweise Informationen zu relevanten Studien zu ermitteln, wurden weitere Quellen beziehungsweise Suchtechniken berücksichtigt.

##### **4.2.1.3.1 Studiengruppen und Fachgesellschaften**

Es erfolgte eine schriftliche Anfrage bei folgenden Fachgesellschaften und Studiengruppen:

- AWMF
- Cochrane Pregnancy and Childbirth Group

##### **4.2.1.3.2 Weitere Suchtechniken**

Systematische Übersichten wurden hinsichtlich weiterer relevanter Studien beziehungsweise Dokumente gesichtet.

##### **4.2.1.4 Autorenanfragen**

Es wurden Anfragen an Autorinnen und Autoren gestellt, falls Informationen, die einen relevanten Einfluss auf die Bewertung zu erwarten ließen, den vorliegenden Studiendokumenten nicht oder nur ungenau zu entnehmen waren.

#### **4.2.2 Selektion relevanter Studien**

##### **Selektion relevanter Studien beziehungsweise Dokumente aus den Ergebnissen der bibliografischen Recherche**

Die durch die Suche in bibliografischen Datenbanken identifizierten und zu screenenden Treffer wurden in einem 1. Schritt anhand ihres Titels und, sofern vorhanden, Abstracts in Bezug auf ihre potenzielle Relevanz bezüglich der spezifischen Einschlusskriterien (siehe Tabelle 2) bewertet. Als potenziell relevant erachtete Dokumente wurden in einem 2. Schritt anhand ihres Volltextes auf Relevanz geprüft. Beide Schritte erfolgten durch 2 Reviewerinnen

unabhängig voneinander. Diskrepanzen wurden durch Diskussion zwischen den beiden Reviewerinnen aufgelöst.

### **Selektion relevanter Studien bzw. Dokumente aus weiteren Suchquellen**

Informationen aus den folgenden Suchquellen wurden von 2 Reviewerinnen unabhängig voneinander in Bezug auf ihre Relevanz bewertet:

- Studienregister
- Informationen von Studiengruppen und Fachgesellschaften

Informationen aus den folgenden Suchquellen wurden von 1 Reviewerin auf Studien gesichtet, die diese dann in Bezug auf ihre Relevanz bewertete; eine 2. Reviewerin überprüfte den gesamten Prozess inklusive der Bewertungen:

- identifizierte systematische Übersichten

Sofern in einem der genannten Selektionsschritte Diskrepanzen auftraten, wurden diese jeweils durch Diskussion zwischen den beiden Reviewerinnen aufgelöst.

## **4.3 Informationsextraktion und -bewertung auf Studienebene**

Alle für den Evidenzbericht relevanten Ergebnisse wurden hinsichtlich ihrer Ergebnis-sicherheit, bestehend aus dem Verzerrungspotenzial und der Genauigkeit der Ergebnisse, überprüft. Für jeden für die Entscheidung kritischen Endpunkt wurde eine endpunkt-spezifische studienübergreifende Aussage zur Qualität der Evidenz in 4 Abstufungen bezüglich der jeweiligen Aussagesicherheit getroffen und diese in eine der Kategorien hohe, moderate, niedrige und sehr niedrige Qualität der Evidenz entsprechend der GRADE-Guideline eingeteilt [23,24].

### **4.3.1 Datenextraktion**

Alle für die Evidenzbewertung notwendigen Informationen wurden aus den Unterlagen zu den eingeschlossenen Studien in standardisierte Tabellen extrahiert. Die Evidenztabellen orientierten sich an Vorlagen der AWMF.

Die Extraktion erfolgte durch 1 Reviewerin und wurde von einer 2. Reviewerin auf Grundlage der Primärstudien kontrolliert.

### **4.3.2 Bewertung des Verzerrungspotenzials der Ergebnisse**

Das Verzerrungspotenzial der Ergebnisse wurde für jede in die Evidenzbewertung eingeschlossene Studie bewertet, und zwar separat für jeden Endpunkt. Dazu wurden insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Aspekte, die das Verzerrungspotenzial beeinflussen, systematisch extrahiert und bewertet [25]:

**A: Aspekte des Verzerrungspotenzials auf Studienebene**

- Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Verblindung der Gebärenden sowie der behandelnden Person
- ergebnisgesteuerte Berichterstattung

Um den Vorgaben für Evidenzberichte der Leitliniengruppe der AWMF und der GRADE-Arbeitsgruppe zu entsprechen, sollten insbesondere auch folgende Aspekte bewertet werden:

- früher Abbruch bei hohen Behandlungseffekten
- Rekrutierungsunterschiede in den Clustern der clusterrandomisierten Studien und Vergleichbarkeit der Gruppen beziehungsweise Berücksichtigung prognostisch relevanter Faktoren (bei Cluster-RCTs) [26]

**B: Aspekte des Verzerrungspotenzials auf Endpunktebene**

- Verblindung der endpunkterhebenden Personen
- Umsetzung des Intention-to-treat(ITT)-Prinzips
- ergebnisgesteuerte Berichterstattung

Anhand dieser Aspekte wurde das Verzerrungspotenzial zusammenfassend als gering, mittel oder hoch eingestuft. Ein geringes Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Bei einem mittleren oder hohen Verzerrungspotenzial ist davon auszugehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern oder sogar stark verändern würden [25].

Für die Bewertung eines Endpunkts wurde zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter (A) aufgeführten Aspekte als gering, mittel oder hoch eingestuft. Das Verzerrungspotenzial wurde als gering beurteilt, wenn dieses für alle genannten Aspekte gering ist. Eine Einteilung in die Kategorie mittel oder hoch erfolgte, wenn es Einschränkungen in 1 oder mehreren Aspekten gab, die das Vertrauen in die Effektschätzungen herabsetzen oder sogar stark herabsetzen. Falls die Einstufung als mittel oder hoch erfolgte, wurde das Verzerrungspotenzial für den Endpunkt auch als mittel oder hoch bewertet. Ansonsten fanden die unter (B) genannten endpunktspezifischen Aspekte Berücksichtigung.

**4.3.3 Bewertung des Evidenzniveaus aller eingeschlossenen Studien**

Zusätzlich erfolgte auf Wunsch der AWMF eine studienspezifische endpunktübergreifende Bewertung der Evidenz jeder eingeschlossenen Studie nach den Oxford-Centre-for-Evidence-based-Medicine(OCEBM)-Kriterien [27]. Diese Bewertung basiert auf dem Studiendesign

und hat bei Studienlimitationen, einer fehlenden Genauigkeit der Effekte und Indirektheit eine Abwertung zur Folge. Zum Beispiel wurden RCTs mit hohem Verzerrungspotenzial vom Evidenzniveau 2 auf das Evidenzniveau 3 abgewertet.

#### **4.4 Informationssynthese und -analyse**

Die Informationen wurden für die kritischen Endpunkte einer Informationssynthese und -analyse unterzogen [28]. Da Evidenz für weniger als 7 kritische Endpunkte vorlag, wurde die Evidenz der klinisch wichtigen, aber für die Empfehlung nicht entscheidenden Endpunkte ebenfalls in die Informationssynthese eingeschlossen. Es war geplant, nach Möglichkeit über die Gegenüberstellung der Ergebnisse der Einzelstudien hinaus die unten beschriebenen Verfahren einzusetzen. Eine abschließende zusammenfassende Bewertung der Informationen erfolgte darüber hinaus in jedem Fall.

##### **4.4.1 Gegenüberstellung der Ergebnisse der Einzelstudien**

Die Ergebnisse zu den in den Studien berichteten Endpunkten wurden im Bericht vergleichend beschrieben.

Es war geplant, in bestimmten Fällen auf eine Darstellung einzelner Ergebnisse bei unzureichender Erfassung des Endpunkts bei allen Gebärenden in der jeweiligen Studie beziehungsweise unvollständiger Berichterstattung zu verzichten. Dies sollte insbesondere zutreffen, wenn ein hoher Anteil der Gebärenden nicht in der Auswertung enthalten ist. Es war geplant, dass Ergebnisse in der Regel nicht in die Evidenzbewertung einfließen sollten, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Gebärenden basieren, das heißt, wenn der Anteil der Gebärenden, die nicht in der Auswertung berücksichtigt wurden, größer als 30 % war [29]. In der Literatur werden zum Teil bereits Auswertungen, in denen 20 % der Studienteilnehmenden nicht berücksichtigt werden, als nicht mehr aussagekräftig betrachtet [29]. Ausnahmen von dieser Regel waren zum Beispiel dann geplant, wenn aus logistischen Gründen für ganze Zentren (ganze Randomisierungsböcke oder Cluster) keine Daten erhoben wurden und dies bereits bei der Studienplanung vorgesehen war [30].

Ergebnisse sollten auch dann nicht in die Evidenzbewertung einbezogen werden, wenn der Unterschied in den Anteilen nicht berücksichtigter Gebärender zwischen der Interventions- und der Vergleichsgruppe größer als 15 Prozentpunkte war.

##### **4.4.2 Metaanalysen**

Es war geplant, sofern die eingeschlossenen Studien hinsichtlich der Fragestellung und relevanter Charakteristika vergleichbar waren, die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammenzufassen. Für die statistische Auswertung sollten primär die Ergebnisse aus ITT-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden. Es war geplant, dass die Auswahl der Modelle für Metaanalysen gemäß den Kriterien erfolgen würde, die in den aktuellen Allgemeinen Methoden des IQWiG

[31] genannt sind. Metaanalysen mit zufälligen Effekten sollten bei 5 oder mehr Studien primär mithilfe der Knapp-Hartung-Methode durchgeführt werden. Bei einer geringeren Anzahl von Studien sollte gegebenenfalls ein Modell mit festem Effekt berechnet oder eine qualitative Zusammenfassung vorgenommen werden. Falls die für eine Metaanalyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorlagen, war geplant, diese nach Möglichkeit aus den vorhandenen Informationen eigenständig zu berechnen beziehungsweise näherungsweise zu bestimmen.

Für stetige Variablen war geplant, die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges'  $g$ , als Effektmaß einzusetzen. Bei binären Variablen sollten Metaanalysen primär anhand des relativen Risikos durchgeführt werden. In begründeten Ausnahmefällen sollten auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen war geplant, ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und von den verfügbaren Daten zu verwenden [32].

Es war geplant, die Effektschätzer und Konfidenzintervalle (KIs) aus den Studien mittels Forest Plots zusammenfassend darzustellen. Anschließend sollte die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes  $I^2$  und des statistischen Tests auf Vorliegen von Heterogenität [33] erfolgen. Es war geplant, bei nicht bedeutsamer Heterogenität der Ergebnisse der Studien den gemeinsamen (gepoolten) Effekt inklusive KI darzustellen. Bei bedeutsamer Heterogenität sollte stattdessen das Prädiktionsintervall dargestellt und die Ergebnisse sollten nur in begründeten Ausnahmefällen gepoolt werden. Außerdem sollte untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu sollten methodische Faktoren (siehe Abschnitt 4.4.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.4.5), zählen.

#### **4.4.3 Aussagen zur Qualität der Evidenz**

Für jeden berichteten kritischen und wichtigen Endpunkt wurde nach den GRADE-Kriterien eine endpunktspezifische studienübergreifende Aussage zum Vertrauen in die Effektschätzungen (Qualität der Evidenz) in 4 Abstufungen bezüglich der jeweiligen Aussagesicherheit getroffen und diese in eine der Kategorien hohe, moderate, niedrige und sehr niedrige Qualität der Evidenz eingeteilt [23,24]. In die Bewertung gingen das Design der eingeschlossenen Studien, Studienlimitationen, das Risiko von Publikationsbias, die Genauigkeit und Konsistenz der Effekte, die Übertragbarkeit und gegebenenfalls Aspekte zur Aufwertung der Qualität der Evidenz ein.

Bei einer hohen Qualität der Evidenz kann mit hoher Wahrscheinlichkeit davon ausgegangen werden, dass das Ergebnis der vorliegenden Studien nahe am wahren Interventionseffekt liegt. Diese Sicherheit nimmt mit geringerer Qualität der Evidenz ab.

Die Bewertung erfolgte durch 2 Reviewerinnen unabhängig voneinander. Diskrepanzen wurden durch Diskussion zwischen den beiden Reviewerinnen aufgelöst. Ergebnisse auf der Grundlage von RCTs starteten aufgrund ihres Designs bei einer hohen Qualität der Evidenz,

Ergebnisse aus Beobachtungsstudien sollten bei einer niedrigen Qualität beginnen [23,24]. Diese Qualität konnte bei einem hohen Verzerrungspotenzial, einem hohen Risiko von Publikationsbias, einer fehlenden Genauigkeit der Effekte, Inkonsistenz oder einer eingeschränkten Übertragbarkeit abgewertet und bei sehr großen Effekten, einer Dosis-Wirkungs-Beziehung oder dann, wenn alle potenziellen Confounder zu einer Effekterhöhung führen würden, aufgewertet werden. Alle vorgenommenen Ab- oder Aufwertungen der Qualität der Evidenz wurden durch Fußnoten gekennzeichnet.

Im GRADE-Ansatz basiert die Stärke von Leitlinienempfehlungen auf dem Vertrauen, dass „die erwünschten Effekte einer Behandlungsstrategie die unerwünschten Effekte insgesamt überwiegen“ [34]. Daher war eine Gesamtbeurteilung des Vertrauens in die vorliegenden Effektschätzungen notwendig. Diese ergab sich aus der niedrigsten Qualität der Evidenz der kritischen Endpunkte [35,36].

#### **A: Abwertung bei Studienlimitationen (Verzerrungspotenzial)**

Die Qualität der Evidenz wurde aufgrund von schwerwiegenden oder sehr schwerwiegenden Limitationen um 1 oder 2 Kategorien herabgestuft, wenn starke Limitierungen in 1 oder mehreren der unter Abschnitt 4.3.2 genannten Kriterien das Vertrauen in die Effektschätzungen beeinträchtigen. Die Qualität der Evidenz wurde beispielsweise um 1 Kategorie herabgestuft, wenn die Effektschätzungen hauptsächlich auf Studien mit mittlerem Verzerrungspotenzial basieren oder wenn in einer Studie ein hohes Verzerrungspotenzial für einen Aspekt oder ein mittleres Verzerrungspotenzial für mehrere Aspekte das Vertrauen in die Effektschätzungen beeinträchtigt. Die Qualität der Evidenz wurde beispielsweise um 2 Kategorien herabgestuft, wenn die Effektschätzungen hauptsächlich auf Studien mit hohem Verzerrungspotenzial basieren oder wenn in einer Studie ein hohes Verzerrungspotenzial für 1 oder mehrere Aspekte das Vertrauen in die Effektschätzungen stark beeinträchtigt [25,37].

#### **B: Abwertung bei Publikationsbias**

Die Qualität der Evidenz wurde um 1 Kategorie herabgestuft, wenn ein Evidenzkörper ein hohes Risiko für einen Publikationsbias aufweist. Kriterien für die Annahme eines Publikationsbias umfassten einen Evidenzkörper, der vorrangig auf kleinen, kommerziell gesponserten Studien beruht, einen beobachteten Zusammenhang zwischen den Effektschätzungen und der Studiengröße (oder der Genauigkeit der Effekte) oder relevante Unterschiede in den Ergebnissen veröffentlichter und nicht veröffentlichter Studien zeigt. Weitere Kriterien sind Studienabbrüche auf der Grundlage von hohen Effektschätzungen oder Ergebnisse, die auf kleinen Fallzahlen und wenigen Ereignissen basieren [38,39].

#### **C: Abwertung bei fehlender Genauigkeit der Effekte**

Die Qualität der Evidenz wurde wegen schwerwiegender oder sehr schwerwiegender Ungenauigkeiten der Effekte um 1 oder 2 Kategorien herabgestuft. Als primäres Kriterium wurde hierfür das KI herangezogen. Unterschieden sich die klinischen Schlussfolgerungen aus der oberen beziehungsweise unteren Grenze des KI (stark), wurde eine Abwertung wegen

(sehr) schwerwiegender fehlender Genauigkeit in Betracht gezogen. Bei kleinen Studien mit großen Effekten auf der Grundlage kleiner Ereigniszahlen konnte die Unterschreitung der Fallzahl für eine adäquat gepowerte Studie („optimal information size“) die Ursache einer Abwertung der Effektstärke wegen schwerwiegender Ungenauigkeiten sein [40,41].

#### **D: Abwertung bei inkonsistenten (heterogenen) Effekten (Inkonsistenz)**

Die Qualität der Evidenz sollte bei schwerwiegender unerklärter Heterogenität (Inkonsistenz) zwischen den Studienergebnissen um 1 oder 2 Kategorien herabgestuft werden. Kriterien für die Einschätzung einer möglichen Heterogenität der Studienergebnisse konnten eine große Varianz und unterschiedliche Richtungen der Effektschätzungen der Einzelstudien, eine geringe oder fehlende Überlappung der KIs und statistische Kriterien für eine bedeutsame Heterogenität wie ein geringer p-Wert oder ein hoher I<sup>2</sup>-Wert sein (Abschnitt 4.4.2). Bei einer hohen, durch die in Abschnitt 4.4.5 definierten Subgruppen erklärten Heterogenität sollte das Poolen der Effekte auf die betreffenden Subgruppen beschränkt werden [42,43].

#### **E: Abwertung bei eingeschränkter Übertragbarkeit (Indirektheit)**

Die Qualität der Evidenz wurde bei schwerwiegenden oder sehr schwerwiegenden Einschränkungen in der Übertragbarkeit auf die Zielpopulation um 1 oder 2 Kategorien herabgestuft werden. Eine eingeschränkte Übertragbarkeit konnte auf Abweichungen zwischen den PICO-Kriterien der Studien und den definierten Einschlusskriterien basieren. Beispielsweise wurden Abweichungen hinsichtlich des Zeitpunkts des Einschlusses der Gebärenden in die Studien oder der Versorgungsverhältnisse und der Qualifikationen der Hebammen beziehungsweise Entbindungspfleger zwischen Deutschland und den Ländern, in denen die Studien durchgeführt wurden, erwartet [44,45].

