Royal Australasian College of Surgeons

Stratifying features for diagnosing hypertrophic stenosis on ultrasound: a diagnostic accuracy study

Toby I. Vinycomb *, Keith Vanhaltren†, Maurizio Pacilli *‡, Michael Ditchfield†§ and Ramesh Mark Nataraja *‡

*Department of Paediatric Surgery, Monash Children's Hospital, Melbourne, Victoria, Australia
†Department of Radiology, Monash Children's Hospital, Melbourne, Victoria, Australia
‡Departments of Paediatrics and Surgery at the School of Clinical Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia and
§Departments of Radiology and Paediatrics, Monash University, Melbourne, Victoria, Australia

Key words
diagnostic imaging, pyloric stenosis, pyloromyotomy, ultrasonography.

Correspondence
Associate Professor Ramesh Nataraja, Department of Paediatric Surgery, Monash Children's Hospital & Monash University, 246 Clayton Road, Clayton, Melbourne, VIC, Australia.
Email: ram.nataraja@monashhealth.org

T. I. Vinycomb MBBS, BMedSc (Hons);
K. Vanhaltren BRadMedImag (Hons), MMU;
M. Pacilli MBBS (Hons), MD(Res);
M. Ditchfield MBBS, FRANZCR;
R. M. Nataraja MBBS, FRACS (Paeds).

The paper was presented at the Royal Australian College of Surgeons Annual Scientific Congress as an oral presentation.

Accepted for publication 10 March 2022.

doi: 10.1111/ans.17649

Abstract

Background: Our aims were to determine if the diagnostic threshold for diagnosing hypertrophic pyloric stenosis (HPS) on ultrasound scan (USS) should be adjusted based on birth weight (BW), current weight (CW), gestational age (GA), chronological age (CA) or corrected gestational age (CGA).

Methods: All patients who underwent either an USS and pyloromyotomy (Group 1) or an USS for possible HPS (Group 2) at our tertiary centre between July 2013 and June 2019 were identified. Ideal threshold values are identified by measuring Youden's Index (J = sensitivity + specificity − 1; higher is better). Mean maximum Youden's Index for stratified results was compared to that for combined results.

Results: Two hundred and eighty-four patients were included (142 patients in both Group 1 and Group 2). Combined maximum Youden's Index for all patients was 0.92 for pyloric canal thickness (PMT) and 0.87 for pyloric canal length (PCL). Mean maximum Youden's Index was higher when patients were stratified by GA, CGA, BW or CW, and equivalent for CA. For pyloric canal length (PCL), mean maximum Youden's Index was lower for all variables when stratified compared to combined. There was no visual trend observed in the diagnostic thresholds between groups.

Conclusion: Stratifying USS PMT diagnostic thresholds values based on age and weight is statistically more accurate than a single threshold in diagnosing HPS. However, the lack of visual correlation indicates a larger dataset is required to validate these results.

Introduction

The diagnosis of hypertrophic pyloric stenosis (HPS) has evolved with the almost universal use of ultrasound as diagnostic tool in Australia.[1,2] Many potential ultrasound measurements have been assessed in the diagnosis of HPS. However, initial research has demonstrated that pyloric muscle thickness (PMT) and pyloric canal thickness (PCL) are the best discriminators for diagnosing HPS.[3]

There has been a limited evidence base for determining if gestation, age (corrected age or corrected gestational age) or weight (birth weight or current weight) should alter the diagnostic criteria.

ANZ Journal of Surgery published by John Wiley & Sons Australia, Ltd on behalf of Royal Australasian College of Surgeons.

Gestation and corrected gestational age may be independent factors as there is evidence that infants of earlier gestation, present at an older age.[4] Previous studies that examined the possible confounding factors of age and weight on pyloric measurements have only observed for trends[1,5–8] or used a subjective cut off values (e.g. premature versus term gestation) to make dichotomous comparisons.[5,9] No studies have attempted to determine ideal USS cut off values for HPS using age or weight as a stratifying factor.

Therefore, the aim of this study was to determine if the ultrasound criteria of HPS should be adjusted based on:

(1) Gestation,
(2) Chronological age (CA),
(3) Corrected gestational age (CGA),
(4) Birth weight (BW), or
(5) Current weight (CW).

Methods

Study design

A cohort selection cross-sectional diagnostic accuracy study from July 2010 to June 2019 comparing the USS of infants with HPS to those without HPS. An infant was deemed to have had HPS if they underwent a pyloromyotomy at our institution (laparoscopic or open) with the intraoperative findings were consistent with HPS.

