

Development and Psychometric Validation of the CaregiverCPP Scale for Caregivers of Children With Central Precocious Puberty: A Multicenter Study in China

Xiao Yu

Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Shuxia Ding

Women and Children's Hospital of Ningbo University

Wenli Lu

Ruijin Hospital

Jing Chen

Women and Children's Hospital, Xiamen University

Shuangxi Cheng

Dongguan Maternal and Child Health Care Hospital

Wei Gu

Children's Hospital of Nanjing Medical University

Qiao Wang

Beijing Children's Hospital

Gaohui Zhu

Children's Hospital of Chongqing Medical University

Yan Sun

Shandong Provincial Hospital Affiliated to Shandong First Medical University

Xiaoping Luo

xpluo@tjhu.tjmu.edu.cn

Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Research Article

Keywords: central precocious puberty, caregiver-reported outcome measure, clinical outcome assessment, psychometric validation, content validity, COSMIN, item response theory

Posted Date: July 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9991907/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background: Central precocious puberty (CPP) requires long-term, family-managed care; however, existing generic instruments assessing family impact or caregiver burden do not adequately capture CPP-specific concerns, including treatment-regimen convenience, disease knowledge, decision-making drivers, psychosocial challenges, and financial strain. This study aimed to develop and psychometrically evaluate CaregiverCPP, a disease-specific caregiver-reported outcome measure designed to assess caregivers' experiences in managing CPP.

Methodology: We conducted a multicenter, mixed-methods study at eight tertiary pediatric endocrinology centers in China in accordance with the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) principles. Phase I involved item generation and content validation through a literature review, caregiver interviews, two rounds of Delphi expert consultation, and cognitive interviews. In Phase II, 1,476 primary caregivers were randomly assigned to exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) samples. Reliability was assessed in terms of internal consistency and test–retest reliability. Item response theory and differential item functioning analyses were used to evaluate measurement precision and preliminary measurement invariance, respectively.

Results: The final 38-item CaregiverCPP comprised six domains: treatment burden and convenience, psychosocial and daily-life impact, disease knowledge and understanding, treatment preferences and decision-making drivers, attitudes and economic burden, and caregiving practice. CFA supported the six-factor structure and demonstrated good model fit (RMSEA = 0.049; CFI = 0.951). Internal consistency was excellent for the total scale (Cronbach's α = 0.912), and test–retest reliability was high (intraclass correlation coefficient = 0.891). Item discrimination parameters were high (a = 1.51–2.79), indicating good measurement precision. No meaningful differential item functioning by child sex was detected ($\Delta R^2 < 0.02$).

Conclusions: CaregiverCPP provides initial evidence of content validity, structural validity, reliability, measurement precision, and sex-related measurement invariance as a disease-specific caregiver-reported outcome measure for CPP. The instrument may support the standardized assessment of CPP-related caregiving experiences in clinical research and family-centered care. Further studies should evaluate its convergent validity, responsiveness, interpretability, and minimal important difference.

Background

Central precocious puberty (CPP) is caused by the premature activation of the hypothalamic–pituitary–gonadal axis, which initiates the process of puberty earlier than expected. Beyond its implications for linear growth and final adult height, CPP poses complex clinical-management issues and profound psychosocial implications. Recent multinational studies indicate an increased incidence of CPP, with an evident trend of “increased diagnosis and accelerated progress” in the context of lifestyle disturbances associated with the COVID-19 pandemic, underscoring the non-negligible public health and family-level implications of this condition [1–6].

From the patient and family perspective, the impact of CPP extends beyond growth velocity and the development of secondary sexual characteristics. Existing research has shown that CPP—particularly idiopathic CPP—is associated with reduced health-related quality of life and a higher risk of behavioral and emotional problems in affected children. The family system and the physical and mental well-being of caregivers are directly linked to the child’s quality of life, supporting the importance of a “family-unit” approach for long-term management, particularly given the requirement for regular gonadotropin-releasing hormone agonist (GnRHa) therapy [1, 7–11].

Existing quantitative methods to characterize the caregiver–patient dyad in CPP are inadequate. General measures such as the PedsQL Family Impact Module and the Zarit Burden Interview (ZBI), although broadly used across chronic conditions, tend to capture only generic constructs and may overlook details specific to CPP, including disease and treatment knowledge, treatment preferences and risk–benefit profiles, real-world management practices, and treatment-related financial burden. Moreover, the absence of a rigorously validated, CPP-specific caregiver-reported outcome instrument hinders accurate quantification of family burden, evaluation of interventions, and generation of interpretable findings for clinicians and decision-maker [8, 12, 13]. Regulators and the academic community increasingly emphasize patient- and caregiver-focused outcomes in clinical evaluation. The U.S. Food and Drug Administration (FDA) Patient-Reported Outcome (PRO) guidance explicitly requires that measurement instruments be supported by strong evidence of content validity and psychometric properties to support claims of clinical benefit. The Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative similarly emphasizes content validity in the development and application of PROMs, and provides transparent evidence-synthesis and reporting standards now widely adopted internationally [14–19].

Against this background, the present study aimed to develop and validate a disease-specific instrument for caregivers of children with CPP—the CaregiverCPP—in accordance with COSMIN guidelines. The aim was to provide a systematic tool to capture the caregiving experience in CPP and to evaluate its reliability, structural validity, and measurement invariance. The instrument is intended to serve as a standardized outcome measure for clinical practice, patient education, family-centered care, and future clinical research.

Methods

Study design

We adopted a multi-phase, mixed-methods design to develop and psychometrically evaluate the CaregiverCPP, a disease-specific caregiver-reported outcome measure for primary caregivers of children with CPP. The study protocol followed COSMIN guidelines and combined content validation (Phase I) with psychometric validation (Phase II). The study was conducted and reported in accordance with the COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures (version 2.0) [17] and additionally referenced the STROBE checklist where relevant.

