



## CLINICAL RESEARCH ARTICLE

# Concordance of expert clinicians' interpretations of the newborn's true physiological state

Jelena Zestic<sup>1</sup>, Helen Liley<sup>1,2</sup> and Penelope Sanderson<sup>1</sup>

**BACKGROUND:** Many physiological aspects of the neonatal transition after birth are unobservable because relevant sensors do not yet exist, compromising clinicians' understanding of a neonate's physiological status. Given that a neonate's true physiological state is currently unavailable, we explored the feasibility of using clinicians' degree of concordance as an approximation of the true physiological state.

**METHODS:** Two phases of structured interviews were conducted. In Phase 1 ( $N = 8$ ) and Phase 2 ( $N = 12$ ), we presented neonatal experts with eight graphical trajectories of real newborns' heart rate and oxygen saturation values in the first 10–15 min after birth. We elicited the participants' interpretations of potential underlying physiological conditions that could explain the vital sign patterns.

**RESULTS:** The global differential diagnosis data for each phase produced the same pattern of results: (1) four trajectories produced a substantial degree of concordance among clinicians (61–80%) and (2) four trajectories produced a strong degree of concordance among clinicians (81–100%).

**CONCLUSIONS:** It is possible to achieve a strong degree of concordance among neonatal experts' interpretations of newborn trajectories. Thus, using the degree of concordance as an approximation of the neonate's true physiological state in resuscitation after birth may be a promising direction to explore for cognitive aid design.

*Pediatric Research* (2022) 91:1222–1230; <https://doi.org/10.1038/s41390-021-01565-9>

## IMPACT:

- Differential diagnoses with a good degree of concordance among expert neonatal clinicians could potentially substitute in part for the direct measurement of key physiological and anatomical variables of the neonatal transition, which is currently unavailable.
- The concordance of clinicians' judgements or inferences with regards to the true physiological state of the newborn during resuscitation after birth has never been explored.
- The findings provide a crucial first step toward using consensus of neonatal experts' judgements in the design of a cognitive aid to support clinicians' management of the newborns who require resuscitation after birth.

## INTRODUCTION

For the neonatal transition from intrauterine life to extrauterine life to occur, a series of complex interrelated physiological adjustments must take place to change how key organ systems function.<sup>1,2</sup> If any of these physiological adjustments is unsuccessful, the newborn may experience life-threatening consequences such as hypoxia and ischaemia.<sup>3</sup> At present, sensors indicating the state of the crucial physiological and anatomical adjustments that support the neonatal transition are unavailable, minimal or compromised.<sup>1–5</sup> Furthermore, monitoring devices currently used and recommended in contemporary resuscitation practice measure vital signs, but they do not convey the physiological and anatomical state of the newborn—the clinician must infer the state from the vital signs. As a result, the newborn is often viewed as a 'black box'; clinicians cannot know for certain what is happening physiologically at any moment in the first 10 min after birth.<sup>1–3</sup> However, such knowledge is vital not only for monitoring

the infant but also for determining underlying conditions and appropriate treatment actions. Thus, until sensors capable of measuring specific physiological and anatomical parameters related to the neonatal transition are developed and integrated into practice, the challenge is to find the best approximation of the neonate's true physiological state that could inform clinicians about physiological adjustments that are currently unobservable. We use the term "true physiological state" to describe the rapidly changing and interrelated state of the infant's lung aeration, organ blood flows, fetal shunts and other variables, including both those visible and invisible to the resuscitation team.

To the best of our knowledge, a newborn's true physiological state during transition has largely been explored using animal participants and highly invasive equipment.<sup>6–8</sup> Where non-invasive measurements have been used on human infants, the newborns were delivered via caesarean section without transition complications and/or congenital malformations.<sup>9–11</sup> Thus, it is

<sup>1</sup>The University of Queensland, St Lucia, QLD, Australia and <sup>2</sup>Mater Mothers' Hospital and Mater Research Institute, Brisbane, QLD, Australia  
Correspondence: Jelena Zestic (jelena.zestic@uqconnect.edu.au)

Received: 17 December 2020 Revised: 23 March 2021 Accepted: 26 April 2021  
Published online: 28 June 2021

unclear whether such equipment is robust in the resuscitation of a newborn with transition complications and/or congenital malformations. However, an effective first approximation to this information may be the similarity of expert neonatal clinicians' inferences about what is happening on the basis of indirect information—a property that has been called 'concordance'.<sup>12,13</sup> For instance, during newborn resuscitation, clinicians can observe the outputs of monitoring instruments such as the neonate's heart rate and oxygen saturation and clinical signs such as skin colour, tone and respiratory effort. If there is a high degree of concordance among expert clinicians in their interpretations of the signs, then analysts could use their combined interpretations as a suitable approximation of the neonate's physiological and anatomical progress during the transition. Notably, the concordance of expert neonatal clinicians' judgements, inferences, and decision-making during the newborn transition has never been explored. However, concordance-based approaches have been used in other clinical contexts where the true physiological state is unavailable, and to validate differential diagnoses provided by diagnostic support algorithms.<sup>14–16</sup>

The rationale for this study is if sufficient concordance can be achieved among expert clinicians' interpretations of newborns' vital signs during the transition, then this information—the diagnosis or diagnoses—could potentially substitute for the sensors that are currently unavailable, minimal or compromised, and inform the design of a cognitive aid. Cognitive aids are tools that guide a user through the performance of a task with the purpose of reducing task errors or omissions and improving speed and fluidity of performance.<sup>17,18</sup> Cognitive aids can be in the form of cards, posters or desktop or hand-held computer-based help systems. A successful cognitive aid will improve the user's comprehension of the situation and their selection of actions. Recent evidence indicates that cognitive aids designed for anaesthesia, acute care, emergency department resuscitation and neonatal resuscitation can improve clinicians' task performance.<sup>19–23</sup> However, few cognitive aids have suggested diagnoses to the end-user.<sup>24</sup> Thus, evaluating the concordance of clinicians' diagnoses as a potential indicator of the neonate's physiological status during resuscitation is a crucial first step toward the design of a cognitive aid that could supplement existing resuscitation algorithms by suggesting diagnoses. The cognitive aid could help clinicians monitor and understand the physiological state of the infant, guiding their treatment decisions where the algorithms do not offer clear direction, as with atypical responses to resuscitation interventions or less common physiological conditions, including lung hypoplasia, congenital diaphragmatic hernia and congenital heart defects.

