12-1-2016

Inter-rater reliability of post-arrest cerebral performance category (CPC) scores.

Anne V. Grossestreuer

Benjamin S. Abella

Kelsey R. Sheak

Marisa J. Cinousis

Sarah M. Perman

See next page for additional authors

Authors
Anne V. Grossestreuer, Benjamin S. Abella, Kelsey R. Sheak, Marisa J. Cinousis, Sarah M. Perman, Marion Leary, Douglas J. Wiebe, and David F. Gaieski

HHS Public Access
Author manuscript
Resuscitation. Author manuscript; available in PMC 2018 June 06.

Published in final edited form as:
Resuscitation. 2016 December ; 109: 21–24. doi:10.1016/j.resuscitation.2016.09.006.

Inter-rater reliability of post-arrest cerebral performance category (CPC) scores☆

Anne V. Grossestreuer[a,b,c,*,1], **Benjamin S. Abella**[a], **Kelsey R. Sheak**[a], **Marisa J. Cinousis**[a], **Sarah M. Perman**[d], **Marion Leary**[a,e], **Douglas J. Wiebe**[b], and **David F. Gaieski**[f]

[a]Center for Resuscitation Science, Department of Emergency Medicine, University of Pennsylvania, Philadelphia, PA, United States

[b]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, United States

[c]Leonard Davis Institute of Healthcare Economics, University of Pennsylvania, Philadelphia, PA, United States

[d]Department of Emergency Medicine, University of Colorado School of Medicine, Aurora, CO, United States

[e]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[f]Department of Emergency Medicine, Thomas Jefferson University, Philadelphia, PA, United States

Abstract

Purpose—Cerebral Performance Category (CPC) scores are often an outcome measure for post-arrest neurologic function, collected worldwide to compare performance, evaluate therapies, and formulate recommendations. At most institutions, no formal training is offered in their determination, potentially leading to misclassification.

Materials and Methods—We identified 171 patients at 2 hospitals between 5/10/2005 and 8/31/2012 with two CPC scores at hospital discharge recorded independently – in an in-house quality improvement database and as part of a national registry. Scores were abstracted retrospectively from the same electronic medical record by two separate non-clinical researchers. These scores were compared to assess inter-rater reliability and stratified based on whether the score was concordant or discordant among reviewers to determine factors related to discordance.

☆A Spanish translated version of the summary of this article appears as Appendix in the final online version at http://dx.doi.org/10.1016/j.resuscitation.2016.09.006.

*Corresponding author at: Center for Resuscitation Science, Department of Emergency Medicine, University of Pennsylvania, 423 Guardian Drive, Blockley Hall, Room 408, Philadelphia, PA 19104, United States. agrosses@upenn.edu (A.V. Grossestreuer).
[1]Tel.: +1 630 989 5236; fax: +1 215 662 3953

Conflict of interest statement
Dr. Grossestreuer, Ms. Sheak, Ms. Cinousis, and Dr. Wiebe report no conflicts of interest related to this study.

Appendix A. Supplementary data
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.resuscitation.2016.09.006.

Results—Thirty-nine CPC scores (22.8%) were discordant (kappa: 0.66), indicating substantial agreement. When dichotomized into "favorable" neurologic outcome (CPC 1–2)/"unfavorable" neurologic outcome (CPC 3–5), 20 (11.7%) scores were discordant (kappa: 0.70), also indicating substantial agreement. Patients discharged home (as opposed to nursing/other care facility) and patients with suspected cardiac etiology of arrest were statistically more likely to have concordant scores. For the quality improvement database, patients with discordant scores had a statistically higher median CPC score than those with concordant scores. The registry had statistically lower median CPC score (CPC 1) than the quality improvement database (CPC 2); $p < 0.01$ for statistical significance.

Conclusions—CPC scores have substantial inter-rater reliability, which is reduced in patients who have worse outcomes, have a non-cardiac etiology of arrest, and are discharged to a location other than home.

