Experimental paper

Similarity of expert clinicians' rank order of differential diagnoses in a newborn resuscitation context

Jelena Zestic [a,*], Helen G. Liley [a,b], Penelope M. Sanderson [a]

Abstract

Background: We tested principles that could lead to a future cognitive aid that offers an interpretation of the newborn's physiological state during resuscitation after birth. Using concordance among experts' interpretations of newborn vital sign patterns as an approximation for an algorithm that could provide an interpretation of the newborn's state, we explored the reliability and generalisability of experts' interpretations.

Methods: Twelve neonatal experts viewed eight pairs of graphical trajectories showing newborns' heart rate and oxygen saturation records supplemented with differential diagnoses elicited previously from other experts. Each pair of trajectories included one trajectory on which the original differential diagnoses had been based, and a similar but novel trajectory to which the original differential diagnoses were now generalised. For each trajectory, experts ranked the differential diagnoses according to their likelihood. We calculated how similar the new experts' ranking was to the original experts' ranking for both original and novel trajectories. We used descriptive categories to interpret the strength of the similarity.

Results: For the original and novel trajectories, the experts' rank ordering of differential diagnoses was mostly moderately to substantially similar to the original rank ordering by the original participants. There were mostly small differences in similarity scores between the paired original and novel trajectories; fewer than 25% of the participants suggested an alternative differential diagnosis.

Conclusions: The concordance of experts' interpretations could serve as an approximation of the newborn's physiological state, and the interpretations could be generalised. The results may justify pursuing an algorithm to underpin a cognitive aid.

Keywords: Newborn, Resuscitation, Concordance, Experts, Cognitive aid, Decision support

Introduction

The neonatal transition from intrauterine life to extrauterine life depends on a series of interrelated physiological processes that must take place to change how every organ system functions in the body; see Zestic et al.[1] for a description of some key early components of this transition. Supporting the newborn transition is challenging for clinicians, largely because devices to measure many of the critical physiological variables are not routinely available, either because the relevant sensors do not exist or because they are impractical for routine clinical use.[1–3] Consequently, the clinician's grasp of a neonate's physiological status may be imperfect. This exacerbates the difficulties of clinical decision-making, particularly where the infant presents with an uncommon underlying physiological condition or an unexpectedly poor response to resuscitation intervention(s).[4,5]

Tools to support clinical task performance and decision-making, including neonatal resuscitation algorithms, use signs such as tone, breathing effort, heart rate and oxygen saturation (SpO_2) as indicators of the newborn's physiological progress and the effectiveness of resuscitation intervention(s). However, current tools in clinical use do not offer differential interpretations of the physiological state of the patient or patient-specific guidance based on that interpretation.[6–8] A more versatile and clinically useful tool could be a cognitive aid capable of providing a plausible interpretation of the newborn's current state from available vital signs. However, this would require the aid to be based on an artificial intelligence (AI) algorithm that can recognise clinically significant patterns of vital signs as resuscitation unfolds and offer an interpretation of those patterns.

The training and optimisation of an AI algorithm—particularly a machine learning algorithm—that offers an interpretation of the infant's physiological state would require a large neonatal vital sign database. The AI system could learn clinically meaningful patterns on one section of the data, and then test its performance on the section of the data not used for learning.[9] However, the above approach is resource-intensive and suitable databases do not currently exist. We therefore conjectured that an approximation of the newborn's physiological state could be provided from neonatal experts' diagnostic interpretations of patterns in newborn vital signs after birth. The experts' interpretations could substitute for an AI algorithm that offers an interpretation of the newborn's state.

We previously explored whether there is good concordance (agreement) across neonatal experts in how they interpret patterns in newborn vital signs after birth (Phase 1 and Phase 2 in Fig. 1).[10] If experts agree, then their interpretations could be taken as an approximation to the newborn's physiological state. In Phases 1 and 2, neonatal experts observed graphical trajectories showing newborns' heart rate and SpO_2 records after birth. We elicited the experts' interpretations of the range of physiological explanations that could apply to the vital sign patterns. Their interpretations were strongly concordant, suggesting that the interpretations could underpin an algorithm for a future cognitive aid that displays experts' interpretations for other clinicians to consider during newborn resuscitation, pending maturation of AI-based solutions or sensing technologies. However, it is important first to establish whether experts agree on the clinical meaning of the vital sign patterns, and whether their interpretations generalise to similar vital sign patterns.

