

Clinical Risk Score for Central Precocious Puberty Among Girls with Precocious Pubertal Development: A Cross Sectional Study

Jingyu You

Department of Pediatrics Endocrinology and Inborn Metabolic Diseases, Children's Hospital, Fudan University

Xianying Cheng

Children's Hospital of Fudan University

Xiaoqing Li

Department of Pediatric Endocrinology and Inborn Metabolic Diseases, Children's Hospital, Fudan University

Mingqin Li

Obstetrics and Gynecology Hospital, Fudan University

Li Yao

Obstetrics and Gynecology Hospital, Fudan University

Feihong Luo

Department of Pediatric Endocrinology and Inborn Metabolic Diseases, Children's Hospital, Fudan University

Ruoqian Cheng

Department of Pediatric Endocrinology and Inborn Metabolic Diseases, Children's Hospital, Fudan University

Li Xi

Department of Pediatric Endocrinology and Inborn Metabolic Diseases, Children's Hospital, Fudan University

Jiangfeng Ye (✉ jiangfeng_ye@hotmail.com)

KK Women's and Children's Hospital <https://orcid.org/0000-0001-6963-2450>

Research article

Keywords: central precocious puberty, GnRH stimulation test, risk score model

Posted Date: October 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-74181/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Endocrine Disorders on April 20th, 2021. See the published version at <https://doi.org/10.1186/s12902-021-00740-7>.

Abstract

Background The gold standard for the diagnosis of central precocious puberty (CPP) is gonadotropin-releasing hormone (GnRH) or GnRH analogs (GnRHa) stimulation test. But the stimulation test is time-consuming and costly. Our objective was to develop a risk score model with readily available features.

Methods A cross sectional study based on the electronic medical record system including 627 girls with precocious puberty were conducted in the Children's Hospital, Fudan University, Shanghai, China from January 2010 to August 2016. Patients were randomly split into the training (n=314) and validation (n=313) sample. In the training sample, variables associated with CPP ($P<0.2$) in univariate analyses were introduced in a multivariable logistic regression model and selected using a forward stepwise analysis. A risk score model was built with the scaled coefficients of the model and tested in the validation sample.

Results CPP was diagnosed in 54.8% (172/314) and 55.0% (172/313) of patients in the training and validation sample respectively. The CPP risk score model included variables of age at onset of puberty, basal luteinizing hormone (LH) concentration, largest ovarian volume, and uterine volume. The C-index was 0.85 (95% CI: 0.81-0.89) for the training sample and 0.86 (95% CI: 0.82-0.90) for the validation sample. Two cut-off points were selected to delimitate a low- (<10 points), median- (10-19 points), and high-risk (≥ 20 points) group.

Conclusions A risk score model developed among girls with precocious pubertal development had moderate discrimination to stratify CPP risk, which could help make decisions on the need for GnRH (GnRHa) stimulation test.

Background

Precocious puberty, defined as the onset of pubertal development before age 8 years in girls and 9 years in boys [1], has a prevalence of 0.43% in China and 0.01%-0.02% in America girls [2, 3]. The early onset of puberty may impair children's normal physical and psychosocial development [4–6]. However, only cases of central precocious puberty (CPP) may need a gonadotropin-releasing hormone analogs (GnRHa) therapy [1]. Pubertal development with no activation of the hypothalamic-pituitary-gonadal axis (HPGA), i.e. peripheral precocious puberty (PPP), will regress or stop progressing without treatment, which accounted for at least 50% of cases of precocious puberty. In addition, with increased awareness of the importance of early treatment of CPP, more females with subtle signs of precocious puberty were evaluated [7]. Therefore, to distinguish CPP from PPP and benign variants of sexual precocity is of great importance.

The gold standard for diagnosis of CPP is gonadotropin-releasing hormone (GnRH) or GnRHa stimulation test [1, 7]. But the stimulation test is time-consuming and costly [8]. To avoid the testing of the stimulated luteinizing hormone (LH) and follicle-stimulating hormone (FSH) concentration, baseline LH has been suggested to be used for diagnosis [9]. However, its generalization is limited by variability among studies

and the small sample size of previous studies [8–13]. Pelvic ultrasonography as a part of the initial diagnostic evaluation of CPP is convenient [14–16]. But ovarian and uterine volume has a substantial overlap among girls in prepubertal and pubertal stage [7]. In addition, enlarged ovaries and uterus volumes are end-organ effects of gonadotropin stimulation, which suggested that it was a highly specific but less sensitive indicator for CPP [16].

The objective of this study was to develop and validate a risk score model to stratify the risk of CPP based on readily available clinical features and pelvic ultrasonography. The risk score could be used for making decisions on the need for further GnRH (GnRHa) stimulation test.

Methods

Study Population

We performed a cross-sectional study based on the electronic medical record system (EMRS) in the Children's Hospital, Fudan University, Shanghai, China. The EMRS systematically collected information on patient's demographics, medical history, results of physical examination and laboratory test, radiology images, diagnosis, and treatment for each time they visited the hospital. The study population was extract from the database including all patients who came to the hospital from January 2010 to August 2016. We included patients according to inclusion criteria as follow: (1) girls with a diagnosis of precocious puberty; (2) age of 8 years old or less on her first visit the hospital; (3) hormone profile (including GnRHa stimulation test) and pelvic ultrasonography was performed in the Children's Hospital, Fudan University; (4) pelvic ultrasonography was performed within one week of the GnRHa stimulation test. Patients with secondary precocious puberty, central nervous system lesion (congenital or acquired), and ovarian cyst were excluded for its possible effects on the HPGA.

This study was approved by the Ethics Committees of the Children's Hospital, Fudan University, Shanghai, China.

Gold standard

GnRHa stimulation test was used as the gold standard diagnosis of CPP [1, 7]. Patients with stimulated peak LH \geq 5 IU/L, and peak LH-to-FSH ratio \geq 0.6 were diagnosed as CPP [1, 8–10, 17]. Details of the GnRHa stimulation test have been published elsewhere [18]. LH and FSH concentration was measured using electrochemiluminescence assay (COBASE 602, Roche, Switzerland). The limit of detection (LOD) of LH and FSH was 0.2 IU/L. Stimulated LH, basal and stimulated FSH concentration was above the LOD in all participants. Basal LH level was below the LOD in 1.8% (11/627) patients [2.5% (8/314) and 1.0% (3/313) in the training and validation sample, respectively].