The primary purpose of this study was to explore whether there was sufficient concordance of experts' diagnoses that they could act as a substitute for the currently unavailable direct measurement of crucial physiological and anatomical parameters in the neonatal transition.

## METHODS

### Context and participants

This two-phase structured interview study was conducted with senior neonatologists from several tertiary institutions in Australia and New Zealand. In *Phase 1*, eight senior neonatal consultants were recruited from two tertiary hospitals in Brisbane, Queensland, Australia ( $n = 7$  and  $n = 1$ , respectively). The participants in *Phase 1* were recruited via a personal approach from one of the authors (H.L.) who is a senior neonatologist in a tertiary hospital in Brisbane, Queensland, Australia. All of the participants in *Phase 1* regularly participated in newborn resuscitation and several participants provided newborn resuscitation training to registrars and midwives.

In *Phase 2*, 12 senior neonatal consultants were recruited from eight tertiary hospitals either via personal approach by the same

author from *Phase 1* or via two other neonatologists in different tertiary hospitals. The geographic distribution of participants was as follows: Queensland (three hospitals,  $n = 6$ ); New South Wales (one hospital,  $n = 2$ ); Victoria ( $n = 1$ ); South Australia ( $n = 1$ ); Western Australia ( $n = 1$ ); and New Zealand ( $n = 1$ ). All participants in *Phase 2* were involved in newborn resuscitation research and practice and several participants were also involved in the development of neonatal resuscitation guidelines in Australia and New Zealand.

The study was given ethical approval by Mater Misericordiae Ltd (approval 53861) and by the Human Research Ethics Committee at The University of Queensland (approval 2019002697). Participants were required to have at least two years of experience as a senior fellow or consultant in newborn medicine. All participants provided written informed consent.

### Design and procedure

The structured walkthrough interview technique is commonly used in human–computer interaction studies to elicit knowledge from subject matter experts engaged in complex tasks and it can be tailored for a specific project domain.<sup>25</sup> In both *Phase 1* and *Phase 2*, participants were presented with eight neonatal trajectories that the authors had extracted and presented as timeline graphics from the Dawson et al.<sup>26</sup> database, which documents the real-time heart rate and oxygen saturation values of ~465 neonates in the first 10–15 min after birth. The 25th and 75th percentile boundaries were inferred by Dawson et al.<sup>26</sup> in 2010 from the database of neonatal trajectories. Further information was also available for each neonatal trajectory, including gestation, birth weight, mode of delivery and supplemental oxygen. The trajectories chosen for the present studies reflected the most common patterns in neonatal transitions found in the database. During the interview, the voice and hands/cursor of the participant were captured on video to preserve the rich level of detail in the participant's responses and to help the researcher subsequently transcribe the interviews.

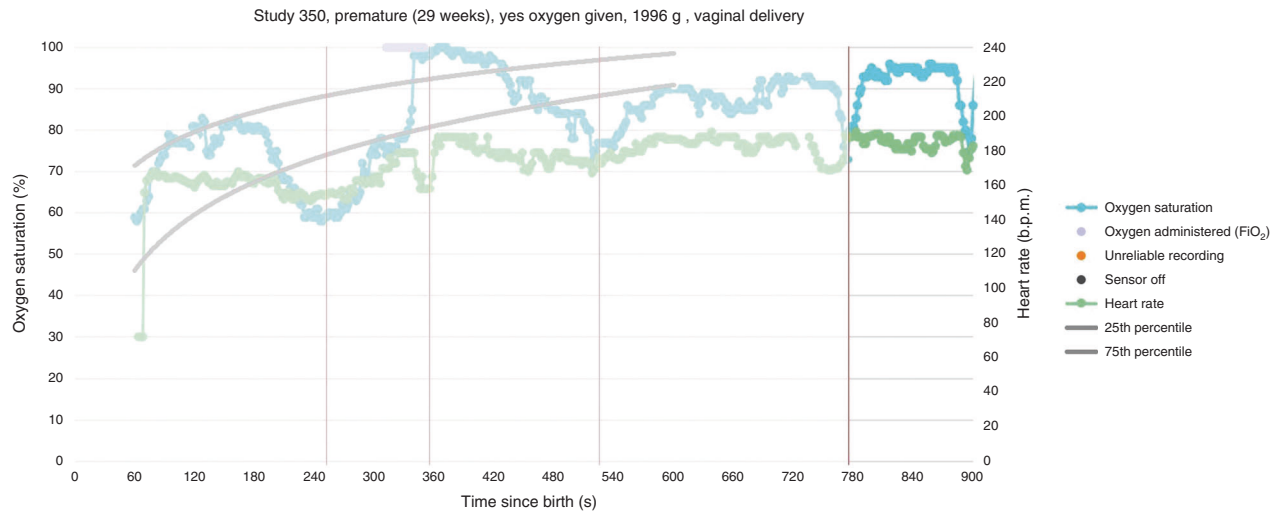
*Phase 1*: The researcher conducted the structured walkthrough interviews in person and individually with each participant. The researcher first introduced the participant to the format of the neonatal trajectories and gave them a brief outline of the key interview questions.

The researcher showed the first neonatal trajectory on a laptop computer. Each trajectory was successively revealed in a series of snapshots, where each snapshot took the trajectory up to a potentially clinically meaningful moment in the resuscitation (see Fig. 1). As a result, the number of snapshots varied between the eight trajectories. At each snapshot, the remainder of the trajectory was occluded to avoid bias that might result if the participant could see the entire trajectory.

During the presentation of each snapshot, the researcher allowed the participant to freely articulate their thoughts, further stimulating the discussion by asking the participant key questions to elicit their opinion of potential underlying physiological diagnoses or physiological states at that point. The researcher did not suggest diagnoses or physiological states or offer any affirmation or contradiction of the participant's suggestions. Below are some examples of the key questions used in the interview:

- What is the concerning sign/s that you see in the trend at this point in time?
- Given the gestational age of the neonate, what are the most likely underlying conditions?
- For each condition/s, why are the sign/s representative of the condition/s you have identified?