Therefore, in this study, Phase 3 in Fig. 1, we performed a *conceptual* test of the feasibility of a real-time decision-support algorithm to help determine whether it is worthwhile pursuing the resource-intensive AI methods that would be needed to develop a cognitive aid that provides a plausible interpretation of the newborn's current state during resuscitation after birth. In this study, the experts' interpretations are a proxy for the AI algorithm, and the AI is a proxy for the currently unavailable sensors. The first purpose of the study was to test the concordance of a new group of experts' physiological explanations with those of the original experts. Specifically, we compared how new participants rank order differential diagnoses for specific heart rate and SpO_2 patterns with how the original experts in Phases 1 and 2 rank ordered their differential diagnoses. The second purpose was to test the concordance of new experts' physiological explanations with those of the original experts, now on novel but similar heart rate and SpO_2 patterns. Specifically, we tested whether the rank order of differential diagnoses for specific heart rate and SpO_2 patterns from Phases 1 and 2 would generalise to similar but novel vital sign patterns. If participants' rank ordering is concordant, then experts' interpretations could be reliable and generalisable, opening the way to developing a cognitive aid.

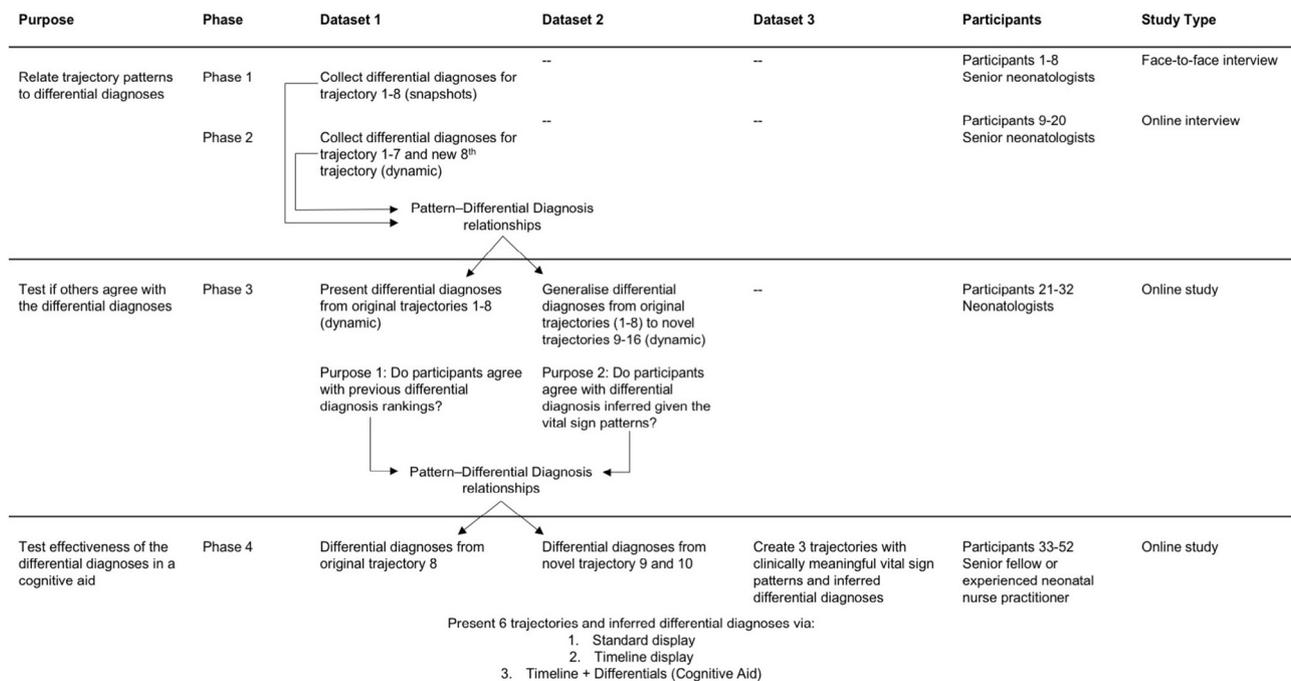| Purpose	Phase	Dataset 1	Dataset 2	Dataset 3	Participants	Study Type
Relate trajectory patterns to differential diagnoses	Phase 1	Collect differential diagnoses for trajectory 1-8 (snapshots)	--	--	Participants 1-8 Senior neonatologists	Face-to-face interview
	Phase 2	Collect differential diagnoses for trajectory 1-7 and new 8th trajectory (dynamic)	--	--	Participants 9-20 Senior neonatologists	Online interview
		→ Pattern–Differential Diagnosis relationships				
Test if others agree with the differential diagnoses	Phase 3	Present differential diagnoses from original trajectories 1-8 (dynamic)	Generalise differential diagnoses from original trajectories (1-8) to novel trajectories 9-16 (dynamic)	--	Participants 21-32 Neonatologists	Online study
		Purpose 1: Do participants agree with previous differential diagnosis rankings?	Purpose 2: Do participants agree with differential diagnosis inferred given the vital sign patterns?			
		→ Pattern–Differential Diagnosis relationships ←				
Test effectiveness of the differential diagnoses in a cognitive aid	Phase 4	Differential diagnoses from original trajectory 8	Differential diagnoses from novel trajectory 9 and 10	Create 3 trajectories with clinically meaningful vital sign patterns and inferred differential diagnoses	Participants 33-52 Senior fellow or experienced neonatal nurse practitioner	Online study
		Present 6 trajectories and inferred differential diagnoses via: 1. Standard display 2. Timeline display 3. Timeline + Differentials (Cognitive Aid)				