Participants and eligibility

Primary caregivers were defined as family members (parents or grandparents) bearing major responsibility for the child's daily care and medical decision-making. A multicenter cross-sectional design was used to maximize representativeness and generalizability, recruiting participants across eight tertiary-level pediatric endocrinology centers in different geographical regions of China. Inclusion criteria were: (i) children with a confirmed clinical diagnosis of CPP; (ii) children aged 2–16 years; (iii) caregivers ≥ 18 years of age with adequate Chinese functional literacy; and (iv) voluntary participation with informed consent. Exclusion criteria were: (i) children with severe chronic illness, significant neurological disorders, or other conditions likely to substantially influence the caregiving experience; and (ii) caregivers with diagnosed intellectual impairment, significant mental health problems, or any other condition that would prevent independent completion of the questionnaire.

Scale development and content validity

Content validity was established using a three-step COSMIN-guided process: item-pool generation, expert consultation, and cognitive interviews. First, an initial item pool was generated by integrating a systematic literature review, semi-structured caregiver interviews, and clinical expert input, organized around a conceptual framework of five domains—illness perception, treatment attitude, care practices, psychological impact, and financial burden—rated on a five-point Likert scale. Two rounds of Delphi consultation were then performed with a multidisciplinary expert panel (pediatric endocrinology, nursing, and psychology), with items evaluated on importance and clarity to derive a pilot version. Finally, cognitive interviews using a think-aloud technique were conducted with caregivers of varying educational backgrounds to evaluate comprehensibility and response processes, ensuring the appropriateness of the instrument for the target population.

Statistical analysis

After content validity was established, psychometric properties were evaluated using a cross-sectional, multicenter sample. In accordance with COSMIN recommendations for evaluation of structural validity, the full sample was randomly split into two independent subsamples stratified by age and sex: Sample A for item analysis and exploratory factor analysis (EFA), and Sample B for confirmatory factor analysis (CFA). Item analysis included evaluation of distributional properties, discrimination, and item–total correlations, integrated with prior content validity evidence.

Structural validity was evaluated using an exploration–verification approach. EFA was performed on Sample A to identify the latent factor structure, allowing for inter-factor correlations; CFA on Sample B was used to test and compare competing models. Internal consistency was assessed using Cronbach's α and McDonald's ω . Item response theory (IRT) analyses using a graded response model (GRM) were used to estimate item discrimination and information functions. Differential item functioning (DIF) analyses tested measurement invariance across major subgroups. Test–retest reliability was evaluated

in a subsample of caregivers (target $n = 80-120$) reassessed approximately two weeks after baseline, assuming no substantial clinical change in the interval; intraclass correlation coefficients (ICCs) were estimated for total and subscale scores. All analyses were performed in R version 4.4.1 and Python version 3.10.

Results

Scale development and content validity

The initial item pool comprised 57 candidate items. Two rounds of Delphi consultation ($n = 10$ experts) screened and refined items to achieve the predefined Item-Content Validity Index (I-CVI) threshold of ≥ 0.78 . In Round 1, items containing technical language or referring to the diagnostic phase were identified as ambiguous. In Round 2, structural improvements were made—notably the decomposition of compound constructs, unification of first-person voice (“I”), and refinement of the financial-burden wording. Subsequent cognitive interviews revealed variability in interpretation of medical terminology (“bone age”) and frequency qualifiers (“frequently”). After incorporating glossarial annotations and explicit time-frame references, the pilot version of the CaregiverCPP comprised 38 core items (Additional file 1).

Characteristics of the study population

Following age-range screening (2–16 years), 1,476 primary caregivers were retained for analysis. The sample was randomly split into two equal subsamples for EFA and CFA ($n = 738$ each). Most caregivers were mothers, with the majority aged 30–39 years, and more than half holding a university degree or above. Most patients were girls and were receiving standard treatment at the time of survey administration (Table 1). The two subsamples were well balanced on key continuous variables (child age, caregiver age, total scale score), with no statistically significant differences (all $P > 0.05$; Table 1).

Table 1
Baseline characteristics and assessment of comparability between subsamples (N = 1,476)

Characteristic	Total (N = 1,476)	Sample A: EFA (n = 738)	Sample B: CFA (n = 738)	P value
Caregiver relationship, n (%)				0.72
Mother	919 (62.3)	454 (61.5)	465 (63.0)	
Father	439 (29.7)	220 (29.8)	219 (29.7)	
Other†	118 (8.0)	64 (8.7)	54 (7.3)	
Caregiver age group, n (%)				0.98
< 30 years	52 (3.5)	26 (3.5)	26 (3.5)	
30–39 years	863 (58.5)	429 (58.1)	434 (58.8)	
40–49 years	489 (33.1)	246 (33.3)	243 (32.9)	
≥ 50 years	72 (4.9)	37 (5.0)	35 (4.7)	
Education level, n (%)				0.96
Middle school or below	106 (7.2)	54 (7.3)	52 (7.0)	
High school / vocational school	175 (11.9)	84 (11.4)	91 (12.3)	
Junior college (associate degree)	308 (20.9)	155 (21.0)	153 (20.7)	
Undergraduate (bachelor's)	746 (50.5)	372 (50.4)	374 (50.7)	
Master's or above	141 (9.6)	73 (9.9)	68 (9.2)	
Monthly household income (CNY), n (%)				0.61
< 5,000	204 (13.8)	105 (14.2)	99 (13.4)	
5,001–10,000	503 (34.1)	245 (33.2)	258 (35.0)	
10,001–20,000	447 (30.3)	227 (30.8)	220 (29.8)	
20,001–30,000	190 (12.9)	91 (12.3)	99 (13.4)	
> 30,000	132 (8.9)	70 (9.5)	62 (8.4)	
Child sex, n (%)				0.34
Note. Data are presented as n (%) or mean ± SD. Comparisons used the χ^2 or Fisher exact test for categorical variables and the independent-samples t test or Wilcoxon rank-sum test for continuous variables. All tests were two-sided; P < 0.05 indicates statistical significance. † “Other” includes grandparents and other legal primary caregivers. CFA, confirmatory factor analysis; EFA, exploratory factor analysis; SD, standard deviation.				