The researcher repeated this process for the remaining snapshots within the neonatal trajectory. Once the entire trajectory had been revealed, the researcher asked the participant to provide



**Fig. 1 An example of a neonatal trajectory in Phase 1 being gradually revealed in snapshots during the interview.** The graph depicts several snapshots—each snapshot is defined by the boundaries of the red lines. At each snapshot, the researcher would administer the same interview questions, but in relation to the new portion of trajectory that had been revealed.

a global differential diagnosis for the neonatal trajectory. The process outlined above was repeated for each neonatal trajectory. At the conclusion of the interview, the researcher provided an overview of the study's rationale to the participant and informed them that a transcript of the interview would be provided within one or two days if they wished to review it.

**Phase 2:** The researcher again conducted structured walk-through interviews individually with each participant, but the interviews were now exclusively online due to COVID-19 pandemic restrictions. The researcher first introduced the participant to the format of the neonatal trajectories and gave them a brief outline of the key interview questions. Seven of the newborn trajectories were the same across Phase 1 and Phase 2—the eighth trajectory from Phase 1 was replaced in Phase 2 because participants had inferred that no underlying physiological derangements were present.

A browser-based application presented trajectories to the participant at ten times real-time speed (see Fig. 2). In contrast to Phase 1, where snapshots were chosen in advance by the researcher, in Phase 2 the participant was asked to pause the progression of the trajectory when they recognised a pattern in the vital signs that were clinically meaningful to them. During the pause, the researcher first allowed the participant to articulate their thoughts, and where necessary, further stimulated the discussion by asking the participant the same key questions as in Phase 1 to elicit their opinion of potential underlying physiological diagnoses or physiological states at that point. The researcher did not suggest diagnoses or physiological states or offer any affirmation or contradiction of the participant's suggestions. The participant and researcher repeated this process until the entire trajectory had been revealed. The researcher then asked the participant to provide a global differential diagnosis for the neonatal trajectory. The above process was repeated for all trajectories and the study was concluded as described for Phase 1. Figure 3 summarises the key procedural similarities and differences between the Phase 1 and Phase 2 interviews.

The selection of trajectories, the division of each trajectory into snapshots, and the construction of questions for the interviews were all completed with the input of a senior neonatologist (author H.L.).

#### Measures and data analysis method

The measures used were one or more of the differential diagnoses provided by each participant, in order of likelihood. For Phase 1

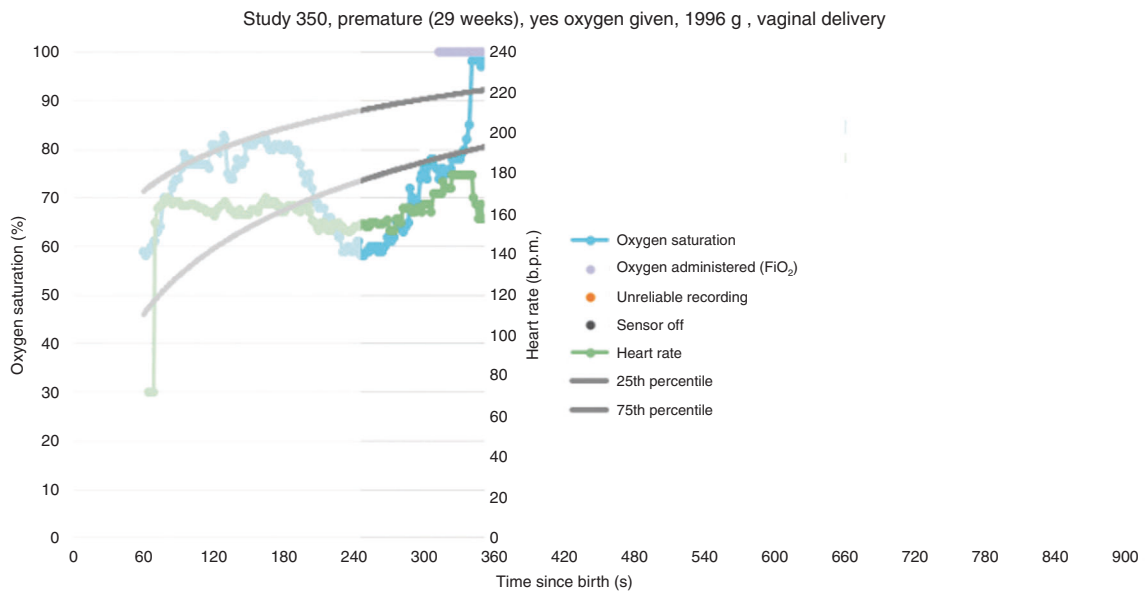
and Phase 2 separately, we analysed (1) the concordance among clinicians' differential diagnoses within and between the snapshots of a neonatal trajectory and (2) the concordance among clinicians' differential diagnoses when they reflected on the entire neonatal trajectory—the global differential diagnosis concordance. Because the study was largely qualitative, we could not apply typical quantitative measurements of agreement or concordance such as Cohen's kappa or Fleiss' kappa to the data without violating a number of their assumptions.<sup>27,28</sup> Nevertheless, as described below, we aggregated the participants' responses and interpreted them using Landis and Koch's<sup>29</sup> interpretation of the level of agreement (typically used for kappa scores).

**Phase 1:** For each snapshot of a trajectory, the number of participants who hypothesised that a specific underlying condition(s) could explain the pattern in vital signs was counted, then divided by the total number of participants and expressed as a percentage. The degree of concordance was divided into four categories based on Landis and Koch's<sup>29</sup> interpretations of the reliability of agreement:

1. Strong: >81% of participants hypothesised the same differential diagnosis for a given snapshot of the trajectory.
2. Substantial: 61–80% of participants hypothesised the same differential diagnosis for a given snapshot of the trajectory.
3. Moderate: 41–60% of participants hypothesised the same differential diagnosis for a given snapshot of the trajectory.
4. Weak: <40% of the participants hypothesised the same differential diagnosis for a given snapshot of the trajectory.

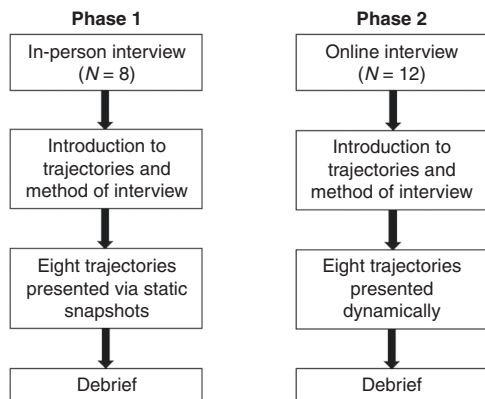
To explore the degree of agreement once the participants had seen the entire trajectory (global differential diagnosis concordance), the above categorisation scheme was applied.