Keywords

Heart arrest; Outcome measures; Epidemiology; Neurologic signs and symptoms

Introduction

Some cardiac arrest survivors have neurologic deficits as a result of the hypoxic-ischemic insult of cardiac arrest and subsequent reperfusion injury.[1] The Utstein guidelines recommend measuring the neurologic function of cardiac arrest survivors for research purposes using either a Cerebral Performance Category (CPC) score or a modified Rankin scale (mRS).[2] The CPC score (Table 1), a five point scale that ranges from good cerebral performance (1) to brain death (5),[2,3] is often used and is commonly dichotomized into "good" (CPC 1–2) versus "poor" (CPC 3–5) outcome when used as a primary or secondary outcome in research studies and quality improvement.[4–6]

Although CPC scores are a recommended data element and collected at centers worldwide for use in comparing outcomes, evaluating therapies, and formulating recommendations, no formal standardized training is offered in calculating CPC scores.[7] These scores are assessed by persons with varying backgrounds and levels of clinical training, and the extent of inter-reviewer reliability in this context has not been fully evaluated. As poor interviewer reliability may produce misclassification and bias, this study sought to assess the degree of inter-rater reliability when assessing post-arrest CPC scores at discharge retrospectively from the electronic medical record and to identify factors that may be associated with discordance across ratings.

Methods

The investigation was a retrospective cohort study of persons who survived in-hospital and out-of-hospital cardiac arrest, were discharged from one of two hospitals, and had CPC classification in two distinct databases between May 2005 and September 2012. We utilized two sources of data for this study, an in-house quality improvement database and a national cardiac arrest registry. Cardiac arrest in both data sources was defined as a loss of pulse with subsequent chest compressions. The in-house quality improvement database is maintained by research personnel to monitor cardiac arrest patients at two hospitals within the University of Pennsylvania Health System, the Hospital of the University of Pennsylvania and Penn Presbyterian Medical Center. The national registry, the Penn Alliance for Therapeutic Hypothermia (PATH), is an internet-based registry established by the University of Pennsylvania. Potentially available to any US hospital, it supports the tracking of patients who experience cardiac arrest and receive cardiopulmonary resuscitation. This study had research approval from the University of Pennsylvania Institutional Review Board.

Prior to September 2012, 171 patients with cardiac arrest who survived to hospital discharge (when CPC score was measured) were entered into both the quality improvement database and the national registry by two separate non-clinically trained research personnel. Research personnel entered data retrospectively using the same electronic medical record (EMR), with the first shared cardiac arrest occurring in May 2005 and the last in August 2012. The EMR contained all information from the patient's cardiac arrest hospitalization, including nursing notes, vital signs, consultations, pathology/imaging reports, medications, and discharge summaries. Data abstractors had all information in the EMR available; items such as age, initial rhythm, and suspected etiology of arrest as well as results from neurologic examinations (if performed) and measurements of functional status were potentially present during CPC assessment. Patients who did not survive to hospital discharge were not included because they are automatically assigned a CPC score of 5.

This situation created a "natural experiment" whereby the CPC scores abstracted by one researcher on a cardiac arrest patient could be compared to those abstracted by another researcher on the same patient using the same data source for the same purposes. Although there were multiple research personnel abstracting data from the EMR for both the registry and the quality improvement database, there was no overlap between the two.

Statistical analysis

Scores were compared to assess inter-rater reliability, using a Cohen's kappa statistic[9] Patients then were stratified based on whether the score was concordant or discordant among reviewers and analyzed to determine factors related to discordance, including demographics, arrest characteristics, post-arrest therapy, and discharge information. Descriptive statistics used proportions, means and standard deviations, medians and interquartile ranges, and histograms to determine the proportion or prevalence and distribution of each variable. Normally distributed continuous data were analyzed using Student's *t*-test with either equal or unequal variance, depending on the distribution of the data. Non-parametric continuous data were analyzed using a Wilcoxon rank-sum test, and proportions were tested using either a Chi-square or Fisher's exact test, depending on the number of observations in each category. A Wilcoxon matched-pairs signed-rank test was used to compare median CPC scores between the two data sources. No multivariate analysis was performed.

