

# Newborn Hip Screenings at 4 to 8 Weeks Are Optimal in Predicting Referral and Treatment Outcomes: A Retrospective Review

Eric C. Lussier<sup>1</sup>, Wei-Te Lei<sup>2,3</sup>, Yi-Ting Sun<sup>1</sup>, Hui-Wen Chen<sup>4,5</sup>, Tung-Yao Chang<sup>1</sup>, Chia-Hsieh Chang<sup>6\*</sup>

<sup>1</sup>Taiji Clinic, Taiwan

<sup>2</sup>Department of Pediatrics, Hsinchu MacKay Memorial Hospital, Taiwan

<sup>3</sup>Graduate Institute of Clinical Medical Sciences, College of Medicine, Chang Gung University, Taiwan

<sup>4</sup>Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Taiwan

<sup>5</sup>Taiwan Adventist Hospital, Taiwan

<sup>6</sup>Chang Gung Memorial Hospital, Taiwan

Email: \*chiahchang@gmail.com

**How to cite this paper:** Lussier, E.C., Lei, W.-T., Sun, Y.-T., Chen, H.-W., Chang, T.-Y. and Chang, C.-H. (2020) Newborn Hip Screenings at 4 to 8 Weeks Are Optimal in Predicting Referral and Treatment Outcomes: A Retrospective Review. *Open Journal of Pediatrics*, 10, 332-346.  
<https://doi.org/10.4236/ojped.2020.102034>

**Received:** May 14, 2020

**Accepted:** June 26, 2020

**Published:** June 29, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Optimal DDH screening timing and whether adding risk profiles could aid in detecting treatment outcome were investigated. Risk factors were employed to supplement ultrasound findings in flagging cases for follow-up. Initial screening results and harness treatment outcomes concordance were compared at different screening ages and screening protocols. Using clinical decision to supplement ultrasound screening allowed to accurately flag all 12 DDH treated cases upon initial visit. Clinical decision correctly identified cases that would have otherwise been missed ( $n = 2$ ). However, doing so increased the rate of false positive cases at all time points of initial screening. Initial screens were more accurate for predicting treatment outcomes when using ultrasound only if done after 28 days [ $\leq 28$  days (88.1%) vs. 29 - 56 days (98.5%), OR = 7.16,  $p < 0.001$ ] or ultrasound with clinical decision [ $\leq 28$  days (86.4%) vs. 29 - 56 days (95.7%), OR = 3.00,  $p < 0.001$ ]. In contrast, screening after 56 days failed to marginally improve accuracy compared to screens done between 29 - 56 days, regardless of the screening protocol employed. Two important trade-offs emerged. First, when choosing timing of initial screening, optimal accuracy and harness treatment schedule should both be considered. Second, when considering whether to use a more conservative risk profile to supplement ultrasound findings, treatment accuracy and the ability to efficiently detect cases requiring harness treatment should both be considered. We provide evidence for performing an initial DDH ultrasound screen between 4 and 8 weeks (29 - 56 days), while employing clinical decision to aid

in determining cases that require further follow-up evaluation.

## Keywords

DDH, Graf Classification, Pediatric Orthopedics, Ultrasound, Screening

---

## 1. Introduction

Developmental dysplasia of the hip (DDH) is a preventable congenital pediatric orthopedic hip condition which affects between 1.5 - 30/1000 live births [1]. DDH is defined as an abnormal relationship between the femur and the acetabulum and ranges from subluxation of the hip joint to complete dislocation [2]. Detection of DDH at the time of walking poses a high risk for corrective hip reduction surgery [3]. Screening early allows for timely, non-invasive treatment for the prevention DDH and associated long-term complications [4].

Currently, ultrasound (US) screening is the best technique for detecting DDH [5]. To reduce rates of surgery, various targeted or universal US screening programs have been implemented. To date, there has yet to be a consensus on screening method and timing that takes into consideration screening accuracy and costs [6] [7]. Targeted screening has traditionally been promoted as a cost-effective method, but has also been shown to effectively decrease surgery rates [8]. A 10-year prospective study by Paton *et al.* [9] found US alone did not provide sufficient predictive value. Although physical exams and risk factors were useful for targeted screening when screening for more severe hip types, the use of risk factors was not supported as an effective tool for selective screening [10]. Universal screening programs have gained support and can provide a reduction in the number of referred cases to orthopedic specialists and early detection of late presenting cases [11] [12]. In Austria, Thaller *et al.* [13] revealed that universal screening became cost-effective, when compared to selective screening, and when considering long-term healthcare costs and falsely identified screened cases [14]. Despite evidence showing that universal US is an effective screening method [15], means of improving screening accuracy and reducing unnecessary follow-up should be considered.

Screening timing is an important factor in accurately assessing the risk of developing DDH [16]. The UK's National Health Services Newborn and Infant Physical Examination (NIPE) and American Academy of Orthopedic Physicians both suggest that targeted screening be employed to detect high-risk DDH cases and screening should occur between 6 and 8 weeks [17], which is largely based on expert opinion [18]. Concerns for delayed screening are merited, with screening later than three months possibly reducing reliability and increasing missed cases [19]. While non-invasive treatment is not typically possible later than 6 months due to advanced hip dysplasia [4]. Severely abnormal hips may even have worse treatment outcomes when initiated after 8 weeks [20], yet treatment of hips has been suggested to be initiated between 8 to 12 weeks of age [21].

Ideally, initial screening age should balance between peak screening accuracy and treatment initiation.

Since optimal screening time, risk profiling and their performance in screening for DDH are not sufficiently understood, further research is needed to support the current recommended guidelines. Our study aimed to compare the screening performance when: 1) only using US findings, and 2) using risk factors to supplement US findings, and comparing performance at different screening age time points in order to find an optimal timing for initiating DDH ultrasound screening.