**Phase 2:** Given that each participant could 'pause' the progression of a trajectory wherever they deemed appropriate, the researcher visually identified clusters of pause points for each neonatal trajectory (see Fig. 4) where >80% of participants identified and commented on the pattern. Once the clusters were determined, the researcher counted the number of participants who hypothesised each potential differential diagnosis for that particular cluster, then divided by the total number of participants who responded and expressed the final value as a percentage. The concordance of responses was mapped onto one of the four categories for the strength of concordance for each of the clusters and for the global differential diagnosis concordance.



Time: 351

**Fig. 2** An example of a neonatal trajectory in Phase 2 being gradually revealed by the browser-based application (10x real speed). The shaded portion of the trajectory depicts the participant's first 'pause', where the researcher would administer the interview questions. The unshaded portion of the trajectory depicts the participant continuing to reveal the trajectory, anticipating a second 'pause'.



**Fig. 3** A high-level comparison of the procedural similarities and differences between Phase 1 and Phase 2 of the structured interview study. In Phase 1 the participants completed the interview in person, whereas in Phase 2 the participants completed the interview online. Furthermore, in Phase 1 the trajectories were presented to the participants as static snapshots, whereas in Phase 2 the trajectories were presented to the participants dynamically (in real time).

**RESULTS**

In Phase 1 and Phase 2, the concordance within and between the snapshots or clusters of each neonatal trajectory was variable. For example, Fig. 5 depicts a neonatal trajectory that produced moderate to substantial degrees of concordance in Phase 1 and Phase 2. Although participants in Phase 2 identified fewer meaningful patterns than had been predetermined by the researchers in Phase 1, there was still some degree of overlap in the pattern of vital signs that were discussed. As Fig. 5 shows, in Phase 1 participants consistently hypothesised that respiratory distress syndrome could be driving the pattern of the vitals. In contrast, participants in Phase 2 were more likely to identify

symptoms of a definitive diagnosis, including not breathing and apnoea (which both relate to respiratory issues in the newborn). In both cases, levels of concordance were in the moderate to substantial ranges.

Figure 6 depicts a neonatal trajectory that produced a substantial to strong degree of concordance within each trajectory snapshot or cluster in Phase 1 and Phase 2 (further examples are in the Supplemental Materials). The degree of overlap in vital sign patterns identified and discussed in both interview phases was strong. Furthermore, in Phase 1 participants hypothesised that respiratory distress syndrome could be driving the vital sign patterns with a strong degree of agreement within and between each snapshot. However, although the agreement was strong within each snapshot in Phase 2, participants were inclined to attribute the vital sign patterns initially to extraneous influences, before agreeing strongly that the newborn was physiologically stressed (a consequence of respiratory distress).

Thus, although trajectories in Phase 1 and Phase 2 produced similar degrees of concordance within snapshots of a trajectory, there were differences between Phase 1 and Phase 2. Table 1 summarises the variability in the strength of concordance across the snapshots of each trajectory in Phase 1 and Phase 2.

In contrast to the above, the global differential diagnosis concordance was stronger and more consistent across all trajectories in Phase 1 and Phase 2 (refer to Table 2). For example, in Fig. 5 there was substantial agreement that respiratory distress syndrome was the underlying physiological condition driving the pattern in the neonate's vital signs in both Phase 1 and Phase 2. Similarly, in Fig. 6 all participants in Phase 1 and Phase 2 agreed that respiratory distress syndrome and intervention-related influences were the most likely diagnoses for the neonatal trajectory.

In summary, the global differential diagnosis data for each of the eight newborn trajectories produced a similar pattern of results for Phase 1 and Phase 2. Four trajectories produced a substantial degree of concordance among clinicians and four trajectories produced a strong degree of concordance among



**Fig. 4 An example of visual clustering for a neonatal trajectory in Phase 2.** Blue horizontal lines beneath the neonatal trajectory represent the ‘snapshot’ or meaningful pattern that participants were referring to when making diagnostic interpretations. Blue and yellow vertical lines beneath the neonatal trajectory represent the timestamp for the start of the ‘snapshot’ or meaningful pattern that participants were referring to when making interpretations. Clusters were identified via the following criteria: (1) ten or more participants overlapped on a start timestamp for a meaningful pattern (density of the vertical lines) and (2) ten or more participants overlapped on the snapshot or meaningful pattern driving their interpretation of the vital signs (length of the horizontal line).

clinicians (see Table 2). Furthermore, of the seven trajectories used in both Phase 1 and Phase 2, two trajectories produced a substantial degree of concordance among clinicians in both Phase 1 and 2. Similarly, two trajectories produced a strong degree of concordance in both Phase 1 and 2 (see Table 2).

**DISCUSSION**

The primary purpose of this study was to investigate the degree of concordance among expert neonatal clinicians’ interpretations of the vital sign patterns of neonates within the first 10 min after birth. If sufficient concordance is observed, then the interpretations could act as a substitute for the currently unavailable direct measurement of crucial physiological and anatomical parameters in the neonatal transition. We found that expert clinicians’ judgements of patterns in newborn vital signs appear mostly to agree. Furthermore, when clinicians consider the entire newborn transition, they produce a high-level interpretation of the potential physiological status of the infant with an even stronger degree of concordance. The implication is that differential diagnoses with a good degree of concordance could potentially substitute in part for the direct measurement of key physiological and anatomical variables, which is currently unavailable, and they could guide the design of a cognitive aid as well as supporting training.