Fig. 1 – Overview of the research program. Phase 1 and Phase 2 involved eliciting the experts' diagnostic interpretations of vital sign patterns exhibited by newborns in the first 10–15 minutes after birth—the eight original neonatal trajectories were extracted from the Dawson et al.[11] database. The present study (Phase 3) builds on the previous phases by (1) testing if other experts agree with the Phase 1 and 2 experts' diagnostic interpretations of the original eight trajectories, and (2) testing if the Phase 1 and 2 experts' diagnostic interpretations generalise to eight novel trajectories (also extracted from the Dawson et al.[11] database). In Phase 4 we plan to explore the effectiveness of the differential diagnosis suggestions when they are presented via a prototype cognitive aid.

Methods

Context and participants

Participants were recruited by personal approach from one of the authors who is a senior neonatologist in a tertiary hospital in Queensland, Australia. Ethical approval was given 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' experience as a senior fellow, consultant, or neonatal nurse practitioner in newborn medicine, and to perform or supervise newborn resuscitation on a regular basis. All participants provided written informed consent.

Trajectory development

The neonatal trajectories were drawn from the Dawson et al.[11] database, which documents the real-time heart rate and SpO_2 values of approximately 465 neonates in the first 15 minutes after birth. No 'true' diagnosis was available for each trajectory. However, additional information was available for each trajectory—gestation, birth weight, mode of delivery, and supplemental oxygen. The 25th and 75th percentile boundaries for SpO_2 targeting were inferred by Dawson et al.[11] in 2010 from the database of neonatal trajectories.

In the present study, Phase 3 in Fig. 1, participants were presented with 16 trajectories from the Dawson et al.[11] database in the form of timeline graphics (Fig. 2). Eight of the 16 trajectories were the same (original) trajectories that had been presented to the original participants in Phase 1 and 2. The other eight trajectories were novel; each was chosen because it resembled one of the original trajectories but had not previously been presented to participants.

During the study, when each original trajectory was played and a clinically meaningful vital sign pattern appeared, the timeline graphic displayed the original participants' rank ordered differential diagnoses for that pattern (Fig. 2). For each novel trajectory, we applied the rank ordered differential diagnoses to vital sign patterns that were the closest match to patterns in the original trajectory. Thus, there were eight pairs of trajectories presented (Supplementary Data 1).

Procedure

The 'RAND' Excel function was used to randomly allocate half of the participants to first observing the original trajectories and the other half to first observing the novel trajectories, with the order of presentation of trajectories also randomised within each kind of trajectory. Due to COVID-19 pandemic restrictions, the researcher conducted the study with each individual participant online. The researcher introduced the participant to the format of the trajectories, and then demonstrated how to interact with the browser-based application that gradually presented the trajectories to the participant at 10 times real-time speed (Fig. 2).

When the trajectory was being presented, at certain points an auditory cue would sound, and the trajectory would freeze (preventing the progression of the trajectory). How often and when the trajectories froze varied between the eight pairs of trajectories. At each freeze, a list of two or three differential diagnoses would appear underneath the trajectory (Fig. 2). The participant rank ordered the
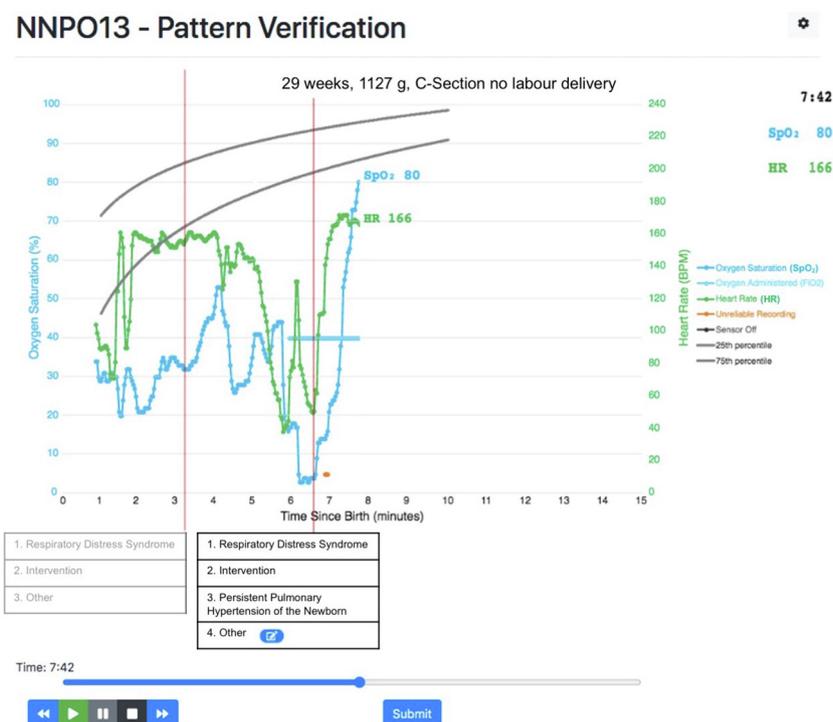