Characteristic	Total (N = 1,476)	Sample A: EFA (n = 738)	Sample B: CFA (n = 738)	P value
Female	1,069 (72.4)	542 (73.5)	527 (71.4)	
Male	407 (27.6)	196 (26.5)	211 (28.6)	
Currently receiving treatment, n (%)				0.22
Yes	1,190 (80.6)	586 (79.4)	604 (81.9)	
No	286 (19.4)	152 (20.6)	134 (18.1)	
Child age (years), mean $\pm$ SD	9.49 $\pm$ 2.11	9.48 $\pm$ 2.13	9.50 $\pm$ 2.09	0.86
Age at diagnosis (years), mean $\pm$ SD	8.42 $\pm$ 1.86	8.41 $\pm$ 1.77	8.42 $\pm$ 1.95	0.91
Caregiver age (years), mean $\pm$ SD	38.60 $\pm$ 5.77	38.31 $\pm$ 5.84	38.89 $\pm$ 5.69	0.09
Total scale score, mean $\pm$ SD	114.66 $\pm$ 21.5	114.63 $\pm$ 21.7	114.70 $\pm$ 21.3	0.93
Note. Data are presented as n (%) or mean $\pm$ SD. Comparisons used the χ^2 or Fisher exact test for categorical variables and the independent-samples t test or Wilcoxon rank-sum test for continuous variables. All tests were two-sided; P < 0.05 indicates statistical significance. † “Other” includes grandparents and other legal primary caregivers. CFA, confirmatory factor analysis; EFA, exploratory factor analysis; SD, standard deviation.				

Item analysis and exploratory factor analysis (Sample A)

In Sample A (n = 738), item analysis demonstrated well-balanced response distributions and strong discrimination without notable floor or ceiling effects (Additional file 1, Tables S2–S3). Parallel analysis showed that the observed eigenvalues for the first six factors exceeded the 95th-percentile eigenvalues from random permutations; consistent with this, the scree plot identified the sixth factor as the structural inflection point, supporting a six-factor solution (Fig. 1). With Promax oblique rotation, most items showed substantial primary loadings on their assigned factors (typically ≥ 0.40), with negligible cross-loading, supporting structural specificity. Item 17 showed a suboptimal primary loading and was retained within its content domain pending confirmatory evaluation. The factor assignments and primary loadings are summarized in Table 2; the full factor-loading matrix and communalities are provided in Additional file 1, Table S2.

Table 2

Exploratory factor analysis of the CaregiverCPP: primary factor assignments and loadings (Sample A, n = 738)

No.	Item description	Factor domain	Loading
1	Understanding of the potential causes of CPP	F3 Disease perception & information	0.54
2	Understanding of the diagnostic process for CPP	F3 Disease perception & information	0.73
3	Understanding of the role of “bone age” in diagnosis and treatment	F3 Disease perception & information	0.67
4	Understanding of the primary goal of GnRHa (“suppression shot”) therapy	F3 Disease perception & information	0.61
5	Awareness of potential impacts on future psychological health, beyond height outcomes	F3 Disease perception & information	0.6
6	Knowledge that the main difference between formulations (e.g., monthly vs. 3-monthly) lies in frequency and convenience	F3 Disease perception & information	0.65
7	Awareness of common, mild side effects during treatment (e.g., local redness or pain)	F3 Disease perception & information	0.73
8	Knowledge of appropriate management for mild local reactions following injection	F3 Disease perception & information	0.67
9	Belief that GnRHa therapy is necessary for children with CPP	F5 Treatment attitude & financial burden	0.67
10	Concern regarding potentially unknown risks of the medication	F5 Treatment attitude & financial burden	0.57
11	Belief that therapy is critical for improving final adult height	F5 Treatment attitude & financial burden	0.66
12	Perceived stress related to routine management tasks (e.g., timely injections, follow-ups)	F5 Treatment attitude & financial burden	0.59
13	Regular monitoring and recording of growth indicators in the past month	F6 Caregiving practices & active management	0.55
14	Active information seeking regarding CPP via internet or books in the past month	F6 Caregiving practices & active management	0.63

Note. Primary factor loadings derived from EFA using Promax oblique rotation. The complete factor-loading matrix and communalities are provided in Additional file 1, Table S2. † Item 17 showed a primary loading of 0.28, below conventional thresholds, but was the only item directly assessing caregivers’ management of and coping with fear during GnRHa injections. Given its unique content coverage and clear clinical relevance confirmed by expert review, the item was retained to preserve content validity. CPP, central precocious puberty; GnRHa, gonadotropin-releasing hormone agonist.

No.	Item description	Factor domain	Loading
15	Attempting to discuss the condition and treatment with the child in an age-appropriate manner	F6 Caregiving practices & active management	0.63
16	Strict adherence to the doctor's schedule to ensure every treatment session in the past month	F4 Treatment preferences & decision drivers	0.61
17	Perceived challenge in helping the child overcome fear of injections in the past month	F2 Psychosocial & daily-life impact	0.28†
18	Feeling depressed or anxious due to the child's condition in the past month	F2 Psychosocial & daily-life impact	0.8
19	Engaging in deep conversations with others regarding the condition in the past month	F2 Psychosocial & daily-life impact	0.48
20	Feeling that family life was disrupted by the child's condition in the past month	F2 Psychosocial & daily-life impact	0.67
21	Significant worry about suboptimal final height due to the condition in the past month	F2 Psychosocial & daily-life impact	0.82
22	Significant worry about unknown long-term side effects of treatment in the past month	F2 Psychosocial & daily-life impact	0.86
23	Significant worry about the child being bullied or isolated at school due to early development	F2 Psychosocial & daily-life impact	0.73
24	Significant worry that the child is psychologically unprepared for early development	F2 Psychosocial & daily-life impact	0.82
25	Perception of examination fees (e.g., ultrasound, bone age, labs, MRI) as a major household expense	F5 Treatment attitude & financial burden	0.47
26	Financial strain caused by transportation and accommodation costs for frequent hospital visits	F5 Treatment attitude & financial burden	0.47
27	Perception that total treatment costs exceed the family's financial capacity	F5 Treatment attitude & financial burden	0.4
28	Effectiveness in controlling sexual development	F4 Treatment preferences & decision drivers	0.65
29	Maximizing improvement in future adult height	F4 Treatment preferences & decision drivers	0.65

Note. Primary factor loadings derived from EFA using Promax oblique rotation. The complete factor-loading matrix and communalities are provided in Additional file 1, Table S2. † Item 17 showed a primary loading of 0.28, below conventional thresholds, but was the only item directly assessing caregivers' management of and coping with fear during GnRHa injections. Given its unique content coverage and clear clinical relevance confirmed by expert review, the item was retained to preserve content validity. CPP, central precocious puberty; GnRHa, gonadotropin-releasing hormone agonist.