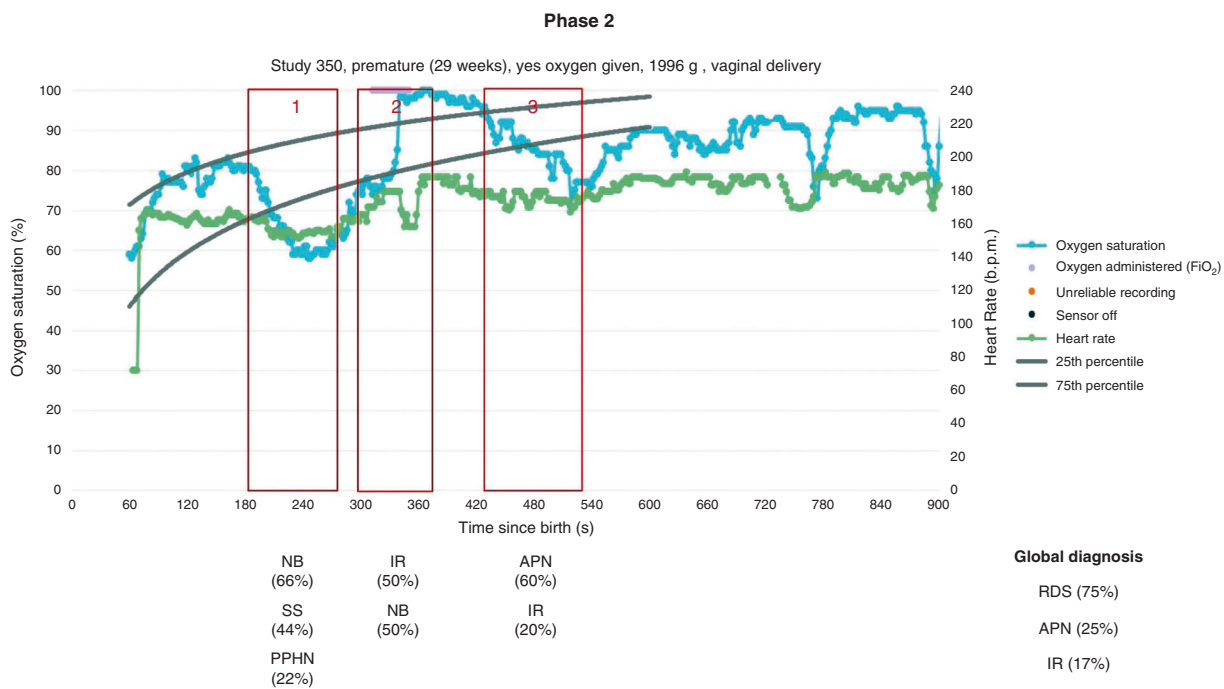
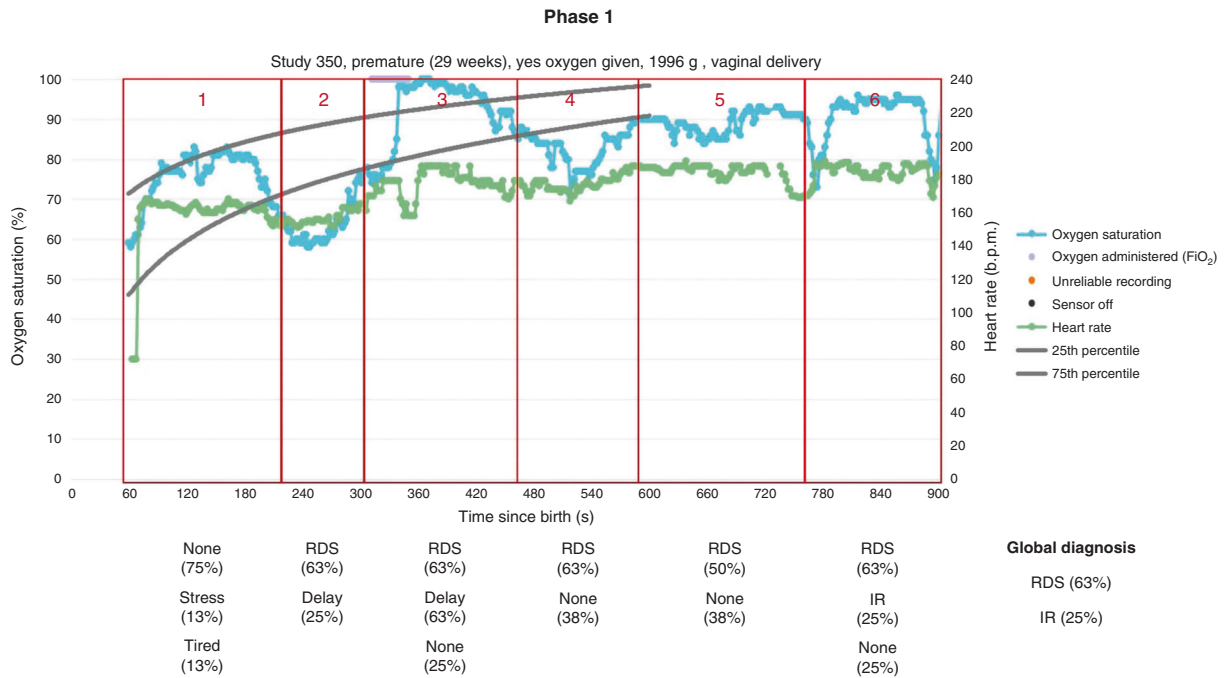
These results provide support for theories of how expertise manifests in such complex and dynamic environments. Expert diagnosis depends on a person’s ability to perceive meaningful cues in the environment, such as a series of clinical signs co-occurring, that subsequently trigger associations in memory. The cues lead the person to certain expectancies concerning the behaviour of the system and to specific actions.<sup>30,31</sup> Similarly, the meaningful patterns identified by the clinicians in our study could be driven by the repertoire of meaningful cues in memory,

drawing clinicians’ attention to features of the newborn’s vital signs that are diagnostic of a particular underlying physiological condition. Consequently, clinicians can rapidly assess and appropriately respond to events in the transition.<sup>30</sup>

Both the literature and the present findings suggest that interpretive concordance among clinicians is an appropriate foundation for the design of a potential cognitive aid. A cognitive aid would guide the end-user towards expert clinicians’ interpretations (i.e. differential diagnoses) of patterns in newborn vital signs that signal important aspects of the physiological transition. A more accurate understanding by clinicians of the physiological status of the newborn at any point in the first 10 min after birth may help them make the most appropriate interventions—especially if the newborn’s physiological state differs substantially from the state that the resuscitation algorithms assume is ‘most likely’.