Results

Between 5/2005 and 8/2012, there were 418 cardiac arrest survivors at the two study institutions. In the 171 cardiac arrest survivors with two recorded CPC scores (41%) analyzed, the mean age was 56 ± 17 years, 62% were male, 44% had an initial shockable rhythm of ventricular fibrillation or pulseless ventricular tachycardia (VF/pVT), 62% had an out-of-hospital cardiac arrest (OHCA), 82% had a witnessed cardiac arrest, 66% had a suspected cardiac etiology of arrest, 78% were treated with targeted temperature management (TTM), and 46% were discharged directly to home from the hospital.

Thirty-nine CPC scores (22.8%) were recorded differently between data sources (Table 2a). Cohen's kappa was 0.66, indicating substantial agreement. When dichotomized into "good" neurologic outcome (CPC 1–2) or "poor" neurologic outcome (CPC 3–5), only 20 (11.7%) scores were discordant (Table 2b), with a corresponding kappa of 0.70, also indicating substantial agreement. When stratified based on whether scores were concordant or discordant, patients discharged home (as opposed to nursing/other care facility) and patients with suspected cardiac etiology of arrest were statistically more likely to have concordant scores. For the quality improvement database, patients with discordant scores had a statistically less favorable median CPC score than those with concordant scores. The registry had statistically more favorable median CPC score than the quality improvement database (Table 3; $p < 0.01$ for statistical significance).

Examining inter-rater reliability by patient demographics or arrest characteristics, we found that the Cohen's kappa dropped to moderate agreement in patients with asystole as their initial rhythm, in-hospital arrests, unwitnessed arrests, arrests with non-cardiac etiology, patients not treated with TTM, and patients not discharged home. When the CPC score was dichotomized, the Cohen's kappa dropped to fair agreement in patients with asystole as their initial rhythm, patients not treated with TTM, and patients discharged home. There was moderate agreement in males and in-hospital arrests. However, the Cohen's kappa for the dichotomized CPC score rose to almost perfect agreement in the patients treated with TTM and females (Supplemental Table 1).

Two hundred and forty-seven patients (59%) did not have CPC scores recorded in both the registry and the quality improvement database between 5/2005 and 8/2012. Patients with multiple scores were younger ($p = 0.005$), had a higher proportion of treatment with TTM ($p < 0.001$), were more likely to have an OHCA ($p = 0.003$), and had a longer duration of arrest ($p < 0.001$) than patients with only a recorded score in one data source (Supplemental Table 2).

Discussion

This study found substantial agreement in terms of inter-rater reliability for CPC scores assessed at hospital discharge by non-clinical research personnel from a retrospectively analyzed EMR. However, almost one-quarter of scores were recorded differently, and over 10 percent of patients were classified differently with regard to a poor versus a good outcome. Patient factors, particularly the location to which the patient was discharged and the suspected etiology of arrest, were associated with discordant assessment. Also, there was unidirectional bias in the results, with the registry reporting more favorable outcomes overall than the quality improvement database. Although the substantial agreement between CPC scores indicates that the CPC score is a reasonable measure of neurologic outcome in post-arrest patients, especially given its ability to be assessed retrospectively in an efficient and low-cost manner, the variability found in this study may be helpful to note when interpreting and powering studies using CPC scores as an outcome.

This investigation expands the work of Ajam et al, who measured inter-rater reliability for CPC scores in patients presenting with a VF arrest.[10] The results in both studies were similar: CPC scores were variable across raters, with unidirectional differences. The current study included all arrest rhythms, and was able to look at factors that might influence the concordance or discordance of CPC results; that is, whether there were patient or arrest factors that made it easier or harder to assess neurologic status using this scale. With regard to initial pulseless rhythm, we found that inter-rater reliability dropped dramatically in patients with an initial rhythm of asystole, who tended to have worse outcomes.

Assessments of the utility of the CPC score have been performed. In 1996, when the CPC was compared to the validated Functional Status Questionnaire in patients who had OHCA, it was not found to be well correlated.[11] The researchers also concluded that CPC did not adequately measure post-arrest quality of life. More recently, the CPC score was found to accurately identify health-related quality of life, as measured by the Health Utilities Index (HUI).[12] Researchers in that study recommended that the HUI be used in conjunction with the CPC, not as a replacement, as the CPC attempts to measure overall functional outcome while the HUI addresses specific domains such as quality of life.