## **2. Methods**

### **2.1. Sample**

Data was retrospectively reviewed from a DDH registry database and was retrieved from all screening visits conducted between January 1, 2017 and December 31, 2017. Our sample consisted of all newborns attending one of nine medical centers or clinics for both self-referred and referred post-natal check-ups. Referred cases may have been referred to one of the nine participating medical centers/clinics due to indication signs of DDH risk factors or simply for a scheduled check-up. Any cases that had, a previous US performed at another site, or that were older than 6 months of age at initial screen, were excluded from our sample. The nine participating screening centers make up the Taiwanese Screening and Audit System for Developmental Dysplasia of the Hip (TSAS-DDH). The TSAS-DDH is a network of pediatric orthopedic specialists, general pediatric physicians, obstetricians, radiological technicians, pediatric nurses, and public health professionals that perform universal ultrasound screening for DDH, as well as holding training, validating data collection, and analyzing data, to develop an evidence-based screening program that can improve DDH screening outcomes in Taiwan and internationally. Informed consent was obtained from parents/guardians for newborns to participate in the screening registry. The database saw 3018 newborns recruited into the US screening program. Ethics approval was granted by the Institutional Review Board of Chang Gung Medical Foundation (IRB#: 201800670B0).

### **2.2. Hip Ultrasound**

Hip US were assessed using Graf's classification technique, measuring Alpha and Beta angle from a standard coronal section of the hip as previously described [22]. The angles are coded as follows: Ia, Ib, IIa, IIb, IIc, D, III, and IV. Exact Alpha and Beta angle measurements and hip types have been published elsewhere [23]. Hip examinations were performed by trained DDH US operators and newborns were given a Graf classification code for both hips. A single code was later assigned to a newborn based on the hip with the most severe Graf type.

### **2.3. Screening Protocol**

A universal US hip screening protocol with preliminary PE test and risk factors

performed prior to turning 6 months of age after birth was implemented. One or more of the PE tests (Barlow test, Ortolani test, limited abduction of the hip, Galeazzi sign, or Allis sign) were performed. DDH risk factors included: family history of DDH, breech delivery, primiparous pregnancy, oligohydramnios, and any postural deformities of the neck or lower extremities. These risk factors are based on previous established factors [19] [24], and that had complete records in our database. Newborns were classified as *Negative Screened Hips* if a case had: both Ia/Ib type hips, a negative PE test, and without DDH risk factors. Conversely, newborns were categorized as *Negative Ultrasound Hips with Clinical Decision (CD)* if screened as Ia/Ib type with: a positive PE, or presence of a risk factor. These cases were flagged for follow-up one month later to confirm initial screening findings. Type IIa/IIb hip were considered *Positive Ultrasound Hips* and were followed-up to monitor hip development. Newborns with IIc, D, III, or IV types were assigned to the *Dysplastic Ultrasound Hips* group and required transfer to a pediatric orthopedic specialist for confirmation, and if required, DDH harness treatment. After following-up, cases still screening positively or inconclusively, were: referred to a pediatric orthopedic specialist, and were monitored until returning to an Ia/Ib hip type, or requiring harness treatment. This screening protocol used CD to support US Graf classification results (*US with CD*), and was compared to a simulated protocol (had we not employed CD in our decision making), relying only on US Graf classification (*US Only*). Despite there currently being no “Golden Standard” outcome for DDH research [6], over-treatment is common in US DDH screening [25], thus harness treatment for DDH was chosen as the primary outcome. A case nurse followed-up by telephone after 1 year to ensure all cases had not developed late DDH. Cases that had a final US screen which was positive and were unreachable, were deemed to be lost to follow-up.

#### 2.4. Variables

Timing of initial screening was reported in 3 time intervals ( $\leq 28$  days, 29 - 56 days,  $> 56$  days), due to previous studies showing that screening after 28 days can significantly impact screening accuracy [26], while screening after 8 weeks has been shown to have poorer surgical outcomes [20]. Only initial scans performed before 6 months were included, since performing surgery after 6 months is usually less effective, or not possible [4]. Descriptive data was collected including: sex (male and female), gestational number (singleton, multiple), term of birth (preterm or term), and screening physician’s background (general practitioner and pediatric orthopedic specialist). The primary outcome for the study was whether Pavlik’s harness treatment was required (Yes, No) and was coded as a binary variable (accurate or inaccurate) in relation to initial screen. Initial screening outcomes were categorized based on Graf hip classification and presence of risk factors and screening accuracy was coded on concordance with treatment outcome.

## 2.5. Statistical Analysis

Initial screening outcome and final screening outcomes were summarized by group percentages and counts. Descriptive data and screening outcomes were categorized by initial screening age groups compared by Fisher's Exact Test ( $\chi^2$ ) of independence. The rate of harness treatment was further compared by initial Graf classification group and age at screening. We present comparisons of the simulated *US Only* protocol, and our actual *US with CD* protocol. Screening performance was given by reporting the stative counts and column percentages of false positive (FP), false negative (FN), true positive (TP) and true negative (TN) and accuracy. Univariate comparison by initial screening age was done for each using Fisher Exact Test ( $\chi^2$ ). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), negative likelihood ratio (NLR) were also reported to further demonstrate screening performance.

A multivariate logistic regression analysis was conducted comparing the likelihood of an initial screening outcome of accurately predicting whether harness treatment was required. The independent variable of interest was timing of initial screening. We used a pairwise logistic regression model which made contrasted between each time interval. All models controlled for the following covariates: gender, multiple fetal numbers, premature birth, and physician's background. Odds ratio (OR) and 95% confidence interval (95% CI) were reported. Moreover, Z-test of two proportions was performed to compare between screening protocols at each age interval.

Finally, a Receiver Operator Characteristics (ROC) curves were developed and the area under the curve (*AUC*), 95% confidence interval (95% *CI*) and p-values (*p*) were reported for performance of initial screens in predicting a treatment outcome stratified by age at initial screen, for each screening protocol. False negative rate was also calculated at each age and for both screening protocols. An *a priori* two-tailed cut-off of significance of  $p < 0.05$  was established for all analysis. Analysis was conducted using SPSS version 22.0 software.

## 3. Results

**Figure 1** demonstrates our DDH screening program experience. Overall, 3018 newborns were screened, with most being *Negative Screened Hips* (88.5%,  $n = 2671$ ), while 11.5% ( $n = 347$ ) were screened positive. Among screened positive cases, 18.1% ( $n = 63$ ) cases had *Negative Ultrasound Hips with CD*, 74.4% ( $n = 258$ ) cases had *Positive Ultrasound Hips*, and 7.5% ( $n = 26$ ) had *Dysplastic Ultrasound Hips*. Among 347 cases that required further evaluation to finalize outcome, 12 cases (3.5%) required harness treatment for DDH, 299 cases (86.2%) had a negative follow-up, and 36 cases (10.4%) were lost to follow-up.