Fig. 2 – An example of a neonatal trajectory being gradually presented (10 times speed) via the browser-based application. The red lines represent the 'freeze' locations where a list of differential diagnoses would appear underneath the trajectory. The portion of the trajectory presented after the second freeze depicts the trajectory continuing to be presented to the participant at 10 times speed, and the participant anticipating a possible third freeze.

differential diagnoses from most likely to least likely. They could move a differential diagnosis to a higher or lower position in the list by using the browser's drag-and-drop function. However, they could not remove any differential diagnoses from the list. Participants could also enter an alternative differential diagnosis in the 'Other' textbox and use the drag-and-drop function to move the 'Other' option to a higher or lower position in the list.

Participants submitted their rank ordered differential diagnoses by clicking a 'Submit' button underneath the trajectory. They then clicked the 'Play' button to continue revealing the trajectory. Their rank ordered differential diagnoses remained visible but became faded when the next freeze was encountered (Fig. 2). This process was repeated at every freeze until the entire trajectory had been revealed.

The researcher (who was not a clinician) did not suggest any rank order or comment on the participant's rank order. Participants were unaware that other expert clinicians had provided the original rank order of differential diagnoses.

Measures and data analysis method

The data collected were the rank order of the differential diagnoses provided by each participant at each freeze. We analysed (1) the similarity between each participant's rank order of diagnoses in Phase 3 and the original rank order from Phases 1 and 2 of the differential diagnoses for each freeze in a trajectory, (2) the difference in the averaged similarity score at each freeze between the paired original and novel trajectories in Phase 3, and (3) the number of participants who provided a response to the 'Other' option, including what their response was, and where their 'Other' response was ranked in the list of differential diagnoses for each freeze in a trajectory.

The similarity between any two rank ordered lists (or 'strings') of items such as differential diagnoses can be measured using string similarity metrics that calculate the number of edit steps required to transform one string into another.[12] However, given that we preferred to measure whether each item in a participant's string was

in the same position as the original string, we created our own string similarity metric and interpreted the similarity score using categories representing the closeness of match (see Supplementary Data 2 for a description of the a priori and post hoc similarity score calculations).

For strings with four items, we identified five categories of possible similarity scores.

1. Strong (identical) – 0–0.999
2. Substantial – 1–1.999
3. Moderate – 2–3.999
4. Weak – 4–5.999
5. Very Weak (completely dissimilar) – 6–8

For strings with three items, we identified three categories of possible similarity scores.

1. Strong (identical) – 0–0.999.
2. Moderate – 1–1.999.
3. Weak (completely dissimilar) – 2–4.

The difference in the averaged similarity score at each freeze between the paired original and novel trajectories was interpreted using the categories below.

1. Small difference – 0–0.999
2. Moderate difference – 1–1.999
3. Large difference – >2

Results

Twelve experienced neonatal care providers were recruited from six tertiary institutions in Australia and New Zealand. The geographic distribution and number of participants was as follows: Queensland (2 hospitals, $n = 7$); New South Wales ($n = 1$); Victoria (2 hospitals,
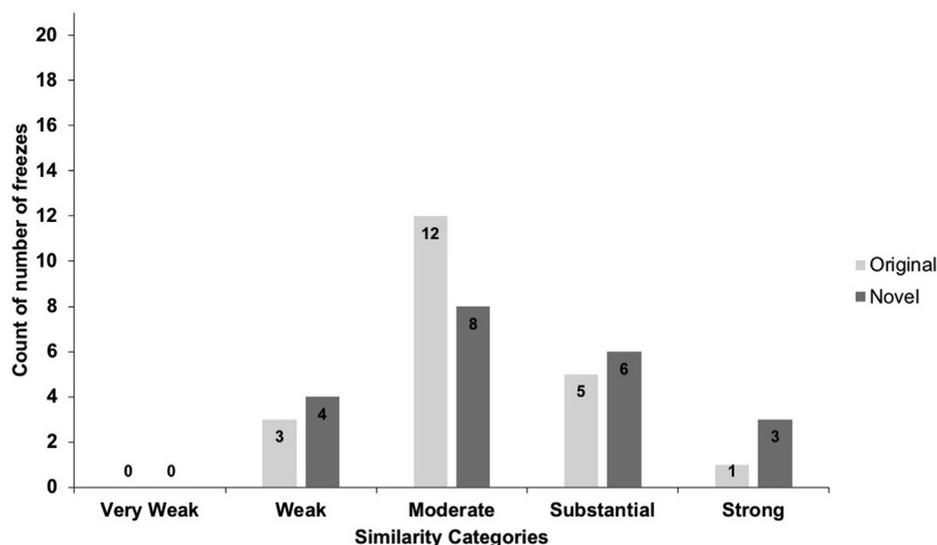