Our primary outcomes were to:

(1) Evaluate the correlation between PMT and PCL, and gestation, CA, CGA, BW and CW;
(2) Determine the accuracy of PMT and PCL for diagnosing HPS when stratified based on gestation, CA, CGA, BW and CW.

Sample size calculation

A power calculation could not be performed to establish the ideal number of infants that need to be included in a diagnostic study without a pre-test probability.[10] As there is no routine screening test and no published data on the number of patients who have HPS as a proportion of all patients referred for an USS from all referral sources, a pre-test probability could not be objectively determined. A previous study examining the outcomes of children with HPS at our institution identified a mean of 15 cases per year between 2010 and 2015.[2] Therefore, we decided to collect data on infants from July 2010 to June 2019 for a total of approximately 135 infants with HPS. This was to be matched by a similar number of pyloric USS in infants who did *not* have HPS, for a total of 270 infants.

Case selection

An infant was determined to have HPS if they underwent an operation for HPS as identified by the Medicare Benefits Scheme (MBS)[11] billing code. The coding is performed by the operating surgeon or surgical registrar/trainee. Every procedure performed at our centre has an operative code electronically entered at the time of the operation. The code used for this procedure is unique (MBS code 43930: HYPERTROPHIC PYLORIC STENOSIS; pyloromyotomy for). There are no other MBS codes that include 'pyloromyotomy' or 'pyloric' in their text minimizing potential identification errors.

Control participants were identified by having been referred for an USS for the investigation of possible HPS and did not undergo pyloromyotomy. There were identified using the hospital picture archiving and communication system (PACS). A search was done of the keywords of 'hypertrophic pyloric stenosis', 'pyloric Stenosis' or 'pylorus' over the study period. Infants who had undergone a pyloromyotomy or with indications not related to HPS were excluded from the control group.

The specific inclusion criteria for each group were:

- Infants with HPS:
  (1) Chronological age ≤17 weeks (4 months);
  (2) Had at least one USS performed;
  (3) The radiology report contained at least pyloric muscle thickness or pyloric canal length, or the images were available to measure these values;
  (4) Underwent a pyloromyotomy at with the intraoperative findings being HPS.
- Infants without HPS:
  (1) Chronological age ≤17 weeks (4 months);
  (2) Underwent an USS during the study period for the investigation of HPS;
  (3) The radiology report contained at least one of: pyloric muscle thickness or pyloric canal length, or these values could be measured from the available images;
  (4) Did not undergo a pyloromyotomy.

Chronological age ≤17 weeks was chosen as we had (predominantly premature) infants present up to a chronological age of 117 days in a previous study that included our centre.[2] Potential participants were excluded if their medical history was not available; or they had previous abdominal surgery. For each financial year, we used a random computer-generated number sequence to match patients with HPS to an equal number of infants that met the study criteria but did not have HPS.

Data collection

Demographic data and clinical information were collected from participant's scanned medical record (SMR) which included external institution's notes and USS reports. Internal institutional USS were reviewed through our internal radiology application Vue Motion Version 12.1 (Rochester, NY: Carestream Health, Inc.). The highest reported or observed value of each PMT and PCL was used for analysis.

Data analysis

Data analysed included demographic, clinical and USS results. For infants with multiple USS the last USS prior to surgery was used for the analysis. Data normality was assessed with a Shapiro–Wilk test. Normally distributed data is presented as mean, standard deviation (SD) and 95% confidence interval (CI) and not normally distributed data as presented as median, inter-quartile range (IQR) and

range. Comparison between Group 1 and 2 utilized a Mann–Whitney U test or student t-tests as appropriate. Youden's Index (J = Sensitivity + Specificity – 1) was used as the comparator between cut-off values. Reported cut-off values are the number at which a test becomes positive. Tests were analysed in 0.1 mm intervals.

Group sizes for each variable (e.g. weight brackets for calculating optimal cut-off values for current weight) were calculated by allocating a minimum of 20 infants to each group. Gestation was rounded down to the nearest week; chronological age and corrected gestational age were rounded down to the nearest day; birth weight and current weight were rounded to the nearest 10 g and were stratified in to 500 g groups.

Data was analysed using SPSS Version 26.0 (Armonk, NY: IBM Corp). Youden's Index and contingency table statistics was calculated using Microsoft Excel Version 16.3 (Redmond, WA; Microsoft Corp). A P-value <0.05 was determined to be statistically significant for all tests.