Descriptive data were compared for all cases by initial screening age (**Table 1**). The majority ( $n = 1941$ ) were screened within 28 days of age, while 722 were screened between 29 days and 56 days of age, and 355 were screened after 56 days of age. Physician's background ( $\chi^2 = 133.608$ ,  $p < 0.001$ ), initial Screen

Outcome ( $X^2 = 126.297$ ,  $p < 0.001$ ) and screening program outcome ( $X^2 = 15.095$ ,  $p < 0.01$ ) all differed significantly by screening age. In contrast, newborn sex, gestational number, and preterm birth did not differ significantly between screening ages.

Among the 2982 cases that were not lost to follow-up, the rate of treatment among Graf classification groups and by age at initial screening were compared (Table 2). Among *Negative Screened Hips*, none required harness treatment (0/2761). Among *Negative Ultrasound Hips with CD*, 3.2% (2/62) required harness treatment with both cases being screened > 56 days. Among *Positive Ultrasound Hips*, 2.7% (6/223) required harness treatment, with 4 being screened  $\leq$  28 days, and 2 being screened between 29 - 56 days. Among initially *Dysplastic Ultrasound Hips*, 15.4% (4/26) required harness treatment, with all cases being screened  $\leq$  28 days.

**Table 1.** Baseline descriptive data for DDH screened newborns (n = 3018).

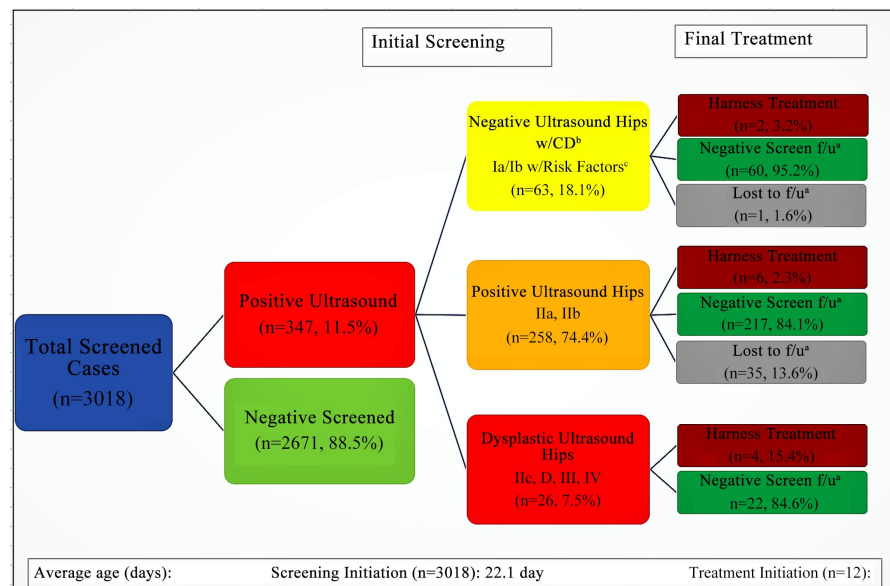
Variables	Age of Newborn at Initial Screening (days)								$(X^2)^b$	Sig.
	Total		$\leq 28$ (n = 1941)		29 - 56 (n = 722)		$> 56$ (n = 355)			
	Count	%	Count	%	Count	%	Count	%		
<b>Newborn sex</b>									0.70	...
Female	1503	49.8%	973	50.1%	350	48.5%	180	50.7%		
Male	1515	50.2%	968	49.9%	372	51.5%	175	49.3%		
<b>Gestational Number</b>									3.72	...
Singleton	2940	97.4%	1885	97.1%	704	97.5%	351	98.9%		
Multiple	78	2.6%	56	2.9%	18	2.5%	4	1.1%		
<b>Preterm Birth</b>									1.15	...
No	2972	98.5%	1908	98.3%	713	98.8%	351	98.9%		
Yes	46	1.5%	33	1.7%	9	1.2%	4	1.1%		
<b>Physician Background</b>									133.61	***
General Pediatric Practitioner	2795	92.6%	1718	88.5%	722	100.0%	355	100.0%		
Pediatric Orthopedic Specialist	223	7.4%	223	11.5%	0	0.0%	0	0.0%		
<b>Initial Screen Outcome</b>									126.30	***
Negative Screened Hips (Ia/Ib)	2671	88.5%	1639	84.4%	687	95.2%	345	97.2%		
Negative Ultrasound (Ia/Ib) with CD <sup>a</sup>	63	2.1%	34	1.8%	20	2.8%	9	2.5%		
Positive Ultrasound Hips (IIa/IIb)	258	8.5%	242	12.5%	15	2.1%	1	0.3%		
Dysplastic Ultrasound Hips (IIc, III, IV, D)	26	0.9%	26	1.3%	0	0.0%	0	0.0%		
<b>Screening Program Outcome</b>									15.10	**
Neg. Screened or Neg. Screened fl u	2970	98.4%	1899	97.8%	718	99.4%	353	99.4%		
Harness Treatment	12	0.4%	8	0.4%	2	0.3%	2	0.6%		
Lost to fl u	36	1.2%	34	1.8%	2	0.3%	0	0.0%		

Notes. a. CD = clinical decision; b. Fisher's Exact Test: Sig. (\*\*p < 0.01, \*\*\*p < 0.001).

**Table 2.** Rate of harness treatment by initial age at screening and hip type (n = 2982).

Graf Hip Type at Initial Screening	Age at Initial Screening						Total	
	≤28 days		29 - 56 days		>56 days		Treated/ Total Graf Type	Rate of Treatment (%)
Negative Screened Hips (Ia/Ib)	0/1639	0.0%	0/687	0.0%	0/345	0.0%	0/2671	0.0%
Negative Ultrasound Hips (Ia/Ib) w/CD <sup>a</sup>	0/33	0.0%	0/20	0.0%	2/9	22.2%	2/62	3.2%
Positive Ultrasound Hips (IIa/IIb)	4/209	1.9%	2/13	15.4%	0/1	0.0%	6/223	2.7%
Dysplastic Ultrasound Hips (IIc/D/III/IV)	4/26	15.4%	0/0	0.0%	0/0	0.0%	4/26	15.4%
<b>Total</b>	<b>8/1907</b>	<b>0.4%</b>	<b>2/720</b>	<b>0.3%</b>	<b>2/355</b>	<b>0.6%</b>	<b>12/2982</b>	<b>0.4%</b>

Notes. a. CD = clinical decision.