Pelvic Ultrasound Evaluation

Transabdominal ultrasonography was performed on patients utilizing a curvilinear 2–7 MHz probe. All pelvic ultrasonograms were obtained with Philips IU22 ultrasound units equipped with duplex/color-flow Doppler broad bandwidth transducers (Phillips, Netherlands). The pediatric radiologist had no information on the results of the GnRHa stimulation test. Ovarian volume for each side was calculated based on the ellipse volume formula: $0.5233 \times \text{length} \times \text{depth} \times \text{breadth}$. Average ovarian volume was calculated as: $(\text{right ovary volume} + \text{left ovary volume}) / 2$. The largest and smallest ovarian volume was defined as the larger and smaller volume between the right and left ovary volume. Uterine volume was calculated according to the same ellipse volume formula. The values of sonographic characteristics were stratified into categories (ovarian volume: <1 mL, 1 – <2 mL, and ≥ 2 mL; uterine length: <3 cm, 3 – <4 cm, and ≥ 4 cm; uterine volume: <3 mL, 3 – <4 mL, and ≥ 4 mL; uterine configuration with the thickness of endometrial stripe: <0.2 cm and ≥ 0.2 cm) [16].

Medical history, physical examination and bone age

A complete medical history and results of the physical examination were abstracted from the database. Breast and pubic hair development was assessed according to the Tanner staging criteria [1]. The bone age (BA) was measured using the Greulich Pyle (GP) method [19].

Statistical Analysis

A random sample of one half of the patients was obtained to develop a clinical prediction model (training sample), leaving the other half of the patients for validation (validation sample). We first compared the clinical characteristics and pelvic ultrasonography between the training and validation sample using a quantitative (t test or Wilcoxon rank sum test) or qualitative (χ^2 test) test as appropriate. Then we built crude logistic regression models to evaluate the association between potential predictors and CPP. A total of 30 variables contained information on medical history, progression of pubertal manifestations, basal hormone level, and pelvic ultrasonography were selected a priori according to previous studies (See Additional file 1 [Additional Table 1]) [1, 7, 16]. Predictors with a P value of less than 0.20 were entered into a multivariable logistic regression model. The prediction model was selected using forward stepwise analysis ($P = 0.05$ include, $P > 0.10$ removed). Performance of the selected model was assessed using C-index, calibration based on Hosmer-Lemeshow test [20]. We performed the internal validation using bootstrap resampling [21].

A risk score model based on the final logistic regression model was derived using the method proposed by Sullivan et al [22]. In the risk score system, estimate risks calculated based on point totals were approximate to the prediction of the logistic regression model. The statistical methods are described in more detailed in Additional file 2. The performance of risk score was measured using C-index, calibration, sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR-) [20]. High- and low-risk cut points for the CPP risk score were determined by consensus of a team of two experienced pediatric endocrinologists, two pediatric radiologists, and an epidemiologist. Validation was performed in

the other half of the patients. Performance of the CPP risk score model in the validation sample was measured as well [20].

Statistical analyses were performed using SAS statistical software version 9.2 (SAS Institute Inc., Cary, NC, USA).

Results

Patient characteristics

A total of 735 patients met the inclusion criteria. Patients with pineal cyst (n = 34), Rathke's cleft cyst (n = 33), McCune Albright syndrome (n = 12), congenital adrenal hyperplasia (n = 10), and ovarian cyst (n = 19) were excluded. Finally, 627 patients were included and randomly separated into the training sample (n = 314) and validation sample (n = 313) (Fig. 1).

The mean age of the participants was 7.5 years [95% confidence interval (CI), 7.4–7.7 years]. The average disease duration was 1.0 years (95% CI, 1.0-1.2 years). CPP was diagnosed in 54.8% (172/314) and 55.0% (172/313) of patients in the training and validation sample respectively. Patients did not show significant difference of clinical and pelvic ultrasonography characteristics in the training and validation sample with an exception of the proportion of patients with a family history of CPP. Detailed description was shown in Table 1.

Table 1
Clinical and ultrasonography characteristics of patients with premature sexual development in training and validation samples

	Training (n=314)	Validation (n=313)	P Value
Central precocious puberty (%)	172 (54.8)	172 (55.0)	0.9649
Clinical characteristics			
Age at onset of puberty [mean (SD), year]	6.5 (1.6)	6.5 (1.6)	0.8132
Chronological age [mean (SD), year]	7.5 (1.6)	7.6 (1.7)	0.9460
Bone age [mean (SD), year]	9.6 (7.5)	8.9 (2.3)	0.1415
Bone age/ Chronological age (SD)	1.2 (0.2)	1.2 (0.2)	0.8146
Duration of disease [mean (SD), year]	1.1 (0.9)	1.0 (0.8)	0.7547
Family history of CPP (%)	5 (1.6)	0 (0.0)	0.0250
Tanner stage for breast development			
Left (%)			0.5959 ^a
I	9 (2.9)	9 (2.9)	
II	200 (64.5)	211 (68.5)	
III	98 (31.6)	87 (28.3)	
IV	3 (1.0)	1 (0.3)	
Right (%)			0.4052 ^a
I	12 (3.9)	8 (2.6)	
II	196 (63.2)	211 (68.5)	
III	99 (31.9)	88 (28.6)	
IV	3 (1.0)	1 (0.3)	
Tanner stage for pubic hair development (%)			
I	277 (88.2)	282 (90.1)	0.1860 ^a
II	32 (10.2)	31 (9.9)	
III	4 (1.3)	0 (0.0)	
IV	1 (0.3)	0 (0.0)	
Height [mean (SD), m]	130.0 (11.6)	129.9 (12.6)	0.8986
Weight [mean (SD), kg]	28.4 (6.8)	28.8 (7.3)	0.5405