Human research and ethics

An ethics exemption (reference: RES-19-0000-808Q) was obtained from our institutional Human Research and Ethics Committee (HREC).

Results

We identified 176 infants underwent a pyloromyotomy. Thirty-four were excluded: 27 did not undergo an USS, five had no medical records available, one was incorrectly coded and one had a previous pyloromyotomy for HPS (initial operation was performed at an external hospital). Therefore, 142 infants were included in the final analysis (median 14 cases/year, range 12–20). There was a statistical difference in proportion of sex between the groups; (males, 97/142 (56%) with HPS versus 121/142 (85%) without HPS, $P < 0.001$).

Table 1 Correlation between PMT or PCL, and variables

	Infants with HPS		Infants without HPS	
	n	Pearson's r	*n*	Pearson's r
PMT and:				
Gestation	120	0.038	133	−0.054
CGA	119	0.329**	125	0.009
CA	142	0.348**	125	0.125
BW	121	0.044	113	0.008
CW	137	0.155	112	0.047
PCL and:				
Gestation	120	0.079	125	−0.157
CGA	119	0.181*	118	−0.154
CA	142	0.124	118	−0.029
BW	121	0.192*	107	0.015
CW	137	0.201*	104	−0.051

*P<0.05 level (2-tailed); **P<0.01 level (2-tailed).

Table 2 Overall mean cut-off values

	PMT (Youden's Index, J)	PCL (Youden's Index, J)
Combined overall optimal cut-off *(PMT≥3 mm, PCL≥14.5 mm)*	*0.94*	*0.94*
Gestation	0.95	0.90
Chronological age	0.94	0.91
Corrected gestational age	0.96	0.89
Birth weight	0.96	0.89
Current weight	0.95	0.90

Correlation

In infants with HPS, PCL had a weak positive correlation with corrected age, birth weight or current weight, and PMT had a moderate positive correlation corrected gestational age and chronological age (Table 1). In infants without HPS, there was no correlation between either PMT or PCL, and the pre-defined variables (Table 1).

Cut-off values stratified by variable

Using maximum Youden's Index for all ultrasounds, ideal cut-off for PMT was 3.0 mm (J = 0.92) and PCL was 14.5 mm (J = 0.87). Combining a PMT of 3.0 mm and a PCL of 14.5 mm provided the optimal cut-off value (J = 0.94, *combined overall optimal cut-off*).

Compared to the combined overall optimal cut-off, the mean maximum Youden's index for PMT was equal or higher for all stratifying variables (Table 2). On the other hand, mean maximum Youden's index for PCL was lower for all stratifying variables (Table 2).

Stratifying by gestation (Table 3), the maximum Youden's Indices for all PMT groups except gestation ≥40 weeks, and a gestation of 39 weeks for PCL were equal or better than the *combined overall optimal cut-off*. The mean maximum Youden's Index value for PMT was 0.95 and PCL was 0.90.

Stratifying by corrected age (Table 4), the maximum Youden's Indices for all corrected age groups except ≤2 weeks for PMT, and corrected age groups of 0–4 weeks or better than the *combined overall optimal cut-off*. Mean maximum Youden's Index for PMT was 0.94 and PCL was 0.91.

Table 3 Cut-off values adjusted for gestation

Weeks (gestation)	*n*	mm	Maximum Youden's Index
PMT			
≤37	69	3.0	0.94
38	57	3.0	0.97
39	50	3.0	0.96
≥40	69	3.2	0.92
PCL			
≤37	64	15.7	0.82
38	55	15.0	0.84
39	46	15.0	1.00
≥40	73	15.5	0.95

Table 4 Cut-off values adjusted for chronological age (CA)

Weeks (CA)	*n*	mm	Maximum Youden's Index
PMT			
≤2	41	3.0	0.80
3	44	3.4	0.94
4	42	3.0	1.00
5	39	3.0	0.95
6–7	48	2.7	0.96
≥8	61	3.5	0.98
PCL			
≤2	38	15.5	1.00
3	42	15.3	0.94
4	40	14.9	0.97
5	39	15.7	0.85
6–7	47	13.5	0.85
≥8	61	14.5	0.85

Stratifying by corrected gestational age (Table 5), the maximum Youden's Indices for corrected gestational age for all groups except 41–42 weeks for PMT were equal or better than the *combined overall optimal cut-off*. There were no maximum Youden's Indices for PCL that were better the previous optimal cut-off. Mean maximum Youden's Index for PMT was 0.96 and PCL was 0.89.