**Figure 1.** Initial screening and final treatment outcomes of cases in the DDH newborn screening program (n = 3018). Notes. a. f/u = follow-up; b. CD = clinical decision; c. Clinical decision are based on negative screened cases with risk factors including: family history of DDH, having had a breech delivery, primiparous birth, oligohydramnios, and postural deformities of neck/lower extremities.

Screening performance for both *US Only* and *US with CD* screening protocols were described in **Table 3**. When using *US Only*, FP decreased significantly from 11.9%, to 1.5%, and 0.3%, as screening age increased (≤28 days, 29 - 56 days, >56 Days, respectively). Two FN cases were found after 56 days, and the TN rate increased with age (87.7%, 98.2%, and 99.2%, respectively). Sensitivity was perfect except in the last screening age group due to no cases being correctly identified after 56 days, while specificity improved from 88.0%, 98.5%, and



**Table 3.** Initial screening performance for detecting harness treatment outcome between *US Only* and *US with CD* (n = 2982).

	US Only ( <i>Ultrasound Graf Classification Only</i> )				US with CD ( <i>Ultrasound Graf Classification with CD</i> )				Sig.
	≤28 Days (n = 1907)	29 - 56 Days (n = 720)	>56 Days (n = 355)	Sig. ( $\chi^2$ ) <sup>b</sup>	≤28 Days (n = 1907)	29 - 56 Days (n = 720)	>56 Days (n = 355)	( $\chi^2$ ) <sup>b</sup>	
<b>False Positive (FP)</b>	227 (11.9%)	11 (1.5%)	1 (0.3%)	109.01 ***	260 (13.6%)	31 (4.3%)	7 (2.0%)	79.39 ***	
<b>False Negative (FN)</b>	0 (0.0%)	0 (0.0%)	2 (0.6%)	14.81 ***	0 (0.0%)	0 (0.0%)	0 (0.0%)	... ...	
<b>True Positive (TP)</b>	8 (0.4%)	2 (0.3%)	0 (0.0%)	1.67 ...	8 (0.4%)	2 (0.3%)	2 (0.6%)	0.52 ...	
<b>True Negative (TN)</b>	1672 (87.7%)	707 (98.2%)	352 (99.2%)	104.97 ***	1639 (85.9%)	687 (95.4%)	346 (97.5%)	77.05 ***	
<b>Sensitivity<sup>d</sup></b>	100.0%	100.0%	0.0%	n/a <sup>c</sup>	100.0%	100.0%	100.0%	n/a <sup>c</sup>	
<b>Specificity<sup>e</sup></b>	88.0%	98.5%	99.7%	n/a <sup>c</sup>	86.3%	95.7%	98.0%	n/a <sup>c</sup>	
<b>Positive Predictive Value (PPV)<sup>f</sup></b>	3.4%	15.4%	0.0%	n/a <sup>c</sup>	3.0%	6.1%	22.2%	n/a <sup>c</sup>	
<b>Negative Predictive Value (NPV)<sup>g</sup></b>	100.0%	100.0%	99.4%	n/a <sup>c</sup>	100.0%	100.0%	100.0%	n/a <sup>c</sup>	
<b>Positive Likelihood Ratio (PLR)<sup>h</sup></b>	8.37	65.27	0.00	n/a <sup>c</sup>	7.30	23.16	50.43	n/a <sup>c</sup>	
<b>Negative Likelihood Ratio (NLR)<sup>i</sup></b>	0.00	0.00	1.01	n/a <sup>c</sup>	0.00	0.00	0.00	n/a <sup>c</sup>	
<b>Accuracy<sup>j</sup></b>	88.1%	98.5%	99.2%	93.62 ***	86.4%	95.7%	98.0%	79.39 ***	

Notes. a. CD = clinical decision; b. Fisher's Exact Test: Sig. (\*p < 0.05; \*\*\*p < 0.001); c. n/a = Fisher's Exact Test is not applicable; d. Sensitivity: TP/TP + FN; e. Specificity: TN/TN + FP; f. PPV: TP/TP + FP; g. NPV: TN/TN + FN; h. PLR: sensitivity/1-specificity; i. NLR: 1-sensitivity/specificity; j. Accuracy: TP + TN/FP + FN + TP + TN.

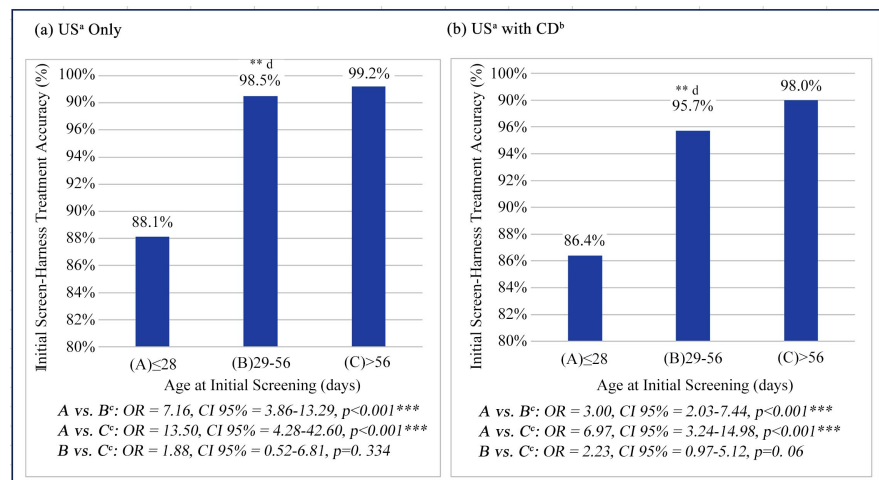
99.7%, respectively. Likewise, the PPV improved from 3.4% (≤28 Days) to 15.4% (29 - 56 Days). NPV was perfect except after 56 days (99.4%). PLR increased from before 28 days (8.37) to between 29 - 56 days (65.27). Screening accuracy followed a similar trend as specificity with significant increases between screening age groups of 87.9%, 98.2%, and 98.9%, respectively. In comparison, when using *US with CD* a similar trend was observed with some key differences. FP rate decreased significantly from 13.6%, to 4.3% and 2.0% as screening age increased. Due to the use of CD in determining follow-up, both cases that screened FN when using *US Only*, were correctly screened due to their risk profiles. However, using CD was a more conservative approach, resulting in an increased FP rate, and decreased TN rate. The TN rate increased between age groups from 85.9%, 95.4%, and 97.5%, while specificity increased with age (86.3%, 95.7% and 98.0%, respectively). Sensitivity and NPV were perfect at all screening ages. Conversely, PPV and PLR both increased as the screening age increased. Accuracy also followed a similar trend increasing with each age group 86.1%, 95.4%, and 97.7%, respectively.