BMI [mean (SD), kg/m ²]	16.7 (2.2)	16.9 (2.4)	0.3691
LH [Median (IRQ), IU/L]			
Baseline	0.43 (0.17, 1.07)	0.43 (0.18, 1.02)	0.6364
Stimulated	10.45 (5.01, 21.36)	9.55 (4.80, 24.97)	0.9503
FSH [mean (SD), IU/L]			
Baseline	3.7 (2.3)	3.6 (2.2)	0.6839
Stimulated	16.7 (8.0)	17.0 (7.8)	0.6182
LH/FSH [Median(IRQ)]			
Baseline	0.14 (0.08, 0.29)	0.14 (0.07, 0.30)	0.5607
Stimulated	0.75 (0.38, 1.46)	0.72 (0.32, 1.45)	0.7394
Estradiol [Median (IRQ), pg/mL]	15.0 (8.0, 29.5)	14.0 (7.0, 26.8)	0.4737
HCG [Median (IRQ), IU/L]	0.08 (0.00, 0.23)	0.05 (0.00, 0.21)	0.4128
Prolactin [mean (SD), ng/mL]	10.7 (7.7)	9.7 (5.8)	0.1042
DHS [mean (SD), µg/dL]	53.3 (38.3)	53.5 (39.6)	0.9594
Testosterone [Median (IRQ), ng/dL]	2.4 (0.0, 16.3)	0.0 (0.0, 15.1)	0.3793
Cortisol [mean (SD), µg/dL]	8.3 (4.9)	8.3 (4.7)	0.9896
ACTH [Median (IRQ), pg/mL]	24.2 (17.8, 33.0)	24.5 (18.0, 34.0)	0.7214
Total triiodothyronine [mean (SD), ng/dL]	139.8 (24.3)	136.5 (25.1)	0.0982
Free triiodothyronine [mean (SD), pg/mL]	3.9 (0.6)	3.8 (0.6)	0.8073
Total thyroxine [mean (SD), µg/dL]	9.0 (1.8)	9.1 (1.8)	0.3588
Free thyroxine [mean (SD), ng/dL]	1.0 (0.2)	1.0 (0.2)	0.6164
TSH [mean (SD), µIU/mL]	2.4 (1.4)	2.2 (1.2)	0.1302
Pelvic sonogram			
Ovarian volume			
Average ovarian (%)	1.9 (0.9)	1.9 (0.9)	0.6001
<1mL	40 (12.7)	44 (14.1)	0.8866
1-<2mL	157 (50.0)	155 (49.5)	
≥2mL	117 (37.3)	114 (36.4)	
Largest ovarian (%)			

<1mL	31 (9.9)	34 (10.9)	0.8914
1-<2mL	142 (45.2)	137 (43.8)	
≥2mL	141 (44.9)	142 (45.4)	
Smallest ovarian (%)			
<1mL	67 (21.3)	73 (23.3)	0.8138
1-<2mL	160 (51.0)	153 (48.9)	
≥2mL	87 (27.7)	87 (27.8)	
Uterine			
Length			
<3cm	278 (88.5)	287 (91.7)	0.1774 ^a
3-<4cm	36 (11.5)	25 (8.0)	
≥ 4cm	0 (0.0)	1 (0.3)	
Volume			
<3mL	244 (77.7)	251 (80.2)	0.7287
3-<4mL	42 (13.4)	36 (11.5)	
≥4mL	28 (8.9)	26 (8.3)	
Endometrium visible (%)	40 (12.7)	40 (12.8)	0.9878
HCG: human chorionic gonadotropin; DHS: dehydroepiandrosterone sulfate; ACTH: adrenocorticotrophic hormone; TSH: thyroid - stimulating hormone.			
^a Calculated using Fisher's exact test			

Training sample

The crude relationship between potential predictors and the diagnosis of CPP was showed in Additional file 1 (Additional Table 1). A total of 21 variables with a *P* value of less than 0.20 were entered into the multiple variable logistic regression model. After a forward stepwise selection, a final model including four predictors (age at onset of puberty, basal LH, largest ovarian volume, and uterine volume) was selected (Additional file 1 [Additional Table 2]). The variance inflation factor was less than 2.0 for all predictors, indicating there was no linear relationship between predictors. The C-index was 0.86 [95% CI, 0.82–0.90; Fig. 2]. Hosmer-Lemeshow test demonstrated goodness of fit for the prediction model (*P*=

0.49). The calibration plot showed an intercept of -0.01, and a slope of 1.01 (Additional file 1 [Additional Fig. 1]). A bootstrap analysis (resampling the model 300 times) showed a corrected C-index of 0.86.

Table 2
Determine points associated with each of the categories of each risk factor

Predictors	Categories	Reference value (W_{ij})	β_i	$\beta_i (W_{ij} / W_{iREF})$	Points _{ij} = $\beta_i (W_{ij} / W_{iREF}) / B^a$
Age at onset of puberty (years)	< 1	0.5 = W_{1REF}	0.45	0	0
	1 - < 2	1.5		0.45	1
	2 - < 3	2.5		0.90	3
	3 - < 4	3.5		1.35	4
	4 - < 5	4.5		1.80	6
	5 - < 6	5.5		2.25	7
	6 - < 7	6.5		2.70	8
	7 - < 8	7.5		3.15	10
Basal LH (IU/L)	< 0.2	0.1 = W_{2REF}	1.63	0	0
	0.2 - < 0.4	0.3		0.326	1
	0.4 - < 0.6	0.5		0.652	2
	0.6 - < 0.8	0.7		0.978	3
	0.8 - < 1.0	0.9		1.304	4
	1.0 - < 1.2	1.1		1.630	5
	1.2 - < 1.4	1.3		1.956	6
	1.4 - < 1.6	1.5		2.282	7
	1.6 - < 1.8	1.7		2.608	8
	1.8 - < 2.0	1.9		2.934	9
	2.0 - < 2.2	2.1		3.260	10
	2.2 - < 2.4	2.3		3.586	11
2.4 - < 2.6	2.5		3.912	12	

^a We define the constant B for the points system (the number of regression units that will correspond to one point) as the increase in risk of CPP associated with a 0.2 (IU/L) increase in basal LH: $B = 0.2 * 1.63 = 0.326$

Points associated with each category of each risk factor are computed by: $Points_{ij} = \beta_i (W_{ij} / W_{iREF}) / B$ and rounded to the nearest integer.