Stratifying by birth weight (Table 6), the maximum Youden's Indices for all birth weight groups except 2.51–3.0 kg for PMT, and birth weight groups 3.01–3.5 kg, and >4 kg for PCL, were equal or better than the *combined overall optimal cut-off*. Mean maximum Youden's Index for PMT was 0.96 and PCL was 0.89.

Stratifying by current weight (Table 7), all maximum Youden's Indices for all groups except a CW >4 kg for PMT, and ≤2.5 kg and 4.01–4.5 kg for PCL, were equal or better than the *combined overall optimal cut-off*. Mean maximum Youden's Index for PMT was 0.95 and PCL was 0.90.

Discussion

Changes in the normal and abnormal pyloric measurements based on both age or weight on face value appear logical. The average birthweight of babies in Australia in 2018 was 3.323 kg[12] and babies on average are expected to double their weight by 4 months of age.[13] It would be expected that a proportion of this weight would be the growth of all the intra-abdominal organs including the pylorus. However, there has been mixed evidence on which factors may correlate with HPS. Potential factors that could influence pyloric size include gestation,[5–7] corrected age,[5,6,9] corrected gestational age[1,5,7] and current weight.[1,5–8]

Table 5 Cut-off values adjusted for corrected gestational age (CGA)

Weeks (CGA)	*n*	mm	Maximum Youden's index
PMT			
≤40	44	3.0	1.00
41–42	59	3.0	0.89
43–44	66	3.1	0.94
45–46	38	3.0	0.96
≥47	38	3.5	1.00
PCL			
≤40	41	16.0	0.85
41–42	54	15.0	0.88
43–44	64	15.1	0.91
45–46	39	15.0	0.92
≥47	40	17.0	0.93

Table 6 Cut-off values adjusted for birth weight (BW)

Weight (kg)	*n*	mm	Maximum Youden's Index
PMT			
≤2.5	28	3.7	0.97
2.51–3.0	53	3.0	0.92
3.01–3.5	75	3.0	0.97
3.51–4.0	48	3.0	0.96
>4	29	2.7	1.00
PCL			
≤2.5	26	15.6	0.85
2.51–3.0	53	16.0	0.88
3.01–3.5	73	14.3	0.97
3.51–4.0	46	15.5	0.79
>4	30	15.1	0.95

There was no trend between these stratifying features and PMT or PCL in infants without HPS. In infants with HPS however, we identified a moderate positive correlation between PMT and CA or CGA. These findings would suggest that there is no pyloric muscle growth in healthy infants less than 4 months of age without HPS. However, older infants with HPS have a larger PMT. This may indicate that older infants have a larger pyloric lumen that tolerates a greater degree of pyloric muscle hypertrophy (i.e. increased PMT) prior to presenting with clinical obstruction.

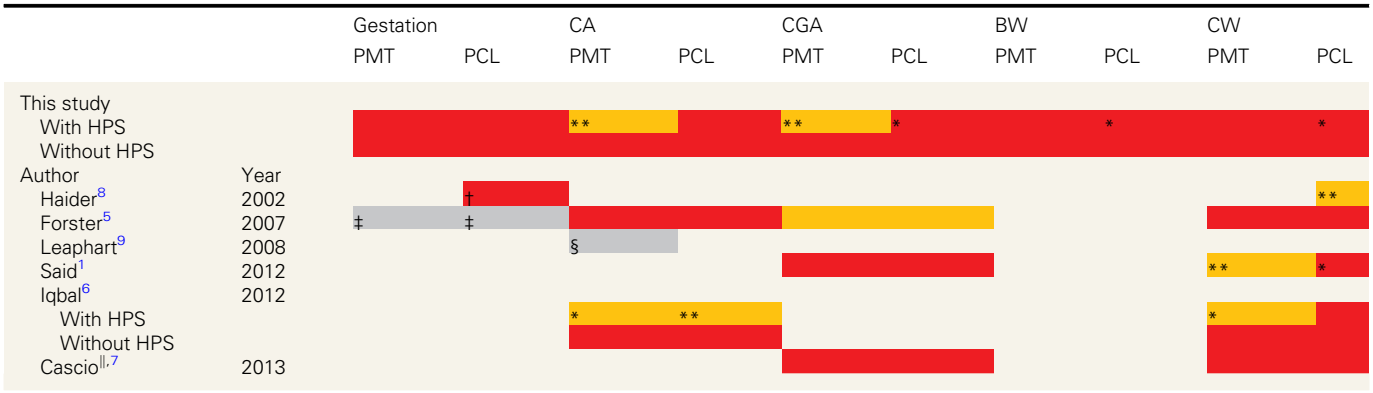
Pyloric canal length demonstrated a statistically significant weak correlation between PCL, and CA, BW as well as CW in infants with HPS. Again, these correlations were not present in infants without HPS. It is unclear what the significance of these correlations may be. Overall, compared to the current literature, our results support the limited evidence that there is largely no strong correlation between weight or age, and PMT or PCL (Table 8).

