Published in final edited form as:

J Neurosurg Pediatr. 2014 June ; 13(6): 679–684. doi:10.3171/2014.3.PEDS13503.

External Validation of the Chicago Chiari Outcome Scale

Chester K. Yarbrough, MD¹, Jacob K. Greenberg, BA¹, Matthew D. Smyth, MD¹, Jeffrey R. Leonard, MD¹, Tae Sung Park, MD¹, and David D. Limbrick, MD PhD¹

¹Department of Neurosurgery, St. Louis Children's Hospital, Washington University School of Medicine, St. Louis, MO

Abstract

Introduction—Historically, assessment of clinical outcomes following surgical management of Chiari malformation type 1 (CM-1) has been challenging due to the lack of a validated instrument for widespread use. The Chicago Chiari Outcome Scale (CCOS) is a novel system intended to provide a less subjective evaluation of outcomes for patients with CM-1. The goal of this study was to externally validate the performance of the CCOS.

Methods—Patients undergoing surgery for CM-1 from 2001 to 2012 were reviewed (n=292). Inclusion criteria for this study were: 1) patients receiving primary posterior fossa decompression; 2) at least 5.5 months of post-operative clinical follow-up; and 3) patients ≥ 18 years of age at the time of surgery. Outcomes were evaluated using the CCOS, along with a “gestalt” impression of whether patients experienced significant improvement after surgery. A subgroup of consecutive patients undergoing operations from 2008 to 2010 (n=118) was selected for analysis of interrater reliability (n=73 meeting inclusion/exclusion criteria). In this subgroup, gestalt and CCOS were independently scored by two reviewers, and interrater reliability was assessed using the intraclass correlation coefficient (ICC) and kappa (κ) statistic.

Results—The median CCOS was 14, and 67% of patients had improved gestalt scores after surgery. Overall, the CCOS was effective at identifying patients with improved outcome after surgery (AUC=0.951). The interrater reliability of the CCOS (ICC=0.71) was high, though the reliability of the component scores ranged from poor to good (ICC 0.23 to 0.89). The functionality subscore demonstrated a low ICC and did not add to the predictive ability of the logistic regression model (Likelihood Ratio = 1.8, p=0.18). When analyzing gestalt outcome, there was moderate agreement between raters (κ=0.56).

Conclusions—In this external validation study, the CCOS was effective at identifying patients with improved outcomes and proved more reliable than our gestalt impression of outcome. However, certain component subscores (functionality and non-pain symptoms) were found to be less reliable, and may benefit from further definition in score assignment. In particular, the functionality subscore does not add to the predictive ability of the CCOS, and may be

To whom correspondence should be addressed: Chester K. Yarbrough MD, Department of Neurosurgery, Washington University School of Medicine, 660 S. Euclid Avenue, Campus Box 8057, St. Louis, Missouri 63110, yarbroughc@wudosis.wustl.edu, Phone: (314)362-3695, Fax: (314)362-2107.

Disclosures: We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

unnecessary. Overall, we found the CCOS to be an improvement over the previously utilized assessment of outcome at our institution.

Keywords

Chiari Malformation; Posterior Fossa Decompression

INTRODUCTION

Chiari malformation Type I (CM-1) is a congenital deformity characterized by herniation of the cerebellar tonsils through the foramen magnum,¹ and results in a variety of signs and symptoms, ranging from headache to brainstem compromise.^{2,12} Management of CM-1, both operative and non-operative, comprises a meaningful portion of any pediatric neurosurgery practice. Many treatment strategies have been described, though the optimal treatment remains unclear.^{6-8,15} A full discussion of the various operative techniques is outside the scope of this study.³

Assessment of patient outcomes is complicated by deficiency of validated outcome measures specific to CM-1. One group has published on patient-reported outcome measures utilizing a patient questionnaire,¹⁴ though the length of questionnaires involved likely precludes use in the pediatric population. Other groups have published case series with outcomes split into “better, unchanged, or worse” categories.^{5,9,13,17,18} A similar assessment algorithm has been in use at our institution, though this classification is highly subjective and poorly suited for assessing outcomes within or between individuals. More recently, the Chicago Chiari Outcome Scale (CCOS) was developed to address outcomes with a more complete approach to the variety of symptoms patients may experience (see Table 1).⁴ The CCOS was then retrospectively applied to the authors’ patients, and a more detailed analysis on predictors of outcome using the CCOS was performed.^{4,11} This study constitutes the first validation of the CCOS in an external pediatric neurosurgical population.