Predictors	Categories	Reference value (W_{ij})	β_i	$\beta_i (W_{ij} - W_{iREF})$	Points _{ij} = $\beta_i (W_{ij} - W_{iREF}) / B^a$
	≥ 2.6	3.0		4.727	14
Largest ovarian volume	1	$1 = W_{3REF}$	0.66	0	0
	2	2		0.66	2
	3	3		1.32	4
Uterine volume	1	$1 = W_{4REF}$	0.85	0	0
	2	2		0.85	3
	3	3		1.70	5
^a We define the constant B for the points system (the number of regression units that will correspond to one point) as the increase in risk of CPP associated with a 0.2 (IU/L) increase in basal LH: $B = 0.2 * 1.63 = 0.326$					
Points associated with each category of each risk factor are computed by: $\text{Points}_{ij} = \beta_i (W_{ij} - W_{iREF}) / B$ and rounded to the nearest integer.					

Points were assigned to each category for each predictor (Table 2). The total risk score with a range of 0 to 33 linearly correlated with the risk estimate of CPP ($r = 0.96$, $P < 0.0001$, Table 3). The proportion of patients diagnosed as CPP for each risk score value was showed in Table 3. C-index for the risk score system was 0.85 (95% CI, 0.81–0.89; Fig. 2). Calibration plot showed an intercept of -0.02, and a slope of 1.02 (Additional file 1 [Additional Fig. 1]). Two cutoff points were selected to stratify CPP risk (low risk: < 10 points; medium risk: 10–19 points; high risk: ≥ 20 points). The proportion of CPP patient in the low-, medium-, and high-risk population was 10% (4/40), 49.8% (100/201), and 93.2% (68/73), respectively. For low-risk population (cutoff point = 10), the sensitivity was 97.8% (95% CI, 95.3% – 99.5%), specificity was 24.8% (95% CI, 18.2% – 32.6%), the LR- was 0.09 (95% CI, 0.02–0.20), and negative predictive value was 90.2% (95% CI, 79.2% – 97.7%). For high-risk population (cutoff point = 20), the specificity was 96.6% (95% CI, 92.9% – 99.2%), sensitivity was 39.6% (95% CI, 32.0% – 46.2%), the LR + was 12.0 (95% CI, 5.49–48.9), and the positive predictive value was 93.3% (95% CI, 86.5% – 98.4%; Table 4).

Table 3

Points system and associated risks for central precocious puberty in the training and validation sample

CPP risk category	Points total	Estimate of risk (95% CI) <i>a</i>	No. with CPP/ Total No. of patients in training sample (%)	No. with CPP/ Total No. of patients in validation sample (%)
Low risk	0	0.01 (0.01, 0.03)	0/0 (-)	0/1 (0.0)
	1	0.01 (0.01, 0.04)	0/1 (0.0)	0/3 (0.0)
	2	0.02 (0.01, 0.05)	0/3 (0.0)	0/0 (-)
	3	0.03 (0.01, 0.07)	0/4 (0.0)	0/6 (0.0)
	4	0.04 (0.02, 0.08)	0/2 (0.0)	0/5 (0.0)
	5	0.05 (0.02, 0.10)	1/3 (33.3)	0/1 (0.0)
	6	0.07 (0.04, 0.13)	0/6 (0.0)	0/3 (0.0)
	7	0.09 (0.05, 0.16)	0/3 (0.0)	0/3 (0.0)
	8	0.12 (0.08, 0.20)	1/9 (12.5)	0/6 (0.0)
	9	0.16 (0.11, 0.24)	2/9 (22.2)	0/11 (0.0)
Medium risk	10	0.21 (0.15, 0.29)	4/19 (21.1)	3/19 (15.8)
	11	0.27 (0.21, 0.35)	5/21 (23.8)	10/19 (52.6)
	12	0.34 (0.27, 0.42)	7/28 (25.0)	13/37 (35.1)
	13	0.42 (0.35, 0.49)	10/24 (41.7)	7/19 (36.8)

$$^a \hat{p} = \frac{1}{1 + \exp(-\sum_{i=0}^P \beta_i X_i)}, \quad \sum_{i=0}^P \beta_i X_i = -6.42 + 0.45 * 0.5 + 1.63 * 0.1 + 0.66 * 1 + 0.85 * 1 + B * (\text{Point total})$$

We define the constant B for the points system (the number of regression units that will correspond to one point) as the increase in risk of CPP associated with a 0.2 (IU/L) increase in basal LH: $B = 0.2 * 1.63 = 0.326$

CPP risk category	Points total	Estimate of risk (95% CI) <i>a</i>	No. with CPP/ Total No. of patients in training sample (%)	No. with CPP/ Total No. of patients in validation sample (%)
	14	0.50 (0.43, 0.57)	16/29 (55.2)	17/35 (48.6)
	15	0.58 (0.51, 0.65)	14/25 (56.0)	17/32 (53.1)
	16	0.66 (0.58, 0.72)	15/19 (79.0)	14/14 (100.0)
	17	0.72 (0.65, 0.79)	9/11 (81.8)	14/17 (82.4)
	18	0.78 (0.70, 0.85)	13/15 (86.7)	5/6 (83.3)
	19	0.83 (0.76, 0.89)	7/10 (70.0)	12/12 (100.0)
High risk	20	0.87 (0.80, 0.92)	9/10 (90.0)	6/6 (100.0)
	21	0.91 (0.84, 0.95)	8/9 (88.9)	7/8 (87.5)
	22	0.93 (0.87, 0.96)	3/3 (100.0)	5/6 (83.3)
	23	0.95 (0.90, 0.98)	6/8 (75.0)	4/5 (80.0)
	24	0.96(0.92, 0.98)	10/11 (90.9)	6/7 (85.7)
	25	0.97 (0.93, 0.99)	5/5 (100.0)	3/3 (100.0)
	26	0.98 (0.95, 0.99)	8/8 (100.0)	5/5 (100.0)
	27	0.99 (0.96, 0.99)	2/2 (100.0)	4/4 (100.0)
	28	0.99 (0.97, 1.00)	4/4 (100.0)	3/3 (100.0)

$$a \hat{p} = \frac{1}{1 + \exp(-\sum_{i=0}^P \beta_i X_i)}, \quad \sum_{i=0}^P \beta_i X_i = -6.42 + 0.45 * 0.5 + 1.63 * 0.1 + 0.66 * 1 + 0.85 * 1 + B * (\text{Point total})$$