No.	Item description	Factor domain	Loading
30	Probability of treatment-induced weight gain	F1 Treatment burden & convenience	0.55
31	Type and probability of potential side effects (e.g., local reactions)	F1 Treatment burden & convenience	0.71
32	Minimizing hospital visits	F1 Treatment burden & convenience	0.76
33	Minimal interference with school, sports, and daily activities	F1 Treatment burden & convenience	0.71
34	Lower injection frequency (e.g., 3-monthly vs. monthly)	F1 Treatment burden & convenience	0.8
35	The child's level of fear regarding different injection methods	F1 Treatment burden & convenience	0.74
36	Lower cost of the treatment regimen	F1 Treatment burden & convenience	0.65
37	Professional recommendation from the doctor	F4 Treatment preferences & decision drivers	0.66
38	Alignment between injection frequency and follow-up frequency	F4 Treatment preferences & decision drivers	0.48
Note. Primary factor loadings derived from EFA using Promax oblique rotation. The complete factor-loading matrix and communalities are provided in Additional file 1, Table S2. † Item 17 showed a primary loading of 0.28, below conventional thresholds, but was the only item directly assessing caregivers' management of and coping with fear during GnRHa injections. Given its unique content coverage and clear clinical relevance confirmed by expert review, the item was retained to preserve content validity. CPP, central precocious puberty; GnRHa, gonadotropin-releasing hormone agonist.			

Confirmatory factor analysis (Sample B)

The six-factor structure derived from Sample A was confirmed in Sample B (n = 738). Standardized loadings ranged from 0.38 to 0.86, with all items loading on their hypothesized factors (Table 3). The measurement model demonstrated good fit: $\chi^2/df = 2.778$, RMSEA = 0.049, CFI = 0.951, and TLI = 0.947 (Table 3). The standardized path coefficients and factor inter-correlations are illustrated in Fig. 2.

Table 3
Six-factor structure and confirmatory factor analysis (CFA) model-fit indices of the CaregiverCPP
(Samples A and B)

Factor / fit index	Items / observed value	EFA loadings (Sample A) / Criterion	CFA loadings (Sample B) / Result
Panel A. Factor structure			
F1 Treatment burden & convenience	q30–q36	0.55–0.80	0.65–0.80
F2 Psychosocial & daily-life impact	q17–q24	0.28–0.86	0.51–0.86
F3 Disease knowledge & understanding	q1–q8	0.54–0.73	0.61–0.73
F4 Preference & decision drivers	q16, q28–q29, q37–q38	0.48–0.66	0.48–0.66
F5 Attitudes & economic burden	q9–q12, q25–q27	0.40–0.67	0.38–0.67
F6 Caregiving practice	q13–q15	0.55–0.63	0.74–0.76
Panel B. CFA model-fit indices			
χ^2/df	2.778	< 5 acceptable; < 3 good	Good
RMSEA	0.049	< 0.08 acceptable; < 0.06 good	Good
CFI	0.951	> 0.90 acceptable; > 0.95 good	Good
TLI	0.947	> 0.90 acceptable; > 0.95 good	Acceptable
Note. Panel A shows the range of primary loadings from EFA (Sample A, n = 738) and the standardized loadings from CFA (Sample B, n = 738; see also Fig. 2). Panel B presents the CFA model-fit indices. Cut-offs follow established guidelines for structural equation modeling [20–22]. CFI, comparative fit index; df, degrees of freedom; RMSEA, root mean square error of approximation; TLI, Tucker–Lewis index; χ^2 , chi-square.			

Internal consistency and reliability

In the full sample (N = 1,476), the CaregiverCPP showed excellent internal consistency, with Cronbach's $\alpha = 0.912$. Subscale reliability was also strong (Cronbach's $\alpha = 0.796$ – 0.878 ; McDonald's $\omega = 0.772$ – 0.883 ; Table 4). Test–retest reliability, evaluated in a convenience subsample of 100 caregivers reassessed approximately two weeks after baseline, was excellent for the total scale (ICC = 0.891) and good to excellent for the subscales (ICC = 0.778–0.876). These results support the construct reliability of both the total scale and its subscales.

Table 4
Internal consistency and test–retest reliability of the CaregiverCPP (N = 1,476)

Factor	Domain	No. of items	Cronbach's α	McDonald's ω	ICC (test–retest)
Total	Total scale	38	0.912	—	0.891
F1	Treatment burden & convenience	7	0.873	0.873	0.852
F2	Psychosocial & daily-life impact	8	0.878	0.883	0.876
F3	Disease knowledge & understanding	8	0.851	0.856	0.821
F4	Preference & decision drivers	5	0.847	0.856	0.803
F5	Attitudes & economic burden	7	0.799	0.772	0.785
F6	Caregiving practice	3	0.796	0.798	0.778

Note. Internal consistency (Cronbach's α and McDonald's ω) was estimated on the aggregate sample (N = 1,476). Test–retest reliability was estimated on a subsample (n = 100) reassessed approximately two weeks after baseline, using a two-way random-effects ICC model with absolute agreement for single ratings. ICC values > 0.75 are considered good and > 0.90 excellent [23–25]. ICC, intraclass correlation coefficient.