There is evidence from contexts such as emergency department resuscitation that a cognitive aid can improve clinicians’ speed and fluidity of diagnosis and treatment in resuscitation.<sup>17,19,21,23</sup> A cognitive aid may also reduce the risk of errors of omission because it encourages the resuscitation team to consider and discuss differential diagnoses that they may have overlooked.<sup>17,18</sup> Nevertheless, more research is needed (a) to determine the feasibility of relying on experts’ knowledge and opinion as a substitute for the direct measurement of crucial but unavailable physiological and anatomical parameters in the neonatal transition, (b) to create the physical form of the cognitive aid so that the presentation of information is most effective at the time and in the location that it is likely to be used and (c) to determine the impact of such a cognitive aid on the management of the infant in the first 10 min after birth.

Much healthcare research shows that machine learning algorithms can extract key features in physiological data to assist with diagnosis (see ref. <sup>32</sup> for a review). However, no such algorithm

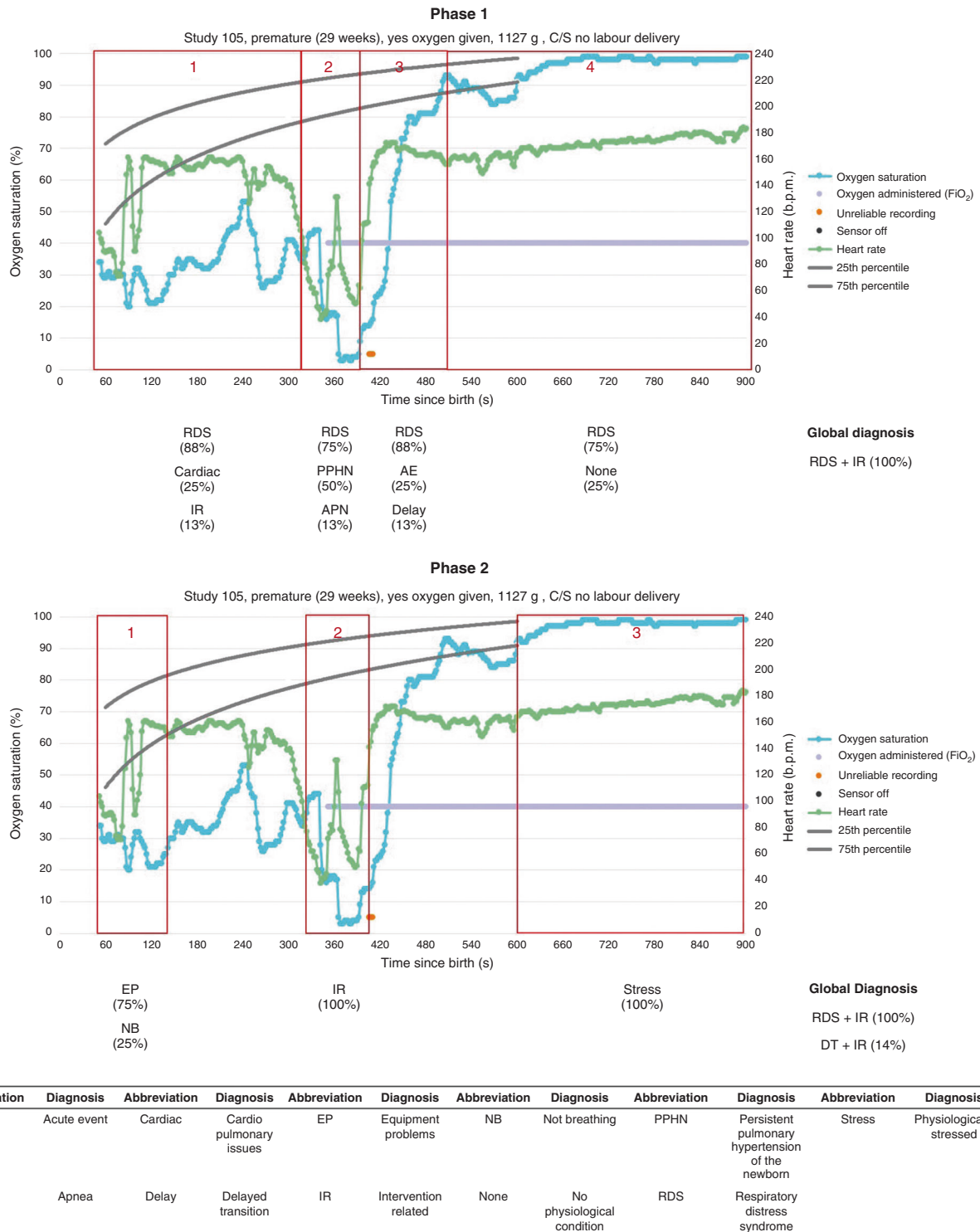


Abbreviation	Diagnosis	Abbreviation	Diagnosis	Abbreviation	Diagnosis	Abbreviation	Diagnosis	Abbreviation	Diagnosis
APN	Apnea	IR	Intervention related	None	No physiological condition	RDS	Respiratory distress syndrome	Stress	Physiologically stressed
Delay	Delayed transition	NB	Not breathing	PPHN	Persistent Pulmonary Hypertension of the newborn	SS	Stiff lungs and secretions	Tired	Neonate tiring

**Fig. 5 The strength of concordance for neonatal trajectory #2.** The trajectory produced a moderate-substantial degree of concordance among clinicians within/between each of the snapshots in the trajectory.

exists for the context of newborn resuscitation. Therefore, further research is needed to determine the practicality and feasibility of a newborn resuscitation cognitive aid supported by a real-time decision-support algorithm. First, the cognitive aid would need to be underpinned by an algorithm that can recognise—in real time—

significant patterns anywhere in the first 10 min after birth and find an interpretation of the pattern. Second, the algorithm would need to take into account any interventions already administered to the newborn during the resuscitation and adjust the identification and interpretation of patterns appropriately.



**Fig. 6 The strength of concordance for neonatal trajectory #4.** The trajectory produced a substantial-strong degree of concordance among clinicians within/between each of the snapshots in the trajectory.

This study has several limitations, despite the promising initial findings. Most of the limitations stem from the constraints related to (a) interviewing the specific expert population we were seeking, and (b) using a relatively constrained database of neonatal vital sign patterns in the first 10–15 min after birth.

First, we recruited a relatively small sample of experts in both phases of the study. Although the sample size is typical of recent

human factors contributions to medical device design for neonatal resuscitation,<sup>33</sup> a larger sample size may have improved the reliability of the findings. Nevertheless, the sample size was suitable for our goal of exploring concordance among experts' responses.