#### **F: Aufwertung**

Die Qualität der Evidenz sollte in methodisch guten Beobachtungsstudien und nicht randomisierten Interventionsstudien bei großen beziehungsweise sehr großen Effekten um 1 oder 2 Kategorien heraufgestuft werden. Eine Heraufstufung um 1 Kategorie war bei großen Effekten geplant, wenn das relative Risiko zwischen 2 und 5 beziehungsweise zwischen 0,2 und 0,5 liegt, ohne dass plausible Confounder diesen Effekt verringern könnten. Eine Heraufstufung um 1 weitere Kategorie sollte bei sehr großen Effekten erfolgen, wenn das relative Risiko über 5 beziehungsweise unter 0,2 liegt, keine relevanten Verzerrungsrisiken vorliegen und die Effekte präzise (mit hinreichend geringen KIs) geschätzt werden können. Eine weitere Bedingung für eine Heraufstufung der Qualität der Evidenz war eine Dosis-Wirkungs-Beziehung oder eine mögliche Verstärkung des beobachteten Ergebnisses bei Berücksichtigung aller potenziellen Confounder oder Verzerrungsquellen [46,47].

#### **4.4.4 Sensitivitätsanalysen**

Zur Einschätzung der Robustheit der Ergebnisse waren Sensitivitätsanalysen hinsichtlich methodischer Faktoren geplant. Die methodischen Faktoren sollten sich aus den im Rahmen der Informationsbeschaffung und -bewertung getroffenen Entscheidungen, zum Beispiel der

Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder der Wahl des Effektmaßes, bilden. Derartige Sensitivitätsanalysen sollten unabhängig von gegebenenfalls weiteren Analysen, mit denen die Ergebnissicherheit eines beobachteten Effekts bewertet wird, erfolgen.

Es wurde angenommen, dass das Ergebnis solcher Sensitivitätsanalysen die für einen beobachteten Effekt abgeleiteten Aussagen zur Qualität der Evidenz beeinflussen kann. Ein als nicht robust eingestuft Effekt kann laut den GRADE-Guidelines dazu führen, dass die Qualität der Evidenz für den betrachteten Endpunkt aufgrund von Inkonsistenz um 1 oder 2 Kategorien abgewertet wird [28,42].

#### **4.4.5 Subgruppenmerkmale und andere Effektmodifikatoren**

Es war geplant, die Ergebnisse hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, zu untersuchen. Dies konnten direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Interventionen sein. Im Gegensatz zu den in Abschnitt 4.4.4 beschriebenen methodischen Faktoren für Sensitivitätsanalysen bestand hier das Ziel, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Für einen Nachweis unterschiedlicher Effekte sollte die auf einem Homogenitäts- beziehungsweise Interaktionstest basierende statistische Signifikanz Voraussetzung sein. In die Untersuchung von Effektmodifikatoren sollten die vorliegenden Ergebnisse aus Regressionsanalysen, die Interaktionsterme beinhalten, und aus Subgruppenanalysen einbezogen werden. Außerdem sollten eigene Analysen in Form von Metaregressionen oder Metaanalysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren erfolgen. Es war geplant, folgende Faktoren bezüglich einer möglichen Effektmodifikation in die Analysen einzubeziehen:

- Alter der Gebärenden
- Parität (Erst- vs. Mehrgebärende)
- Einschluss von Schwangeren mit besonderen Risiken (z. B. Zustand nach traumatischem Geburtserlebnis, vorheriger Kaiserschnitt, Mehrlinge, psychosoziale Belastung, Adipositas der Gebärenden, Migrationshintergrund)

Es war geplant, sofern sich aus den verfügbaren Informationen weitere mögliche Effektmodifikatoren ergeben hätten, diese ebenfalls begründet einzubeziehen. Jedoch konnten für keine der oben beschriebenen Subgruppen ausreichend Informationen für eine Subgruppenanalyse extrahiert werden.

Bei Identifizierung möglicher Effektmodifikatoren war eine Präzisierung der für die beobachteten Effekte abgeleiteten Aussagen zur Qualität der Evidenz geplant. Beispielsweise konnten Ergebnisse für die Subgruppen getrennt berechnet und die Qualität der Evidenz für spezielle Subgruppen von Gebärenden eingeschränkt beurteilt werden.

#### 4.4.6 Evidenzprofil

Für die Fragestellung wurde die Evidenz für alle kritischen und wichtigen Endpunkte aus allen vorliegenden Studien in einem Evidenzprofil zusammengefasst [28,48-50]. Dieses enthält für die Fragestellung und gegebenenfalls die in Abschnitt 4.4.5 genannten Subgruppen folgende Informationen:

- die Anzahl der Studien und der Designs
- eine detaillierte Bewertung der Qualität der Evidenz zu den Punkten Studienlimitationen, Inkonsistenz, Indirektheit, fehlende Genauigkeit und anderen Faktoren einschließlich des Publikationsbias
- die Summe der Anzahl der Ereignisse und der eingeschlossenen Gebärenden in den Interventions- und Vergleichsgruppen aller eingeschlossenen Studien
- die (relativen) Effektschätzungen für die beobachteten Interventionen mit ihren jeweiligen KIs
- Den Unterschied zwischen den Risiken mit und ohne Intervention [51] (und die entsprechenden KIs). Dieser basiert auf dem Populationsrisiko und der relativen Effektschätzung und ihrem KI. Die typische Häufigkeit des Endpunkts in der deutschen Bevölkerung (Populationsrisiko) basiert auf repräsentativen Daten der stationären Versorgung in Deutschland [5]. Bei dichotomen Endpunkten sollte die Ausprägung, in Abhängigkeit von der Häufigkeit der beschriebenen Ereignisse, auf 100 Geburten bei einem Populationsrisiko über 10 %, auf 1000 bei einem Populationsrisiko unter 10 % oder für sehr seltene Ereignisse auf 10 000 Geburten bezogen werden. Das Ziel war die Angabe ganzer Zahlen.
- Die Einteilung der Qualität der Evidenz nach den in Abschnitt 4.4.3 genannten GRADE-Kriterien in 1 der Kategorien hoch, moderat, niedrig und sehr niedrig. Die Gründe der Auf- oder Abwertung werden in Fußnoten nachvollziehbar dargestellt [35,52].
- Kommentare wie relevant abweichende Ergebnisse aus den in Abschnitt 4.4.4 genannten Sensitivitätsanalysen, die in Fußnoten nachvollziehbar dargestellt werden sollten

Untersuchte eine der eingeschlossenen Studien nicht alle der vorher festgelegten Endpunkte, wurde dies in den Tabellen beschrieben. Es war geplant, die Ergebnisse aus Sensitivitätsanalysen zu nennen, wenn sie zu stark abweichenden Effektschätzungen führen. Wenn Metaanalysen nicht möglich waren, wurde die vorhandene Evidenz für diese Endpunkte dargestellt.

## **5 Ergebnisse**

### **5.1 Umfassende Informationsbeschaffung**

#### **5.1.1 Bibliografische Datenbanken**

Abbildung 1 zeigt das Ergebnis der systematischen Literaturrecherche in den bibliografischen Datenbanken und der Studienselektion gemäß den Kriterien für den Studieneinschluss.

Die Suchstrategien für die Suche in bibliografischen Datenbanken finden sich in Anhang A.1. Die letzte Suche fand am 28.09.2017 statt.

Die Referenzen der als Volltexte geprüften, aber ausgeschlossenen Treffer finden sich mit Angabe des jeweiligen Ausschlussgrundes in Abschnitt 10.3.

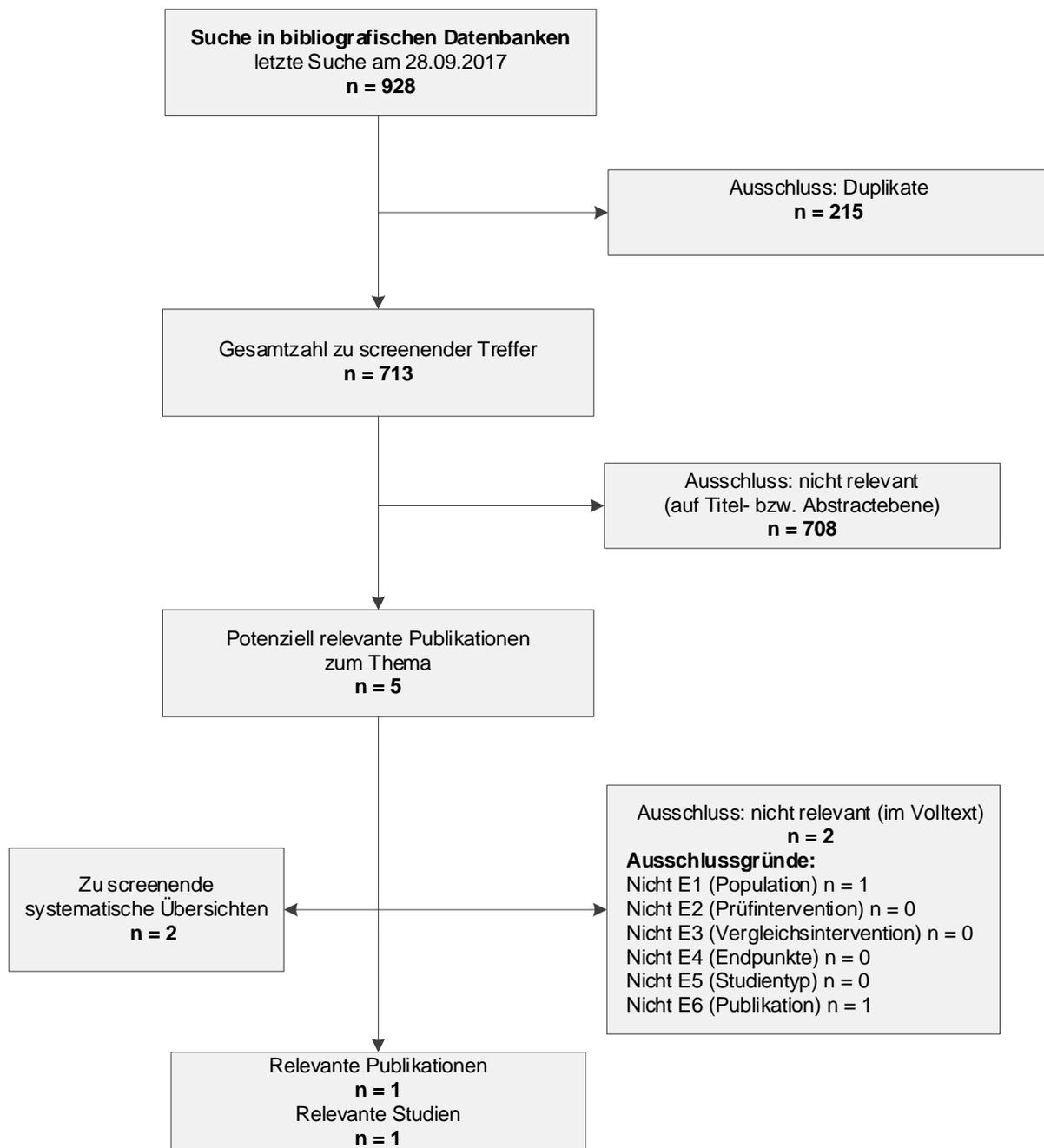


Abbildung 1: Ergebnis der bibliografischen Recherche und der Studienselektion

### 5.1.2 Studienregister

Durch die Suche in Studienregistern wurden keine relevanten Studien beziehungsweise Dokumente identifiziert.

In den Studienregistern wurden 2 laufende und 2 abgeschlossene Studien ohne berichtete Ergebnisse identifiziert (siehe Abschnitt 5.1.5).

Die Suchstrategien für die Suche in Studienregistern finden sich in Anhang A.2. Die letzte Suche in Studienregistern fand am 29.09.2017 statt.

### **5.1.3 Weitere Informationsquellen und Suchtechniken**

Über weitere Informationsquellen und Suchtechniken identifizierte relevante Studien beziehungsweise Dokumente werden nachfolgend nur dargestellt, wenn sie nicht bereits über die primären Informationsquellen gefunden wurden.

#### **5.1.3.1 Studiengruppen und Fachgesellschaften**

Die schriftlich erfolgten Anfragen bei den Fachgesellschaften und Studiengruppen ergaben keine relevanten Studien beziehungsweise Dokumente, die nicht über andere Rechenschritte identifiziert werden konnten.

#### **5.1.3.2 Anwendung weiterer Suchtechniken**

Im Rahmen der Informationsbeschaffung wurden systematische Übersichten identifiziert – die entsprechenden Referenzen finden sich in Abschnitt 10.2. Die Referenzlisten dieser systematischen Übersichten wurden gesichtet.

Es fanden sich keine relevanten Studien beziehungsweise Dokumente, die nicht über andere Rechenschritte identifiziert werden konnten.

#### **5.1.3.3 Autorenanfragen**

Für die vorliegende Evidenzbewertung wurden Autorenanfragen versendet (siehe Tabelle 3). Die Informationen aus den eingegangenen Antworten sind in die Studienbewertung eingeflossen.

Tabelle 3: Übersicht über Autorenanfragen

Studie / Studienregister-ID	Inhalt der Anfrage	Antwort eingegangen ja / nein	Inhalt der Antwort
Hodnett 2009 [53]	<ul style="list-style-type: none"> <li>▪ Frage, ob weitere für diesen Rapid Report relevante Endpunkte erhoben wurden, Frage nach der Gesamtanzahl der Antworten auf die Einzelfragen zum Geburtserleben sowie nach Angaben zum Gestationsalter der Schwangeren.</li> <li>▪ Frage nach weiteren bekannten relevanten Studien</li> </ul>	ja	<ul style="list-style-type: none"> <li>▪ Ergebnisse zu weiteren Endpunkten liegen nicht vor. Die Ergebnisse aller erhobenen Endpunkte werden in der Publikation beschrieben. Keine Antwort auf die Frage nach Angaben zum Gestationsalter der Teilnehmerinnen.</li> <li>▪ Verweis auf eine anlaufende Studie in Deutschland (Ayerle 2017)</li> </ul>
Ayerle 2017 [54]	<ul style="list-style-type: none"> <li>▪ Frage nach allgemeinen Informationen zum methodischen Vorgehen und zum aktuellen Status</li> </ul>	ja	<ul style="list-style-type: none"> <li>▪ Beginn der Studie: 01.10.2017</li> <li>▪ Beginn der Rekrutierung: 01.04.2018</li> <li>▪ geplantes Studienende: 30.09.2020</li> <li>▪ Studienprotokoll wurde zur Verfügung gestellt</li> </ul>
IRCT138810 203034N1 [55]	<ul style="list-style-type: none"> <li>▪ Frage, ob bereits Ergebnisse aus der abgeschlossenen Studie vorliegen, die für diesen Bericht zur Verfügung gestellt werden können</li> <li>▪ Frage nach weiteren bekannten relevanten Studien</li> </ul>	nein	-
IRCT201312 073034N13 [56]	<ul style="list-style-type: none"> <li>▪ Frage, ob bereits Ergebnisse aus der abgeschlossenen Studie vorliegen, die für diesen Bericht zur Verfügung gestellt werden können</li> <li>▪ Frage nach weiteren geeigneten Studien</li> </ul>	nein	-
Manesh 2015 [57]	<ul style="list-style-type: none"> <li>▪ Laut den Angaben in der Publikation bestand eine große Differenz zwischen der SSW der Interventions- und Vergleichsgruppe (<math>37,7 \pm 1,7</math> vs. <math>35,4 \pm 1,7</math>), die nicht dem Ergebnis des t-Tests entsprach. Frage nach differenzierten Angaben zur SSW der Teilnehmerinnen</li> <li>▪ Frage, ob weitere für diesen Rapid Report relevante Endpunkte erhoben wurden</li> <li>▪ Frage nach weiteren geeigneten Studien</li> </ul>	nein	-
NCT024783 85 [58]	<ul style="list-style-type: none"> <li>▪ Frage, ob bereits Ergebnisse aus der noch laufenden Studie vorliegen, die für diesen Bericht zur Verfügung gestellt werden können</li> </ul>	nein	-

### 5.1.3.4 Zusätzliche relevante Studien bzw. Dokumente

Es wurde folgende relevante Studie beziehungsweise folgendes Dokument identifiziert, die nicht über andere Rechenschritte gefunden werden konnte (Tabelle 4):

Tabelle 4: Zusätzlich identifizierte relevante Studie bzw. identifiziertes Dokument

Studie	Verfügbare Dokumente [Zitat]
Aburas 2017	Vollpublikation [59]

### 5.1.4 Resultierender Studienpool

Durch die verschiedenen Suchschritte konnten insgesamt 2 relevante Studien (2 Dokumente) identifiziert werden (siehe Tabelle 5). Die entsprechenden Referenzen finden sich in Abschnitt 10.1.

Tabelle 5: Studienpool der Evidenzbewertung

Studie	Verfügbare Dokumente	
	Vollpublikation (in öffentlich zugänglichen Fachzeitschriften)	Ergebnisbericht aus Studienregistern beziehungsweise Eintrag Studienregister
Aburas 2017	ja [59]	nein
Hodnett 2009	ja [53]	nein

### 5.1.5 Studien ohne berichtete Ergebnisse

In Tabelle 6 sind alle durch die Informationsbeschaffung identifizierten Studien ohne bisher berichtete Ergebnisse dargestellt. Alle Autorinnen und Autoren wurden kontaktiert.

Tabelle 6: In der Informationsbeschaffung identifizierte Studien ohne berichtete Ergebnisse

Studie	Dokumentart, ggf. Studienregister-ID, Zitat	Studientyp	Geplante Fallzahl	Status (ggf. geplantes Studienende)
Ayerle 2017	Projekt-Homepage [54], vertrauliches Studienprotokoll [60]	RCT	4000	laufend (30.09.2020)
NCT02478385	ClinicalTrials.gov [58]	RCT	656	laufend (12/2017)
IRCT138810203034N1	Iranian Registry of Clinical Trials [55]	RCT	100	abgeschlossen (22.05.2011)
IRCT201312073034N13	Iranian Registry of Clinical Trials [56]	RCT	130	abgeschlossen (20.02.2016)

Im Oktober 2017 hat die Durchführung einer multizentrischen randomisierten kontrollierten Studie in 12 geburtshilflichen Einrichtungen in Sachsen-Anhalt, Nordrhein-Westfalen,

Thüringen, Sachsen und Berlin zum Effekt der Geburtsumgebung auf den Geburtsmodus und das Wohlbefinden von Frauen am Geburtstermin (BE-UP) begonnen (Ayerle 2017). Diese Studie soll ab April 2018 über 23 Monate insgesamt 4000 Erst- und Mehrgebärende mit einer Einlingsschwangerschaft zwischen der vollendeten 37. und 41 + 6 SSW, die sich in der aktiven EP befinden und eine vaginale Geburt anstreben, rekrutieren. Untersucht werden soll, inwieweit eine alternative Gestaltung des Gebärraumes im Vergleich zum Standardkreißaal einen Einfluss auf den Geburtsmodus hat. Dieser Gebärraum enthält kein oder ein verdecktes Kreißbett, eine Matratze auf dem Boden, verschiedene Kissenarten, einen Sitzsack, einen Gebärstuhl, einen Tisch und Stühle einschließlich einer Snackbar, Poster mit abgebildeten aufrechten Geburtspositionen, einen DVD-Bildschirm mit Naturszenen und ausreichend Platz für Bewegung. Primärer Endpunkt dieser Studie ist eine vaginale Geburt, sekundäre Endpunkte umfassen die maternale Selbstbestimmung, das Auftreten höhergradiger Dammriss (III° / IV°), eine Episiotomie, einen hohen mütterlichen Blutverlust  $\geq 1000$  ml, schwere maternale Morbidität und einen kritischen Zustand des Neugeborenen sowie schwerwiegende geburtshilfliche Komplikationen.