When we adjusted PMT or PCL cut-off values based on age or weight, the increased mean Youden's index indicates we should consider adjust cut-off values of PMT (but not PCL) based on an infant's gestation, CA, CGA, BW, CW. However, the validity of this result needs to be weighed against the lack of strong statistical

Table 7 Ideal cut-off values stratified by current weight (CW)

Weight (kg)	*n*	mm	Maximum Youden's index
PMT			
≤2.5	27	3.0	0.94
2.51–3	50	3.0	0.97
3.01–3.5	57	3.3	0.95
3.51–4	48	3.0	0.96
4.01–4.5	30	3.2	1.00
>4.5	37	2.8	0.91
PCL			
≤2.5	24	15.0	0.94
2.51–3	47	15.6	0.84
3.01–3.5	55	15.1	0.92
3.51–4	45	14.3	0.86
4.01–4.5	30	15.1	0.94
>4.5	40	13.5	0.88

Table 8 Correlation between infant factors and pyloric measurements



White – test not performed, red – no/weak non-statistically significant correlation (*r*-value: −0.29 to 0.29), orange – moderately positive (*r*-value: 0.3 to 0.49), grey – comparative testing only, **P*<0.05, ***P*<0.01; [†]preterm infants only; [‡]positive difference between preterm to term infants (*t*-tests, both *P*<0.001); [§]positive difference between infants ≤21 days to >21 days (*t*-test, *P*<0.05); [‖]no difference between preterm and term infants (multivariate regression analysis, $P = 0.626$).

correlation or a visual trend in cut-off measurements. For example, the mean Youden's Index for PMT stratified by gestation was 0.95. This Youden's Index was higher than the Youden's Index identified by combining PMT (≥3.0 mm) and PCL (≥14.5 mm). However, only babies with a gestation of ≥40 weeks had a change in PMT cut-off measurement from 3.0 to 3.2 mm ($J = 0.92$, Table 3). This is likely to have occurred for two reasons – both of which are related to lower number of data points in this analysis.

Firstly, even though minimum group numbers were then used for each stratified group, a reduction in data points is associated with a reduced significance in the calculation of the maximum Youden's Index. The curvature of the line is reliant on the number of data points, with the higher number of distinct data points increasing the smoothness of the curve.

Secondly, small changes in the reported measurements when there are a low number of data points may change the ideal cut-off. The measurements of PMT and PCL are done to 1/10th of a millimetre. A change of up to 0.2 to 0.3 mm could be explained by a change in probe position or compression on the abdomen in a moving alert infant. Inter-observer variability has not been studied in the diagnosis of HPS but could also be a contributing factor in small changes in reported US measurements. For example, a baby born at 41 weeks that goes on to develop HPS has their PMT measured by as 3.0 mm by the ultrasonographer, but then 3.2 mm by the radiologist. This may change the Youden's Index for the ≥40 weeks' gestation group if that was the discriminating USS. This variation is overcome by having larger number of infants in each group.

We did not re-measure ultrasounds to standardize results in our study despite evidence there is only moderate interobserver agreement on measurements.[14] This was done for a number of reasons. It reduces spectrum bias as is favourable against the Quality Assessment of Studies of Diagnostic Accuracy (QUADAS) guidelines.[15] There is also evidence that a number of other non-controlled factors can influence measurements including hydration.[16] Finally, we hoped to match findings to those 'in the real world' but using ultrasounds reported from a variety of sources.

Pyloric muscle thickness stratified by CA and CGA demonstrated a mean Youden's Index of 0.94 and 0.96, equal or greater than the Youden's Index for the combined PMT/PCL cut-off ($J = 0.94$). However, unlike the moderate correlation demonstrated in these groups there was no incremental positive trend in the ideal cut-off values. For example, when PMT was stratified by CA, ideal PMT cut-off dropped from 3.0 mm at 5 weeks, to 2.7 mm at 6 to 7 weeks, and again increased to 3.5 mm at ≥8 weeks CA (Table 4). There was similar lack of visual trends amongst the groups that had a significant weak positive correlation (CA, BW or CW, and PCL).