Second, given the constraints on participants' time, we could present only eight neonatal trajectories in each phase of the study—one trajectory for each kind of common neonatal transition pattern.

**Table 1.** Summary of the strength of concordance across the snapshots of each trajectory in Phase 1 and Phase 2 of the interview study.

Trajectory	Phase 1					Phase 2				
	Number of snapshots	Weak (<40%)	Moderate (41–60%)	Substantial (61–80%)	Strong (81–100%)	Number of Snapshots	Weak (<40%)	Moderate (41–60%)	Substantial (61–80%)	Strong (81–100%)
1	6	2	2	1	1	4	0	1	1	2
2	6	0	1	5	0	3	0	2	1	0
3	5	0	1	3	1	3	0	0	3	0
4	4	0	0	2	2	3	0	0	1	2
5	3	0	0	1	2	2	0	0	2	0
6	3	0	0	3	0	3	0	1	0	2
7	3	0	2	1	0	4	0	1	2	1
8	2	0	0	1	1	4	3	0	0	1

**Table 2.** Global differential diagnosis concordance among clinicians for each neonatal trajectory in Phase 1, Phase 2 and overall.

Global differential diagnosis	Concordance		
	Phase 1	Phase 2	Overall
<b>Trajectory #1</b>			
Persistent pulmonary hypertension of the newborn	86% <sup>a</sup>	71% <sup>b</sup>	73% <sup>b</sup>
Respiratory distress syndrome	38%	0%	20%
Delayed transition	0%	29%	13%
<b>Trajectory #2</b>			
Respiratory distress syndrome	63% <sup>b</sup>	75% <sup>b</sup>	69% <sup>b</sup>
Intervention related	25%	17%	25%
Apnoea throughout	0%	25%	13%
<b>Trajectory #3</b>			
Delayed transition + intervention related	63% <sup>b</sup>	83% <sup>a</sup>	71% <sup>b</sup>
Respiratory distress syndrome	50%	33%	43%
<b>Trajectory #4</b>			
Respiratory distress + intervention related	100% <sup>a</sup>	100% <sup>a</sup>	100% <sup>a</sup>
Delayed transition + intervention related	0%	14%	6%
<b>Trajectory #5</b>			
Respiratory distress syndrome	100% <sup>a</sup>	88% <sup>a</sup>	94% <sup>a</sup>
Persistent pulmonary hypertension of the newborn	13%	13%	13%
<b>Trajectory #6</b>			
Respiratory distress syndrome	75% <sup>b</sup>	70% <sup>b</sup>	72% <sup>b</sup>
Persistent pulmonary hypertension of the newborn	13%	20%	17%
<b>Trajectory #7</b>			
Distressed + delayed transition	75% <sup>b</sup>	100% <sup>a</sup>	79% <sup>b</sup>
Respiratory distress + pulmonary hypertension	25%	0%	14%

Note: Only the seven trajectories presented in both Phase 1 and 2 are presented for comparison.

<sup>a</sup>Strong global differential diagnosis concordance among clinicians for the specific phase.

<sup>b</sup>Substantial global differential diagnosis concordance among clinicians for the specific phase.

Consequently, it was not possible to determine whether clinicians' agreement would be consistent across many patterns of the same apparent kind. In Phase 1, the trajectories were not presented to the participant in real time, which may have compromised their ability to recognise meaningful patterns as they unfolded and interpret them. In addition, the trajectories were selected from a database that was collected in 2005,<sup>26</sup> so their clinical courses may not be typical of infants resuscitated using contemporary guidelines. In future research, we will seek more recent neonatal resuscitation databases and present more than one example of commonly encountered trajectories closer to real time.

Third, the neonatal trajectories displayed very limited vital sign information—namely, heart rate, oxygen saturation, maximum percentage of oxygen administered, and pulse oximetry signal quality. At the time these recordings were collected, devices such as electrocardiogram monitors and respiratory function monitors were not in routine use. Apart from gestation, birth weight and mode of delivery, the database did not provide information about maternal and fetal history, other intervention administration, and clinical signs of the newborn. The lack of such information may have contributed to the variability in concordance observed within and between the snapshots of each neonatal trajectory. In future research, we hope to access more recent neonatal databases that capture all this information—and perhaps in video form as well as numerical. Nonetheless, the lack of information was partly an advantage. Our participants were not at risk of bias relating to presumptions from maternal or fetal history, but instead were making judgements based solely on objective measures available to them.

Fourth, we were unable to consider the baseline rate for each differential diagnosis. Participants were making judgements and inferences independently of one another, and like the investigators, were not aware of the true physiological state of the neonate. Even so, because some differential diagnoses are more likely than others, the degree of concordance that we found could be inflated. Next steps should be to redesign the interviews to satisfy the assumptions of more robust measures of inter-rater reliability, such as Fleiss' kappa.<sup>28</sup>

## CONCLUSIONS

We conducted an analysis of the concordance of expert neonatal clinicians' interpretation of the patterns of vital signs exhibited by neonates within the first 10 min after birth. Our results showed a strong degree of agreement among neonatal experts. This is a small but important step towards using a majority or consensus of neonatal experts' judgements in the design of a cognitive aid. To further validate this method, future



research should seek more recent neonatal databases, present several examples of commonly encountered trajectories in close to real time, and utilise more formal measures of agreement. Nevertheless, this method could be used in other areas where clinicians do not have access to the true physiological state of the patient. Ultimately, the findings will help us design a cognitive aid to support clinicians' management of the newborns who require resuscitation after birth—possibly enhancing the health of this vulnerable population.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of Jennifer Dawson for providing us with access to the neonatal database, and Isaac Salisbury for assisting in the extraction of the neonatal trajectories from the database. Support for this research was provided by J.Z.'s research higher degree (HDR) support funds at The University of Queensland and her Australian Government Research Training Programme (RTP) Scholarship. The project was also supported by The University of Queensland School of Psychology Strategic funds allocation to P.S.

## AUTHOR CONTRIBUTIONS

Substantial contributions to conception and design, acquisition of data or analysis and interpretation of data: J.Z., H.L. and P.S. Drafting the article or revising it critically for important intellectual content: J.Z., H.L. and P.S. Final approval of the version to be published: J.Z., H.L. and P.S.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41390-021-01565-9>.