In Dänemark wird monozentrisch eine weitere randomisierte Studie durchgeführt (NCT02478385), die den Einfluss der Geburtsumgebung auf das Geburtserleben und relevante Geburtsendpunkte untersucht. Diese Studie plante, von Mai 2015 bis Dezember 2017 insgesamt 656 Erstgebärende mit einer Einlingsschwangerschaft zwischen der vollendeten 37. und 42. SSW zu rekrutieren. Auch diese Studie vergleicht eine alternative Gestaltung des Gebärraumes mit der derzeitigen Standardgestaltung. Diese Gestaltung bietet den werdenden Eltern die Möglichkeit der Selbstregulierung von Geräusch- und Lichteffekten. Primärer Endpunkt dieser Studie ist der Einsatz von Wehenmitteln.

## **5.2 Charakteristika der in die Bewertung eingeschlossenen Studien**

### **5.2.1 Studiendesign und Studienpopulationen**

Es wurden 2 RCTs zur Bewertung einer modifizierten Umgebungsgestaltung eingeschlossen (Aburas 2017, Hodnett 2009). Die Charakteristika sind in Tabelle 7 dargestellt.

Die Studien wurden von März 2007 bis April 2008 beziehungsweise Februar bis Mai 2014 durchgeführt. Bei Hodnett 2009 handelt es sich um eine multizentrische Machbarkeitsstudie aus Kanada, während Aburas 2017 die Studienteilnehmerinnen in einem Zentrum in den USA rekrutierte. Insgesamt wurden 122 Gebärende eingeschlossen, von denen 55 einem modifizierten Gebärraum und 57 einem Gebärraum, der dem üblichen Standard in der jeweiligen Studie entsprach, zugewiesen wurden. In Aburas 2017 fehlen Informationen zu 10 randomisierten, aber nicht eingeschlossenen Teilnehmerinnen, ohne das hierfür Gründe angegeben werden. In Hodnett 2009 werden für 1 randomisierte Teilnehmerin keine Ergebnisse berichtet und als Grund eine widerrufenen Einverständniserklärung angegeben.

In Hodnett 2009 wurden Frauen unter der Geburt bei bevorstehender Aufnahme in den Kreißaal mit einer Einlingsschwangerschaft am Termin, einem Fetus in Schädellage sowie

ohne Kontraindikationen für eine vaginale Geburt eingeschlossen. Ausgeschlossen wurden Frauen aufgrund medizinisch bedingter Indikationen oder von Präferenzen, die die Mobilität beeinträchtigen könnten, z. B. bei einem Wunsch nach unmittelbarer Periduralanästhesie (PDA). Da die Studie in Kanada durchgeführt wurde, wurden systembedingt weiterhin Frauen, die von einer Hebamme oder Doula<sup>1</sup> betreut wurden, ausgeschlossen, da laut den Autorinnen und Autoren der Effekt auf das Verhalten der Krankenschwestern untersucht werden sollte. In Aburas 2017 finden sich bezüglich der Einschlusskriterien nur die Angaben, dass ausschließlich Frauen älter als 18 Jahre mit ausreichenden Englischkenntnissen und auch Frauen mit einem vorherigen Kaiserschnitt, sofern sie eine vaginale Geburt anstrebten, eingeschlossen wurden. Es werden keine Informationen berichtet, aus denen hervorgeht, zu welchem Zeitpunkt die Frauen eingeschlossen wurden, jedoch beschreiben die Autorinnen und Autoren, dass der Einfluss der Intervention auf Frauen während der Geburt untersucht werden sollte. Daher wurde die Studie bezüglich der Population als geeignet bewertet. Ausgeschlossen wurden Frauen, die einen primären Kaiserschnitt planten, wenn ein Kaiserschnitt bereits voraussehbar war oder wenn die Frau bereits mehr als einen Kaiserschnitt aufwies. Frauen mit einem Zustand nach einer Frühgeburt, bekannten fetalen Anomalien oder Frauen mit einem generellen höheren Risiko, ohne nähere Differenzierung durch die Autorinnen und Autoren, wurden ebenfalls ausgeschlossen.

In Hodnett 2009 betrug der Anteil der Erstgebärenden 71 %. In Aburas 2017 werden von insgesamt 40 Teilnehmerinnen mit Angaben zu Baselinecharakteristika 23 (58 %) als Erstgebärende beschrieben, Angaben zu weiteren 10 Teilnehmerinnen, von denen Ergebnisse berichtet werden, fehlen. In beiden Studien fehlen Informationen zur genauen SSW der Teilnehmerinnen. Ausgehend von einer normalen Schwangerschaftsdauer von 40 Wochen (40 + 0 SSW) und einer normalverteilten Abweichung vom Geburtstermin, nach der nur 17,5 % aller Frauen mehr als 10 Tage vor dem errechneten Geburtstermin gebären [8], wurde auf Grundlage der genannten Einschlusskriterien davon ausgegangen, dass mindestens 80 % der Teilnehmerinnen nach der vollendeten 37. SSW gebären.

Beide Studien untersuchten eine modifizierte Umgebungsgestaltung im Vergleich zum üblichen Kreißaal. In Hodnett 2009 gehörten zur modifizierten Umgebungsgestaltung umfangreiche Komponenten. Als Hauptmerkmal wird beschrieben, dass das Kreißbett aus dem Raum entfernt und durch eine Doppelmatratze mit großen Kissen in einer Ecke ersetzt wurde. Dies sollte den Teilnehmerinnen eine freie Positionswahl und einen engen Kontakt mit den Betreuungspersonen ermöglichen sowie den Routineeinsatz einer elektronischen Überwachung der fetalen Herzfrequenz limitieren. Zudem wurde ein Hinweisschild mit der Bitte, vor dem Betreten des Raumes anzuklopfen, angebracht, medizinische Geräte entfernt oder abgedeckt, Möbel, die eine aufrechte Position unterstützen sollten, bereitgestellt, beruhigende Naturbilder auf eine weiße Wand projiziert, das Licht gedimmt sowie eine große Musikauswahl zur Verfügung gestellt. Es wurde eine Auskultation der fetalen Herzfrequenz

---

<sup>1</sup> Begleitperson ohne medizinische Ausbildung, die eine werdende Mutter vor, während und nach der Geburt emotional und physisch begleitet

vorgenommen und eine elektronische Herzfrequenzüberwachung erfolgte nur, wenn diese als medizinisch notwendig erachtet wurde.

Der Standard-Gebärraum enthielt ein Gebärbett, das zentral im Raum positioniert war, kein Hinweisschild, eine Infusionsstange, angeschaltete Deckenbeleuchtung, klinikübliche Umgebungsgeräusche und die Herzfrequenz des Fetus wurde als Routinekontrolle durchgeführt, in der Regel mittels kontinuierlicher elektronischer Herzfrequenzmessung. Die Komponenten sind in Tabelle 7 detailliert beschrieben.

In Aburas 2017 bestand die modifizierte Umgebungsgestaltung aus einer geringen Veränderung in Form von Naturbildern, die über einen gesonderten Fernschirmschirm im Gebärraum angezeigt wurden. Die Naturbilder zeigten Bäume, Blumen, Wasser oder andere natürliche Umgebungen. Die Teilnehmerinnen konnten mittels Fernbedienung die Darstellung je nach Wunsch steuern, andernfalls wechselte der Bildschirm alle 30 Sekunden das Bild. Im Standard-Gebärraum war kein zusätzlicher Fernschirmschirm mit Naturbildern vorhanden. Beide Gruppen konnten darüber hinaus einen normalen Fernseher nutzen.

Bei Hodnett 2009 handelt es sich um eine multizentrisch durchgeführte Studie, die in 2 Krankenhäusern durchgeführt wurde, die sich laut Angabe in der Publikation sowohl in der Größe als auch in der Einstellung des Personals gegenüber einer modifizierten Umgebungsgestaltung deutlich unterschieden. 65 % der Gesamtanzahl der Teilnehmerinnen gebaren in dem Krankenhaus mit einer positiveren Einstellung zur modifizierten Umgebungsgestaltung. Die von Aburas 2017 veröffentlichte Studie wurde in 8 von 17 Kreißsälen des Geburtszentrums eines Universitätsklinikums durchgeführt.

Tabelle 7: Charakteristika der eingeschlossenen Studien: Evidenztabelle

Quelle / Studientyp / Rekrutierungsland / Rekrutierungszeitraum / Ziel	Ein- / Ausschlusskriterien	Charakteristika <sup>a</sup> der Population	Intervention <sup>b</sup> (Interventionsgruppe) vs. Vergleichsintervention (Vergleichsgruppe)	Verzerrungspotenzial auf Studienebene <sup>c</sup>	Evidenzniveau nach OCEBM
<p><b>Hodnett 2009</b> RCT, multizentrisch</p> <p>Kanada 03/2007–04/2008</p> <p>Machbarkeitsstudie zur umfangreichen modifizierten Umgebungsgestaltung</p>	<ul style="list-style-type: none"> <li>▪ Eingeschlossen wurden Frauen mit einer Einlingsschwangerschaft in Schädellage am Termin unter der Geburt vor der Aufnahme in den Kreißsaal ohne Kontraindikation für eine vaginale Geburt.</li> <li>▪ Ausschluss von Frauen mit mobilitätseinschränkenden medizinischen Indikationen (z. B. Komplikationen oder Notwendigkeit einer intravenösen Infusion) oder Präferenzen (z. B. Wunsch nach einer zeitnahen PDA), Frauen unter kontinuierlicher Betreuung durch eine Hebamme oder Doula</li> </ul>	<ul style="list-style-type: none"> <li>▪ N = 62</li> <li>▪ Erstgebärende: 71 %</li> <li>▪ höchster Bildungsgrad:                             <ul style="list-style-type: none"> <li>▫ &lt; Abitur: 3 %</li> <li>▫ Abitur: 8 %</li> <li>▫ akademische Ausbildung: 68 %</li> <li>▫ unbekannt: 21 %</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>▪ Umfangreich modifizierter Gebärraum (n = 31) mit                             <ul style="list-style-type: none"> <li>▫ geschlossener Zimmertür mit Hinweisschild, vor Eintritt anzuklopfen</li> <li>▫ aus dem Gebärraum entferntem Kreißbett, das bei Notwendigkeit zurückgeholt werden kann</li> <li>▫ breiter Matratze für 2 Personen mit bunten Laken und großen Kissen in einer Ecke auf dem Boden</li> <li>▫ abwesenden oder verdeckten medizinischen Geräten</li> <li>▫ einem Stuhl / Pezziball zur Unterstützung des aufrechten / nach vorn geneigten Sitzens im Raum</li> <li>▫ Postern, die aufrechte Gebärhaltungen zeigen</li> <li>▫ einer Routinekontrolle der Herzfrequenz des Fetus über eine Auskultation, kontinuierliche elektronische Herzfrequenzmessung nur, wenn als medizinisch notwendig erachtet</li> <li>▫ einem Overheadprojektor mit eingebautem DVD-Player mit verschiedenen Filmen, die beruhigende Naturbilder auf eine Wand projizieren</li> <li>▫ ausgeschalteter Deckenbeleuchtung und gedimmtem Licht</li> <li>▫ fehlendem Außenfenster</li> </ul> </li> </ul>	<p>mittel (wegen fehlender Planung und Differenzierung zwischen primären und sekundären Endpunkten)</p>	<p>3</p>

(Fortsetzung)

Tabelle 7: Charakteristika der eingeschlossenen Studien: Evidenztabelle (Fortsetzung)

Quelle / Studientyp / Rekrutierungsland / Rekrutierungszeitraum / Ziel	Ein- / Ausschlusskriterien	Charakteristika <sup>a</sup> der Population	Intervention <sup>b</sup> (Interventionsgruppe) vs. Vergleichsintervention (Vergleichsgruppe)	Verzerrungspotenzial auf Studienebene <sup>c</sup>	Evidenzniveau nach OCEBM
			<ul style="list-style-type: none"> <li>▫ einem MP3-Player mit großer Musikauswahl verschiedener Stilrichtungen (nutzbar über Kopfhörer oder Lautsprecher), Variabilität des Tempos korrespondierend zu den Variationen der Schmerzintensität von frühe bis späte Geburt</li> <li>vs.</li> <li>▪ Standardgeburtsraum (n = 31) mit                         <ul style="list-style-type: none"> <li>▫ geschlossener Zimmertür ohne Hinweisschild</li> <li>▫ einem Gebärbett, das zentraler Gegenstand des Raumes ist</li> <li>▫ Standardmöblierung mit Clubsessel, Infusionsstange</li> <li>▫ einem Pezziball auf der Station</li> <li>▫ einer Routinekontrolle der Herzfrequenz des Fetus für gewöhnlich mittels kontinuierlicher elektronischer Herzfrequenzmessung</li> <li>▫ (meistens) mindestens einem Bild an der Wand</li> <li>▫ fehlendem Außenfenster</li> <li>▫ einer Deckenbeleuchtung, es sei denn, die Frau schläft</li> <li>▫ normalen Umgebungsgeräuschen, Musik hängt von dem mitgebrachten Equipment der Frau ab</li> </ul> </li> </ul>		

(Fortsetzung)

Tabelle 7: Charakteristika der eingeschlossenen Studien: Evidenztabelle (Fortsetzung)

Quelle / Studientyp / Rekrutierungsland / Rekrutierungszeitraum / Ziel	Ein- / Ausschlusskriterien	Charakteristika <sup>a</sup> der Population	Intervention <sup>b</sup> (Interventionsgruppe) vs. Vergleichsintervention (Vergleichsgruppe)	Verzerrungspotenzial auf Studienebene <sup>c</sup>	Evidenzniveau nach OCEBM
<p><b>Aburas 2017</b> RCT, monozentrisch</p> <p>USA 02–05/2014</p> <p>visuelle Stimulation zur Verbesserung des Geburtserlebens und klinischer Parameter</p>	<ul style="list-style-type: none"> <li>▪ Eingeschlossen wurden Frauen ≥ 18 Jahre mit ausreichend Englischkenntnissen sowie Frauen mit einem vorherigen Kaiserschnitt, die eine vaginale Geburt anstrebten.</li> <li>▪ Ausgeschlossen wurden Frauen &lt; 18 Jahre, die keine ausreichenden Englischkenntnisse besaßen, einen primären Kaiserschnitt planten, dieser wahrscheinlich vorzuzusehen war oder die bereits mehr als einen Kaiserschnitt aufwiesen. Ausgeschlossen wurden auch Frauen mit einem Zustand nach Frühgeburt, bekannten fetalen Anomalien oder Frauen mit einem generellen höheren Risiko.</li> </ul>	<ul style="list-style-type: none"> <li>▪ N = 60<sup>d</sup></li> <li>▪ Erstgebärende: 58 %</li> <li>▪ Alter:                             <ul style="list-style-type: none"> <li>▫ &lt; 20: 10 %</li> <li>▫ 20–29: 65 %</li> <li>▫ 30–39: 22 %</li> <li>▫ 40–49: 2,5 %</li> </ul> </li> <li>▪ verheiratet: 72 %</li> <li>▪ Single: 25 %</li> <li>▪ geschieden: 2,5 %</li> </ul>	<ul style="list-style-type: none"> <li>▪ Darstellung von Naturbildern (n = 24)                             <ul style="list-style-type: none"> <li>▫ Naturbilder von Bäumen, Blumen, Wasser oder anderen natürlichen Umgebungen, dargestellt auf einem Studien-Fernsehbildschirm mit Zugang zum normalen Fernseher</li> <li>▫ Teilnehmerinnen konnten Darstellung mittels Fernbedienung je nach Wunsch steuern, ansonsten wechselte das Bild alle 30 Sekunden</li> </ul> </li> <li>vs.</li> <li>▪ keine Darstellung von Naturbildern (n = 26)                             <ul style="list-style-type: none"> <li>▫ ausschließlich Zugang zum normalen Fernseher</li> </ul> </li> </ul>	<p>hoch (wegen fehlender Angaben zur Randomisierung, selektiven Berichtens signifikanter Ergebnisse und intransparenten Teilnehmerinnenflusses)</p>	<p>3</p>
<p>a: Extrahiert wurden alle berichteten Informationen zu den Faktoren Alter, Parität, Einlinge / Mehrlinge, SSW, Adipositas, vorheriger Kaiserschnitt, psychosoziale Belastung, Migrationshintergrund, Sozialstatus, PDA und Risikoprofil.                      b: Die Angaben enthalten die Beschreibung der modifizierten Umgebung.                      c: ausführliche Beschreibung siehe Abschnitt 5.3                      d: Studiencharakteristika werden für 40 Studienteilnehmerinnen beschrieben.</p> <p>N: Anzahl randomisierter Gebärender insgesamt; n: Anzahl randomisierter Gebärender je Interventionsgruppe, für die Ergebnisse vorliegen; OCEBM: Oxford Centre for Evidence-based Medicine; PDA: Periduralanästhesie; RCT: Randomized controlled Trial; SSW: Schwangerschaftswoche</p>					

## 5.2.2 Übersicht der bewertungsrelevanten Endpunkte

Tabelle 8 zeigt die Übersicht der verfügbaren Daten zu den von der Leitliniengruppe festgelegten, für die Entscheidung kritischen 10 maternalen und 5 neonatalen Endpunkten aus den eingeschlossenen Studien. Die Endpunkte Spontangeburt, vaginal-operative Geburt und Kaiserschnitt wurden aufgrund der gegenseitigen Abhängigkeit voneinander zu dem Endpunkt Geburtsmodus zusammengefasst.