We define the constant B for the points system (the number of regression units that will correspond to one point) as the increase in risk of CPP associated with a 0.2 (IU/L) increase in basal LH: B = 0.2*1.63 = 0.326

CPP risk category	Points total	Estimate of risk (95% CI) <i>a</i>	No. with CPP/ Total No. of patients in training sample (%)	No. with CPP/ Total No. of patients in validation sample (%)
	29	0.99 (0.97, 1.00)	3/3 (100.0)	1/1 (100.0)
	30	0.99 (0.98, 1.00)	2/2 (100.0)	1/1 (100.0)
	31	1.00 (0.98, 1.00)	6/6 (100.0)	7/7 (100.0)
	32	1.00 (0.99, 1.00)	0/0 (-)	0/0 (-)
	33	1.00 (0.99, 1.00)	2/2 (100.0)	8/8 (100.0)
$a \hat{p} = \frac{1}{1 + \exp(-\sum_{i=0}^p \beta_i X_i)}, \quad \sum_{i=0}^p \beta_i X_i = -6.42 + 0.45 * 0.5 + 1.63 * 0.1 + 0.66 * 1 + 0.85 * 1 + B * (\text{Point total})$				
<p>We define the constant B for the points system (the number of regression units that will correspond to one point) as the increase in risk of CPP associated with a 0.2 (IU/L) increase in basal LH: B = 0.2 * 1.63 = 0.326</p>				

Table 4
 Predictive ability of the risk score system for CPP in training and validation samples

	Training Sample (n = 314)	Validation Sample (n = 313)
AUC (95%CI)	0.85 (0.81, 0.89)	0.86 (0.82, 0.90)
Calibration	a=-0.02, b = 1.02	a=-0.02, b = 1.06
Cutoff point = 10		
Sensitivity (% , 95%CI)	97.8 (95.3, 99.5)	100.0 (-)
Specificity (% , 95%CI)	24.8 (18.2, 32.6)	27.7 (20.2, 34.9)
Positive likelihood ratio (95%CI)	1.30 (1.20, 1.46)	1.38 (1.25, 1.54)
Negative likelihood ratio (95%CI)	0.09 (0.02, 0.20)	0.0 (-)
Positive predictive value (% , 95%CI)	61.2 (55.9, 67.0)	62.6 (56.8, 68.6)
Negative predictive value (% , 95%CI)	90.2 (79.2, 97.7)	100.0 (-)
Cutoff point = 20		
Sensitivity (% , 95%CI)	39.6 (32.0, 46.2)	34.8 (27.8, 42.1)
Specificity (% , 95%CI)	96.6 (92.9, 99.2)	97.3 (94.2, 100.0)
Positive likelihood ratio (95%CI)	12.0 (5.49, 48.9)	12.6 (5.44, 48.8)
Negative likelihood ratio (95%CI)	0.63 (0.55, 0.71)	0.67 (0.59, 0.75)
Positive predictive value (% , 95%CI)	93.3 (86.5, 98.4)	93.9 (86.6, 100.0)
Negative predictive value (% , 95%CI)	56.9 (50.8, 62.9)	55.2 (48.8, 61.4)

Validation Sample

There were 313 patients in the validation sample. C-index for the logistic regression model and risk score model was both 0.86 (95% CI, 0.82% – 0.90%) (Fig. 2). Calibration plot of the observed frequency of CPP patients against the predicted probability of CPP showed an intercept of -0.02, and a slope of 1.06, suggesting acceptable calibration (Additional file 1 [Additional Fig. 1]). The total risk score in the validation sample ranged from 0 to 33. The proportion of CPP patient in the low-, medium-, and high-risk population was 0.0% (0/39), 53.3% (112/210), and 93.8% (60/64), respectively (Table 4). The test characteristics retained in the validation sample (Table 4).

Discussion

GnRH (GnRHa) stimulation test is the gold standard for diagnosis of CPP. But it is time-consuming and costly [1, 7]. In this study, we developed a risk score system (4 items with a 33 - point total scale) containing information on age at onset of puberty, basal LH concentration, and pelvic sonography for the prediction of CPP. The risk score model performed well in both training and validation sample (C-index of 0.85 and 0.86, respectively). Two cut-off points of the risk score were selected to classify patients into the low-, median-, and high-risk for CPP. The low-risk cut-off point (< 10 points) defined a model with a sensitivity of 97.8% and a LR- of 0.09; the high-risk cut-off point (\geq 20 points) defined a model with a specificity of 96.6% and a LR + of 12.0. The stratification of the risk score would give support on making decisions on the need for further diagnosis test.

All variables in the prediction model have been demonstrated to be associated with CPP in previous studies [1, 7]. Thelarche is the first sign of puberty [23]. Premature thelarche occurred before age 2 years usually regressed completely, while it generally led to early puberty when it occurred after age 2 years old [24]. LH concentration is the most valuable parameter for diagnosis of CPP. Various cut-off points of basal LH ranging from 0.1 to 1.5 IU/L had been used to evaluate the activation of HPAG, which resulted in a sensitivity and specificity ranging from 60–100% [8–13, 25]. The wide variations had hampered the definition of cut-off point of basal LH to discriminate CPP. In addition, basal LH elevates after the increase of stimulated LH, which suggested that basal LH was an indicator with a high specificity but low sensitivity [1, 9, 16]. Our findings agreed with previous studies and confirmed that CPP patients had a high risk score resulted from an elevated basal LH concentration and associated enlarged ovarian volume.