Fig. 3 – Count of the total number of freezes in each similarity score category for the original trajectories and the novel trajectories in Phase 3. For both the original and novel trajectories, participants' rank ordering of differential diagnoses was mostly moderately to substantially similar to the original rank ordering in Phase 1 and 2.

Table 1 – Summary of the count and the percentage of freezes categorised as having a small, moderate, or large difference in the averaged similarity score between the paired original and novel trajectories in Phase 3.

Pair	Number of Freezes	Small Difference < 0.999	Moderate Difference 1–1.999	Large Difference > 2
1	3	3	0	0
2	3	1	2	0
3	2	2	0	0
4	3	2	1	0
5	1	0	1	0
6	2	1	1	0
7	4	3	1	0
8	3	2	1	0
Count Total	21	14	7	0
Percentage Total	n/a	67%	33%	0%

n = 3); and New Zealand (n = 1). In Phase 1 and 2 of the research project, sample sizes of 12 participants were deemed sufficient to characterise concordance.

Across all the 42 freezes tested, 36% of the participants' rank orderings of the differential diagnoses were categorised as strongly or substantially similar to the original rank ordering of the differential diagnoses in Phase 1 and 2 (Fig. 3). Moreover, 48% of the participants' rank orderings of the differential diagnoses were categorised as moderately similar to the original rank ordering of the differential diagnoses in Phase 1 and 2. Importantly, only 17% of the freezes were categorised as having weak similarity, and there were no freezes in the very weak similarity category.

A similar pattern of results was found within the Phase 3 original and novel trajectories, taken separately (Fig. 3). For the original trajectories, the rank ordering of differential diagnoses showed strong similarity to the original rank ordering for 5% of the freezes, substantial similarity for 24% of the freezes and moderate similarity for 57% of the freezes. For the novel trajectories, the rank ordering showed strong similarity to the original rank ordering for 14% of the freezes, substantial similarity for 29% of the freezes and moderate similarity for 38% of the freezes.

Reflecting the above, 67% of the freezes were categorised as having a small difference in the averaged similarity score between the paired original and novel trajectories in Phase 3 (Table 1). The remaining 33% of freezes produced a moderate difference between the paired original and novel trajectories; there were no freezes that produced a large difference.

The number of participants who provided an 'Other' response for each freeze in a trajectory varied; however, the count of participants doing so was similar between the two trajectories in each pair (Table 2). For 67% of the freezes, 0–33% of the participants included an 'Other' response. When participants did provide an 'Other' response, usually one to three further differential diagnoses were given. Moreover, if an 'Other' response was provided for both the original and novel trajectories in a pair, the 'Other' differential diagnoses often overlapped (Table 2). We also found variability in the rank that participants assigned to their 'Other' responses; 'Other' was not always ranked in first position (Table 2 and Table 3).

Discussion

The purpose of this study was to determine whether experienced neonatal clinicians assigned differential diagnoses to specific vital

sign patterns that were similar to the differential diagnoses of other experts, and to determine whether those differential diagnoses could generalise to similar but novel vital sign patterns. Using similarity scores, we found evidence of moderate to strong concordance between the clinicians' rank ordering of differential diagnoses and the original rank order from Phases 1 and 2 of the research project for both original and novel trajectories. We also found mostly small differences in the averaged similarity score at each freeze between the paired original and novel trajectories; one quarter of the participants (or fewer) provided an additional diagnosis.

Implications

The findings provide further evidence that expert clinicians' interpretations of vital sign patterns are sufficiently concordant that they are likely to reflect the true physiological state of the newborn. Importantly, the original expert group in Phases 1 and 2 predominantly included senior consultant clinicians who were recognised experts in the physiological and anatomical processes of the neonatal transition, developing newborn resuscitation guidelines, or teaching newborn resuscitation. In contrast, the recruitment of the expert group in the current study, Phase 3, focused on clinicians who had both recent and frequent hands-on involvement in leading teams and performing newborn resuscitation. Thus, the concordance between the two groups also suggests that the interpretations of those who are expert and focused on the physiology of resuscitation resonate well with those who are performing it frequently, and vice versa. The findings also suggest that expert clinicians' interpretations could generalise to other similar vital sign patterns, opening the way to developing a cognitive aid. An aid that supports clinicians' understanding of the possible physiological status of the newborn during the post-natal transition may guide the clinician to the most appropriate intervention. This is especially critical if the newborn presents with transition complications and/or congenital malformations that resuscitation algorithms do not explicitly consider.