Zu den Endpunkten maternale und fetale Mortalität, höhergradige Dammmrisse (III° / IV°), hoher mütterlicher Blutverlust  $\geq 1000$  ml, Stillbeziehung bei initialem Stillwunsch, teilweise / vollständige Plazentaretention, subjektive Gesundheitseinschätzung, schwere metabolische Acidose, Basendefizit  $\geq -16$  mmol/l und Verlegung auf die Neugeborenen-Intensivstation wurden keine Daten berichtet. Der Apgar-Score wurde in Hodnett 2009 in die Kategorien Apgar-Score nach 5 Minuten  $< 7$  und Apgar-Score nach 5 Minuten  $\geq 7$  eingeteilt und in Aburas 2017 liegen ausschließlich Mittelwertangaben vor, sodass zum kritischen Endpunkt Apgar-Score nach 5 Minuten  $< 6$  keine Ergebnisse vorlagen. Die vorliegenden ergänzenden Informationen werden dargestellt, aber nicht in die Evidenzbewertung einbezogen.

Aus diesem Grund verbleiben Informationen zu 2 für die Entscheidung kritischen Endpunkten für die GRADE-Bewertung: Geburtsmodus und Geburtserleben. In Aburas 2017 liegen zum Geburtsmodus ausschließlich Informationen zum Kaiserschnitt vor, da eine Differenzierung zwischen einer Spontangeburt und einer vaginal-operativen Geburt nicht vorgenommen wurde. Das Geburtserleben wurde in beiden Studien mit einem nicht validierten Fragebogen erhoben. Die vorliegenden Ergebnisse werden als ergänzende Informationen in die Bewertung einbezogen.

Es konnten Ergebnisse zu 1 klinisch wichtigen, aber für die Empfehlung nicht entscheidenden Endpunkt (Wehenmittel) extrahiert werden (Hodnett 2009). Zu den weiteren Endpunkten (Episiotomie, Amniotomie, Kristellerhilfe, Infektionen, leichte, mittelgradige oder fortgeschrittene Acidose, Ikterus, Anämie) wurden keine Daten berichtet.

Tabelle 8: Matrix der kritischen und klinisch wichtigen Endpunkte

Studie	Kritische Endpunkte										Klinisch wichtige Endpunkte												
	Maternal										Neonatal												
	Geburtsmodus																						
	Mortalität	Spontangeburt	Vaginal-operative Geburt	Kaiserschnitt	Höhergradige Dammrisse (III° / IV°)	Hoher mütterlicher Blutverlust ≥ 1000 ml	Stillbeziehung bei initialem Stillwunsch	Teilweise / vollständige Plazentaretention	Geburtserleben	Subjektive Gesundheitseinschätzung	Mortalität	Apgar-Score nach 5 Minuten < 6	Schwere metabolische Acidose	Basendefizit ≥ -16 mmol/l	Verlegung auf die NICU	Episiotomie	Wehenmittel	Amniotomie	Kristellerhilfe	Infektionen	Leichte, mittelgradige, fortgeschrittene Acidose	Ikterus	Anämie
Aburas 2017	-	-	-	■	-	-	-	-	● <sup>a</sup>	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Hodnett 2009	-	■	■	■	-	-	-	-	■ <sup>a</sup>	-	-	-	-	-	-	■	-	-	-	-	-	-	-

● primärer Endpunkt; ■ weitere / sekundäre Endpunkte; -: der Endpunkt wurde nicht erhoben  
a: Die Erhebung dieses Endpunkts erfolgte mit nicht validierten Fragebogen und wird daher nur ergänzend dargestellt.  
NICU: Neugeborenen-Intensivstation

### **5.3 Einschätzung des Verzerrungspotenzials**

#### **5.3.1 Einschätzung des Verzerrungspotenzials auf Studienebene**

Für die Bewertung lagen 2 Studien vor. Die Einschätzung des Verzerrungspotenzials auf Studienebene ist in Tabelle 9 dargestellt.

In Hodnett 2009 lagen eine adäquate Erzeugung der Randomisierungssequenz sowie eine verdeckte Gruppenzuteilung vor. Eine Verblindung der Teilnehmerinnen und der behandelnden Personen war generell nicht möglich. Diese fehlende Verblindung wurde bei der Beurteilung des Verzerrungspotenzials auf Endpunktebene in Abschnitt 5.3.2 berücksichtigt. Die ergebnisunabhängige Berichterstattung wurde als unklar bewertet, da keine Differenzierung der geplanten Endpunkte in primäre und sekundäre Endpunkte vorliegt. In der Folge wurde das Verzerrungspotenzial auf Studienebene für Hodnett 2009 als mittel bewertet.

In Aburas 2017 ist die Erzeugung der Randomisierungssequenz ebenso wie die verdeckte Gruppenzuteilung unklar, da Angaben zur Erzeugung der Zuteilungssequenz fehlen und keine ausreichenden Informationen zur Verdeckung vorliegen. Zudem beschreiben die Autorinnen und Autoren in Bezug auf die Limitationen der Studie Probleme während der Randomisierung, ohne diese näher zu erläutern. Eine Verblindung der Teilnehmerinnen und der behandelnden Personen war generell nicht möglich. Diese fehlende Verblindung wurde bei der Beurteilung des Verzerrungspotenzials auf Endpunktebene in Abschnitt 5.3.2 berücksichtigt. Die ergebnisunabhängige Berichterstattung wurde als nicht gegeben bewertet, da für den Endpunkt Geburtserleben selektiv Subskalen mit als statistisch signifikant beschriebenen Ergebnissen berichtet wurden. Zudem liegen sonstige Aspekte vor, die mit Verzerrungen der Studienergebnisse einhergehen. So liegt ein intransparenter Teilnehmerinnenfluss mit relevanten Dateninkonsistenzen vor. Nach der Randomisierung wurden Daten von 10 Teilnehmerinnen aus der Analyse ausgeschlossen, ohne dass für dieses Vorgehen eine Begründung angegeben wird. Zudem liegen zu insgesamt nur 40 Teilnehmerinnen Informationen zu den Charakteristika der Studienteilnehmerinnen vor. In der Folge wurde das Verzerrungspotenzial auf Studienebene für die Studie Aburas 2017 als hoch bewertet.

Tabelle 9: Bewertung des Verzerrungspotenzials auf Studienebene

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung der Teilnehmerinnen	Verblindung der behandelnden Personen	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial auf Studienebene
Aburas 2017	unklar	unklar	nein	nein	nein	nein	hoch <sup>a</sup>
Hodnett 2009	ja	ja	nein	nein	unklar	ja	mittel <sup>b</sup>

a. Angaben zur Erzeugung der Zuteilungssequenz und zur Umsetzung einer adäquaten Verdeckung fehlen, selektives Berichten von signifikanten Ergebnissen sowie intransparenter Teilnehmerinnenfluss  
b: Machbarkeitsstudie ohne Fallzahlplanung und ohne Differenzierung zwischen primären und sekundären Endpunkten

### 5.3.2 Einschätzung des Verzerrungspotenzials auf Endpunktebene

#### Geburtsmodus

Das Verzerrungspotenzial für den Geburtsmodus basiert auf dem Verzerrungspotenzial auf Studienebene und endpunktspezifischen Abwertungen. Diese liegen begründet in der fehlenden Verblindung der behandelnden Personen in beiden Studien und einer fehlenden ITT-Auswertung sowie der unklaren ergebnisunabhängigen Berichterstattung in 1 Studie. Das Verzerrungspotenzial für den Geburtsmodus wurde folglich in beiden Studien als hoch bewertet (siehe Tabelle 13 im Anhang).

#### Geburtserleben

Das Verzerrungspotenzial für das Geburtserleben für die ergänzend dargestellten mit nicht validierten Messinstrumenten erfassten Informationen basiert auf dem Verzerrungspotenzial auf Studienebene und endpunktspezifischen Abwertungen. Diese liegen begründet in der fehlenden Verblindung der behandelnden und endpunkterhebenden Personen sowie in Problemen in der Anwendung des ITT-Prinzips in beiden Studien. In 1 Studie wird die Anzahl der vorliegenden Fragebogen berichtet, die aber nicht der Gesamtanzahl der vorliegenden Antworten auf die Einzelfragen entspricht. In der 2. Studie liegt keine adäquate Umsetzung des ITT-Prinzips vor, da von 60 randomisierten Teilnehmerinnen Fragebogen von nur 42 Teilnehmerinnen ausgewertet wurden. Zusätzlich ist in dieser Studie eine ergebnisabhängige Berichterstattung wahrscheinlich, da nur als statistisch signifikant beschriebene Ergebnisse ausgewählter Subskalen berichtet werden. In beiden Studien werden nicht validierte Fragebogen verwendet. Das Verzerrungspotenzial für die ergänzend dargestellten Informationen zum Geburtserleben wurde folglich in beiden Studien als hoch bewertet (siehe Tabelle 14 im Anhang).

## Wehenmittel

Das Verzerrungspotenzial für die Effektschätzung des Endpunkts Wehenmittel wurde, bedingt durch das Verzerrungspotenzial auf Studienebene und einer fehlenden Verblindung der behandelnden Personen in 1 Studie, als hoch bewertet (siehe Tabelle 15 im Anhang).

### 5.4 Ergebnisse zu kritischen Endpunkten

Die nachfolgende Tabelle 10 gibt einen Überblick über die Ergebnisse der Evidenz zu den kritischen Endpunkten für die eingeschlossenen Studien. Aufgrund der unterschiedlichen Intensität der Modifizierungsmaßnahmen konnten die Ergebnisse nicht metaanalytisch zusammengefasst werden.

Tabelle 10: Zusammenfassung der Ergebnisse zu den kritischen Endpunkten

Endpunkt	Ergebnisse
Geburtsmodus	<ul style="list-style-type: none"> <li>▪ Umfangreiche Modifizierung:               <ul style="list-style-type: none"> <li>▫ Spontangeburt: RR: 0,87; 98,3 %-KI: [0,61; 1,24]</li> <li>▫ vaginal-operative Geburt: RR: 1,72; 98,3 %-KI: [0,33; 8,85]</li> <li>▫ Kaiserschnitt: RR: 1,38; 98,3 %-KI: [0,25; 7,71]</li> </ul> </li> <li>▪ Geringe Modifizierung:               <ul style="list-style-type: none"> <li>▫ Spontangeburt: k. A.</li> <li>▫ vaginal-operative Geburt: k. A.</li> <li>▫ Kaiserschnitt: RR: 0,72; 98,3 %-KI: [0,09; 5,77]</li> </ul> </li> </ul>
Geburtsleben	<ul style="list-style-type: none"> <li>▪ Umfangreiche Modifizierung:               <ul style="list-style-type: none"> <li>▫ Auf der Grundlage eines nicht validierten Fragebogens konnte kein klinisch relevanter Effekt für die Interventions- und Vergleichsgruppe festgestellt werden.</li> </ul> </li> <li>▪ Geringe Modifizierung:               <ul style="list-style-type: none"> <li>▫ Auf der Grundlage eines nicht validierten Fragebogens konnte kein klinisch relevanter Effekt für die Interventions- und Vergleichsgruppe festgestellt werden.</li> </ul> </li> </ul>
k. A.: keine Angabe; KI: Konfidenzintervall; RR: relatives Risiko	

#### 5.4.1 Geburtsmodus

Bei den Ergebnissen zu den Ereignissen Spontangeburt, vaginal-operative Geburt und Kaiserschnitt handelt es sich um die 3 möglichen Ausprägungen des nominalen Endpunkts Geburtsmodus. Da diese Ausprägungen wie 3 separate binäre Endpunkte ausgewertet wurden, ergab sich aufgrund ihrer gegenseitigen Abhängigkeit voneinander die Notwendigkeit, das Signifikanzniveau der Analysen für multiples Testen zu adjustieren. Hierzu wurde die Bonferroni-Methode verwendet und das 98,3 %-KI berechnet. Insgesamt lagen zum Endpunkt Geburtsmodus Daten aus 2 Studien vor, wobei in Aburas 2017 wegen der fehlenden Differenzierung zwischen Spontangeburten und vaginal-operativen Geburten ausschließlich Informationen zum Kaiserschnitt vorliegen (siehe Tabelle 16 im Anhang).

### **Spontangeburt**

Informationen zur Häufigkeit einer Spontangeburt werden in 1 Studie berichtet (siehe Tabelle 16 im Anhang). Insgesamt 46 (75,4 %) der eingeschlossenen 61 Frauen gebären spontan: 70,0 % der Gebärenden in der Gruppe mit der Prüfintervention verglichen mit 80,6 % der Frauen in der Gruppe mit der Vergleichsintervention (RR: 0,87; 98,3 %-KI: [0,61; 1,24]).

### **Vaginal-operative Geburt**

Informationen zur Häufigkeit einer vaginal-operativen Geburt werden in 1 Studie berichtet (siehe Tabelle 16 im Anhang). Insgesamt 8 (13,1 %) der eingeschlossenen 61 Frauen wiesen eine vaginal-operative Geburt auf: 16,7 % der Gebärenden in der Gruppe mit der Prüfintervention verglichen mit 9,7 % der Frauen in der Gruppe mit der Vergleichsintervention (RR: 1,72; 98,3 %-KI: [0,33; 8,85]).

### **Kaiserschnitte**

Informationen zur Häufigkeit eines Kaiserschnitts werden in 2 Studien berichtet (siehe Tabelle 16 im Anhang). In Hodnett 2009 erhielten insgesamt 7 (11,5 %) der eingeschlossenen 61 Frauen einen Kaiserschnitt: 13,3 % der Gebärenden in der Gruppe mit der Prüfintervention verglichen mit 9,7 % der Frauen in der Gruppe mit der Vergleichsintervention (RR: 1,38; 98,3 %-KI: [0,25; 7,71]). In Aburas 2017 erhielten insgesamt 5 (10,0 %) der 50 Frauen, von denen Ergebnisse vorliegen, einen Kaiserschnitt: 8,3 % der Gebärenden in der Gruppe mit der Prüfintervention verglichen mit 11,5 % der Frauen in der Gruppe mit der Vergleichsintervention (RR: 0,72; 98,3 %-KI: [0,09; 5,77]).

### **5.4.2 Geburtserleben**

Alle Informationen zum Geburtserleben basieren auf einer Erhebung mit nicht validierten Fragebogen und werden daher ergänzend dargestellt.

In Hodnett 2009 wurden ausschließlich Informationen zur Wirkung der verglichenen Gebärräume und zur Unterstützung durch das Pflegepersonal berichtet, wobei die Anzahl der ausgefüllten Fragebogen, nicht aber die Anzahl der Antworten auf die Einzelfragen zur Wirkung der Gebärräume berichtet wurde (siehe Tabelle 17 und Tabelle 18 im Anhang).

In Aburas 2017 wurden Informationen zur Qualität der Betreuung aus der Teilnehmerinnenperspektive erhoben. Dabei werden Mittelwerte des Gesamtscores ohne Angaben zur beobachteten Variabilität berichtet, sodass eine eigene Berechnung von nicht berichteten Konfidenzintervallen nicht möglich war. Es werden ausschließlich die Ergebnisse aus Subskalen berichtet, die in der Veröffentlichung als statistisch signifikant beschrieben werden, auch wenn dies tatsächlich nur auf 2 der 3 berichteten Subskalen zutrifft (siehe Tabelle 19 im Anhang).

Damit bleibt unklar, inwieweit eine alternative Gestaltung des Gebärraumes das Geburtserleben beeinflussen kann.

### 5.4.3 Apgar-Score nach 5 Minuten < 6

Für den Apgar-Score nach 5 Minuten < 6 liegen ausschließlich ergänzende Informationen in 2 Studien vor (siehe Tabelle 20 im Anhang). In Hodnett 2009 liegen Informationen zu einem Apgar-Score nach 5 Minuten < 7 vor. Insgesamt 1 Neugeborenes der Gruppe mit der Prüfintervention wies einen Apgar-Score nach 5 Minuten < 7 auf. In Aburas 2017 werden ausschließlich die Mittelwerte der Apgar-Scores nach 5 Minuten berichtet, sodass ein Rückschluss auf die Anzahl der Neugeborenen mit einem Wert < 6 nicht möglich ist.

## 5.5 Ergebnisse zu wichtigen Endpunkten

Die nachfolgende Tabelle 11 gibt einen Überblick über die Ergebnisse der Evidenz zu den wichtigen Endpunkten. Informationen zu wichtigen Endpunkten werden in 1 Studie für 1 Endpunkt berichtet.

Tabelle 11: Zusammenfassung der Ergebnisse zu den wichtigen Endpunkten

Endpunkt	Ergebnisse
Wehenmittelgabe	<ul style="list-style-type: none"> <li>▪ Umfangreiche modifizierte Umgebungsgestaltung:               <ul style="list-style-type: none"> <li>▫ Wehenmittel während der Geburt: RR: 0,59; 95 %-KI: [0,36; 0,97]</li> </ul> </li> </ul>
KI: Konfidenzintervall; RR: relatives Risiko	

### 5.5.1 Wehenmittel

Informationen zur Häufigkeit der Gabe von Wehenmitteln werden in 1 Studie berichtet (siehe Tabelle 22 im Anhang). Insgesamt 33 (54,1 %) der eingeschlossenen 61 Frauen erhielten eine Wehenmittelgabe (Oxytocin intravenös): 40,0 % der Gebärenden in der Gruppe mit der Prüfintervention verglichen mit 67,7 % der Frauen in der Gruppe mit der Vergleichsintervention (RR: 0,59; 95 %-KI: [0,36; 0,97]).

## **6 Evidenzprofil für die kritischen und wichtigen Endpunkte**

Im Folgenden wird das Evidenzprofil für die modifizierte Umgebungsgestaltung des Kreißsaals verglichen mit der Standardumgebung beschrieben.