Item characteristics and measurement precision

To evaluate item-level measurement precision, a GRM under the IRT framework was fitted to the final sample. Item discrimination (a) parameters and threshold (b) parameters were estimated, and the test- and item-information functions were plotted to characterize the instrument's precision (Additional file 1, Table S3 and Figures S2–S7). Discrimination parameters ranged from 1.51 to 2.79, indicating strong item discrimination with respect to the latent trait. The information functions concentrated maximum precision around the mid-range of the latent trait dimension, indicating that the scale is most informative for moderate levels of caregiving burden, disease understanding, and treatment preferences (Additional file 1, Table S3 and Figures S2–S7).

Measurement invariance across key subgroups

To evaluate fairness across subgroups, we conducted DIF analyses for child sex. A nested ordinal logistic regression approach was used to test both uniform and non-uniform DIF, with a conservative cut-off of $\Delta R^2_{\text{DIF}} \geq 0.020$ [26, 27]. ΔR^2 values for all items were well below this threshold (maximum 0.017; Additional file 1, Table S4 and Figure S8), supporting measurement invariance across child sex and indicating that between-group differences in scale scores reflect genuine differences rather than measurement artifacts.

Discussion

Principal findings

Using a large multicenter real-world dataset, we developed and validated the CaregiverCPP—a disease-specific instrument for caregivers of children with CPP. The instrument demonstrates a robust six-dimensional structure that spans the spectrum of caregiving experience from burden to proactive management, with strong construct validity, internal consistency, and test–retest reliability. To our knowledge, this work fills a notable gap in pediatric endocrinology: the absence of a disease-specific caregiver-reported outcome measure capable of quantifying the chronic, context-dependent demands of CPP management.

Comparison with previous studies

In terms of disease context, recent epidemiological work, including studies during and after the COVID-19 pandemic, has documented increased CPP incidence and care-utilization, making rigorous measurement of family caregiving burden not only a methodological priority but a pressing practical need [28, 29]. Systematic reviews highlight regional heterogeneity in pandemic-era CPP epidemiology and underscore the need for granular assessment of family and caregiving experience to explain such heterogeneity. High-quality reviews also emphasize that decision-making in CPP care involves complex trade-offs among height gain, psychosocial impact, treatment regimens, and follow-up—features that generic instruments are not designed to capture [1, 27].

From a measurement-tools perspective, currently used instruments are largely generic. The PedsQL™ Family Impact Module (FIM) captures the effect of pediatric chronic illness on parental health-related quality of life and family functioning; although valuable for cross-condition comparison, its domains center on “general functioning and emotional consequences,” leaving CPP-specific themes such as the convenience of injections and visits, preferences and decision drivers, and disease-specific knowledge under-addressed [30]. The ZBI has long served as a standard for caregiver burden, with extensive evidence on its validity and short forms; however, its primary use has been in geriatric and long-term-care contexts, and direct adaptation to pediatric endocrinology risks measurement gaps because it captures generic stress while overlooking the disease-management processes and decisional conflicts characteristic of CPP [31–33].

Mechanisms and innovation

A key innovation of this study is the development of the first disease-specific assessment framework for the CPP caregiving experience. The CaregiverCPP operationalizes a continuous process model linking “objective burden” to “subjective appraisal” and ultimately to “behavioral adaptation,” providing a mechanistic lens on how caregivers experience and respond to CPP care demands. First, the treatment-convenience and economic-attitude domains capture friction costs inherent to the healthcare system: the long follow-up periods and injection regimens generate cumulative stress, and this medical friction is often an underlying driver of variable adherence. Second, the disease-knowledge and decision-driver

domains reflect the specific challenges of CPP treatment and empirically support the health-literacy–decision-conflict model, illustrating the additional cognitive load that uncertainty places on caregivers. Third, the psychosocial-impact and proactive-management domains capture broader social dimensions of the disease: CPP can be understood not only as an endocrine condition but as a social phenomenon entangled with peer pressure and psychosomatic interactions [34–36]. Together, these domains describe a transition from passive to proactive management—that is, toward active resource mobilization—which may attenuate psychological strain. This process-and-mechanism framework offers a novel theoretical perspective on family adaptability in pediatric chronic care.

Clinical and research implications

Clinically, the CaregiverCPP supports precision in caregiver-centered nursing. A profile of “high treatment burden” combined with “low disease understanding” functions as an alert prompting clinicians to optimize care pathways and decision support, whereas a “high psychosocial impact” profile indicates a need for multidisciplinary psychological support. For research, the instrument provides a relevant secondary endpoint for trials, observational studies, and real-world evidence, enabling investigators to evaluate how different treatment regimens or educational programs affect not only physiological outcomes (e.g., bone-age progression, growth velocity) but also the spillover impact on family quality of life. The instrument also aligns with the U.S. FDA Patient-Focused Drug Development initiative by enabling a two-tier “patient–caregiver” outcome-evaluation framework.

Strengths and limitations

The study has several methodological strengths. A multicenter collaborative design involving eight major hospitals across geographically and economically diverse regions of China, combined with a large sample (N = 1,476), enhances representativeness across health-care payment systems, socioeconomic contexts, and cultural backgrounds, and provides statistical power for advanced psychometric analyses (including DIF). The real-world–based development pathway supports ecological content validity, and the large-sample exploratory–confirmatory design strengthens the robustness of the resulting structure.

Several limitations should be noted. First, recruitment from tertiary hospitals may have over-sampled caregivers with higher educational attainment and better access to healthcare resources, introducing potential socioeconomic selection bias and limiting generalizability to primary-care or community populations. Second, because the sample was drawn from a single cultural context, cross-cultural applicability requires further measurement-invariance testing in populations with different healthcare systems and languages. Third, as a cross-sectional study, longitudinal responsiveness, causal relationships with treatment-related changes, and the minimal clinically important difference (MCID) require evaluation in future cohort studies. Finally, additional analyses of concurrent validity against widely used instruments (e.g., FIM, ZBI) are needed to quantify the incremental explanatory value of this disease-specific measure beyond generic tools.