Much healthcare research demonstrates that machine learning algorithms can be trained and optimised—with or without human input—to detect and classify clinically significant relationships between data from numerous physiological parameters.[13–18] Thus, algorithms may also have the potential to guide newborn resuscitation practice in the not-to-distant future. However, until more sensors become available in the delivery room to either (a) directly provide the newborn's physiological state or (b) provide a greater volume and variety of physiological and anatomical data from which algorithms can independently detect and interpret clinically significant

Table 2 – Tabulation of 'Other' options chosen in Phase 3, presenting the count of participants who provided a response to the 'Other' option, what their response was, and where their 'Other' option was ranked in the list of differential diagnoses for each freeze in each neonatal trajectory pair. Numbers in parentheses are total count of participants providing an 'Other' response for the freeze noted.

Original Trajectory	Count of participants	'Other' Rank Position	Novel Trajectory	Count of participants	'Other' Rank Position
PAIR 1					
Freeze 1	(3)		Freeze 1	(3)	
Poor Trace	1	1	Suboptimal intervention	1	4
Suboptimal intervention	1	3	Delayed Transition	1	1
Acute event	1	1	Persistent Pulmonary Hypertension	1	3
Freeze 2	(8)		Freeze 2	(9)	
Suboptimal intervention	3	2, 3	Suboptimal intervention	4	1, 2, 3
Asphyxia	2	1, 3	Delayed Transition	2	1, 4
Apnoea	1	1	Asphyxia	1	1
Meconium Aspiration Syndrome	1	3	Meconium Aspiration Syndrome	1	1
Cardiac related	1	1	Airway	1	3
Freeze 3	(3)		Freeze 3	(4)	
Acute event	3	1, 2	Transient Tachypnoea of the Newborn	2	1
			Asphyxia	1	3
			Meconium Aspiration Syndrome	1	1
PAIR 2					
Freeze 1	(8)		Freeze 1	(3)	
Apnoea	4	1, 2	Suboptimal intervention	2	2
Suboptimal intervention	3	2, 3	Acute event	1	2
Acute event	1	1			
Freeze 2	(6)		Freeze 2	(4)	
Suboptimal intervention	6	1, 3	Suboptimal intervention	4	1, 2
Freeze 3	(4)		Freeze 3	(5)	
Suboptimal intervention	3	2, 4	Suboptimal intervention	5	1, 2, 3
Delayed Transition	1	3			
PAIR 3					
Freeze 1	(4)		Freeze 1	(3)	
Suboptimal intervention	2	2, 4	Suboptimal intervention	3	3
Asphyxia	1	1			
Poor trace	1	1			
Freeze 2	(11)		Freeze 2	(9)	
Suboptimal intervention	7	1, 2, 3	Suboptimal intervention	6	1, 2, 3
Apnoea	2	1, 3	Apnoea	3	1
General anaesthetic	1	2	Delayed Transition	1	3
Poor trace	1	3			
PAIR 4					
Freeze 1	(2)		Freeze 1	(2)	
Delayed Transition	1	1	Delayed Transition	2	2, 3
Persistent Pulmonary Hypertension	1	3			
Freeze 2	(3)		Freeze 2	(1)	
Acute event	3	1, 2	Poor trace	1	2
Freeze 3	(1)		Freeze 3	(3)	
Delayed Transition	1	1	Delayed Transition	1	2
			Persistent Pulmonary Hypertension	1	4
			Apnoea	1	2
PAIR 5					
Freeze 1	(5)		Freeze 1	(1)	
Poor trace	3	1, 4	Delayed Transition	1	1
Delayed Transition	1	2			
Transient Tachypnoea of the Newborn	1	3			

Table 2 (continued)

Original Trajectory	Count of participants	'Other' Rank Position	Novel Trajectory	Count of participants	'Other' Rank Position
PAIR 6					
Freeze 1	(4)		Freeze 1	(7)	
Poor trace	2	1, 3	Apnoea	3	1, 2, 3
Acute event	1	1	Poor trace	3	1
Delayed Transition	1	3	Acute event	1	1
Asphyxia	1	3			
Freeze 2	(3)		Freeze 2	(1)	
Apnoea	1	1	Acute event	1	2
Delayed Transition	1	4			
Poor trace	1	3			
PAIR 7					
Freeze 1	(5)		Freeze 1	(4)	
Early cord clamping	1	1	Poor trace	2	1
Delayed Transition	1	1	Delayed Transition	1	1
Secondary Apnoea	1	1	Infection	1	3
Suboptimal intervention	1	4			
General anaesthetic	1	1			
Freeze 2	(6)		Freeze 2	(5)	
Suboptimal intervention	3	1, 2	Suboptimal intervention	4	1, 2, 3
Asphyxia	1	3	Asphyxia	1	2
Infection	1	4			
Suboptimal intervention	1	2			
Freeze 3	(3)		Freeze 3	(4)	
Suboptimal intervention	3	1, 2	Suboptimal intervention	2	1, 4
			Infection	1	1
			Asphyxia	1	4
Freeze 4	(4)		Freeze 4	(4)	
Delayed Transition	2	2	Suboptimal intervention	2	1, 3
Transient Tachypnoea of the Newborn	1	2	Delayed Transition	1	3
Meconium Aspiration Syndrome	1	1	Infection	1	2
PAIR 8					
Freeze 1	(5)		Freeze 1	(5)	
Poor trace	3	1	Respiratory Distress Syndrome	5	1, 2, 3
Respiratory Distress Syndrome	2	2, 3			
Freeze 2	(2)		Freeze 2	(1)	
Apnoea	2	2, 3	Persistent Pulmonary Hypertension	1	3
Freeze 3	(3)		Freeze 3	(3)	
Suboptimal intervention	3	2, 3	Intervention adjustment	3	1, 2, 3