Tabelle 12: GRADE-Evidenzprofil

Frage: modifizierte Umgebungsgestaltung verglichen mit der Standardumgebung bei Gebärenden am Termin ab der aktiven Eröffnungsphase											
Setting: klinische Betreuung unter der Geburt											
Bewertung der Qualität der Evidenz							Anzahl der Ereignisse / Gebärende		Wirkung		Qualität der Evidenz
Anzahl der Studien	Studien-design	Studien-limitationen	Inkon-sistenz	Indirekt-heit	Fehlende Genauig-keit	Andere Fak-toren	Modi-fizierte Umge-bungs-gestaltung	Standard-umgebung	Relatives Risiko [KI]	Absolut [KI]	
<b>Maternale Mortalität – keine Angaben</b>											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Geburtsmodus</b>											
<b>Spontangeburt</b>											
1	RCT	schwer-wiegend <sup>a</sup>	nicht relevant	nicht schwer-wiegend <sup>b</sup>	schwer-wiegend <sup>c</sup>	starker Verdacht auf Pub-likations-bias <sup>d</sup>	21 / 30 (70,0 %)	25 / 31 (80,6 %)	RR: 0,87; 98,3 %-KI: [0,61; 1,24]	8 weniger pro 100 [von 24 weniger bis 15 mehr] <sup>e</sup>	⊕○○○ SEHR NIEDRIG
<b>Vaginal-operative Geburt</b>											
1	RCT	schwer-wiegend <sup>a</sup>	nicht relevant	nicht schwer-wiegend <sup>b</sup>	sehr schwer-wiegend <sup>f</sup>	starker Verdacht auf Pub-likations-bias <sup>d</sup>	5 / 30 (16,7 %)	3 / 31 (9,7 %)	RR: 1,72; 98,3 %-KI: [0,33; 8,85]	5 mehr pro 100 [von 5 weniger bis 55 mehr] <sup>e</sup>	

(Fortsetzung)

Tabelle 12: GRADE-Evidenzprofil (Fortsetzung)

Bewertung der Qualität der Evidenz							Anzahl der Ereignisse / Gebärende		Wirkung		Qualität der Evidenz
Anzahl der Studien	Studien-design	Studien-limitationen	Inkon-sistenz	Indirekt-heit	Fehlende Genauig-keit	Andere Fak-toren	Modi-fizierte Umge-bungs-gestaltung	Standard-umgebung	Relatives Risiko [KI]	Absolut [KI]	
<b>Kaiserschnitt</b>											
<b>Umfangreiche Modifizierung:</b>											
2	RCT	schwer-wiegend <sup>a</sup>	nicht relevant	nicht schwer-wiegend <sup>b</sup>	sehr schwer-wiegend <sup>f</sup>	starker Verdacht auf Pub-likations-bias <sup>d</sup>	4 / 30 (13,3 %)	3 / 31 (9,7 %)	RR: 1,38; 98,3 %-KI: [0,25; 7,71]	12 mehr pro 100 [von 24 weniger bis 215 mehr] <sup>e</sup>	
<b>Geringe Modifizierung:</b>											
	RCT	sehr schwer-wiegend <sup>g</sup>	nicht relevant	schwer-wiegend <sup>h</sup>	sehr schwer-wiegend <sup>f</sup>	starker Verdacht auf Pub-likations-bias <sup>d</sup>	2 / 24 (8,3 %)	3 / 26 (11,5 %)	RR: 0,72; 98,3 %-KI: [0,09; 5,77]	9 weniger pro 100 [von 29 weniger bis 153 mehr] <sup>e</sup>	
<b>Höhergradige Dammrisse (III<sup>o</sup>, IV<sup>o</sup>) – keine Angaben</b>											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Hoher mütterlicher Blutverlust ≥ 1000 ml – keine Angaben</b>											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Aufnahme Stillbeziehung bei initialem Stillwunsch – keine Angaben</b>											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Teilweise / vollständige Plazentaretention – keine Angaben</b>											
–	–	–	–	–	–	–	–	–	–	–	–

(Fortsetzung)

Tabelle 12: GRADE-Evidenzprofil (Fortsetzung)

Bewertung der Qualität der Evidenz							Anzahl der Ereignisse / Gebärende		Wirkung		Qualität der Evidenz
Anzahl der Studien	Studien-design	Studien-limitationen	Inkon-sistenz	Indirekt-heit	Fehlende Genauig-keit	Andere Fak-toren	Modi-fizierte Umge-bungs-gestaltung	Standard-umgebung	Relatives Risiko [KI]	Absolut [KI]	
<b>Geburtserleben</b> (ergänzende Informationen)											
<b>Umfangreiche Modifizierung:</b>											
2	RCT	sehr schwer-wiegend <sup>a, i</sup>	nicht relevant	nicht schwer-wiegend <sup>b</sup>	schwer-wiegend <sup>j</sup>	starker Verdacht auf Pub-likations-bias <sup>d</sup>	Raum wirkt beruhigend		⊕○○○ SEHR NIEDRIG		
							<ul style="list-style-type: none"> <li>▪ gesamte / meiste Zeit: 44,8 % vs. 43,3 %</li> <li>▪ andere Angabe: 55,2 % vs. 46,7 %</li> </ul>				
							ausreichende Privatsphäre				
							<ul style="list-style-type: none"> <li>▪ gesamte / meiste Zeit: 72,4 % vs. 73,3 %</li> <li>▪ andere Angabe: 27,6 % vs. 26,7 %</li> </ul>				
							Raum wirkt entspannend				
<ul style="list-style-type: none"> <li>▪ gesamte / meiste Zeit: 51,7 % vs. 33,3 %</li> <li>▪ andere Angabe: 48,3 % vs. 66,7 %</li> </ul>											
Raum wirkt beängstigend											
<ul style="list-style-type: none"> <li>▪ niemals / selten: 65,5 % vs. 83,3 %</li> <li>▪ andere Angabe: 34,5 % vs. 16,5 %</li> </ul>											
Unterstützung durch das Pflegepersonal <sup>k</sup>											
<ul style="list-style-type: none"> <li>▪ MD: 2,90; 95 %-KI: [-2,15; 7,95]</li> </ul>											
<b>Geringe Modifizierung:</b>											
	RCT	sehr schwer-wiegend <sup>g, i</sup>	nicht relevant	schwer-wiegend <sup>h</sup>	schwer-wiegend <sup>j</sup>	starker Verdacht auf Pub-likations-bias <sup>d</sup>	Subskala des QPP-Fragebogens <sup>l</sup>		<ul style="list-style-type: none"> <li>▪ MD: -0,17; 95 %-KI: k. A.; p &gt; 0,05</li> </ul>		

(Fortsetzung)

Tabelle 12: GRADE-Evidenzprofil (Fortsetzung)

Bewertung der Qualität der Evidenz							Anzahl der Ereignisse / Gebärende		Wirkung		Qualität der Evidenz
Anzahl der Studien	Studien-design	Studien-limitationen	Inkon-sistenz	Indirekt-heit	Fehlende Genauig-keit	Andere Fak-toren	Modi-fizierte Umgebungs-gestaltung	Standard-umgebung	Relatives Risiko [KI]	Absolut [KI]	
<b>Subjektive Gesundheitseinschätzung</b> – keine Angaben											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Neonatale Mortalität</b> – keine Angaben											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Apgar-Score nach 5 Minuten &lt; 6</b> – keine Angaben											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Schwere metabolische Acidose (pH &lt; 7,0)</b> – keine Angaben											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Basendefizit ≥ – 16 mmol/l</b> – keine Angaben											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Verlegung auf die Neugeborenen-Intensivstation</b> – keine Angaben											
–	–	–	–	–	–	–	–	–	–	–	–
<b>Wehenmittel</b>											
1	RCT	schwer-wiegend <sup>a</sup>	nicht relevant	nicht schwer-wiegend <sup>b</sup>	nicht schwer-wiegend	starker Verdacht auf Pub-likations-bias <sup>d</sup>	12 / 30 (40,0 %)	21 / 31 (67,7 %)	RR: 0,59; 95 %-KI: [0,36; 0,97]	11 weniger pro 100 [von 17 weniger bis 1 weniger] <sup>e</sup>	⊕⊕○○ NIEDRIG

(Fortsetzung)

Tabelle 12: GRADE-Evidenzprofil (Fortsetzung)

- a: Machbarkeitsstudie ohne Differenzierung zwischen primären und sekundären Endpunkten. Es erfolgte eine Abwertung.
- b: Ausgeschlossen wurden Frauen aufgrund medizinisch bedingter Indikationen oder von Präferenzen, die die Mobilität beeinträchtigen könnten. Die Gruppe Gebärender mit den aufgezählten Ausschlussgründen würde auch im allgemeinen Versorgungsalltag einen Teil der Angebote nicht nutzen können, weswegen keine eingeschränkte Übertragbarkeit vermutet wird. Es erfolgte keine Abwertung.
- c: Das 98,3 %-KI schließt sowohl einen Nutzen als auch einen Schaden ein. Es erfolgte eine Abwertung.
- d: Es wurden 2 registrierte Studien ohne veröffentlichte Ergebnisse identifiziert. Alle identifizierten Studien sind von kleiner oder moderater Studiengröße. Es erfolgte eine Abwertung.
- e: Zur Berechnung der absoluten Wirkung (Risikodifferenz) wurden die Daten aus der Bundesauswertung zum Erfassungsjahr 2016 Geburtshilfe [5] verwendet.
- f: Das 98,3 %-KI schließt sowohl einen großen Nutzen als auch einen großen Schaden ein. Es erfolgte eine Abwertung.
- g: Fehlende Informationen zur Randomisierung, keine ITT-Auswertungen, selektives Ergebnisberichten. Es erfolgte eine Abwertung.
- h: Umfangreiche, nicht genau differenzierte Ausschlusskriterien. Die Teilgruppe der ausgeschlossenen Gebärenden könnte im allgemeinen Versorgungsalltag die Intervention erhalten. Es erfolgte eine Abwertung.
- i: Verwendung eines nicht validierten Fragebogens und fehlende Verblindung der endpunkterhebenden Personen. In 1 Studie (umfangreiche Modifizierung) unklare Anwendung des ITT-Prinzips aufgrund fehlender Angaben zur Gesamtanzahl der vorliegenden Antworten. Die Anzahl der Fragebogen entspricht nicht der Anzahl der Antworten. In 2. Studie (geringe Modifizierung) fehlende Anwendung des ITT-Prinzips sowie Vermutung einer ergebnisabhängigen Berichterstattung. Es erfolgte eine Abwertung.
- j: Es erfolgte eine Abwertung, da die vorliegenden Ergebnisse zu unpräzise sind, um eine Empfehlung für oder gegen die untersuchte Intervention unterstützen zu können.
- k: Skala von 23 bis 69 zur Beschreibung von keiner bis häufiger physischer, emotionaler Unterstützung, Beratung und Unterstützung, höhere Werte beschreiben eine aus Patientenperspektive bessere Unterstützung durch das Pflegepersonal
- l: Skala des QPP-Fragebogens, höhere Werte beschreiben eine aus Patientenperspektive bessere Qualität der Pflege. Angaben zur Variabilität wurden nicht berichtet
- k. A.: keine Angabe; GRADE: Grading of Recommendations, Assessment, Development, and Evaluation (Bewertung von Empfehlungen, Einschätzung, Entwicklung und Auswertung); ITT: Intention to treat; KI: Konfidenzintervall; MD: mittlere Differenz; QPP-Fragebogen: Quality of Care from the Patient's Perspective (Fragebogen zur Erhebung der Qualität der Betreuung aus Patientenperspektive); RR: relatives Risiko

Im Folgenden wird die vorliegende Evidenz zu den 2 kritischen und dem 1 wichtigen Endpunkt der eingeschlossenen Studien zusammenfassend dargestellt. Die gesamte Evidenz zu allen Endpunkten dieser Fragestellung stammt aus 2 Studien (Aburas 2017, Hodnett 2009) mit insgesamt 122 Studienteilnehmerinnen, von denen für 111 Studienteilnehmerinnen Ergebnisse vorliegen.

Das Populationsrisiko für 2 Endpunkte ohne gezielte Intervention basiert auf den Angaben der Bundesauswertung zur stationären Qualitätssicherung zum Erfassungsjahr 2016 des Instituts für Qualitätssicherung und Transparenz im Gesundheitswesen (IQTIG) [5]. Die Evidenzqualität der RCTs wurde aufgrund schwerwiegender Studienlimitationen, schwerwiegender Indirektheit, schwerwiegender fehlender Genauigkeit und des Verdachts auf einen Publikationsbias herabgestuft.

Die **schwerwiegenden Studienlimitationen**, die zur Herabstufung um 1 bis 2 Kategorien führten, umfassten

- fehlende Informationen zur Randomisierung, die sowohl die Erzeugung der Randomisierungssequenz als auch die Verdeckung der Gruppenzuteilung betreffen,
- eine fehlende Differenzierung zwischen primären und sekundären Endpunkten,
- eine selektive Berichterstattung einzelner Items der Subskalen eines Fragebogens,
- eine fehlende oder unklare Verblindung der endpunkterhebenden Personen bei einer subjektiv beeinflussbaren Messung von Endpunkten,
- die Verwendung eines nicht validierten Fragebogens und
- eine nicht adäquate Umsetzung des ITT-Prinzips aufgrund des Ausschlusses von Studienteilnehmerinnen aus den Analysen oder fehlender Angaben zur Gesamtanzahl der vorliegenden Antworten auf die Einzelfragen.

**Schwerwiegende fehlende Genauigkeit** führte zur Abwertung der Evidenz für einen Endpunkt, wenn die Ergebnisse zu unpräzise waren, um eine Entscheidung für oder gegen eine Empfehlung stützen zu können. Primär basierte eine Abwertung um 1 Kategorie auf einem KI, das sowohl einen möglichen Nutzen als auch einen möglichen Schaden einschließt. Eine Abwertung wegen fehlender Genauigkeit erfolgte um 2 Kategorien, wenn das KI des Effekts einen bedeutenden Nutzen und großen Schaden einschließt. Diese geringe Genauigkeit basiert auf der geringen Fallzahl und bei den vaginal-operativen Geburten und Kaiserschnitten (Geburtsmodus) auf der den Erwartungen entsprechenden geringen Anzahl von Ereignissen.

In der Studie mit umfangreichen Modifizierungen der Umgebungsgestaltung wurde nicht wegen **schwerwiegender Indirektheit** abgewertet. Die in dieser Studie vorgenommenen Modifizierungen dienten dem übergeordneten Ziel, eine freie Positionswahl zu fördern und den Routineeinsatz einer elektronischen Überwachung der fetalen Herzfrequenz zu limitieren.

Gebärende wurden daher aufgrund medizinischer Indikationen oder von Präferenzen, die ihre Mobilität beeinträchtigen konnten, aus der Studie ausgeschlossen. Die Gruppe Gebärender mit den aufgezählten Ausschlussgründen würde auch im allgemeinen Versorgungsalltag einen Teil der Angebote nicht nutzen können, weswegen keine eingeschränkte Übertragbarkeit vermutet wird. In der Studie mit geringer Modifizierung führten neben weiteren auch Kriterien wie ein nicht näher bestimmtes, allgemein höheres Risiko zum Ausschluss der Gebärenden. Die Teilgruppe der ausgeschlossenen Gebärenden könnte im allgemeinen Versorgungsalltag jedoch die Intervention erhalten, weswegen für diese Ergebnisse aufgrund schwerwiegender Indirektheit um 1 Kategorie abgewertet wurde.

Es erfolgte **keine Abwertung wegen Inkonsistenz**. Die Ergebnisse der vorliegenden 2 Studien werden getrennt beschrieben, da eine quantitative Zusammenfassung der Ergebnisse aufgrund der unterschiedlichen Charakteristika der Interventionen nicht sinnvoll ist, sodass die Inkonsistenz als nicht relevant beschrieben wurde.

Für die untersuchte Fragestellung konnten in Studienregistern 2 potenziell relevante Studien, durchgeführt im Iran, identifiziert werden, die bereits im Mai 2011 beziehungsweise Februar 2016 abgeschlossen werden sollten, deren Ergebnisse aber nicht veröffentlicht wurden. Zusätzlich wurde 1 weitere Studie desselben Studienkoordinators identifiziert, die aufgrund des nicht zutreffenden Einschlusskriteriums für die Population nicht in diesen Rapid Report eingeschlossen werden konnte. Eine gestellte Autorenanfrage blieb unbeantwortet. Sowohl diese Studien als auch die eingeschlossene Machbarkeitsstudie sind von moderater beziehungsweise kleiner Studiengröße, sodass ein **Verdacht auf einen Publikationsbias** existiert und deshalb um 1 Kategorie abgewertet wurde.

## 6.1 Geburtsmodus

Insgesamt traten in 1 Studie zur umfangreichen Modifizierung der Umgebungsgestaltung in der Interventionsgruppe 21 Spontangeburt, 5 vaginal-operative Geburten sowie 4 Kaiserschnitte bei insgesamt 30 Frauen auf. In der Vergleichsgruppe traten 25 Spontangeburt, 3 vaginal-operative Geburten und 3 Kaiserschnitte bei insgesamt 31 Frauen auf. In der Studie mit geringer Modifizierung traten 2 Kaiserschnitte bei 24 Frauen in der Interventionsgruppe und in der Vergleichsgruppe 3 Kaiserschnitte bei 26 Frauen auf.

Die Qualität der Evidenz wurde für den Endpunkt Geburtsmodus zusammenfassend als sehr niedrig beurteilt (siehe Tabelle 12). Die Qualität der Evidenz zu einer umfangreichen Modifikation wurde aufgrund schwerwiegender Studienlimitationen und des Verdachts auf einen Publikationsbias für den Geburtsmodus und aufgrund schwerwiegender fehlender Genauigkeit für Spontangeburt beziehungsweise sehr schwerwiegender fehlender Genauigkeit für vaginal-operative Geburten und Kaiserschnitte herabgestuft. Grund für die Abstufung aufgrund von Studienlimitationen, die das Vertrauen in die Effektschätzungen in der vorliegenden Machbarkeitsstudie ohne Fallzahlplanung beeinträchtigen, ist eine fehlende Differenzierung zwischen primären und sekundären Endpunkten.

Die weitere Herabstufung basierte auf Ungenauigkeiten der Effektschätzung. Die Effektschätzung für Spontangeburt schließt aufgrund der geringen Fallzahl zum Signifikanzniveau von 1,7 % sowohl eine geringere als auch eine höhere, durch die Intervention bedingte Anzahl von Spontangeburt (RR: 0,87; 98,3 %-KI: [0,61; 1,24]) ein. Die Effektschätzungen für vaginal-operative Geburten und Kaiserschnitte schließen, da erwartungsgemäß insgesamt nur sehr wenige Ereignisse auftraten, sowohl eine durch die Intervention bedingte höhere als auch eine geringere Anzahl von vaginal-operativen Geburten (RR: 1,72; 98,3 %-KI: [0,33; 8,85]) und Kaiserschnitten (RR: 1,38; 98,3 %-KI: [0,25; 7,71]) ein.