Conclusions

The CaregiverCPP is a reliable and practical instrument for quantifying the complex caregiving burden experienced by families of children with CPP. By extending outcome assessment beyond biological indicators, it supports a comprehensive biopsychosocial perspective in pediatric endocrinology. Future work should apply the CaregiverCPP in intervention research to inform and optimize supportive-care strategies in pediatric endocrine disorders.

Abbreviations

CFA

Confirmatory factor analysis

CFI

Comparative fit index

COSMIN

Consensus-Based Standards for the Selection of Health Measurement Instruments

CPP

Central precocious puberty

DIF

Differential item functioning

EFA

Exploratory factor analysis

FDA

U.S. Food and Drug Administration

FIM

PedsQL™ Family Impact Module

GnRHa

Gonadotropin-releasing hormone agonist

GRM

Graded response model

HRQoL

Health-related quality of life

ICC

Intraclass correlation coefficient

I-CVI

Item-content validity index

IRB

Institutional review board

IRT

Item response theory

MCID

Minimal clinically important difference

PROM/PRO

Patient-reported outcome measure / patient-reported outcome

RMSEA

Root mean square error of approximation

SD

Standard deviation

TLI

Tucker–Lewis index

ZBI

Zarit Burden Interview

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (the lead center; approval number TJ-IRB202508001), and subsequently approved or recognized by the institutional review boards (IRBs) of the other participating centers. All research procedures were performed in accordance with the ethical standards of the institutional and national research committees and with the 1964 Declaration of Helsinki and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Funding

This study was supported by the Beijing Top-Doctors Charity Foundation. The funder had no role in the study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit the manuscript for publication.

Author Contribution

XY conceptualized the study, designed the methodology, curated data, performed formal analysis and visualization, and drafted the manuscript. YS and XL conceptualized the study, supervised the work, and reviewed and edited the manuscript. XL acquired funding and administered the project. SD, WL, JC, SC, WG, QW, and GZ contributed to investigation, resources, and data curation, and reviewed and edited the manuscript. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgement

We extend our sincere gratitude to all the children with CPP and their caregivers who participated in this study for their trust, cooperation, and generous contribution of time to medical research. We also acknowledge the research teams at the eight clinical centers for their support and diligent efforts in participant recruitment, data collection, and study coordination: Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology; Women and Children's Hospital, School of Medicine, Xiamen University; Dongguan Maternal and Child Health Care Hospital; The Affiliated Women and Children's Hospital of Ningbo University; Children's Hospital of Nanjing Medical University; Ruijin Hospital, Shanghai Jiao Tong University School of Medicine; Shandong Provincial Hospital Affiliated to Shandong First Medical University; and Beijing Children's Hospital, Capital Medical University. We thank the Beijing Top-Doctors Charity Foundation for its generous support, and the experts who participated in the Delphi consultations and cognitive interviews for their insightful recommendations on instrument development. ChatGPT (OpenAI) was used only for language polishing and formatting support during manuscript preparation. The authors reviewed and edited all AI-assisted output. All scientific content, data analyses, interpretation, and conclusions were produced and verified by the authors, who take full responsibility for the manuscript.

Data Availability

The datasets generated and/or analysed during the current study are not publicly available due to participant privacy and ethical restrictions, but are available from the corresponding author on reasonable request. The CaregiverCPP instrument is provided in Additional file 1.

References

1. Kim SJ, Kim JH, Hong YH, Chung IH, Lee EB, Kang E et al (2023) 2022 Clinical practice guidelines for central precocious puberty of Korean children and adolescents. *Ann Pediatr Endocrinol Metab* 28(3):168–177
2. Kang S, Park MJ, Kim JM, Yuk J-S, Kim S-H (2023) Ongoing increasing trends in central precocious puberty incidence among Korean boys and girls from 2008 to 2020. *PLoS ONE* 18(3):e0283510
3. Prosperi S, Chiarelli F (2023) Early and precocious puberty during the COVID-19 pandemic. *Front Endocrinol (Lausanne)* 13:1107911
4. Umano GR, Maddaluno I, Riccio S, Lanzaro F, Antignani R, Giuliano M et al (2022) Central precocious puberty during COVID-19 pandemic and sleep disturbance: an exploratory study. *Ital J Pediatr* 48(1):60
5. Chioma L, Chiarito M, Bottaro G, Paone L, Todisco T, Bizzarri C et al (2023) COVID-19 pandemic phases and female precocious puberty: the experience of the past 4 years (2019 through 2022) in an Italian tertiary center. *Front Endocrinol (Lausanne)* 14:1132769

6. Zhang X, Xu Y, Yan L, Wang X, Xiong J, Wang F et al (2026) Global prevalence and incidence of precocious puberty: a systematic review and meta-analysis. *BMC Public Health* 26:99. 10.1186/s12889-025-25723-4
7. Klein KO, Soliman AM, Grubb E, Nisbet P (2020) A survey of care pathway and health-related quality of life impact for children with central precocious puberty. *Curr Med Res Opin* 36(3):411–418
8. Yang H, Luo S, Liang X, Lin Q, Cheng T, Zeng L et al (2021) The association between family impact and health-related quality of life of children with idiopathic central precocious puberty in Chongqing, China. *Health Qual Life Outcomes* 19(1):171
9. Raafat S, Abdelaal E, Mowafi I, Khalil M (2024) Assessment of quality of life and behavioral problems in children with central precocious puberty. *Oman Med J* 39(5):e670
10. Dinkelbach L, Grasemann C, Kiewert C, Leikeim L, Schmidt B, Hirtz R (2025) Central precocious puberty and psychiatric disorders. *JAMA Netw Open* 8(6):e2516679
11. Popovic J, Geffner ME, Rogol AD, Silverman LA, Kaplowitz PB, Mauras N et al (2022) Gonadotropin-releasing hormone analog therapies for children with central precocious puberty in the United States. *Front Pediatr* 10:968485
12. Varni JW, Sherman SA, Burwinkle TM, Dickinson PE, Dixon P (2004) The PedsQL Family Impact Module: preliminary reliability and validity. *Health Qual Life Outcomes* 2:55
13. Gratião ACM, Brigola AG, Ottaviani AC, Luchesi BM, Souza ÉN, Rossetti ES et al (2019) Brief version of the Zarit Burden Interview (ZBI) for burden assessment in older caregivers. *Dement Neuropsychol* 13(1):122–129
14. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health (2006) Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. *Health Qual Life Outcomes* 4:79
15. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J et al (2018) COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res* 27(5):1159–1170
16. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB (2018) COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 27(5):1147–1157
17. Gagnier JJ, de Arruda GT, Terwee CB, Mokkink LB (2025) COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures: version 2.0. *Qual Life Res* 34(7):1901–1911. 10.1007/s11136-025-03950-x
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 63(7):737–745
19. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB (2018) COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*