Table 3 – Summary of the count and the percentage of 'Other' differential diagnoses ranked in the 1st, 2nd, 3rd, and 4th position for the freezes in each neonatal trajectory pair. Counts are summed across participants.

Trajectory Pairs	'Other' Rank Position				Total 'Other' Responses
	1st	2nd	3rd	4th	
Pair 1	15	5	7	2	29
Pair 2	12	10	6	2	30
Pair 3	11	8	9	1	29
Pair 4	6	4	2	1	13
Pair 5	3	1	1	1	6
Pair 6	8	2	4	1	15
Pair 7	16	10	5	4	35
Pair 8	6	6	7	0	19
Total Freezes	77	46	41	12	176
Percentage Total	44%	26%	23%	7%	n/a

relationships between data, human expertise will probably be needed to assist in training and optimising algorithms for newborn resuscitation. Therefore, for the domain of newborn resuscitation, further research is needed (a) to explore the feasibility of different AI methods for training and optimising an algorithm using the typically available physiological data, and (b) to determine if it is possible for the algorithm to consider contextual information that could significantly change the interpretation of the newborn's vital signs.

Beyond the performance of an AI algorithm itself, it is also crucial to consider how, when, and where the processes and outputs of such an algorithm would be presented to the clinician. There is evidence from contexts such as emergency department resuscitation that when the physical form of a cognitive aid is designed to anticipate user needs and workflow, the speed of clinicians' comprehension of the patient's physiological state and their selection of actions can be improved.[19–23] In the next phase of the research program (Phase 4 in Fig. 1), we will explore if (and how) a cognitive aid that provides an interpretation of the newborn's state affects management of the infant during resuscitation after birth. More research is also needed to determine the ideal physical form of a newborn resuscitation cognitive aid, and to determine if (and how) a cognitive aid could interact with pre-existing resuscitation algorithms.

Limitations

Despite the encouraging findings, this study has several limitations. First, given constraints on participants' time, we could present only eight neonatal trajectory pairs. Second, the trajectories in the database were collected in 2005 and displayed limited information. Without detailed maternal and fetal history and knowledge of the appearance of the newborn, clinicians lacked some information that they might use in clinical practice for ranking differential diagnoses, potentially increasing the concordance across clinicians. Moreover, the newborns' clinical course may not have been typical of a clinical course in the context of contemporary guidelines and devices. Third, the participants could not remove any differential diagnoses from the original rank ordered list, and they could provide only one response to the 'Other' option. These restrictions could have inflated the degree of similarity that we found. However, across the freezes only 33% of participants on average provided an 'Other' response, which were often iatrogenic.

Conclusions

The findings provide additional support for the feasibility of using concordance of neonatal experts' interpretation of newborn vital sign patterns as an approximation of the newborn's true physiological state. The findings are a further important step towards determining whether experts' interpretations can generalise to other similar neonatal vital signs patterns. Until clinically practical technology emerges that conveys direct information about the neonate's anatomy and physiology during the post-natal transition, the present results may justify the development and training of a algorithm that could underpin a cognitive aid to support clinicians' management of the newborn during resuscitation after birth.

Declarations of Interest

None.

CRediT authorship contribution statement

Jelena Zestic: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Helen G. Liley:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Penelope M. Sanderson:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.resplu.2022.100263.