Für die geringe Modifizierung sind für den Geburtsmodus ausschließlich Ergebnisse für den Kaiserschnitt vorhanden. Die Qualität der Evidenz wurde aufgrund sehr schwerwiegender Studienlimitationen, schwerwiegender Indirektheit, sehr schwerwiegender Ungenauigkeiten der Effektschätzung und des Verdachts auf einen Publikationsbias herabgestuft. Die Gründe für die Abstufung aufgrund von Studienlimitationen sind fehlende Informationen zur Randomisierung, fehlende ITT-Auswertungen und das selektive Berichten von Ergebnissen. Die Effektschätzung für Kaiserschnitte zeigt eine sehr schwerwiegende Ungenauigkeit und schließt sowohl eine durch die Intervention bedingte höhere als auch eine geringere Anzahl von Kaiserschnitten (RR: 0,72; 98,3 %-KI: [0,09; 5,77]) ein. Ein weiterer Grund für die Abstufung der Qualität der Evidenz ist eine schwerwiegende Indirektheit, da Gebärende mit einem nicht näher beschriebenen erhöhten Risiko generell aus der Studie ausgeschlossen wurden.

Die Angaben zum Populationsrisiko basieren auf den Angaben des IQTIG [5]. Demnach erfolgten 2016 in Deutschland 61 pro 100 Geburten spontan, 7 pro 100 vaginal-operativ und 32 pro 100 per Kaiserschnitt.

Eine Veränderung im Geburtsmodus konnte weder für die Häufigkeit von Spontangeburt noch für diejenige vaginal-operativer Geburten oder von Kaiserschnitten nachgewiesen werden.

## **6.2 Geburtserleben**

In den vorliegenden Studien wurden nur Teilaspekte des Geburtserlebens, die die Wirkung der verglichenen Gebärräume, die Unterstützung durch das Pflegepersonal und die Qualität der Pflege aus der Patientenperspektive betreffen, berichtet.

Insgesamt beurteilten 50 Frauen der Interventionsgruppen und 51 Frauen der Vergleichsgruppen ihr Geburtserleben. Die Erfassung basiert auf nicht validierten Fragebogen, sodass ausschließlich ergänzende Informationen vorliegen.

Die Qualität der Evidenz wurde sowohl bei der umfangreichen als auch bei der geringen Modifizierung des Gebärraums als sehr niedrig eingeschätzt (Tabelle 12), da aufgrund sehr schwerwiegender Studienlimitationen, schwerwiegender fehlender Genauigkeit und dem Verdacht auf Publikationsbias herabgestuft wurde. Zudem wurde in der Studie mit geringer

Modifizierung des Gebärraums eine schwere Indirektheit aufgrund der nicht differenziert angegebenen Ausschlusskriterien festgestellt. Eine Teilgruppe dieser ausgeschlossenen Frauen könnte im allgemeinen Versorgungsalltag die Intervention erhalten. Die Gründe für die Abstufung aufgrund von Studienlimitationen, die das Vertrauen in die Ergebnisse beeinträchtigen, umfassen die fehlende Verblindung der endpunkterhebenden Personen und die fehlende Validierung der Fragebogen. In der Studie mit umfangreichen Modifizierungen bleibt zudem die Anwendung des ITT-Prinzips unklar, da Angaben zur Anzahl der vorliegenden Antworten auf die Einzelfragen fehlen. In der Studie mit geringer Modifizierung des Gebärraumes wurde das ITT-Prinzip nicht angewendet, da Antworten ohne Angabe von Gründen fehlen, außerdem wurden ergebnisgesteuert die Ergebnisse ausgewählter Subskalen berichtet. Eine Abwertung aufgrund schwerwiegender fehlender Genauigkeit erfolgte, da die vorliegenden Ergebnisse zu unpräzise sind, um eine Empfehlung für oder gegen die untersuchte Intervention unterstützen zu können.

Für keine der Fragen wurde ein bedeutsamer Unterschied bezüglich des Geburtserlebens zwischen der Interventions- und Vergleichsgruppe beobachtet.

### **6.3 Wehenmittel**

Insgesamt erhielten 12 von 30 Frauen der Interventionsgruppe und 21 von 31 Frauen der Vergleichsgruppe Wehenmittel.

Die Qualität der Evidenz wurde für den Endpunkt Wehenmittel als niedrig bewertet (siehe Tabelle 12), da aufgrund schwerwiegender Studienlimitationen und des Verdachts auf einen Publikationsbias herabgestuft wurde. Der Grund für die Abstufung aufgrund von Studienlimitationen, die das Vertrauen in die Effektschätzungen in der vorliegenden Machbarkeitsstudie ohne Fallzahlplanung beeinträchtigt, umfasst eine fehlende Differenzierung zwischen primären und sekundären Endpunkten.

Die Effektschätzung für den Endpunkt Wehenmittel zeigte einen Nutzen der modifizierten Umgebungsgestaltung mit einer geringeren Anzahl von Wehenmittelgaben in der Interventionsgruppe (RR: 0,59; 95 %-KI: [0,36; 0,97]). Demnach würden, ausgehend von einer Wehenmittelgabe unter der Geburt bei 27 von 100 Geburten [5], aufgrund der Intervention 11 (1 bis 17) Frauen weniger pro 100 Geburten ein Wehenmittel erhalten.

### **6.4 Gesamtbeurteilung der Qualität der Evidenz für alle Endpunkte**

Es konnten Effektschätzungen für insgesamt 2 kritische und 1 wichtigen Endpunkt zusammengetragen werden. Die Qualität der Evidenz wurde für 2 kritische Endpunkte als sehr niedrig (Geburtsmodus, Geburtserleben) und für 1 wichtigen Endpunkt als niedrig (Wehenmittel) bewertet. Die Bewertung des Endpunkts Geburtserleben basiert ausschließlich auf ergänzenden Informationen. Ergebnisse aus RCTs zur Mortalität der Mutter oder des Neugeborenen, zu höhergradigen Dammrissen (III°, IV°), zu hohem mütterlichen Blutverlust  $\geq 1000$  ml, zur Aufnahme einer Stillbeziehung bei initialem Stillwunsch, zur teil-

weisen / vollständigen Plazentaretention, zur subjektiven Gesundheitseinschätzung, zum Apgar-Score nach 5 Minuten  $< 6$ , zu schwerer metabolischer Acidose ( $\text{pH} < 7,0$ ), zum Basendefizit  $\geq -16$  mmol/l und zur Verlegung auf die Neugeborenen-Intensivstation wurden nicht berichtet.

Zusammenfassend wurde die Qualität der Evidenz als sehr niedrig bewertet, sodass folglich das Gesamtvertrauen in die Effektschätzungen als sehr niedrig bewertet wird und der wahre Effekt wahrscheinlich relevant verschieden von der berechneten Effektschätzung ist.

## 7 Diskussion

Ziel einer modifizierten Umgebungsgestaltung des Kreißsaals soll die Schaffung einer Umgebung sein, die den Bedürfnissen und Erwartungen der werdenden Eltern entspricht und das Geburtserleben verbessert [59,61]. Es wird diskutiert, ob die modifizierte Umgebungsgestaltung die Mobilität der Gebärenden fördert und – gemeinsam mit weiteren Maßnahmen – die Gebärende und ihre Begleitpersonen im Treffen aktiver und selbstbestimmter Entscheidungen während des Geburtsverlaufs unterstützt und dies zu einer reduzierten Interventionsrate führen kann. Letzteres entspricht dem nationalen Gesundheitsziel „Gesundheit rund um die Geburt“ [62], in dem die Förderung der natürlichen Geburt und die Verhinderung der Pathologisierung als ausdrückliche Ziele genannt werden. Dieses Gesundheitsziel soll durch 2 interdisziplinär und interprofessionell erarbeitete S3-Leitlinien „Die vaginale Geburt am Termin“ und „Die Sectio caesarea“ unterstützt werden [1,63].

In die Evidenzbewertung zur Modifizierung der Umgebungsgestaltung wurden ausschließlich Ergebnisse zu von der Leitliniengruppe festgelegten maternalen und neonatalen kritischen und wichtigen Endpunkten einbezogen. Lediglich 2 RCTs (Aburas 2017, Hodnett 2009) konnten in die Evidenzbewertung einbezogen werden, aus denen Informationen zu 3 Endpunkten bei insgesamt sehr niedriger Qualität der Evidenz extrahiert werden konnten. Bei 1 dieser Studien handelt es sich um eine Machbarkeitsstudie mit einer Teilnehmerinnenanzahl von insgesamt 62 Gebärenden. Laut den Angaben der Autorinnen und Autoren war diese Studie nicht dafür geplant, abschließende Ergebnisse zu klinisch relevanten Endpunkten zu erzielen. Statistische Tests wurden für rein explorative Zwecke verwendet, um die Durchführung einer großen Wirksamkeitsstudie vorzubereiten. Eine solche Studie wird aktuell in Deutschland durchgeführt [60].

In Studienregistern konnten 2 Studien [55,56], beide im Iran durchgeführt, identifiziert werden, zu denen trotz der Beendigung laut Studienregistereintrag keine publizierten Ergebnisse vorliegen. Zudem wurde 1 Publikation zu einer weiteren Studie [57], ebenfalls im Iran an der Universität in Teheran durchgeführt, identifiziert. Zu dieser Studie wurde eine Autorenanfrage gestellt, die Studie wurde aber nicht in die Evidenzbewertung aufgenommen, da laut den Informationen aus der Publikation über 80 % der Gebärenden vor der 37. SSW gebären. Auffallend ist jedoch, dass alle 3 Studien von den gleichen Studienkoordinatoren begleitet wurden und sich die Teilnehmerinnenzahl und die Beschreibung der Interventionen ähnelten. Für alle 3 Studien sind jedoch unterschiedliche Durchführungszeiträume aufgeführt. Eine Autorenanfrage an die Studienkoordinatoren zur Klärung dieser unstimmigen Angaben blieb unbeantwortet, sodass dies als starker Verdacht auf einen Publikationsbias bewertet wurde.

Die Studienlage zur Fragestellung der Umgebungsgestaltung von Geburtsräumen reicht für eine Bewertung des Effekts nicht aus. Foureur et al. 2010 [64] beschreiben in einem konzeptuellen Modell den Zusammenhang zwischen der Sicherheit und Zufriedenheit mit der Geburtsbetreuung und bezeichnen als übergeordneten Einflussfaktor unter anderem das Design der Geburtsstation. Sie fordern, dass zukünftige Forschungsvorhaben den Zusammenhang zwischen Architektur und den Neurowissenschaften sowie zwischen Kommunikation, Qualität, Risiken

und Sicherheit im Gesundheitswesen untersuchen. Die Bedeutung dieses Zusammenhangs wird von den Ergebnissen einer Befragung von 2000 Müttern in Großbritannien bezüglich ihrer Wünsche und des Einflusses der Geburtsumgebung unterstrichen [61]. Die Autorinnen und Autoren stellten fest, dass den meisten Frauen die Kontrolle über die Raumtemperatur sowie über die Personen, die während der Geburt den Raum betreten, ausreichend Bewegungsfreiheit, gemütliche Möbel oder Sitzgelegenheiten, eine häusliche, nicht klinische Umgebung, nicht belauscht werden zu können und ein Zugang zu Snacks und Getränken wichtig sind. Mehr als die Hälfte der befragten Frauen gab jedoch an, dass sie über diese Faktoren keine Kontrolle hatten.

Ergebnisse aus einer Übersichtsarbeit mit weiter gefasster Fragestellung unterstreichen die allgemeine Bedeutung der Umgebungsgestaltung im Krankenhaus. In dieser Übersichtsarbeit thematisieren Ulrich et al. 2008 [65] allgemein die Umgebungsgestaltung im Krankenhaus und deren mögliche Einflüsse auf relevante Endpunkte. Sie konnten feststellen, dass ein lärmreduzierender Innenausbau sowie die Aussicht auf die Naturgestaltung außerhalb der klinischen Räumlichkeiten subjektiven Stress bei Patientinnen und Patienten deutlich verringern können und der Blick auf die Naturgestaltung zu einer klinisch relevanten Schmerzverringering führen kann. Es wird auf eine randomisierte Studie verwiesen, die eine zusätzliche Förderung der schmerzreduzierenden Wirkung durch das Imitieren von Naturbildern auf Fernsehbildschirmen – vermutlich aufgrund der akustischen und visuellen Ablenkung – feststellen konnte.

In 1 weiteren systematischen Übersicht [66] mit weiter gefasster Fragestellung werden alternative institutionelle Geburtssettings mit konventionellen Settings während der Geburtsbetreuung verglichen. Eingeschlossen wurden 10 randomisierte kontrollierte Studien, von denen 6 Studien vor dem Jahr 2000 durchgeführt wurden und 3 Studien primär verschiedene Versorgungsmöglichkeiten (z. B. Möglichkeit der Durchführung einer PDA oder vaginal-operativen Geburt) oder Betreuungskonzepte (z. B. hebammengeleitete, ärztlich geleitete oder kontinuierliche Betreuung) miteinander verglichen, die teilweise bereits im 2. Trimester der Schwangerschaft begannen. Zur alternativen Gestaltung der Geburtsumgebung wurde ausschließlich die auch in diesen Rapid Report eingeschlossene Studie Hodnett 2009 eingeschlossen. Die Autorinnen und Autoren fordern die Durchführung adäquat gepowerter Studien bezüglich des architektonischen Designs, das die Mobilität und das Gefühl von Beruhigung und Kontrolle fördern soll.

Aktuell werden 2 RCTs in Deutschland und Dänemark mit einem geplanten Einschluss von jeweils 4000 beziehungsweise 656 Gebärenden durchgeführt [54,58]. Da RCTs bei methodisch adäquater und angemessener Durchführung mit der geringsten Ergebnisunsicherheit behaftet sind, wurden in diesem Rapid Report trotz der unzureichenden Evidenzlage keine weiteren Studien mit einem geringeren Evidenzniveau betrachtet, da die aktuell laufenden Studien Informationen zu den kritischen und wichtigen Endpunkten erheben. Die geplante Laufzeit der beiden Studien ist bis Ende Dezember 2017 und bis Ende September 2020, sodass die Fragestellung dann auf der Basis von Evidenz aus großen randomisierten Studien mit guter Übertragbarkeit für die klinikbezogene Versorgung in Mitteleuropa beziehungsweise Deutschland beantwortet werden kann.

## 8 Fazit

Auf der Basis von 2 randomisierten kontrollierten Studien mit einer insgesamt geringen Teilnehmerinnenanzahl bleibt der Einfluss einer modifizierten Umgebungsgestaltung des Kreißsaals auf den Endpunkt Geburtsmodus und den Endpunkt Geburtserleben, zu welchem nur ergänzende Informationen dargestellt werden konnten, unklar. Das Geburtserleben wurde mit nicht validierten Fragebogen erhoben. Für den klinisch wichtigen, aber für die Empfehlung nicht entscheidenden Endpunkt Wehenmittel konnte in 1 Studie eine Auswirkung in Form einer signifikant geringeren Gabe von Wehenmitteln unter der Geburt festgestellt werden. Ob dies mit einem Vor- oder Nachteil für die Mutter und das Neugeborene verbunden ist, bleibt aufgrund der fehlenden Informationen zu weiteren kritischen Endpunkten unklar. Zusammenfassend wurde die Qualität der Evidenz und somit das Gesamtvertrauen in die Effektschätzungen für die vorliegende Fragestellung als sehr niedrig bewertet. Diese geringe Qualität der Evidenz ist vorrangig durch schwerwiegende Studienlimitationen wie fehlenden Angaben zur Randomisierung, einem intransparenten Teilnehmerinnenfluss, einer selektiven Berichterstattung und einer fehlenden Differenzierung zwischen primären und sekundären Endpunkten sowie einer fehlenden Genauigkeit der Effektschätzungen bedingt.

Insgesamt ist zu beachten, dass nur für 3 der vordefinierten Endpunkte Evidenz aus 2 randomisierten kontrollierten Studien vorlag. Aktuell werden 2 potenziell relevante randomisierte kontrollierte Studien mit einer großen Anzahl von Teilnehmerinnen durchgeführt, die Informationen zu den kritischen und wichtigen Endpunkten erheben. Die Ergebnisse dieser Studien lassen eine Verbesserung der Evidenzlage und eine Beantwortung der Fragestellung auf der Basis von hoher Evidenz bei hoher Übertragbarkeit erwarten.

## 9 Literatur

1. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Die vaginale Geburt am Termin: angemeldetes Leitlinienvorhaben [online]. [Zugriff: 01.02.2017]. URL: <http://www.awmf.org/leitlinien/detail/anmeldung/1/II/015-083.html>.
2. Statistisches Bundesamt. Todesursachen bei Säuglingen: gestorbene Säuglinge nach Alter und ausgewählten Todesursachen der ICD-10 für 2015 [online]. [Zugriff: 13.11.2017]. URL: <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Todesursachen/Tabellen/GestorbeneSaeuglinge.html>.
3. Bundesinstitut für Bevölkerungsforschung. Müttersterblichkeit [online]. [Zugriff: 13.11.2017]. URL: <http://www.bib-demografie.de/SharedDocs/Glossareintraege/DE/M/muettersterblichkeit.html>.
4. Zeitlin J, Mohangoo A, Delnord M. European perinatal health report: health and care of pregnant women and babies in Europe in 2010 [online]. [Zugriff: 09.12.2016]. URL: <http://www.europeristat.com/images/doc/Peristat%202013%20V2.pdf>.
5. Institut für Qualitätssicherung und Transparenz im Gesundheitswesen. Bundesauswertung zum Erfassungsjahr 2016: Geburtshilfe; Qualitätsindikatoren [online]. 12.07.2017. URL: [https://www.iqtiq.org/downloads/ergebnisse/bundesauswertung/2016/indirekte\\_verfahren/QSKH\\_16n1-GE BH\\_2016\\_BUAW\\_V02\\_2017-07-12.pdf](https://www.iqtiq.org/downloads/ergebnisse/bundesauswertung/2016/indirekte_verfahren/QSKH_16n1-GE BH_2016_BUAW_V02_2017-07-12.pdf).
6. Kolip P, Nolting HD, Zich K. Kaiserschnittgeburten in Deutschland: starker Anstieg und große regionale Unterschiede [online]. 2012 [Zugriff: 01.02.2017]. URL: [https://faktencheck-gesundheit.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Themenblatt\\_VV\\_FCKaiserschnitt\\_2012.pdf](https://faktencheck-gesundheit.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Themenblatt_VV_FCKaiserschnitt_2012.pdf).
7. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Klassifikation S3-Leitlinien [online]. [Zugriff: 19.10.2016]. URL: <http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung/awmf-regelwerk-01-planung-und-organisation/po-stufenklassifikation/klassifikation-s3.html>.
8. Dudenhausen JW. Praktische Geburtshilfe. Berlin: DeGruyter; 2011.
9. Krahl A, Schnepf W, Zu Sayn-Wittgenstein F. Die Bedeutung der Latenzphase der Geburt: eine historische Analyse. Z Geburtshilfe Neonatol 2016; 220(04): 155-165.
10. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Definitionen der Geburtsphasen: Rapid Report; Auftrag V16-01A. Köln: IQWiG; 2017. (IQWiG-Berichte; Band 517).
11. Deutsches Netzwerk für Qualitätsentwicklung in der Pflege, Verbund Hebammenforschung. Expertinnenstandard Förderung der physiologischen Geburt: Entwicklung, Konsentierung, Implementierung. Osnabrück: Deutsches Netzwerk für Qualitätsentwicklung in der Pflege; 2013.