METHODS

Following approval from the Washington University Institutional Review Board, a retrospective review was performed by searching for all consecutive patients treated for CM-I at Saint Louis Children’s Hospital (SLCH) over a 12-year period from 2001–2012. Inclusion criteria for this study were: 1) patients receiving primary posterior fossa decompression at SLCH between 2001 and 2012; 2) at least 5.5 months of post-operative clinical follow-up; and 3) patients ≥ 18 years of age at the time of surgery. 5.5 months of follow-up was chosen ensure the availability of longer-term data while also capturing patients whose six-month appointment occurred shortly before six months postoperatively. Exclusion criteria were: 1) previous surgical treatment for CM-I prior to operative treatment at SLCH; 2) insufficient chart information to evaluate clinical outcome.

A single reviewer analyzed the charts of all patients meeting inclusion criteria. Clinical outcomes were evaluated using the recently-published CCOS.⁴ As in the initial study describing CCOS, each patient was assigned a gestalt score based on an overall impression of whether a patient benefited significantly from surgery. When assigning gestalt scores, a

patient was considered “improved” if all major symptoms improved to a significant degree or resolved entirely, and overall quality of life improved. If minor symptoms persisted, a patient was still considered improved. A patient was considered “unchanged” if: some or all major symptoms failed to improve to a significant degree after surgery; new CM-1-related symptoms developed after surgery that caused significant impairment; or a patient required repeat surgery for CM-1-related symptoms or syringomyelia, regardless of final outcome. A patient was considered “worse” if overall quality of life decreased as a result of increased severity of major preoperative symptoms or the development of new CM-1-related symptoms postoperatively. When analyzing outcomes for headache symptoms, only occipital headaches and headache types present preoperatively were considered. Both CCOS and gestalt outcome scores were assigned based on each patient’s most recent clinic follow-up visit.

To evaluate interrater reliability, a second reviewer independently analyzed the charts of 73 consecutive patients treated from 2008–2010 and assigned CCOS and gestalt scores. Both reviewers adhered to the guidelines for the CCOS originally published. After all score assignments were made, cases of significantly differing CCOS scores were reviewed to identify potential points of ambiguity in the system.

Statistical Analysis

The ability of the CCOS to discriminate patients with gestalt improved outcome was investigated by examining the area under the receiver operating characteristic (ROC) curve. For this analysis, gestalt improved outcome was used as the gold standard for disease state. The area under the ROC curve represents the probability that a randomly chosen patient will be correctly assigned to the appropriate outcome category. This area ranges from 0.5 (the test being evaluated has no accuracy) to 1.0 (perfect accuracy).¹⁰ The coordinates of the curve were used to find the CCOS score with optimal sensitivity and specificity for identifying patients with good outcome. To determine which components of the CCOS score were most predictive of gestalt outcome, a logistic regression analysis was done, including the four component scores of the CCOS as predictor variables. For both the ROC and logistic regression analyses, “unchanged” and “worse” outcomes were grouped together. Intraclass correlation coefficients (ICC) and kappa statistics were calculated to test the interrater reliability of the CCOS and gestalt impression scores respectively. All statistical analyses were done using SPSS version 21 (IBM, Armonk, NY). P values < 0.05 were considered significant.