Enlarged ovaries and uterus is end-organ effect of gonadotropin stimulation, which occurred in late stage of puberty development (ovary development in stage 3 and uterine development in stage 4) [15, 16]. It was reported that a female with an average ovarian volume less than 2 mL has 75% chance of being prepuberty.[16] A uterine volume of greater than 2 mL has also been used as a cut-off point for the diagnosis of CPP [26]. However, there was substantial overlap in ovarian and uterine volumes between girls in the prepubertal and pubertal stage, which suggested that pelvic ultrasonography alone could not be a sensitive indicator for CPP [16]. We found that largest ovarian volume is the most sensitive pelvic ultrasonography variable. But even the largest ovarian volume could not serve as an indicator independently to discriminate CPP from PPP.

Our study confirmed findings of previous studies and developed a risk score model for CPP including information on both basal LH and pelvic ultrasonography. Based on the stratification of the CPP risk score, we suggest that patients in the high-risk category (\geq 20 points) could be diagnosed as CPP without the need of stimulation test; patients with median-risk (10–19 points) will need a stimulation test; the best policy for patients with a low-risk CPP score (< 10 points) may need a follow-up for the pubertal development.

Strengths of this study included the objective assessment of pelvic ultrasonography. Pelvic ultrasonography was performed within one week of the GnRHa stimulation test. Radiologists had no

information on the result of the diagnosis test. Moreover, a large external validation sample confirmed good predictive performance of the risk score model. The minor difference between the training and validation sample increased the generalizability of the model. To our knowledge, it is the first study that developed and validated a risk score model for the diagnosis of CPP in a large sample.

However, there are several limitations to this study. First, all patients (both training and validation sample) came from the Children's Hospital, Fudan University. Performance of the risk score may vary in different populations, which resulted in the limitation of the generalizability. But the Department of Endocrinology of this tertiary hospital receives patients from all around China. Given the prevalence of precocious puberty was 0.43% in China [2], we may speculate that the current study population has included a heterogeneous patients. Future study would benefit from the assessment of the risk score in other clinical settings. Second, most subjects in the current study were patients with recent onset of puberty. It may not represent the complicated spectrum of precocious puberty. But patients with longer duration of pubertal development may have more difference in hormonal profile and pelvic sonography than the newly onset patients. The inclusion of patients with longer disease duration may not decrease the diagnostic value of the risk score model. Third, LH and FSH concentration was measured using electrochemiluminescence assay with a LOD of 0.2 IU/L in this study. The LH concentration records were extracted from the medical history of the database. Variations among batches could not be avoided. Assay characteristics and interassay variations should be taken into account in clinical utility [1]. Fourth, the variation in the pelvic ultrasonography measurement among radiologists may also introduce bias. However, all radiologists had no information on the results of the stimulation test. The misclassification was not differential, which may result in an underestimation of the performance of the risk score model. Finally, both basal LH and pelvic ultrasonography are indicators of the activation of HPGA in the late stage, which leads to the risk score model with more specificity but less sensitivity. Patients with high-risk score would be a major beneficiary of the risk score model obviating an unnecessary stimulation test; patients with medium-risk of CPP need to perform a diagnostic test promptly, while patients in the low-risk category need following up the progression of pubertal development.

Conclusions

A risk score model developed from girls with precocious pubertal development had moderate test characteristics to stratify CPP risk. This stratification offers the advantage of making decisions on the need for further diagnostic test (GnRH or GnRHa stimulation test). Validations in other clinical settings are needed before the adoption in clinical practice.

Abbreviations

CPP: central precocious puberty; GnRH: gonadotropin-releasing hormone; GnRHa: gonadotropin-releasing hormone analogs; HPGA: hypothalamic-pituitary-gonadal axis; PPP: peripheral precocious puberty; LH: luteinizing hormone; FSH: follicle-stimulating hormone; EMRS: electronic medical record system; LOD: limit of detection; BA: bone age; GP: Greulich Pyle; LR+: positive likelihood ratio; LR-: negative likelihood

ratio; CI: confidence interval; HCG: human chorionic gonadotropin; DHS: dehydroepiandrosterone sulfate; ACTH: adrenocorticotrophic hormone; TSH: thyroid - stimulating hormone

Declarations

Ethics approval and consent to participate

This research is in accordance with Declaration of Helsinki. This study was approved by the Ethics Committees of the Children's Hospital, Fudan University, Shanghai, China (reference number: 2012-130). Informed written consent has been obtained from each participant's parents or legal guardian after full explanation of the purpose and nature of all procedures used.

Consent for publication

Not applicable.

Availability of data and materials

The dataset analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

All authors approved the final manuscript and declared no financial or nonfinancial competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Authors' contributions

Jiangfeng Ye, Li Xi and Feihong Luo conceived this study. Jingyu You, Xianying Cheng, Feihong Luo take responsibility for the data collection. Jiangfeng Ye analyzed and interpreted the data. Jingyu You, Li Xi, and Jiangfeng Ye drafted the manuscript. All the authors provided critical feedback on interpretation of results and on the manuscript draft.

Author details

¹Department of Pediatric Endocrinology and Inborn Metabolic Diseases, Children's Hospital, Fudan University, Shanghai, China.

²Department of Ultrasonography, Children's Hospital, Fudan University, Shanghai, China.

³Institute of Obstetrics and Gynecology, Obstetrics and Gynecology Hospital, Fudan University, Shanghai, China.