Screening accuracy was further analyzed with a multivariate logistic pairwise regression model between initial screen and concordance with the final treatment outcome stratified by age groups of initial screening when using *US Only* (Figure 2(a)) and when using *US with CD* (Figure 2(b)). When analyzing *US Only*, significant increases in accuracy were detected between ≤28 days of age

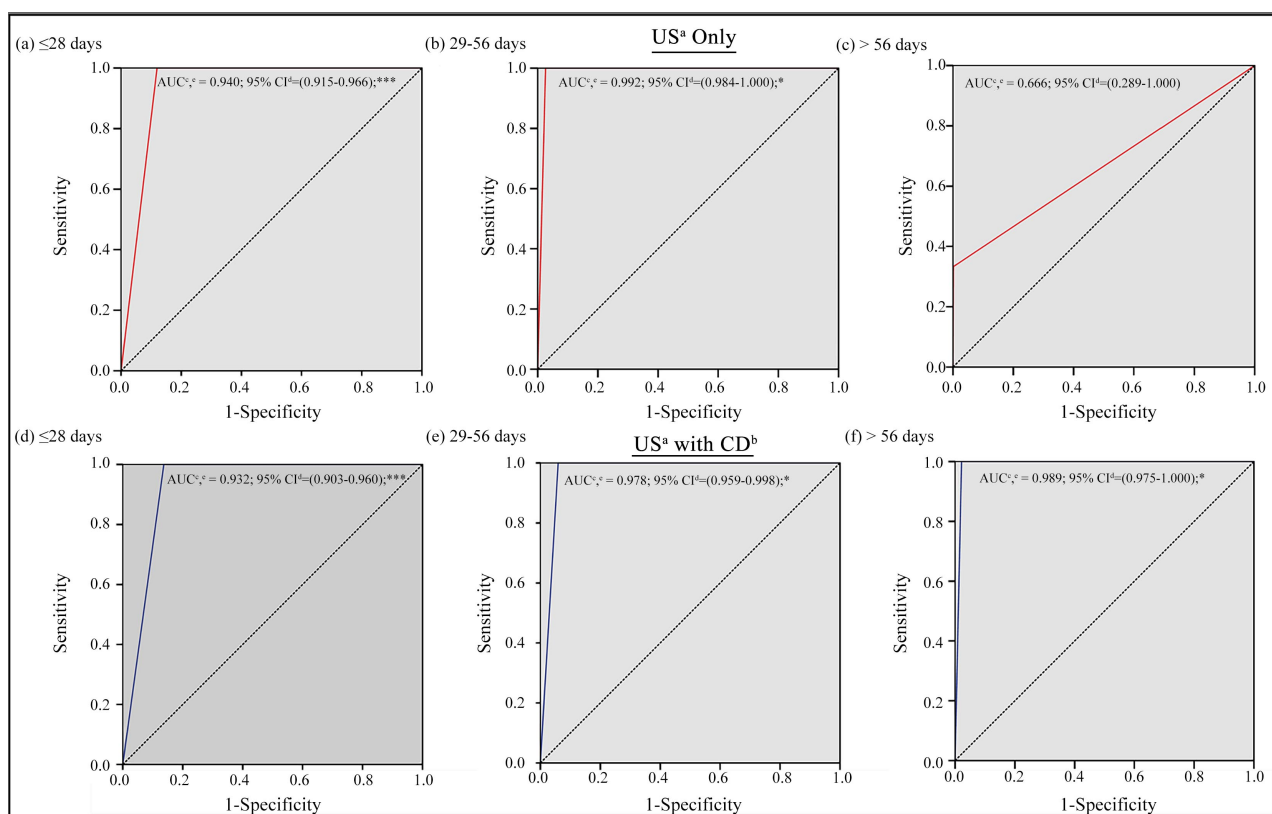


and 29 - 56 days of age groups (88.1% vs. 98.5%; OR = 7.16, 95% CI = 3.86 - 13.29,  $p < 0.001$ ), as well as when comparing screening before 28 days and after 56 days (88.1% vs. 98.9%; OR = 13.50, 95% CI = 4.28 - 42.60,  $p < 0.001$ ). However, when comparing 29 - 56 days and >56 days age groups, increases in accuracy failed to reach significance (OR = 1.88, 95% CI = 0.52 - 6.81,  $p = 0.334$ ). A similar trend was observed when analyzing *US with CD*. When comparing screening before 28 days and screening between 29 - 56 days, a significant increase in accuracy was observed (86.4% vs. 95.7%; OR = 3.00, 95% CI = 2.03 - 4.44,  $p < 0.001$ ). Furthermore, when comparing screening before 28 days and screening after 56 days, accuracy increased significantly (OR = 6.97; 95% CI = 3.24 - 14.98;  $p < 0.001$ ). Similar to when screening with *US Only*, the difference in accuracy failed to reach significance when comparing screening between 29 - 56 days and screening after 56 days (OR = 2.23, 95% CI = 0.97 - 5.12,  $p = 0.06$ ). Protocol screening accuracy was further compared in each screening age group and revealed a significantly higher accuracy at 29 - 56 days ( $z = 2.997$ ,  $p = 0.003$ ) for *US Only* (98.5%) compared to *US with CD* (95.7%). No significant differences were observed between protocols when comparing screening at  $\leq 28$  days or >56 days.

Lastly, results from ROC analysis (Figures 3(a)-(f)) for each screening protocol stratified by age showed a good ability to predict treatment (AUC Range = 0.932 - 0.992), except for when screening with *US Only* after 56 days, which performed poorly (AUC = 0.666, 95% CI = 0.289 - 1.000). Poor performance was attributed to the high false negative rate (FNR) observed in this subgroup (100%) which only relied on ultrasound to determine screening results.