RESULTS

Patient Characteristics

292 consecutive patients received primary CM-1 decompression between 2001 and 2012 at SLCH (Table 2), and 215 patients were available for analysis. A subgroup of consecutive patients undergoing operations from 2008 to 2010 (n=118) was selected for analysis of interrater reliability, with 73 patients meeting all inclusion and exclusion criteria. This period was chosen to maximize the availability of relevant medical record data and provide for adequate follow-up with consecutive patients. Our institution underwent a migration to

the current electronic medical records system prior to 2008. In addition to the interrater analysis, all patients were evaluated by one reviewer using both the gestalt outcome and CCOS.

Patient Outcome

Among the 215 included patients, the median CCOS was 14, and 67% had improved gestalt outcome. This mean follow-up was 37 ± 29 months. Comparison of the CCOS and gestalt outcome revealed a correlation between good gestalt outcome and high CCOS (Figure 1). All patients with a CCOS of 15 or greater had improved gestalt outcome, whereas no patients with a CCOS of 12 or lower had improved outcome.

The ability of the CCOS to discriminate patients with good gestalt outcome was plotted using an ROC curve (Figure 2). The ROC plots sensitivity against (1-specificity), and shows the probability that a randomly chosen patient would be assigned the correct gestalt outcome based on CCOS. The area under the ROC curve was 0.951, suggesting excellent prediction of gestalt outcome. Sensitivities and specificities for the CCOS were also calculated to determine the optimal CCOS value to use as a cutoff score (Table 3). These results showed that a CCOS value of 14 has excellent sensitivity (0.95) and good specificity (0.746) for identifying patients with good gestalt outcome.

To test whether certain components of the CCOS have a stronger relationship than others with gestalt outcome, logistic regression for each subscore was performed (Table 4). The results of this analysis showed increased likelihood of improved gestalt outcome with higher CCOS subscores for all domains except function. Though the confidence intervals for the odds ratios were wide, the pain and non-pain component scores appeared to have a particularly large impact on the probability of a positive surgical outcome compared to the complications score. Given that the functionality subscore was not a significant predictor of gestalt outcome, we also tested a regression model that excluded this component of the CCOS. The Likelihood Ratio statistic was 1.8 ($p=0.18$), indicating that the inclusion of the functionality component did not significantly improve the overall predictive ability of the model.

Reliability Assessment of CCOS

All patients who underwent Chiari decompression between 2008 and 2010 and met inclusion criteria were analyzed by independent reviewers (Table 5). ICCs were calculated, showing moderate to excellent agreement for the composite CCOS (0.71) and subscores for complications (0.89), pain (0.63), and non-pain sign and symptoms (0.58). ICC for the functionality score (0.23) was poor. Review of cases in disagreement showed different interpretations of the functionality score guidelines, as well as inconsistent results due to limited record documentation. Overall, the two reviewers obtained identical total CCOS for 43% of patients. In the patients with differing CCOS, the mean difference between reviewers was 1.4 (range 1–3). The k statistic for gestalt outcome assessment was 0.56, showing moderate agreement. However, after identifying the functionality score as having lower interrater reliability than the overall CCOS and other subscores, the data were reanalyzed excluding the functionality subscore. A new ROC was plotted with an area under

the curve of 0.945, similar when compared with the original ROC curve. Additionally, ICC was calculated at 0.74 when functionality subscore was removed. Subscore ICCs were unchanged.

DISCUSSION

CM-1 is a common referring diagnosis to pediatric neurosurgical practices. Assessment of patient outcomes is complicated by lack of validated, quantifiable outcome measures. A gestalt assessment algorithm of “better, unchanged, or worse” has been in use at our institution and others, though this assessment has obvious weaknesses in its failure to provide quantifiable results. Aliaga *et al* developed the Chicago Chiari Outcome Scale to address these weaknesses.⁴ While the CCOS was applied to patients at the initial authors’ home institution, it has not been externally verified.^{4,11} This study constitutes the first validation of the CCOS by an external pediatric neurosurgical practice.