References

1. Carel JC, Leger J: Clinical practice. Precocious puberty. *The New England journal of medicine* 2008, 358(22):2366-2377.
2. Zhu M, Fu J, Liang L, Gong C, Xiong F, Liu G, Luo F, Chen S: [Epidemiologic study on current pubertal development in Chinese school-aged children]. *Zhejiang da xue xue bao Yi xue ban = Journal of Zhejiang University Medical sciences* 2013, 42(4):396-402.
3. Partsch CJ, Sippell WG: Pathogenesis and epidemiology of precocious puberty. Effects of exogenous oestrogens. *Human reproduction update* 2001, 7(3):292-302.
4. Lakshman R, Forouhi NG, Sharp SJ, Luben R, Bingham SA, Khaw KT, Wareham NJ, Ong KK: Early age at menarche associated with cardiovascular disease and mortality. *The Journal of clinical endocrinology and metabolism* 2009, 94(12):4953-4960.
5. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet Oncology* 2012, 13(11):1141-1151.
6. Xhrouet-Heinrichs D, Lagrou K, Heinrichs C, Craen M, Dooms L, Malvaux P, Kanen F, Bourguignon JP: Longitudinal study of behavioral and affective patterns in girls with central precocious puberty during long-acting triptorelin therapy. *Acta paediatrica (Oslo, Norway : 1992)* 1997, 86(8):808-815.
7. Latronico AC, Brito VN, Carel JC: Causes, diagnosis, and treatment of central precocious puberty. *The lancet Diabetes & endocrinology* 2016, 4(3):265-274.
8. Freire AV, Escobar ME, Gryngarten MG, Arcari AJ, Ballerini MG, Bergada I, Ropelato MG: High diagnostic accuracy of subcutaneous Triptorelin test compared with GnRH test for diagnosing central precocious puberty in girls. *Clinical endocrinology* 2013, 78(3):398-404.
9. Brito VN, Batista MC, Borges MF, Latronico AC, Kohek MB, Thirone AC, Jorge BH, Arnhold IJ, Mendonca BB: Diagnostic value of fluorometric assays in the evaluation of precocious puberty. *The Journal of clinical endocrinology and metabolism* 1999, 84(10):3539-3544.
10. Neely EK, Hintz RL, Wilson DM, Lee PA, Gautier T, Argente J, Stene M: Normal ranges for immunochemiluminometric gonadotropin assays. *The Journal of pediatrics* 1995, 127(1):40-46.
11. Houk CP, Kunselman AR, Lee PA: Adequacy of a single unstimulated luteinizing hormone level to diagnose central precocious puberty in girls. *Pediatrics* 2009, 123(6):e1059-1063.

12. Lee DS, Ryoo NY, Lee SH, Kim S, Kim JH: Basal luteinizing hormone and follicular stimulating hormone: is it sufficient for the diagnosis of precocious puberty in girls? *Annals of pediatric endocrinology & metabolism* 2013, 18(4):196-201.
13. Pasternak Y, Friger M, Loewenthal N, Haim A, Hershkovitz E: The utility of basal serum LH in prediction of central precocious puberty in girls. *European journal of endocrinology* 2012, 166(2):295-299.
14. Herter LD, Golendziner E, Flores JA, Moretto M, Di Domenico K, Becker E, Jr., Spritzer PM: Ovarian and uterine findings in pelvic sonography: comparison between prepubertal girls, girls with isolated thelarche, and girls with central precocious puberty. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine* 2002, 21(11):1237-1246; quiz 1247-1238.
15. Haber HP, Wollmann HA, Ranke MB: Pelvic ultrasonography: early differentiation between isolated premature thelarche and central precocious puberty. *European journal of pediatrics* 1995, 154(3):182-186.
16. Sathasivam A, Rosenberg HK, Shapiro S, Wang H, Rapaport R: Pelvic ultrasonography in the evaluation of central precocious puberty: comparison with leuprolide stimulation test. *The Journal of pediatrics* 2011, 159(3):490-495.
17. Carretto F, Salinas-Vert I, Granada-Yvern ML, Murillo-Valles M, Gomez-Gomez C, Puig-Domingo M, Bel J: The usefulness of the leuprolide stimulation test as a diagnostic method of idiopathic central precocious puberty in girls. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme* 2014, 46(13):959-963.
18. Chi CH, Durham E, Neely EK: Pharmacodynamics of aqueous leuprolide acetate stimulation testing in girls: correlation between clinical diagnosis and time of peak luteinizing hormone level. *The Journal of pediatrics* 2012, 161(4):757-759.e751.
19. WW Greulich, SI. PyleRadiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford University Press: Stanford, California;1959.
20. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW: Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass)* 2010, 21(1):128-138.
21. Harrell F Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY: Springer;2010.
22. Sullivan LM, Massaro JM, D'Agostino RB, Sr.: Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Statistics in medicine* 2004, 23(10):1631-1660.
23. de Vries L, Guz-Mark A, Lazar L, Reches A, Phillip M: Premature thelarche: age at presentation affects clinical course but not clinical characteristics or risk to progress to precocious puberty. *The Journal of pediatrics* 2010, 156(3):466-471.
24. Stanhope R: Premature thelarche: clinical follow-up and indication for treatment. *Journal of pediatric endocrinology & metabolism : JPEM* 2000, 13 Suppl 1:827-830.

25. Carel JC, Eugster EA, Rogol A, Ghizzoni L, Palmert MR, Antoniazzi F, Berenbaum S, Bourguignon JP, Chrousos GP, Coste J *et al*: Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics* 2009, 123(4):e752-762.
26. de Vries L, Horev G, Schwartz M, Phillip M: Ultrasonographic and clinical parameters for early differentiation between precocious puberty and premature thelarche. *European journal of endocrinology* 2006, 154(6):891-898.