**Figure 2.** (a) (b) Pairwise Logistic regression of association between accuracy of initial screening assignment to predict treatment stratified by initial age at screening. Notes. a. US = ultrasound; b. CD = clinical decision; c. Logistic pairwise regression model controlled for: gender, multiple births, premature births, and screening doctor's background: Sig. (\*\*\*) =  $p < 0.001$ ); d. Z-test of 2 Proportions between US Only and US w/CD protocols: 1)  $\leq 28$  days:  $z = 1.513$ ,  $p = 0.13$ ; 2) 29 - 56 days:  $z = 2.997$ ,  $p = 0.003$ ; 3)  $> 56$  days:  $z = 1.274$ ,  $p = 0.203$ : Sig. (\*\*) =  $p < 0.01$ ).



**Figure 3.** (a)-(f) ROC curves of screening performance for predicting harness treatment by age and screening protocol. Notes. a. US = ultrasound; b. CD = clinical decision; c. AUC = area under the curve; d. 95% CI = 95% confidence interval; e. Receiver Operator Characteristic (ROC) test results controlled for: newborn gender, multiple pregnancies, premature birth and physician type; Sig. (\* $p < 0.05$ , \*\*\* $p < 0.001$ ); f. Screening with US Only after 56 days, resulted in 100% false negative rate, due to the presence of 2 FN and no TP cases; all other screening protocols/timing had a FNR = 0.0%.

## 4. Discussion

Using a US screening protocol with PE and risk factor profiles, led to all treatment cases being accurately detected at first screening. Most cases were identified before 28 days ( $n = 8$ ), with two additional cases screened between 29 - 56 days, and two cases screening after 56 days. Using *US Only* to select cases for follow-up would have resulted in two cases being missed that were correctly identified with the *US with CD* protocol. Despite, accuracy being higher when screening after 28 days compared to screening before 28 days, screening later than 56 days failed to marginally increase accuracy. Although employing an *US Only* protocol resulted in better accuracy in the 29 - 56 day range, *US with CD* was able to correctly identify all cases that required harness treatment upon first screening.

We found that supplementing US findings with risk factor profiles to select cases that required further follow-up, resulted in all 12 cases eventually requiring harness treatment being correctly identified upon first screening visit. Current evidence for selected screening protocol has indicated that cases with breech delivery, family history of DDH or clinical hip instability detected from PE are at risk of developing late DDH [19] [24], on which we based our screening protocol

to guide CD by a physician. However, rather than using risk for targeting US screening, risk screening protocol was used as an additional decision-making tool to select negative screened cases that were at risk of late developing DDH. Risk profiling has been shown to be moderately supported by evidence [27], and has also been promoted as a tool for detecting late developing DDH cases [28]. In our application of risk screening protocol, none of the DDH treatment cases were missed upon first evaluation, which provides support of conservatively screening out cases only if they have a negative Graf classification, have a negative PE, and are without risk factors for DDH. Further research is needed to confirm our findings and to develop a consensus on risk factor profile that could supplement US screening and improve accuracy of detecting cases requiring treatment.

US screens done > 28 days were found to be effective in improving screening accuracy. Past studies have found a similar trend in later screens with better screening performance with delayed screening protocols [16] [26] [29]. Delayed screening has gained support due to the fact that dysplastic hips are likely to resolve naturally without intervention [30]. However, we found that improvements in screening accuracy did not marginally improve when screening after 56 days compared to screening between 29 - 56 days. Our findings overlap with the NIPE and the AAOP suggested screens between 6 - 8 weeks [17] [18]. Roovers *et al.* [12] have suggested that screening between 2 - 3 months may be beneficial for reducing overtreatment when screening with US. Furthermore, treatment for DDH >8 weeks is shown to have poorer screening and treatment outcomes [4] [20], since newborn hips undergo maturing (muscle tightness increases and capsular laxity decreases) in the 8 - 12 weeks of age range [31]. In light of possible risks of delaying screening too much, and of our findings of reduced marginal improvements in screening accuracy after 56 days, screening should only be delayed insofar that accuracy is improved by delayed screening and that the effects on screening and surgery performance are limited.

Implementing risk screening into the follow-up clinical decision making revealed an important difference in screening performance. Screening performance was statistically superior when relying on ultrasound Graf classification findings to decide on cases requiring follow-up. Specificity and accuracy were higher in all age groups when assessed in the *US Only* protocol. Screening with risk profiling led to more screened cases being unnecessarily followed-up at all screening age time points. Sensitivity was found to be higher by Roovers *et al.* [12] and Mace *et al.* [32] Although earlier screens in both screening strategies had lower specificity than Mace *et al.* (99.8% to 99.9%), our delayed screenings (>28 days of age) had a similar performance. It is interesting to note that the mean age for cases from Mace *et al.* [32] was around 1 week; contrasting our earlier findings of higher accuracy at later screening ages. The discrepancy in our screening performance was likely due to relying on a more conservative screening approach; leading to higher number of cases being unnecessarily followed-up, and a decrease in accuracy, but with an ability to accurately detect

upon first evaluation all cases that required harness treatment. Roovers *et al.* [12] also found when relying on delayed universal ultrasound Graf classification screening, missed positive cases still persisted (6/1000, 11.5% of DDH cases). Despite this difference when using *US with CD* protocol, screening accuracy was not severely lower with significant decrease in accuracy observed in the 29 - 56 day screening range, while other age groups were not significantly different. Clinical application is important to consider, since cases requiring surgery are potentially the most costly “inaccurately” screened cases due to subsequent treatment complexity and increased costs [14]. A clear trade-off emerged when implementing CD into US screening decision making, where using CD decreased accuracy but correctly identified all cases requiring harness treatment. Further research with a larger number in false negative cases is needed to confirm our findings in this observed trade-off.