12. Fannin M. Domesticating birth in the hospital: “family-centered” birth and the emergence of “homelike” birthing rooms. *Antipode* 2003; 35(3): 513–535
13. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G et al. GRADE guidelines; 2: framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011; 64(4): 395-400.
14. Langer G, Meerpohl JJ, Perleth M, Gartlehner G, Kaminski-Hartenthaler A, Schunemann H. GRADE-Leitlinien; 2: Formulierung der Fragestellung und Entscheidung über wichtige Endpunkte. *Z Evid Fortbild Qual Gesundheitswes* 2012; 106(5): 369-376.
15. Schneider H, Husslein P, Schneifer KTM. *Die Geburtshilfe*. Heidelberg: Springer; 2006.
16. Diedrich K, Holzgreve W, Jonat W, Schultze-Mosgau A, Schneider KTM, Weiss JM (Ed). *Gynäkologie und Geburtshilfe*. Heidelberg: Springer; 2007.
17. Oxford Centre for Evidence-based Medicine. Levels of evidence [online]. In: *Essential Evidence Plus*. [Zugriff: 09.12.2016]. URL: [https://www.essentialevidenceplus.com/product/ebm\\_loe.cfm?show=oxford](https://www.essentialevidenceplus.com/product/ebm_loe.cfm?show=oxford).
18. Seo HJ, Kim SY, Lee YJ, Jang BH, Park JE, Sheen SS et al. A newly developed tool for classifying study designs in systematic reviews of interventions and exposures showed substantial reliability and validity. *J Clin Epidemiol* 2016; 70: 200-205.
19. Rose L, Schmied-Knittel I. Magie und Technik: moderne Geburt zwischen biografischem Event und kritischem Ereignis. In: Villa PI, Moebius S, Thiessen B (Ed). *Soziologie der Geburt: Diskurse, Praktiken und Perspektiven*. Frankfurt am Main: Campus-Verlag; 2011. S. 75-100.
20. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ et al. CONSORT 2010: explanation and elaboration; updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
21. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health* 2004; 94(3): 361-366.
22. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147(8): 573-577.
23. Balshe H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J et al. GRADE guidelines; 3: rating the quality of evidence. *J Clin Epidemiol* 2011; 64(4): 401-406.
24. Meerpohl JJ, Langer G, Perleth M, Gartlehner G, Kaminski-Hartenthaler A, Schunemann H. GRADE-Leitlinien; 3: Bewertung der Qualität der Evidenz (Vertrauen in die Effektschätzer). *Z Evid Fortbild Qual Gesundheitswes* 2012; 106(6): 449-456.

25. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P et al. GRADE guidelines; 4: rating the quality of evidence; study limitations (risk of bias). *J Clin Epidemiol* 2011; 64(4): 407-415.
26. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med* 2009; 6(5): e1000065.
27. Centre for Evidence-Based Medicine. OCEBM levels of evidence [online]. [Zugriff: 26.04.2017]. URL: <http://www.cebm.net/ocebmllevels-of-evidence/>.
28. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R et al. GRADE guidelines; 12: preparing summary of findings tables; binary outcomes. *J Clin Epidemiol* 2013; 66(2): 158-172.
29. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.
30. Lange S. The all randomized/full analysis set (ICH E9): may patients be excluded from the analysis? *Drug Inf J* 2001; 35(3): 881-891.
31. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 5.0. Köln: IQWiG; 2017. URL: <https://www.iqwig.de/download/Allgemeine-Methoden-Version-5-0.pdf>.
32. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.
33. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414): 557-560.
34. Nussbaumer B, Gartlehner G, Kien C, Kaminski-Hartenthaler A, Langer G, Meerpohl JJ et al. Grade Leitlinien; 15: von der Evidenz zur Empfehlung; Determinanten, die Richtung und Stärke einer Empfehlung bestimmen. *Z Evid Fortbild Qual Gesundhwes* 2014; 108(7): 421-431.
35. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P et al. GRADE guidelines; 11: making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013; 66(2): 151-157.
36. Kaminski-Hartenthaler A, Gartlehner G, Kien C, Meerpohl JJ, Langer G, Perleth M et al. GRADE-Leitlinien; 11: Gesamtbeurteilung des Vertrauens in Effektschätzer für einen einzelnen Studienendpunkt und für alle Endpunkte. *Z Evid Fortbild Qual Gesundhwes* 2013; 107(9-10): 638-645.
37. Meerpohl JJ, Langer G, Perleth M, Gartlehner G, Kaminski-Hartenthaler A, Schunemann H. GRADE-Leitlinien; 4: Bewertung der Qualität der Evidenz; Studienlimitationen (Risiko für Bias). *Z Evid Fortbild Qual Gesundhwes* 2012; 106(6): 457-469.

38. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J et al. GRADE guidelines; 5: rating the quality of evidence; publication bias. *J Clin Epidemiol* 2011; 64(12): 1277-1282.
39. Nolting A, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A et al. GRADE Leitlinien; 5: Einschätzung der Qualität der Evidenz; Publikationsbias. *Z Evid Fortbild Qual Gesundhwes* 2012; 106(9): 670-676.
40. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D et al. GRADE guidelines; 6: rating the quality of evidence; imprecision. *J Clin Epidemiol* 2011; 64(12): 1283-1293.
41. Kulig M, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A et al. GRADE Leitlinien; 6: Einschätzung der Qualität der Evidenz; unzureichende Präzision. *Z Evid Fortbild Qual Gesundhwes* 2012; 106(9): 677-688.
42. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M et al. GRADE guidelines; 7: rating the quality of evidence; inconsistency. *J Clin Epidemiol* 2011; 64(12): 1294-1302.
43. Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A, Schunemann HJ. GRADE Leitlinien; 7: Einschätzung der Qualität der Evidenz; Inkonsistenz. *Z Evid Fortbild Qual Gesundhwes* 2012; 106(10): 733-744.
44. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M et al. GRADE guidelines; 8: rating the quality of evidence; indirectness. *J Clin Epidemiol* 2011; 64(12): 1303-1310.
45. Rasch A, Perleth M, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A et al. GRADE Leitlinien; 8: Einschätzung der Qualität der Evidenz; Indirektheit. *Z Evid Fortbild Qual Gesundhwes* 2012; 106(10): 745-753.
46. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P et al. GRADE guidelines; 9: rating up the quality of evidence. *J Clin Epidemiol* 2011; 64(12): 1311-1316.
47. Kien C, Gartlehner G, Kaminski-Hartenthaler A, Meerpohl JJ, Flamm M, Langer G et al. GRADE-Leitlinien; 9: Heraufstufen der Qualität der Evidenz. *Z Evid Fortbild Qual Gesundhwes* 2013; 107(3): 249-255.
48. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA et al. GRADE guidelines; 13: preparing summary of findings tables and evidence profiles; continuous outcomes. *J Clin Epidemiol* 2013; 66(2): 173-183.
49. Langer G, Meerpohl JJ, Perleth M, Gartlehner G, Schunemann H. GRADE-Leitlinien; 12: Erstellen von „summary-of-findings“-Tabellen; dichotome Endpunkte. *Z Evid Fortbild Qual Gesundhwes* 2013; 107(9-10): 646-664.
50. Schell LK, Meerpohl JJ, Gartlehner G, Langer G, Perleth M, Schunemann HJ. GRADE Leitlinien; 13: Erstellen von summary-of-findings-Tabellen und Evidenzprofilen; kontinuierliche Endpunkte. *Z Evid Fortbild Qual Gesundhwes* 2014; 108(5-6): 333-347.

51. Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa RA, Mbuagbaw L et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. *J Clin Epidemiol* 2016; 74: 7-18.
52. Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, Heus P et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016; 74: 28-39.
53. Hodnett ED, Stremler R, Weston JA, McKeever P. Re-conceptualizing the hospital labor room: the PLACE (pregnant and laboring in an ambient clinical environment) pilot trial. *Birth* 2009; 36(2): 159-166.
54. Ayerle G. Be-Up Studie: Geburt aktiv; klinische Studie [online]. 10.11.2017 [Zugriff: 30.11.2017]. URL: <https://be-up-studie.de/>.
55. Vice chancellor for research, Tehran University of Medical Sciences. Effect of multisensory stimulation room on pain of delivery intensity, childbearing self-efficacy and breastfeeding self-efficacy in one of selected hospital in Tehran University of Medical Sciences 1389 [online]. In: Iranian Registry of Clinical Trials. [Zugriff: 24.11.2017]. URL: <http://www.irct.ir/trial/3018>.
56. Vice chancellor for Research, School of Nursing and Midwifery, Iran University of Medical Sciences. Effectiveness of snoezelen room on fear, anxiety and satisfaction of childbirth in primiparous women [online]. In: Iranian Registry of Clinical Trials. [Zugriff: 24.11.2017]. URL: <http://www.irct.ir/trial/3024>.
57. Manesh MJ, Kalati M, Hosseini F. Snoezelen room and childbirth outcome: a randomized clinical trial. *Iran Red Crescent Med J* 2015; 17(5): e18373.
58. Herning Hospital. Birth environment of the future: full text view [online]. In: *ClinicalTrials.gov*. 12.10.2016 [Zugriff: 20.11.2017]. URL: <https://ClinicalTrials.gov/show/NCT02478385>.
59. Aburas R, Pati D, Casanova R, Adams NG. The influence of nature stimulus in enhancing the birth experience. *HERD* 2017; 10(2): 81-100.
60. Ayerle G. Effects of the birthing room environment on vaginal births and client-centred outcomes in women at term planning a vaginal birth: a multicentre RCT (BE-UP); study protocol [unveröffentlicht]. 2017.
61. Newburn M, Singh D. Creating a better birth environment: women's views about the design and facilities in maternity units; a national survey. London: National Childbirth Trust; 2003. URL: [https://www.nct.org.uk/sites/default/files/related\\_documents/BBE\\_report\\_311003.pdf](https://www.nct.org.uk/sites/default/files/related_documents/BBE_report_311003.pdf).

62. Bundesministerium für Gesundheit. Nationales Gesundheitsziel: Gesundheit rund um die Geburt [online]. 01.2017 [Zugriff: 13.12.2017]. URL: [https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5\\_Publikationen/Praevention/Broschueren/Nationales\\_Gesundheitsziel\\_Gesundheit\\_rund\\_um\\_die\\_Geburt.pdf](https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/5_Publikationen/Praevention/Broschueren/Nationales_Gesundheitsziel_Gesundheit_rund_um_die_Geburt.pdf).
63. Deutsche Gesellschaft für Gynäkologie und Geburtshilfe. Die Sectio caesarea: angemeldetes Leitlinienvorhaben [online]. [Zugriff: 29.11.2017]. URL: <http://www.awmf.org/leitlinien/detail/anmeldung/1/ll/015-084.html>.
64. Foureur M, Davis D, Fenwick J, Leap N, Iedema R, Forbes I et al. The relationship between birth unit design and safe, satisfying birth: developing a hypothetical model. *Midwifery* 2010; 26(5): 520-525.
65. Ulrich RS, Zimring C, Zhu X, DuBose J, Seo HB, Choi YS et al. A review of the research literature on evidence-based healthcare design. *HERD* 2008; 1(3): 61-125.
66. Hodnett ED, Downe S, Walsh D. Alternative versus conventional institutional settings for birth. *Cochrane Database Syst Rev* 2012; (8): CD000012.
67. Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. *J Med Libr Assoc* 2006; 94(4): 451-455.
68. Lefebvre C, Manheimer E, Glanville J. Searching for studies [online]. In: Higgings JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions: version 5.1.0*. 03.2011 [Zugriff: 17.02.2017]. URL: [http://handbook.cochrane.org/chapter\\_6/6\\_searching\\_for\\_studies.htm](http://handbook.cochrane.org/chapter_6/6_searching_for_studies.htm).
69. Wong SSL, Wilczynski NL, Haynes RB. Optimal CINAHL search strategies for identifying therapy studies and review articles. *J Nurs Scholarsh* 2006; 38(2): 194-199.
70. Hodnett ED. Pain and women's satisfaction with the experience of childbirth: a systematic review. *Am J Obstet Gynecol* 2002; 186(Suppl 5): S160-S172.

## **10 Studienlisten**

### **10.1 Liste der eingeschlossenen Studien**

#### **Aburas 2017**

Aburas R, Pati D, Casanova R, Adams NG. The influence of nature stimulus in enhancing the birth experience. *HERD* 2017; 10(2): 81-100.

#### **Hodnett 2009**

Hodnett ED, Stremler R, Weston JA, McKeever P. Re-conceptualizing the hospital labor room: the PLACE (pregnant and laboring in an ambient clinical environment) pilot trial. *Birth* 2009; 36(2): 159-166.

### **10.2 Liste der gesichteten systematischen Übersichten**

1. Hodnett ED. Pain and women's satisfaction with the experience of childbirth: a systematic review. *Am J Obstet Gynecol* 2002; 186(Suppl 5): S160-S172.

2. Hodnett ED, Downe S, Walsh D. Alternative versus conventional institutional settings for birth. *Cochrane Database Syst Rev* 2012; (8): CD000012.

### **10.3 Liste der ausgeschlossenen Publikationen zum Thema mit Ausschlussgründen**

#### **Nicht E1 (Population)**

1. Manesh MJ, Kalati M, Hosseini F. Snoezelen room and childbirth outcome: a randomized clinical trial. *Iran Red Crescent Med J* 2015; 17(5): e18373.

#### **Nicht E6 (Publikation)**

1. Homer CSE. Home-like environments for labour and birth: benefits for women and babies. *Evid Based Med* 2013; 18(4): e32.

## Anhang A – Suchstrategien

### A.1 – Bibliografische Literaturrecherche

#### 1. MEDLINE

##### *Suchoberfläche: Ovid*

- Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations September 27, 2017
- Ovid MEDLINE(R) 1946 to September Week 3 2017
- Ovid MEDLINE(R) Daily Update September 27, 2017
- Ovid MEDLINE(R) Epub Ahead of Print September 27, 2017

Es wurden folgende Filter übernommen:

- Systematische Übersicht: Wong [67] – High specificity strategy
- RCT: Lefebvre [68] – Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity-maximizing version (2008 revision)

#	Searches
1	Delivery Rooms/
2	((birth or labor or labour or ambient*) adj3 room*).ab,ti.
3	(birthroom* or birth environment*).ab,ti.
4	(delivery adj3 suite).ab,ti.
5	or/1-4
6	randomized controlled trial.pt.
7	controlled clinical trial.pt.
8	(randomized or placebo or randomly or trial or groups).ab.
9	drug therapy.fs.
10	or/6-9
11	10 not (exp animals/ not humans.sh.)
12	Cochrane database of systematic reviews.jn.
13	(search or MEDLINE or systematic review).tw.
14	meta analysis.pt.
15	or/12-14
16	or/11,15
17	and/5,16
18	..l/ 17 yr=2000-Current
19	18 not (comment or editorial).pt.

## 2. PubMed

### *Suchoberfläche: NLM*

- PubMed – as supplied by publisher
- PubMed – in process
- PubMed – pubmednotmedline

Search	Query
#1	Search (birth room*[TIAB] OR birthroom*[TIAB] OR labor room*[TIAB] OR labour room*[TIAB] OR birth environment*[TIAB])
#2	Search ((birth[TIAB] or labor[TIAB] or labour[TIAB]) AND bedroom[TIAB])
#3	Search (delivery [TIAB] AND suite [TIAB])
#4	Search (#1 OR #2 OR #3)
#5	Search (clinical trial*[TIAB] OR random*[TIAB] OR placebo[TIAB] OR trial[TI])
#6	Search (search[TIAB] OR meta analysis[TIAB] OR MEDLINE[TIAB] OR systematic review[TIAB])
#7	Search (#5 OR #6)
#8	Search (#4 AND #7)
#9	Search (#8 NOT Medline [SB])
#10	Search (#9 AND 2000:2017 [DP])

## 3. Cinahl

### *Suchoberfläche: Ebsco*

Es wurden folgende Filter übernommen:

- Systematische Übersicht: Wong [69] – Small drop in sensitivity with a substantive gain in specificity;
- RCT: Wong [69] – Small drop in sensitivity with a substantive gain in specificity

#	Query
S1	(MH "Facility Design and Construction")
S2	(MH "Interior Design and Furnishings")
S3	TX (birth or labor or labour) N3 room*)
S4	TX delivery N3 suite OR TX ( birthroom* or birth environment* )
S5	S1 OR S2 OR S3 OR S4
S6	MH "Treatment Outcomes+"
S7	MH "Experimental Studies"

#	Query
S8	TX random*
S9	S6 OR S7 OR S8
S10	TX meta analysis
S11	PT review
S12	S10 OR S11
S13	S9 OR S12
S14	S5 AND S13
S15	S14 AND (PY 2000-2017)

#### 4. Embase

##### *Suchoberfläche: Ovid*

- Embase 1974 to 2017 September 27

Es wurden folgende Filter übernommen:

- Systematische Übersicht: Wong [67] – High specificity strategy;
- RCT: Wong [67] – Strategy minimizing difference between sensitivity and specificity

#	Searches
1	Delivery room/
2	*Maternity ward/
3	((birth or labor or labour or ambient*) adj3 room*).ab,ti.
4	(birthroom* or birth environment*).ab,ti.
5	(delivery adj3 suite).ab,ti.
6	or/1-5
7	(random* or double-blind*).tw.
8	placebo*.mp.
9	or/7-8
10	(meta analysis or systematic review or MEDLINE).tw.
11	or/9-10
12	and/6,11
13	..l/ 12 yr=2000-Current
14	13 not medline.cr.
15	14 not (exp animal/ not exp humans/)
16	15 not (Conference Abstract or Conference Review or Editorial).pt.