**Competing interests:** The authors declare no competing interests.

**Consent statement:** Patient consent was not required for this study. Clinician participants provided written informed consent. The study was given ethical approval by Mater Misericordiae Ltd (approval 53861) and by the Human Research Ethics Committee at The University of Queensland (approval 2019002697).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

- Hillman, N. H., Kallapur, S. G. & Jobe, A. H. Physiology of transition from intrauterine to extrauterine life. *Clin. Perinatol.* **39**, 769–783 (2012).
- Hooper, S., Kitchen, J. M., Polglase, R. G., Roehr, C. C. & Te Pas, B. A. The physiology of neonatal resuscitation. *Curr. Opin. Pediatr.* **30**, 187–191 (2018).
- Askin, D. F. Complications in the transition from fetal to neonatal life. *J. Obstet. Gynecol. Neonatal Nurs.* **31**, 318–327 (2002).
- Schmölzer, G. M., Morley, C. J. & Kamlin, O. C. O. F. Enhanced monitoring during neonatal resuscitation. *Semin. Perinatol.* **43**, 151177 (2019).
- Zestic, J., Sanderson, P., Dawson, J. & Liley, H. Defining information needs in neonatal resuscitation with work domain analysis. *J. Clin. Monit. Comput.* **13**, 1–22 (2020).
- Crossley, K. J. et al. Dynamic changes in the direction of blood flow through the ductus arteriosus at birth. *J. Physiol.* **587**, 4695–4704 (2009).
- Hooper, S., Polglase, G. R. & Roehr, C. C. Cardiopulmonary changes with aeration of the newborn lung. *Paediatr. Respir. Rev.* **16**, 147–150 (2015).
- Polglase, G. R. et al. Physiologically based cord clamping stabilises cardiac output and reduces cerebrovascular injury in asphyxiated near-term lambs. *Arch. Dis. Child. Fetal Neonatal Ed.* **103**, F530 (2018).
- Blank, D. A. et al. Respiratory changes in term infants immediately after birth. *Resuscitation* **130**, 105–110 (2018).
- Jeroen, Jv. V. et al. Noninvasive measurements of hemodynamic transition directly after birth. *Pediatr. Res.* **75**, 448–452 (2013).
- van Vonderer, J., Roest, A., Klumper, F., Hooper, S. & Pas, A. The effect of breathing on ductus arteriosus blood flow directly after birth. *Eur. J. Pediatr.* **176**, 1581–1585 (2017).
- Hou, M., Burns, C. M. & Banbury, S. Advances and challenges in intelligent adaptive interface design. In *Contemporary Issues in Systems Science and Engineering* (Zhou, M., Li, H.-X. & Weijnen, M., eds) 367–424 (Wiley, New Jersey, 2015).
- Vicente, K. J. Coherence-and correspondence-driven work domains: Implications for systems design. *Behav. Inform. Technol.* **9**, 493–502 (1990).
- Baig, M. M., Gholam Hosseini, H., Kouzani, A. & Harrison, M. J. Anaesthesia monitoring using fuzzy logic. *J. Clin. Monit. Comp.* **25**, 339–347 (2011).
- Bolon-Canedo, V. et al. Dealing with inter-expert variability in retinopathy of prematurity: a machine learning approach. *Comput. Meth. Prog. Biomed.* **122**, 1–15 (2015).
- Miller, A., Scheinkestel, C. & Steele, C. The effects of clinical information presentation on physicians' and nurses' decision-making in ICUs. *Appl. Ergon.* **40**, 753–761 (2009).
- Grundgeiger, T. et al. Cognitive aids in acute care: investigating how cognitive aids affect and support in-hospital emergency teams. In *Proc. 2019 CHI Conference on Human Factors in Computing Systems* 1–14 (Association for Computing Machinery, 2019).
- Marshall, S. The use of cognitive aids during emergencies in anesthesia: a review of the literature. *Anesth. Analg.* **117**, 1162–1171 (2013).
- Calder, L. A. et al. Healthcare providers' perceptions of a situational awareness display for emergency department resuscitation: a simulation qualitative study. *Int. J. Qual. Health Care* **30**, 16–22 (2018).
- Fuerch, J. H., Yamada, N. K., Coelho, P. R., Lee, H. C. & Halamek, L. P. Impact of a novel decision support tool on adherence to Neonatal Resuscitation Program algorithm. *Resuscitation* **88**, 52–56 (2015).
- Gonzales, M., Henry, J., Calhoun, A. & Riek, L. Visual TASK: a collaborative cognitive aid for acute care resuscitation. In *10th EAI International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health)* 1–8 (Cornell University Library, 2016).
- Roitsch, C., Patricia, K., Hagan, J., Arnold, J. & Sundgren, N. Tablet-based decision support tool improves performance of neonatal resuscitation: a randomized trial in simulation. *Simul. Health* **15**, 243–250 (2020).
- Schild, S. et al. A digital cognitive aid for anesthesia to support intraoperative crisis management: results of the user-centered design process. *JMIR mHealth uHealth.* **7**, e13226 (2019).
- Klüber, S. et al. Experience matters: design and evaluation of an anesthesia support tool guided by user experience theory. In *Proc. 2020 ACM Designing Interactive Systems Conference* 1523–1535 (Association for Computing Machinery, 2020).
- Nielsen, J. Usability inspection methods. In *Computer Human Interaction* 413–414 (Computer Human Interaction Conference Companion, 1994).
- Dawson, J. A. et al. Defining the reference range for oxygen saturation for infants after birth. *Pediatrics* **125**, e1340–e1347 (2010).
- Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
- Fleiss, J. L. Measuring nominal scale agreement among many raters. *Psychol. Bull.* **76**, 378–382 (1971).
- Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
- Ericsson, K. A. & Lehmann, A. C. Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annu. Rev. Psychol.* **47**, 273–305 (1996).
- Klein, G. In *Advances in Man-Machine Systems* (ed. Rouse, W. B.) 47–92 (JAI Press, 1989).
- Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S. & Acharya, U. R. Deep learning for healthcare applications based on physiological signals: a review. *Comput. Meth. Prog. Biol.* **161**, 1–13 (2018).
- Pickup, L. et al. Development of a clinical interface for a novel newborn resuscitation device: human factors approach to understanding cognitive user requirements. *J. Med. Internet Res.* **6**, e12055 (2019).