Our study had some limitations. First, the design was observational and selection for screening age was not randomized. Higher risk cases may have presented for earlier screenings than lower risk cases, which may have introduced bias into our sample selection. A further study comparing consecutive cases, receiving repeated, blind measurements at different screening time points may be needed. Second, although our multi-center healthcare units were varied in settings, the sample chosen may not be generalizable to the general population. Additionally, 36 cases were unreachable for follow-up despite receiving an abnormal hip classification on their last screening visit. Most of these cases ( $n = 35$ ) were Type IIa/IIb hips at initial screening. Cases may have chosen to attend another treatment site outside of our network. Thus, DDH treatment rate may have been underestimated in our sample. However, it is more likely that the cases resolved naturally with maturity, since the overall DDH incidence in our sample was 0.4%, similar to the incidence in the general population [3] [11] [33], and the likelihood that a case was present in the 36 lost cases was low (0.8%). Third, we compared a simulated screening protocol, which used the same populations but removed CD from the study protocol. Lastly, although we found significant differences between screening protocols, only 2 cases were missed (after 56 days), which may have been attributable to physician error. The majority of early screens were performed by general pediatric practitioners, previous studies have shown that screening operator experience plays an important role in screening accuracy, screening early leads to more variability in screening performance [5] and higher rates of over-treatment [6]. We controlled for physician experience and found that regardless of operator experience, delayed screening may improve screening performance. Thus, further research with sufficient sample size may be needed to confirm this finding.

With the ongoing debate between targeted and universal ultrasound screening, a universal screening approach which utilizes both delayed screening and risk profiling to support ultrasound assessment was supported by our findings. Two important trade-offs were identified. First, the timing of initial DDH screening should consider both screening performance and potential of impact

on surgical outcomes of screening too late. Second, choosing whether to employ risk factors to aid in screening should consider both screening performance and the cost of missing cases that require harness treatment. Our findings reveal a window for optimal DDH screening in newborns between 4 - 8 weeks (29 to 56 days), while employing risk profiles to guide follow-up decision making. Employing this screening method ensures the optimizing of screening accuracy, while preventing potential harm from late detection, and missed cases requiring harness treatment.

### Acknowledgements

In addition to the authors that contributed to data collection, we would like to thank the following data contributors to the study: Shwu Meei Wang, Hsin Yun-Tsai, Pi-Ying Tai, Wan-Ting Chang, and Chun-Shan Wu. The study was funded by the Taiwanese Institute of Fetal Medicine (Taiwan Registered Non-profit Organization 1080280906).

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Woodacre, T., Ball, T. and Cox, P. (2016) Epidemiology of Developmental Dysplasia of the Hip within the UK: Refining the Risk Factors. *Journal of Children's Orthopaedics*, **10**, 633-642. <https://doi.org/10.1007/s11832-016-0798-5>
- [2] Noordin, S., Umer, M., Hafeez, K. and Nawaz, H. (2010) Developmental Dysplasia of the Hip. *Orthopedic Reviews*, **2**, 73-78. <https://doi.org/10.4081/or.2010.e19>
- [3] Chang, C.-H., Chiang, Y.-T., Lee, Z.-L. and Kuo, K.N. (2007) Incidence of Surgery in Developmental Dysplasia of the Hip in Taiwan. *Journal of the Formosan Medical Association*, **106**, 462-466. [https://doi.org/10.1016/S0929-6646\(09\)60295-3](https://doi.org/10.1016/S0929-6646(09)60295-3)
- [4] Storer, S.K. and Skaggs, D.L. (2006) Developmental Dysplasia of the Hip. *American Family Physician*, **74**, 1310-1316.
- [5] Dogruel, H., Atalar, H., Yavuz, O. and Sayli, U. (2008) Clinical Examination versus Ultrasonography in Detecting Developmental Dysplasia of the Hip. *International Orthopaedics*, **32**, 415-419. <https://doi.org/10.1007/s00264-007-0333-x>
- [6] Shipman, S.A., Helfand, M., Moyer, V.A. and Yawn, B.P. (2006) Screening for Developmental Dysplasia of the Hip: A Systematic Literature Review for the US Preventive Services Task Force. *Pediatrics*, **117**, e557-e576. <https://doi.org/10.1542/peds.2005-1597>
- [7] Woolacott, N.F., Puhan, M.A., Steurer, J. and Kleijnen, J. (2005) Ultrasonography in Screening for Developmental Dysplasia of the Hip in Newborns: Systematic Review. *BMJ*, **330**, 1413-1418. <https://doi.org/10.1136/bmj.38450.646088.E0>
- [8] Godward, S. and Dezateux, C. (1998) Surgery for Congenital Dislocation of the Hip in the UK as a Measure of Outcome of Screening. *The Lancet*, **351**, 1149-1152. [https://doi.org/10.1016/S0140-6736\(97\)10466-4](https://doi.org/10.1016/S0140-6736(97)10466-4)
- [9] Paton, R., Hinduja, K. and Thomas, C. (2005) The Significance of At-Risk Factors in Ultrasound Surveillance of Developmental Dysplasia of the Hip: A Ten-Year Pros-