20. Hu L, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6(1):1–55
21. Browne MW, Cudeck R (1992) Alternative ways of assessing model fit. *Sociol Methods Res* 21(2):230–258
22. Kline RB (2023) *Principles and practice of structural equation modeling*, 5th edn. Guilford, New York
23. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–428
24. Deng L, Chan W (2017) Testing the difference between reliability coefficients alpha and omega. *Educ Psychol Meas* 77(2):185–203
25. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155–163
26. Baker FB (2001) *The basics of item response theory*, 2nd edn. ERIC Clearinghouse on Assessment and Evaluation, College Park
27. Zevin EL, Eugster EA (2023) Central precocious puberty: a review of diagnosis, treatment, and outcomes. *Lancet Child Adolesc Health* 7(12):886–896
28. Zhang J, Xu J, Tang X, Wu R (2024) Comparison of central precocious puberty frequency before and during COVID-19: a systematic review and meta-analysis. *BMC Endocr Disord* 24(1):219
29. Chioma L, Bizzarri C, Verzani M, Fava D, Salerno M, Capalbo D et al (2022) Sedentary lifestyle and precocious puberty in girls during the COVID-19 pandemic: an Italian experience. *Endocr Connect* 11(2):e210650
30. Lim Y, Lee S, Hong I, Han A (2024) Preliminary reliability and validity of the PedsQL Family Impact Module in parents of children with congenital muscular dystrophy. *J Rare Dis* 3:33. 10.1007/s44162-024-00057-8
31. Domínguez-Vergara J, Santa-Cruz-Espinoza H, Chávez-Ventura G (2023) Zarit Caregiver Burden Interview: psychometric properties in family caregivers of people with intellectual disabilities. *Eur J Investig Health Psychol Educ* 13(2):391–402
32. Cejalvo E, Gisbert-Pérez J, Martí-Vilar M, Badenes-Ribera L (2024) Systematic review following COSMIN guidelines: short forms of the Zarit Burden Interview. *Geriatr Nurs* 59:278–295
33. Mokkink LB, Elsman EB, Terwee CB (2024) COSMIN guideline for systematic reviews of patient-reported outcome measures, version 2.0. *Qual Life Res* 33(11):2929–2939
34. Çoban ÖG, Bedel A, Önder A, Adanır AS, Tuhan H, Parlak M (2021) Psychiatric disorders, peer-victimization, and quality of life in girls with central precocious puberty. *J Psychosom Res* 143:110401
35. Yılmaz İN, Abseyi SN, Şenyazar G, Berberoğlu M, Şıklar Z, Aycan Z (2024) Evaluation of quality of life in children with precocious puberty. *Clin Endocrinol (Oxf)* 100(4):338–342

36. Shen MY, Chiu CF, Chen MY, Yang LY, Okoli CTC, Liu YM (2025) The relationship between health literacy and medical decision conflict among parents of children newly diagnosed with central precocious puberty. *J Pediatr Nurs* 85:362–369