In our statistical assessment of the CCOS, there was a clear correlation between higher CCOS and gestalt outcome (Figure 1). Additionally, two independent raters showed moderate to good agreement in composite score and all subscores of the CCOS except the functionality score (Table 5). In cases where we observed disagreement, the average composite CCOS difference was 1.4, with most common disagreement being one point. This difference could result from retrospective bias, as the clinical charts were not designed to capture fully the CCOS. However, despite some inconsistencies, the composite CCOS showed good interrater agreement relative to gestalt outcome. Further, our logistic regression analysis of the CCOS showed that each subcomponent of the CCOS except for functionality had a strong impact on the likelihood for an improved outcome (Table 4).

Based on the observed inconsistency in the assignment of the functionality subscore and its uncertain impact on the relationship between CCOS and gestalt outcome, we examined the scoring methods and contribution of this subscore to the composite CCOS in detail. Indeed, we found some ambiguity in scoring functionality; specifically it can be difficult to distinguish between subscores 2 (able to work or go to school <50%) and 3 (able to work or go to school >50%), and between subscores 3 (able to work or go to school >50%) and 4 (fully functional). We found that many patients have some minor complaints that may or may not result from CM-1, but cannot be ruled out based on clinical exam or from the documentation available in the clinical chart. The subscore 3, in particular, comprises a large group of potential patients who are not fully symptom-free, but are not completely debilitated either. Delving deeper, we found that the functionality subscore did not contribute significantly to the predictive ability of the logistic regression model; thus, our findings indicate that the CCOS may be improved by clarifying the definition and scoring of the functionality subscore or perhaps by removing it altogether.

An additional area of ambiguity within the CCOS is the headache subscore. Headaches are common in the general population, with yearly prevalence over 50% in adolescents and children.¹⁶ Thus, a large portion of these patients will ultimately have clinical courses complicated by a headache syndrome unrelated to CM-1. To control for this fact in our patient population, we limited recurrent headache syndromes to occipital headaches, post-

tussive or exertional headaches, or headaches similar to that which the patient suffered at initial presentation. Though we feel this is a rational limitation, it invites some ambiguity by introducing examiner's interpretation of what is truly related to CM-1. This factor may be inescapable, as headaches are subjective experiences. Nonetheless, we feel that some attempt to focus the evaluation of headache symptoms is essential.

The Chicago Chiari Outcome Scale is a strong step forward in providing an outcomes assessment tool for treatment of patients with CM-1. However, it does not fully remove subjectivity from the clinical assessment, as different interpretations are possible in at least the headache and functionality subscores. Additionally, we feel that there is some difficulty in adequately separating patients with severe symptoms who have minor improvements and in patients who suffer complications but also show symptomatic improvement. There are relatively few patients with subscores of "1" in any category, and scores for "improved" (median CCOS=15) and "unchanged or worsened" (median CCOS=13) gestalt outcomes show significant clustering at the higher end of the CCOS. Thus, while we are able to validate the CCOS adequately predicting gestalt outcome, additional refinements may be necessary to overcome these challenges.

CONCLUSION

In this paper, we have evaluated the CCOS at our institution and found it be more consistently applied between examiners than the previously utilized gestalt assessment. This score is clearly an improvement over the status quo, and should allow for more subtle analysis of a large and challenging patient population. As currently defined, the functionality subscore allows for more ambiguity in interpretation than the other subscores without augmenting the CCOS's ability to estimate outcomes. Overall, the CCOS represents a major advance in assessing clinical outcomes in CM-1, but this tool may benefit from further refinement, including modifying or eliminating the functionality subscore.

Acknowledgements

This study was supported in part by the Park-Reeves Syringomyelia Research Consortium, and by the Washington University Institute of Clinical and Translational Sciences grants UL1 TR000448 and TL1 TR000449 from the National Center for Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The authors acknowledge Michael Wallendorf, Ph.D. for his suggestions regarding statistical analysis.