- pective Study. *The Journal of Bone and Joint Surgery. British Volume*, **87**, 1264-1266. <https://doi.org/10.1302/0301-620X.87B9.16565>
- [10] Schams, M., Labruyère, R., Zuse, A. and Walensi, M. (2017) Diagnosing Developmental Dysplasia of the Hip Using the Graf Ultrasound Method: Risk and Protective Factor Analysis in 11,820 Universally Screened Newborns. *European Journal of Pediatrics*, **176**, 1193-1200. <https://doi.org/10.1007/s00431-017-2959-z>
- [11] Tong, S.H., Eid, M.A., Chow, W. and To, M.K. (2011) Screening for Developmental Dysplasia of the Hip in Hong Kong. *Journal of Orthopaedic Surgery (Hong Kong)*, **19**, 200-203. <https://doi.org/10.1177/230949901101900214>
- [12] Roovers, E., Boere-Boonekamp, M.M., Castelein, R., Zielhuis, G. and Kerkhoff, T. (2005) Effectiveness of Ultrasound Screening for Developmental Dysplasia of the Hip. *Archives of Disease in Childhood*, **90**, F25-F30.
- [13] Thaler, M., Biedermann, R., Lair, J., Krismer, M. and Landauer, F. (2011) Cost-Effectiveness of Universal Ultrasound Screening Compared with Clinical Examination Alone in the Diagnosis and Treatment of Neonatal Hip Dysplasia in Austria. *The Journal of Bone and Joint Surgery. British Volume*, **93**, 1126-1130. <https://doi.org/10.1302/0301-620X.93B8.25935>
- [14] Woodacre, T., Dhadwal, A., Ball, T., Edwards, C. and Cox, P. (2014) The Costs of Late Detection of Developmental Dysplasia of the Hip. *Journal of Children's Orthopaedics*, **8**, 325-332. <https://doi.org/10.1007/s11832-014-0599-7>
- [15] von Kries, R., Ihme, N., Oberle, D., Lorani, A., Stark, R., Altenhofen, L., et al. (2003) Effect of Ultrasound Screening on the Rate of First Operative Procedures for Developmental Hip Dysplasia in Germany. *The Lancet*, **362**, 1883-1887. [https://doi.org/10.1016/S0140-6736\(03\)14957-4](https://doi.org/10.1016/S0140-6736(03)14957-4)
- [16] Chen, H.-W., Chang, C.-H., Tsai, S.-T., Liu, W.-J., Chua, C., Chen, Y.-Y., et al. (2010) Natural Progression of Hip Dysplasia in Newborns: A Reflection of Hip Ultrasonographic Screenings in Newborn Nurseries. *Journal of Pediatric Orthopaedics B*, **19**, 418-423. <https://doi.org/10.1097/BPB.0b013e328339ecff>
- [17] Public Health England (2016) Newborn and Infant Physical Examination Screening Programme Handbook 2016/17. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/524425/NIPE\\_Programme\\_Handbook\\_2016\\_to\\_2017\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524425/NIPE_Programme_Handbook_2016_to_2017_.pdf)
- [18] American Academy of Orthopaedic Surgeons (2014) Detection and Nonoperative Management of Pediatric Developmental Dysplasia of the Hip in Infants Up to Six Months of Age. <https://pediatrics.aappublications.org/content/136/4/e1141>
- [19] Mulpuri, K., Schaeffer, E.K., Andrade, J., Sankar, W.N., Williams, N., Matheney, T.H., et al. (2016) What Risk Factors and Characteristics Are Associated with Late-Presenting Dislocations of the Hip in Infants? *Clinical Orthopaedics and Related Research*, **474**, 1131-1137. <https://doi.org/10.1007/s11999-015-4668-0>
- [20] Kubo, H., Pilge, H., Holthoff, J.-P., Hufeland, M., Westhoff, B. and Krauspe, R. (2018) Poorer Radiological Outcome after Delayed Diagnosis and Treatment in Human Position in Fettweis Plaster Cast in 93 Unstable Hip Joints Type D, III and IV According to Graf. *Journal of Children's Orthopaedics*, **12**, 1-9.
- [21] Elbourne, D., Dezateux, C., Arthur, R., Clarke, N., Gray, A., King, A., et al. (2002) Ultrasonography in the Diagnosis and Management of Developmental Hip Dysplasia (UK Hip Trial): Clinical and Economic Results of a Multicentre Randomised Controlled Trial. *The Lancet*, **360**, 2009-2017. [https://doi.org/10.1016/S0140-6736\(02\)12024-1](https://doi.org/10.1016/S0140-6736(02)12024-1)
- [22] Graf, R. (1987) The Ultrasound Examination of the Hip. In: *Congenital Dysplasia*



*and Dislocation of the Hip in Children and Adults*, Springer, Berlin, 172-211.

[https://doi.org/10.1007/978-3-642-71038-4\\_14](https://doi.org/10.1007/978-3-642-71038-4_14)

- [23] Graf, R., Scott, S., et al. (2017) Essentials of Infant Hip Sonography. <http://www.graf-hipsonography.com/hipsonography.html>
- [24] Ortiz-Neira, C.L., Paolucci, E.O. and Donnon, T. (2012) A Meta-Analysis of Common Risk Factors Associated with the Diagnosis of Developmental Dysplasia of the Hip in Newborns. *European Journal of Radiology*, **81**, e344-e351. <https://doi.org/10.1016/j.ejrad.2011.11.003>
- [25] Schwend, R.M., Shaw, B.A. and Segal, L.S. (2014) Evaluation and Treatment of Developmental Hip Dysplasia in the Newborn and Infant. *Pediatric Clinics of North America*, **61**, 1095-1107. <https://doi.org/10.1016/j.pcl.2014.08.008>
- [26] Lussier, E.C., Sun, Y.-T., Chen, H.-W., Chang, T.-Y. and Chang, C.-H. (2018) Ultrasound Screening for Developmental Dysplasia of the Hip after 4 Weeks Increases Exam Accuracy and Decreases Follow-Up Visits. *Pediatrics and Neonatology*, **60**, 270-277. <https://doi.org/10.1016/j.pedneo.2018.07.008>
- [27] Cooper, A.P., Doddabasappa, S.N. and Mulpuri, K. (2014) Evidence-Based Management of Developmental Dysplasia of the Hip. *Orthopedic Clinics of North America*, **45**, 341-354. <https://doi.org/10.1016/j.ocl.2014.03.005>
- [28] Hassan, F.O.A. and Shannak, A. (2007) Associated Risk Factors in Children Who Had Late Presentation of Developmental Dysplasia of the Hip. *Journal of Children's Orthopaedics*, **1**, 205-210. <https://doi.org/10.1007/s11832-007-0041-5>
- [29] Rawlings, E., Burnett, M. and Reddan, T. (2017) An Audit of Referral Timeframes for Ultrasound Screening of Developmental Hip Dysplasia (DDH) in Neonates with a Normal Clinical Examination. *Sonography*, **4**, 1-7.
- [30] Lowry, C., Donoghue, V. and Murphy, J. (2005) Auditing Hip Ultrasound Screening of Infants at Increased Risk of Developmental Dysplasia of the Hip. *Archives of Disease in Childhood*, **90**, 579-581. <https://doi.org/10.1136/adc.2003.033597>
- [31] Collins-Sawaragi, Y.C. and Jain, K. (2018) How to Use ... Hip Examination and Ultrasound in Newborns. *Archives of Disease in Childhood*, **103**, 34-40. <https://doi.org/10.1136/archdischild-2014-307942>
- [32] Mace, J. and Paton, R. (2015) Neonatal Clinical Screening of the Hip in the Diagnosis of Developmental Dysplasia of the Hip. *The Bone & Joint Journal*, **97**, 265-269. <https://doi.org/10.1302/0301-620X.97B2.34858>
- [33] Chang, C.H., Chiang, Y.-T., Chen, L. and Kuo, K.N. (2018) The Influence of Health Policy on Early Diagnosis and Surgical Incidence of Developmental Dysplasia of the Hip. *PLoS ONE*, **13**, e0200995. <https://doi.org/10.1371/journal.pone.0200995>