Conclusion

We have identified that PMT and PCL did not have a strong correlation with infant age or weight. There was statistical improvement when cut-off values for PMT or PCL were adjusted for an infant's age or weight. However, the lack of correlation and visual trend indicates that we should not adjust cut-off values for PMT and PCL in the ultrasound diagnosis of HPS.

Acknowledgement

All the staff in the Paediatric Surgery, Emergency Department and Radiology Departments at our Institution. Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

Author contributions

Toby I. Vinycomb: Conceptualization; data curation; formal analysis; methodology; writing – original draft. **Keith Vanhaltren:** Conceptualization; data curation; methodology; resources; writing – review and editing. **Maurizio Pacilli:** Conceptualization; methodology; project administration; supervision; writing – review and editing. **Michael Ditchfield:** Conceptualization; project administration; resources; writing – review and editing. **Ramesh Nataraja:** Conceptualization; methodology; project administration; resources; supervision; writing – review and editing.

Conflict of interest

None declared.

References

1. Said M, Shaul DB, Fujimoto M, Radner G, Sydorak RM, Applebaum H. Ultrasound measurements in hypertrophic pyloric stenosis: don't let the numbers fool you. *Perm. J.* 2012; **16**: 25–7.
2. Vinycomb TI, Laslett K, Gwini SM, Teague W, Nataraja RM. Presentation and outcomes in hypertrophic pyloric stenosis: an 11-year review. *J. Paediatr. Child Health* 2019; **55**: 1183–7.
3. Blumhagen JD, Maclin L, Krauter D, Rosenbaum DM, Weinberger E. Sonographic diagnosis of hypertrophic pyloric stenosis. *AJR Am. J. Roentgenol.* 1988; **150**: 1367–70.
4. Costanzo CM, Vinocur C, Berman L. Prematurity affects age of presentation of pyloric stenosis. *Clin. Pediatr.* 2017; **56**: 127–31.
5. Forster N, Haddad RL, Choroomi S, Dilley AV, Pereira J. Use of ultrasound in 187 infants with suspected infantile hypertrophic pyloric stenosis. *Australas. Radiol.* 2007; **51**: 560–3.
6. Iqbal CW, Rivard DC, Mortellaro VE, Sharp SW, St Peter SD. Evaluation of ultrasonographic parameters in the diagnosis of pyloric stenosis relative to patient age and size. *J. Pediatr. Surg.* 2012; **47**: 1542–7.
7. Cascio S, Steven M, Livingstone H, Young D, Carachi R. Hypertrophic pyloric stenosis in premature infants: evaluation of sonographic criteria and short-term outcomes. *Pediatr. Surg. Int.* 2013; **29**: 697–702.
8. Haider N, Spicer R, Grier D. Ultrasound diagnosis of infantile hypertrophic pyloric stenosis: determinants of pyloric length and the effect of prematurity. *Clin. Radiol.* 2002; **57**: 136–9.
9. Leaphart CL, Borland K, Kane TD, Hackam DJ. Hypertrophic pyloric stenosis in newborns younger than 21 days: remodeling the path of surgical intervention. *J. Pediatr. Surg.* 2008; **43**: 998–1001.
10. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. *J. Clin. Diagn. Res.* 2016; **10**: YE01–6.
11. MBS Online [Website]. Australian Government; 2020 [updated 01/05/2020]. Available from: http://www.mbsonline.gov.au/
12. Australian Institute of Health and Welfare. Australia's mothers and babies 2018—in brief. Canberra 29/05/2020. Report No.: PER 108.
13. WHO Multicentre Growth Reference Study Group. *WHO Child Growth Standards: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age: Methods and Development*. Geneva: World Health Organization, 2006.
14. Calle-Toro JS, Kaplan SL, Andronikou S. Are we performing ultrasound measurements of the wall thickness in hypertrophic pyloric stenosis studies the same way? *Pediatr. Surg. Int.* 2020; **36**: 399–405.
15. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med. Res. Methodol.* 2003; **3**: 25.
16. Starinsky R, Klin B, Siman-Tov Y, Barr J. Does dehydration affect thickness of the pyloric muscle? An experimental study. *Ultrasound. Med. Biol.* 2002; **28**: 421–3.