Abbreviations

CM-1	Chiari Malformation Type 1
CCOS	Chicago Chiari Outcome Scale
ICC	Intraclass correlation coefficient
ROC	Receiver operating characteristic
SLCH	St. Louis Children's Hospital

REFERENCES

1. Aboulezz AO, Sartor K, Geyer CA, Gado MH. Position of cerebellar tonsils in the normal population and in patients with Chiari malformation: a quantitative approach with MR imaging. *J Comput Assist Tomogr.* 1985; 9:1033–1036. [PubMed: 4056132]
2. Aitken LA, Lindan CE, Sidney S, Gupta N, Barkovich AJ, Sorel M, et al. Chiari type I malformation in a pediatric population. *Pediatr Neurol.* 2009; 40:449–454. [PubMed: 19433279]
3. Alden TD, Ojemann JG, Park TS. Surgical treatment of Chiari I malformation: indications and approaches. *Neurosurg Focus.* 2001; 11:E2. [PubMed: 16724812]
4. Aliaga L, Hekman KE, Yassari R, Straus D, Luther G, Chen J, et al. A novel scoring system for assessing Chiari malformation type I treatment outcomes. *Neurosurgery.* 2012; 70:656–664. discussion 664–655. [PubMed: 21849925]
5. Alzate JC, Kothbauer KF, Jallo GI, Epstein FJ. Treatment of Chiari I malformation in patients with and without syringomyelia: a consecutive series of 66 cases. *Neurosurg Focus.* 2001; 11:E3. [PubMed: 16724813]
6. Caldarelli M, Novegno F, Vassimi L, Romani R, Tamburrini G, Di Rocco C. The role of limited posterior fossa craniectomy in the surgical treatment of Chiari malformation Type I: experience with a pediatric series. *J Neurosurg.* 2007; 106:187–195. [PubMed: 17465383]
7. Di X. Endoscopic suboccipital decompression on pediatric Chiari type I. *Minim Invasive Neurosurg.* 2009; 52:119–125. [PubMed: 19650014]
8. Erdogan E, Cansever T, Secer HI, Temiz C, Sirin S, Kabatas S, et al. The evaluation of surgical treatment options in the Chiari Malformation Type I. *Turk Neurosurg.* 2010; 20:303–313. [PubMed: 20669102]
9. Galarza M, Sood S, Ham S. Relevance of surgical strategies for the management of pediatric Chiari type I malformation. *Childs Nerv Syst.* 2007; 23:691–696. [PubMed: 17252266]
10. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982; 143:29–36. [PubMed: 7063747]
11. Hekman KE, Aliaga L, Straus D, Luther A, Chen J, Sampat A, et al. Positive and negative predictors for good outcome after decompressive surgery for Chiari malformation type 1 as scored on the Chicago Chiari Outcome Scale. *Neurol Res.* 2012; 34:694–700. [PubMed: 22781921]
12. Milhorat TH, Chou MW, Trinidad EM, Kula RW, Mandell M, Wolpert C, et al. Chiari I malformation redefined: clinical and radiographic findings for 364 symptomatic patients. *Neurosurgery.* 1999; 44:1005–1017. [PubMed: 10232534]
13. Navarro R, Olavarria G, Seshadri R, Gonzales-Portillo G, McLone DG, Tomita T. Surgical results of posterior fossa decompression for patients with Chiari I malformation. *Childs Nerv Syst.* 2004; 20:349–356. [PubMed: 15022006]
14. Parker SL, Godil SS, Zuckerman SL, Mendenhall SK, Wells JA, Shau DN, et al. Comprehensive assessment of 1-year outcomes and determination of minimum clinically important difference in pain, disability, and quality of life after suboccipital decompression for Chiari malformation I in adults. *Neurosurgery.* 2013; 73:569–581. [PubMed: 23787878]
15. Parker SR, Harris P, Cummings TJ, George T, Fuchs H, Grant G. Complications following decompression of Chiari malformation Type I in children: dural graft or sealant? *J Neurosurg Pediatr.* 8:177–183. [PubMed: 21806360]
16. Stovner L, Hagen K, Jensen R, Katsarava Z, Lipton R, Scher A, et al. The global burden of headache: a documentation of headache prevalence and disability worldwide. *Cephalalgia.* 2007; 27:193–210. [PubMed: 17381554]
17. Tubbs RS, Beckman J, Naftel RP, Chern JJ, Wellons JC 3rd, Rozzelle CJ, et al. Institutional experience with 500 cases of surgically treated pediatric Chiari malformation Type I. *J Neurosurg Pediatr.* 2011; 7:248–256. [PubMed: 21361762]
18. Ventureyra EC, Aziz HA, Vassilyadi M. The role of cine flow MRI in children with Chiari I malformation. *Childs Nerv Syst.* 2003; 19:109–113. [PubMed: 12607030]