## 5. The Cochrane Library

### *Suchoberfläche: Wiley*

- Cochrane Database of Systematic Reviews: Issue 9 of 12, September 2017
- Cochrane Central Register of Controlled Trials: Issue 8 of 12, August 2017

ID	Search
#1	[mh ^"Delivery Rooms "]
#2	((birth or labor or labour) near/3 room*):ti,ab
#3	(birthroom* or birth near/1 environment*):ab,ti
#4	(delivery near/3 suite):ti,ab
#5	#1 or #2 or #3 or #4 Publication Year from 2000, in Cochrane Reviews (Reviews and Protocols) and Trials

## 6. Health Technology Assessment Database

### *Suchoberfläche: Centre for Reviews and Dissemination*

- HTA

Line	Search
1	(MeSH DESCRIPTOR Delivery rooms EXPLODE ALL TREES)
2	((birth OR labor OR labour) AND room*)
3	(birthroom* OR birth environment*)
4	(delivery AND suite)
5	#1 OR #2 OR #3 OR #4
6	((#5) FROM 2000 TO 2017)
7	((#6) IN HTA)

## A.2 – Suche in Studienregistern

### 1. ClinicalTrials.gov

#### *Anbieter: U.S. National Institutes of Health*

- URL: <http://www.clinicaltrials.gov>
- Eingabeoberfläche: Advanced Search

Suchstrategie
( birth room OR birthroom OR labor room OR labour room OR delivery room OR birth environment ) [TREATMENT]

## 2. International Clinical Trials Registry Platform Search Portal

*Anbieter: World Health Organization*

- URL: <http://apps.who.int/trialsearch/>
- Eingabeoberfläche: Standard Search

<b>Suchstrategie</b>
birth room OR birthroom OR labor room OR labour room OR delivery room OR birth environment

## Anhang B – Verzerrungspotenzial auf Endpunktebene

### B.1 – Geburtsmodus

Die folgende Tabelle 13 zeigt die endpunktspezifische Bewertung des Verzerrungspotenzials für den Geburtsmodus.

Tabelle 13: Bewertung des Verzerrungspotenzials auf Endpunktebene: Geburtsmodus

Studie	Verzerrungspotenzial auf Studienebene	Verblindung endpunkterhebende Personen	ITT-Prinzip adäquat umgesetzt	Ergebnisunabhängige Berichterstattung	Fehlen sonstiger Aspekte	Verzerrungspotenzial auf Endpunktebene
Aburas 2017	hoch	unklar	nein	unklar	nein	hoch <sup>a, b</sup>
Hodnett 2009	mittel	unklar	ja	ja	nein	hoch <sup>a</sup>

a: fehlende Verblindung der behandelnden Personen  
b: Ergebnisse zu 10 randomisierten Teilnehmerinnen sind nicht in der Analyse enthalten. Fehlende Differenzierung zwischen Spontangeburt und vaginal-operativer Geburt  
ITT: Intention to treat

### B.2 – Geburtserleben

Die folgende Tabelle 14 zeigt die endpunktspezifische Bewertung des Verzerrungspotenzials für das Geburtserleben.

Tabelle 14: Bewertung des Verzerrungspotenzials auf Endpunktebene: Geburtserleben

Studie	Verzerrungspotenzial auf Studienebene	Verblindung endpunkterhebende Personen	ITT-Prinzip adäquat umgesetzt	Ergebnisunabhängige Berichterstattung	Fehlen sonstiger Aspekte	Verzerrungspotenzial auf Endpunktebene
Aburas 2017	hoch	nein	nein	nein	nein	hoch <sup>a, b</sup>
Hodnett 2009	mittel	nein	nein	ja	nein	hoch <sup>a, c</sup>

a: fehlende Verblindung der behandelnden und endpunkterhebenden Personen sowie Verwendung eines nicht validierten Fragebogens  
b: fehlende Anwendung des ITT-Prinzips, da nur Ergebnisse zu 42 von 60 Teilnehmerinnen vorliegen; Vermutung einer ergebnisabhängigen Berichterstattung, da nur als statistisch signifikant bezeichnete Ergebnisse ausgewählter Subskalen berichtet werden  
c: unklare Anwendung des ITT-Prinzips aufgrund fehlender Angaben zur Gesamtanzahl der vorliegenden Antworten  
ITT: Intention to treat

### B.3 – Wehenmittel

Die folgende Tabelle 15 zeigt die endpunktspezifische Bewertung des Verzerrungspotenzials für die intrapartale Wehenmittelgabe.

Tabelle 15: Bewertung des Verzerrungspotenzials auf Endpunktebene: Wehenmittel intrapartal

Studie	Verzerrungspotenzial auf Studienebene	Verblindung endpunkterhebende Personen	ITT-Prinzip adäquat umgesetzt	Ergebnisunabhängige Berichterstattung	Fehlen sonstiger Aspekte	Verzerrungspotenzial auf Endpunktebene
Hodnett 2009	mittel	unklar	ja	ja	nein	hoch <sup>a</sup>
a: fehlende Verblindung der behandelnden Personen ITT: Intention to treat						

## **Anhang C – Ergebnisse**

In diesem Abschnitt folgt eine tabellarische Darstellung der Ergebnisse zu den kritischen und klinisch wichtigen Endpunkten für die Suche nach vergleichenden Interventionsstudien zur modifizierten Umgebungsgestaltung. Aus der Kategorie der kritischen Endpunkte wurden Ergebnisse zum Geburtsmodus (siehe Tabelle 16) und zum Geburtserleben (siehe Tabelle 17 bis Tabelle 19) berichtet. Zusätzlich werden ergänzende Informationen zum Apgar-Score nach 5 Minuten  $< 6$  siehe (Tabelle 20) dargestellt.

Aus der Kategorie der klinisch wichtigen Endpunkte wurden Ergebnisse zur Wehenmittelgabe berichtet (siehe Tabelle 22).

### **C.1 – Geburtsmodus**

Die folgende Tabelle 16 zeigt die Ergebnisse für den Geburtsmodus.

Tabelle 16: Ergebnisse zum Geburtsmodus

Studie (modifizierte Umgebungsgestaltung)	n <sup>a</sup>	Spontangeburt		Kaiserschnitt		Vaginal-operative Geburt	
		Gebärende mit Ereignis n (%)	RR [95 %-KI] / [98,3 %-KI]	Gebärende mit Ereignis n (%)	RR [95 %-KI] / [98,3 %-KI]	Gebärende mit Ereignis n (%)	RR [95 %-KI] / [98,3 %-KI]
<b>Hodnett 2009</b> (umfangreiche Umgebungsgestaltung)							
Prüfintervention	30 <sup>b</sup>	21 (70,0) <sup>c</sup>	0,87 [0,65; 1,16] <sup>c</sup> /	4 (13,3) <sup>c</sup>	1,38 [0,34; 5,64] <sup>c</sup> /	5 (16,7) <sup>c</sup>	1,72 [0,45; 6,58] <sup>c</sup> /
Vergleichsintervention	31	25 (80,6) <sup>c</sup>	0,87 [0,61; 1,24] <sup>c</sup>	3 (9,7) <sup>c</sup>	1,38 [0,25; 7,71] <sup>c</sup>	3 (9,7) <sup>c</sup>	1,72 [0,33; 8,85] <sup>c</sup>
<b>Aburas 2017</b> (geringere Umgebungsgestaltung)							
Prüfintervention	24 <sup>b</sup>	k. A.	–	2 (8,3) <sup>c</sup>	0,72 [0,13; 3,96] <sup>c</sup>	k. A.	–
Vergleichsintervention	26 <sup>b</sup>	k. A.		3 (11,5) <sup>c</sup>	0,72 [0,09; 5,77] <sup>c</sup>	k. A.	
a: Anzahl randomisierter Teilnehmerinnen. Werte, die auf anderen Zahlen basieren, werden in der entsprechenden Spalte gekennzeichnet, wenn Abweichung relevant							
b: Anzahl ausgewerteter Teilnehmerinnen							
c: eigene Berechnung							
k. A.: keine Angabe; KI: Konfidenzintervall; n: Anzahl der Gebärenden mit Ereignis; RR: relatives Risiko							

## C.2 – Geburtserleben

Tabelle 17 bis Tabelle 19 fassen die berichteten Ergebnisse zum Geburtserleben zusammen. Alle vorliegenden Informationen wurden über nicht validierte Fragebogen erhoben.

Tabelle 17: Ergebnisse Geburtserleben (Teilaspekt Wirkung des Gebärraumes)

Studie Endpunktkategorie	n <sup>a</sup> / n <sup>b</sup>	Geburtserleben (emotionale Wirkung des Gebärraums)		
		Beurteilungen durch Studienteilnehmerinnen, n (%)		
<b>Hodnett 2009</b> (umfangreiche Umgebungsgestaltung)				
<b>Beruhigende Wirkung des Raumes</b>		gesamte / meiste Zeit	andere / keine Antwort <sup>c</sup>	Testergebnisse
Prüfintervention	31 / 29	13 (44,8) <sup>d</sup>	16 <sup>d</sup> (55,2) <sup>d</sup>	k. A.
Vergleichsintervention	31 / 30	13 (43,3) <sup>d</sup>	17 <sup>d</sup> (46,7) <sup>d</sup>	
<b>Ausreichende Privatsphäre</b>		gesamte / meiste Zeit	andere / keine Antwort <sup>c</sup>	
Prüfintervention	31 / 29	21 (72,4) <sup>d</sup>	8 <sup>d</sup> (27,6) <sup>d</sup>	k. A.
Vergleichsintervention	31 / 30	22 (73,3) <sup>d</sup>	8 <sup>d</sup> (26,7) <sup>d</sup>	
<b>Entspannende Wirkung des Raumes</b>		gesamte / meiste Zeit	andere / keine Antwort <sup>c</sup>	
Prüfintervention	31 / 29	15 (51,7) <sup>d</sup>	14 <sup>d</sup> (48,3) <sup>d</sup>	X <sup>2</sup> = 2,42, p = 0,12
Vergleichsintervention	31 / 30	10 (33,3) <sup>d</sup>	20 <sup>d</sup> (66,7) <sup>d</sup>	
<b>Beängstigende Wirkung des Raumes</b>		niemals / selten	andere / keine Antwort <sup>c</sup>	
Prüfintervention	31 / 29	19 (65,5) <sup>d</sup>	10 <sup>d</sup> (34,5) <sup>d</sup>	X <sup>2</sup> = 1,89, p = 0,17
Vergleichsintervention	31 / 30	25 (83,3) <sup>d</sup>	5 <sup>d</sup> (16,5) <sup>d</sup>	
a: Anzahl randomisierter Teilnehmerinnen				
b: Anzahl ausgefüllter Fragebogen. Es liegen keine Informationen zur Anzahl der vorliegenden Antworten vor.				
c: Es wird nur die Anzahl von 2 Kategorien berichtet. Informationen zu weiteren Kategorien liegen nicht vor.				
d: eigene Berechnung bezogen auf die Anzahl ausgefüllter Fragebogen				
k. A.: keine Angabe; KI: Konfidenzintervall; n: Anzahl der Studienteilnehmerinnen mit Antwort				

Tabelle 18: Ergebnisse zum Geburtserleben (Teilaspekt Unterstützung durch das Pflegepersonal)

Studie (modifizierte Umgebungsgestaltung)	n <sup>a</sup>	Wert ± SD	Gruppenunterschied
			Differenz [95 %-KI]; p-Wert Intervention minus Kontrolle
<b>Hodnett 2009</b> (umfangreiche Umgebungsgestaltung)			
Prüfintervention	29	52,2 ± 8,9	2,90 [-2,15; 7,95] <sup>b, c</sup> ; p = 0,255
Vergleichsintervention	30	49,3 ± 10,4	
a: Anzahl der Teilnehmerinnen, die in der Auswertung zur Berechnung des Effektschätzers berücksichtigt wurden			
b: eigene Berechnung			
c: Labour Support Questionnaire [70], höhere Scorewerte beschreiben eine aus Patientenperspektive bessere Unterstützung			
KI: Konfidenzintervall; SD: Standardabweichung			

Tabelle 19: Ergebnisse zum Geburtserleben (Teilaspekt Qualität der Pflege aus der Patientenperspektive)

Studie (modifizierte Umgebungsgestaltung)	n <sup>a</sup> / n <sup>b</sup>	Geburtserleben (Subskala QPP-Fragebogen)	
		Mittelwert (SD)	Gruppenunterschied Differenz [95 %-KI]; p-Wert
<b>Aburas 2017</b> (geringere Umgebungsgestaltung)			
<b>Gesamtscore</b>			
Prüfintervention	24 / 21	4,63 (k. A.)	-0,17 <sup>c</sup> [k. A.]; p > 0,05
Vergleichsintervention	26 / 21	4,46 (k. A.)	
<b>Das Gefühl, dass ich von den Ärzten ernst genommen und respektvoll behandelt wurde (Item 1)</b>			
Prüfintervention	24 / 21	k. A.	-0,381 [-0,756; -0,006]; p = 0,047
Vergleichsintervention	26 / 21	k. A.	
<b>Das Gefühl, dass ich von den Ärzten in einer positiven Art und Weise behandelt wurde (Item 2)</b>			
Prüfintervention	24 / 21	k. A.	-0,283 [-0,652; 0,081]; p = 0,123
Vergleichsintervention	26 / 21	k. A.	
<b>Das Gefühl, dass die Ärzte Mitgefühl zeigten, wenn ich z. B. an Angst oder Schmerzen litt (Item 7)</b>			
Prüfintervention	24 / 21	k. A.	-0,524 [-1,048; 0,000]; p = 0,050
Vergleichsintervention	26 / 21	k. A.	
a: Anzahl ausgewerteter Teilnehmerinnen b: Anzahl ausgefüllter Fragebogen c: eigene Berechnung k. A.: keine Angabe; KI: Konfidenzintervall; n: Anzahl der Studienteilnehmerinnen mit Antwort; QPP-Fragebogen: Quality of Care from the Patient's Perspective (Fragebogen zur Erhebung der Qualität der Betreuung aus Patientenperspektive); SD: Standardabweichung			

### C.3 – Ergänzende Informationen zum Apgar-Score nach 5 Minuten < 6

Tabelle 20 und Tabelle 21 fassen die vorliegenden ergänzenden Informationen zum Auftreten eines geringen Apgar-Scores nach 5 Minuten zusammen. In 1 Studie lagen ausschließlich Informationen zu einem Apgar-Score nach 5 Minuten < 7 und in 1 weiteren Studie zum Mittelwert und zur Standardabweichung vor.

Tabelle 20: Ergänzende Informationen zum Apgar-Score nach 5 Minuten < 6 (Apgar-Score nach 5 Minuten < 7)

Studie (Managementoptionen)	n <sup>a</sup>	Apgar-Score nach 5 Minuten < 7	
		Neugeborene mit Ereignis n (%)	RR [95 %-KI]
<b>Hodnett 2009</b> (umfangreiche Umgebungsgestaltung)			
Prüfintervention	30 <sup>b</sup>	1 (3,3) <sup>c</sup>	3,10 [0,13; 73,16]
Vergleichsintervention	31	0	
a: Anzahl randomisierter Teilnehmerinnen b: Anzahl ausgewerteter Neugeborener c: eigene Berechnung Apgar-Score: Schema zur Beurteilung der postpartalen Adaptation und Vitalität des reifen Neugeborenen nach den Kriterien Atmung, Puls, Grundtonus, Aussehen und Reflexe; KI: Konfidenzintervall; n: Anzahl der Neugeborenen mit Ereignis; RR: relatives Risiko			

Tabelle 21: Ergänzende Informationen für den Apgar-Score nach 5 Minuten

Studie (modifizierte Umgebungsgestaltung)	n <sup>a</sup>	Mittelwert (SD)	Gruppenunterschied
			Differenz [95 %-KI]; p-Wert Intervention minus Kontrolle
<b>Aburas 2017</b> (geringere Umgebungsgestaltung)			
Prüfintervention	24	8,92 (0,408)	k. A.; p = 0,84 <sup>c</sup>
Vergleichsintervention	26	8,65 (0,689)	
a: Anzahl der Teilnehmerinnen, die in der Auswertung zur Berechnung des Effektschätzers berücksichtigt wurden b: eigene Berechnung c: Vergleich zwischen 3 Gruppen Apgar-Score: Schema zur Beurteilung der postpartalen Adaptation und Vitalität des reifen Neugeborenen nach den Kriterien Atmung, Puls, Grundtonus, Aussehen und Reflexe; k. A.: keine Angabe; KI: Konfidenzintervall; n: Anzahl der Gebärenden mit Ereignis; SD: Standardabweichung			

#### C.4 – Wehenmittel

Die folgende Tabelle 22 zeigt die Ergebnisse für die Häufigkeit einer Wehenmittelgabe während der Geburt.

Tabelle 22: Ergebnisse für die Häufigkeit einer Wehenmittelgabe während der Geburt

Studie (modifizierte Umgebungsgestaltung)	n <sup>a</sup>	Wehenmittelgabe	
		Gebärende mit Ereignis n (%)	RR [95 %-KI]
<b>Hodnett 2009</b> (umfangreiche Umgebungsgestaltung)			
Prüfintervention	30 <sup>b</sup>	12 (40,0) <sup>c</sup>	0,59 [0,36; 0,97] <sup>c</sup>
Vergleichsintervention	31	21 (67,7) <sup>c</sup>	
a: Anzahl randomisierter Teilnehmerinnen b: Anzahl ausgewerteter Teilnehmerinnen c: eigene Berechnung KI: Konfidenzintervall; n: Anzahl der Gebärenden mit Ereignis; RR: relatives Risiko			