Author details

[a]School of Psychology, The University of Queensland, Brisbane, Queensland, Australia[b]Mater Mothers' Hospital and Mater Research Institute, Brisbane, Queensland, Australia

REFERENCES

1. Zestic J, Sanderson P, Dawson J, Liley H. Defining information needs in neonatal resuscitation with work domain analysis. J Clin Mont Comput 2021;35:689–710. https://doi.org/10.1007/s10877-020-00526-7.
2. Hooper S, Kitchen JM, Polglase RG, Roehr CC, Te Pas BA. The physiology of neonatal resuscitation. Curr Opin Pediatr 2018;30:187–91. https://doi.org/10.1097/MOP.0000000000000590.
3. Schmölzer GM, Morley CJ, Kamlin OCOF. Enhanced monitoring during neonatal resuscitation. Semin Perinatol 2019;43:151–77. https://doi.org/10.1053/j.semperi.2019.08.006.
4. Askin DF. Complications in the transition from fetal to neonatal life. J Obstet Gynecol Neonatal Nurs 2002;31:318–27. https://doi.org/10.1111/j.1552-6909.2002.tb00054.x.

5. Hillman NH, Kallapur SG, Jobe AH. Physiology of transition from intrauterine to extrauterine life. Clin Perinatol 2012;39:769–83. https://doi.org/10.1016/j.clp.2012.09.009.

6. Bould MD, Hayter MA, Campbell DM, Chandra DB, Joo HS, Naik VN. Cognitive aid for neonatal resuscitation: a prospective single-blinded randomized controlled trial. Br J Anaesth 2009;103:570–5. https://doi.org/10.1093/bja/aep221.

7. 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 Healthc 2020;15:243–50. https://doi.org/10.1097/SIH.0000000000000422.

8. Fuerch JH, Yamada NK, Coelho PR, Lee HC, Halamek LP. Impact of a novel decision support tool on adherence to Neonatal Resuscitation Program algorithm. Resuscitation 2018;88:52–6. https://doi.org/10.1016/j.resuscitation.2014.12.016.

9. François C. What is deep learning? In: Taylor C, Gaines J, Dragosavljevic A, Taylor T, editors. Deep learning with Python. New York: Manning Publications Company; 2018. p. 3–24.

10. Zestic J, Liley H, Sanderson P. Concordance of expert clinicians' interpretations of the newborn's true physiological state. Pediatr Res 2021. https://doi.org/10.1038/s41390-021-01565-9.

11. Dawson JA, Kamlin OC, Vento M, et al. Defining the reference range for oxygen saturation for infants after birth. Pediatrics 2010;125: e1340–7. https://doi.org/10.1542/peds.2009-1510.

12. Levenshtein VI. Binary codes capable of correcting deletions, insertions, and reversals. Dokl Phys 1966;10:707–10.

13. Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Inf Sci 2017;405:81–90. https://doi.org/10.1016/j.ins.2017.04.012.

14. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Comput Biol Med 2018;100:270–8. https://doi.org/10.1016/j.compbiomed.2017.09.017.

15. Längkvist M, Karlsson L, Loutfi A. Sleep stage classification using unsupervised feature learning. Adv Artif Neural Syst 2012;2012:1–9. https://doi.org/10.1155/2012/107046.

16. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: a review. Comput Meth Prog Bio 2018;161:1–13. https://doi.org/10.1016/j.cmpb.2018.04.005.

17. Kia A, Timsina P, Himanshu JN, et al. MEWS++: Enhancing the prediction of clinical deterioration in admitted patients through a machine learning model. J Clin Med 2020;9:343. https://doi.org/10.3390/jcm9020343.

18. Shamout F, Zhu T, Clifton DA. Machine learning for clinical outcome prediction. IEEE Rev Biomed Eng 2020;14:116–26. https://doi.org/10.1109/RBME.2020.3007816.

19. Grundgeiger T, Huber S, Reinhardt D, Steinisch A, Happel O. Cognitive aids in acute care: Investigating how cognitive aids affect and support in-hospital emergency teams. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; 2019 May 4-9; Glasgow, UK. New York: Association for Computing Machinery; 2019. https://doi.org/10.1145/3290605.3300884.

20. Marshall S. The use of cognitive aids during emergencies in anesthesia: a review of the literature. Anesth Analg 2013;117:1162–71. https://doi.org/10.1213/ANE.0b013e31829c397b.

21. Calder LA, Bhandari A, Mastoras G, Day K, Momtahan K, Falconer M. Healthcare providers' perceptions of a situational awareness display for emergency department resuscitation: a simulation qualitative study. Int J Qual Health Care 2018;30:16–22. https://doi.org/10.1093/intqhc/mzx159.

22. Gonzales M, Henry J, Calhoun A, Riek L. Visual TASK: a collaborative cognitive aid for acute care resuscitation. In: Proceedings of 10th EAI International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health); 2016 May 16-19; Cancun, Mexico. New York: Cornell University Library; 2016. https://doi.org/10.4108/eai.16-5-2016.2263328.

23. Schild S, Sedlmayr B, Schumacher A, Sedlmayr M, Prokosch H, St. Pierre M. A digital cognitive aid for anesthesia to support intraoperative crisis management: results of the user-centered design process. JMIR Mhealth Uhealth 2019;7:e13226. https://doi.org/10.2196/13226.