Figure 1. Comparison of CCOS and gestalt outcomes

Clear alignment of higher CCOS with better gestalt outcome is shown, with all patients scoring 15 or 16 on CCOS having improved gestalt outcome. Conversely, all except one patient with CCOS 10 or lower were worse than preoperative status on gestalt assessment.

Figure 2. Receiver operator curve for CCOS

ROC had an area under the curve of 0.951, showing excellent fit of the CCOS to predict outcome on the gestalt assessment.

Table 1
Chicago Chiari Outcome Scale

Details of the CCOS, as reported in Aliaga et al.⁴

Characteristic	Score			
	1	2	3	4
Pain	Worse	Unchanged and Refractory to Medication	Improved or Controlled with Medication	Resolved
Non-pain	Worse	Unchanged or improved but impaired	Improved and unimpaired	Resolved
Functionality	Unable to attend	Moderate impairment (<50% attendance)	Mild impairment (>50% attendance)	Fully functional
Complication	Persistent complication, poorly controlled	Persistent complication, well-controlled	Transient complication	Uncomplicated course

Table 2**Patient Characteristics**

Demographic and outcome characteristics of patients from 2001–2012 rated by a single reviewer.

Characteristic	Value
Mean age in years (range)	10 ± 5.1 (0.5–18.7 years)
Sex F (%)	126 (59)
Mean postoperative follow-up in months (range)	37 ± 29 (5.7–142 months)
Median CCOS (range)	14 (7–16)
Improved gestalt outcome (%)	144 (67)

Table 3**Coordinates of the ROC curve**

Sensitivity and specificity of the CCOS for predicting good gestalt outcome, rated by a single reviewer (n=215). A positive test indicates good gestalt outcome.

Positive if Greater Than or Equal To	Sensitivity	Specificity
16	0.33	1.00
15	0.72	1.00
14	0.95	0.746
13	1.00	0.437
12	1.00	0.225
11	1.00	0.07
10	1.00	0.06
9	1.00	0.03
8	1.00	0.01
7	1.00	0.00

Table 4
Logistic regression of CCOS subscores

A logistic regression analysis was calculated. Higher subscores for complication, pain, and non-pain symptoms were associated with improved gestalt outcome.

Variable	B (SE)	Odds ratio	95% CI of OR
Complications	2.0 (0.5)	7.3	2.9–18.7
Pain	3.9 (0.7)	47.2	12.3–180.6
Non-pain	3.8 (0.7)	43.4	11.7–160.3
Functionality	1.2 (1.0)	3.2	0.5–22.4

Nagelkerke $R^2 = 0.81$

Table 5
Interclass coefficients of CCOS and subscores

ICC was calculated for total score and for all component measures. Moderate to excellent agreement was noted in all except functionality subscore. After reviewing areas of disagreement, the authors noted that both interpretation of the CCOS system as well as interpretation of patient records likely affected this subscore more than others.

Measure	ICC
CCOS	0.71
Complication Score	0.89
Pain Score	0.63
Non-pain Score	0.58
Function Score	0.23

ICC > 0.7 is reliable, by certain conventions