Figures

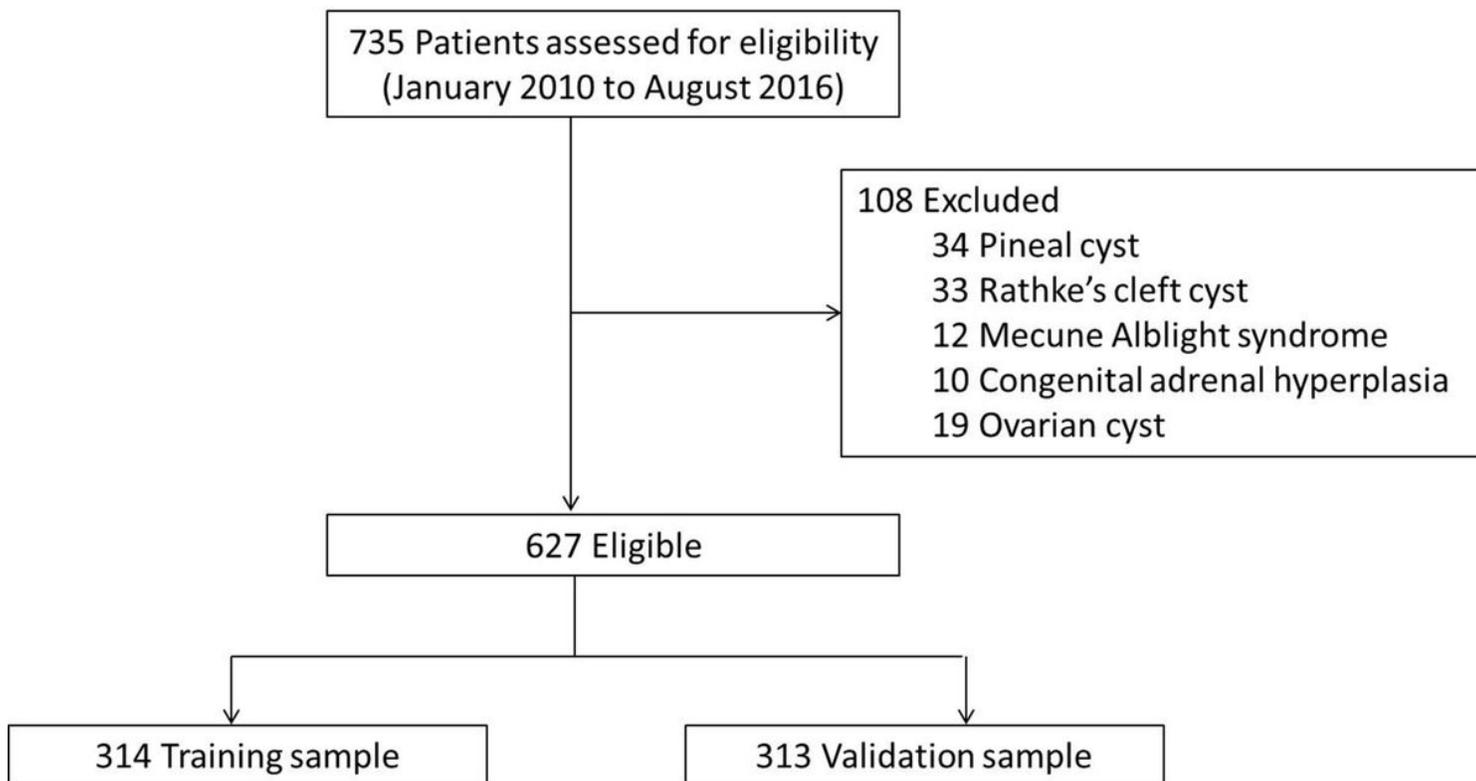


Figure 1

Study subjects flow chart

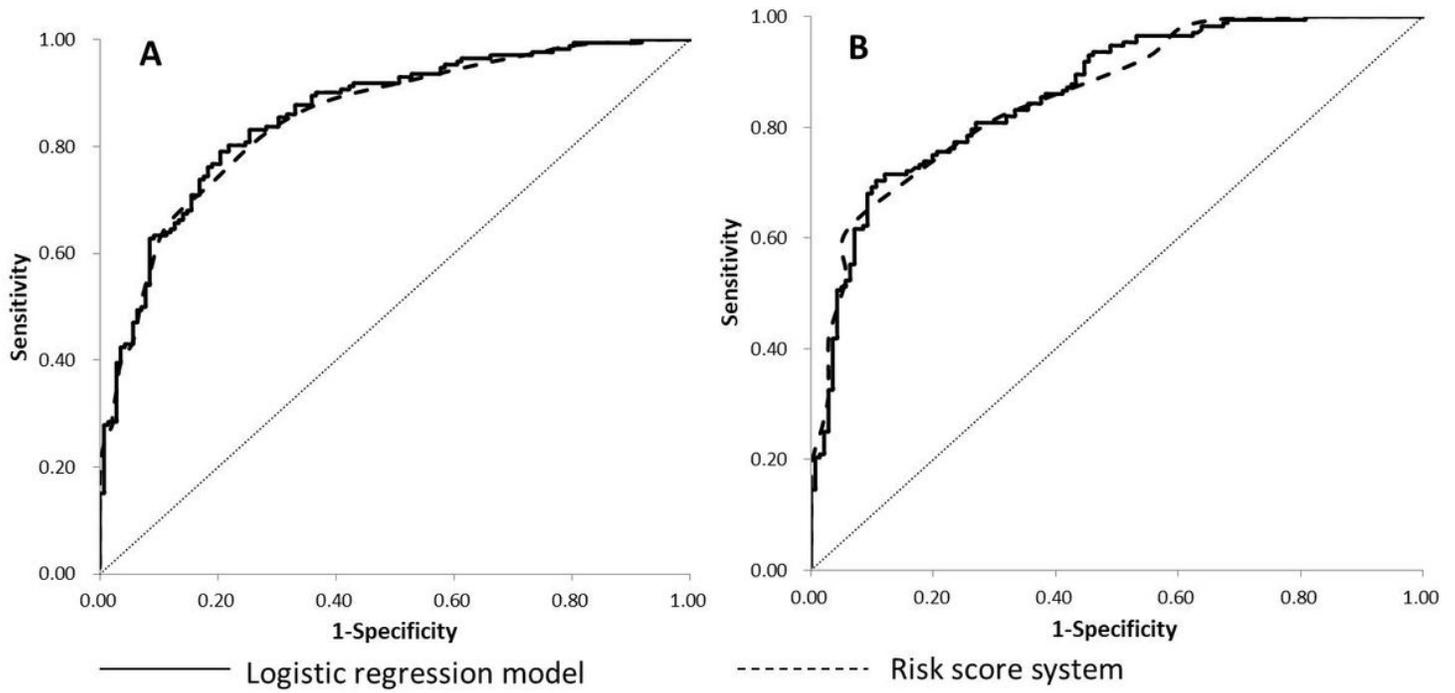


Figure 2

Receiver operating characteristic curves of the prediction model and risk score system. A. Training sample B. Validation sample

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)