Figures

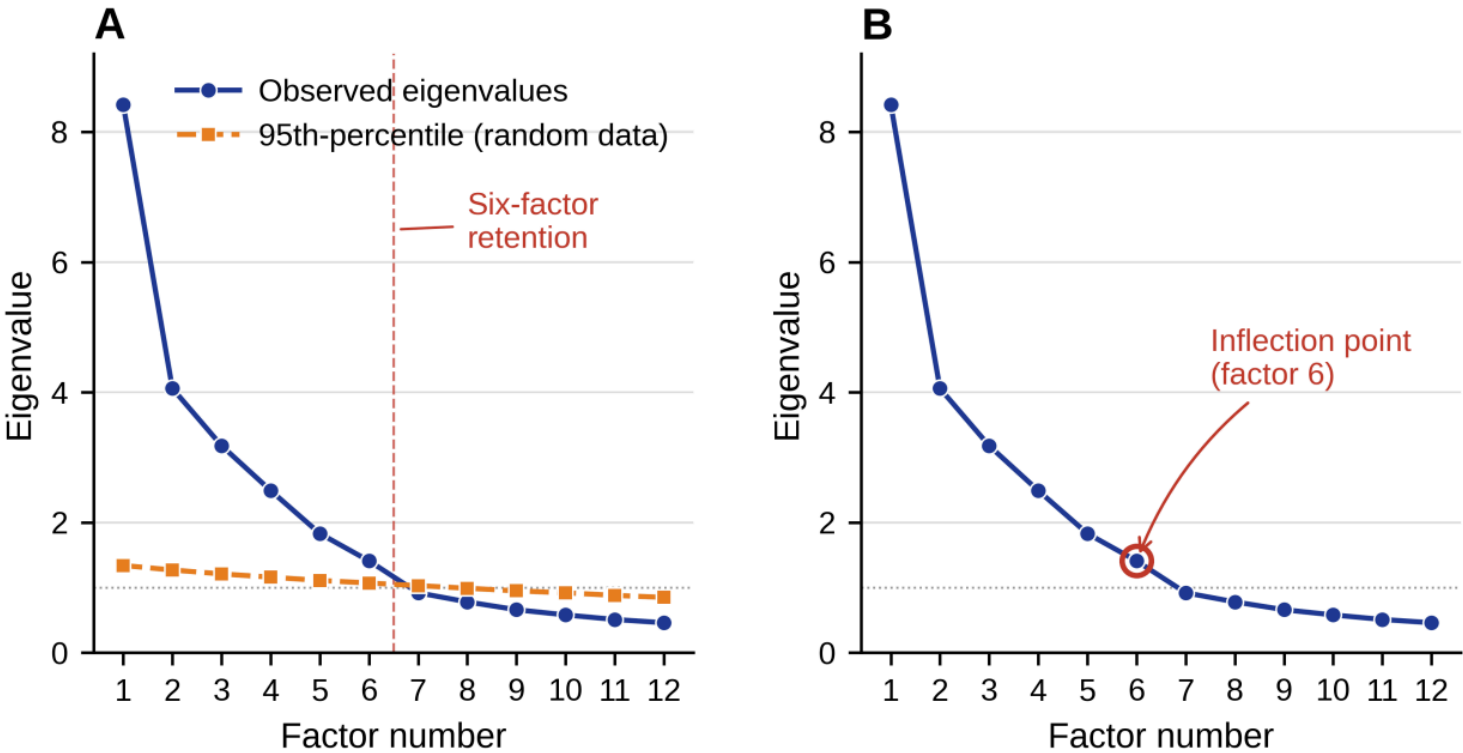
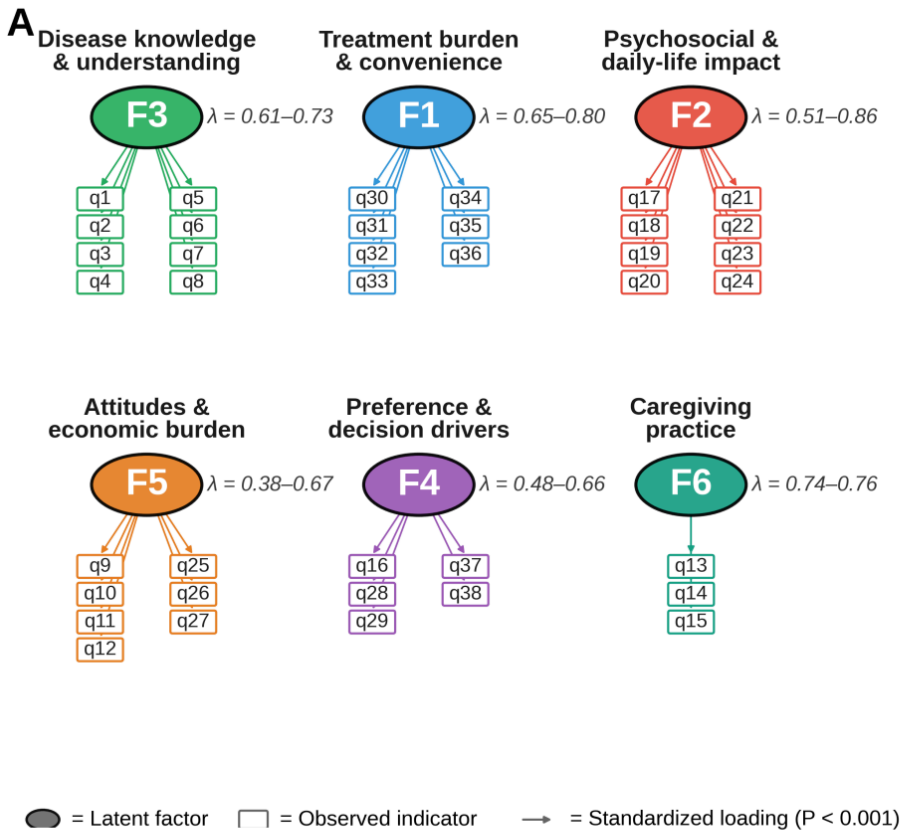


Figure 1

Parallel analysis and scree plot supporting the six-factor structure of the CaregiverCPP (Sample A, n = 738). (A) Parallel analysis: the blue line indicates observed eigenvalues; the orange line indicates the 95th-percentile eigenvalues from randomly permuted data. The first six factors exceed the random benchmark. (B) Scree plot showing an inflection at the sixth factor. Both analyses support a six-factor structure.



B

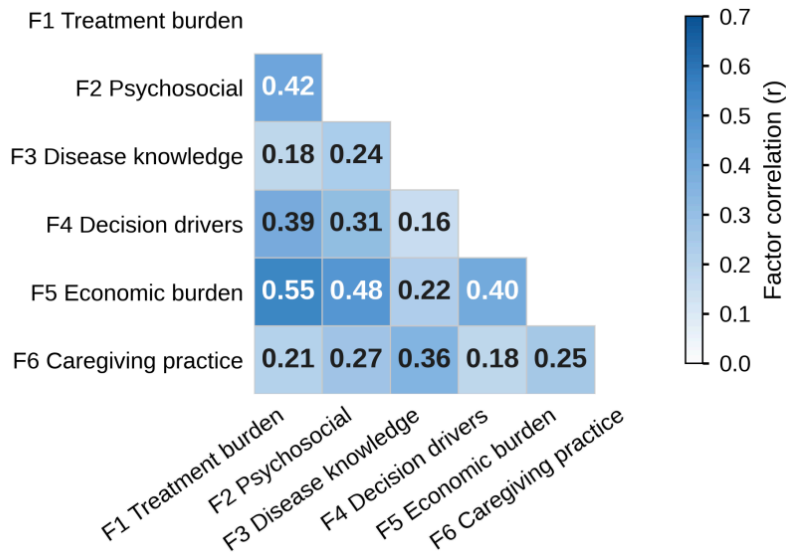


Figure 2

Path diagram of the confirmatory factor analysis (CFA) for the CaregiverCPP (Sample B, n = 738). Latent variables (F1–F6) are represented by circles; observed indicators (q1–q38) by rectangles. Single-headed arrows denote standardized factor loadings, all significant at $P < 0.001$. Curved double-headed arrows represent correlations between latent factors.

Supplementary Files

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

- [03AdditionalFile1SupplementaryMaterials.docx](#)