Another study used the Computerized Assessment for Mild Cognitive Injury (CAMCI) to assess post-arrest cognitive deficits. It was found that many patients with a good CPC/mRS measure had cognitive injury, particularly with regard to executive function and recall, potentially leaving survivors with unacknowledged deficits that might be amenable to rehabilitation.[13] The Cerebral Recovery Index, which is a combination of five quantitative electroencephalography (qEEG) features applicable to patients treated with post-arrest TTM, showed evidence of predictive ability for neurologic outcome, but was validated using the CPC score as a gold standard.[14]

When changes in CPC and mRS scores measured at hospital discharge were compared those measured one month later, it was found that CPC scores at discharge tended to be more favorable than those one month later, leading to concern that neurologic deficits might be missed at discharge.[15] A comparison of CPC score, mRS, and discharge disposition found that while the CPC score and modified Rankin scale had a moderate amount of agreement, neither measure correlated strongly with discharge disposition, which has been occasionally used as a surrogate for the CPC or mRS score.[16] This finding is similar to our finding that discordance in CPC rating was significantly increased in patients who were not discharged home.

An alternative to the CPC score, the Cerebral Performance Categories-Extended (CPC-E), has been proposed as a measure to capture impairment or disability that may be missed by the current CPC score.[17] It also standardizes administration, reducing variability, and provides a more nuanced profile of a patient's impairments and disabilities, which could facilitate timely referrals to rehabilitation or community support services targeted to a particular deficit. Although the CPC-E was developed and validated at a single center, it potentially could be a replacement measure for the CPC score.

Many of these proposed alternatives to the CPC score are likely more costly and time-intensive due to the requirement for an in-person evaluation or follow-up after hospital discharge by research personnel. The latter is problematic given the results of a recent study of OHCAs in which only 56% of cardiac arrest survivors consented to participate in telephone follow-up after hospital discharge, especially as neurologic function differed significantly between those who agreed to participate and those who did not consent.[18] This may make the current CPC alternatives impractical for large-scale adoption, especially with regard to quality improvement efforts, and limit their generalizability.

Despite its limitations, the CPC score at hospital discharge was shown in two studies, one in all patients treated with TTM[19] and one in patients experiencing OHCA,[20] to predict long-term survival outcomes in post-arrest patients, with favorable CPC scores associated with a greater likelihood of long-term survival.[19,20] Recently, the construct validity of the CPC score was assessed with the conclusion that a CPC can be used to measure functional outcome, particularly when administered as a standardized interview.[7] Although 23% of raw CPC scores and 12% of dichotomized CPC were discordant in this study, they still showed substantial agreement in inter-rater reliability overall. Contradicting evidence regarding the utility and accuracy of the use of CPC scores warrants more investigation.

Our study has several limitations. No multivariate analysis was performed on variables potentially associated with discrepancy in CPC assessment, so this represents a preliminary assessment of association. As an uncontrolled experiment that occurred naturally, the patient population receiving double CPC score assessment is not necessarily representative of all cardiac arrest survivors. The study cohort was younger, had a high proportion of treatment with TTM, was more likely to have an OHCA, and had a longer duration of arrest than the total cohorts of patients in the two data sources (Supplemental Table 3). The reason these cases were assessed twice is unknown; it could be a matter of research priorities, the manner in which data collection was approached, chance, or something else. Another limitation is the lack of two single raters for either the registry or the quality improvement database. Although no rater was involved in both data sources, and only non-clinically trained research personnel abstracted the CPC scores, there is not documentation of which abstractor or a precise count of how many abstractors provided the CPC score in each data set. However, having multiple reviewers likely increases the role of chance, making it possible that our values underestimate the true interrater reliability. Finally, discrepancies in CPC score recording could be a function of data entry errors as opposed to differences in CPC assessment.

Conclusions

Inter-reviewer reliability of the CPC is imperfect. This reliability is reduced in patients who have worse outcomes, have a non-cardiac etiology of arrest, and are discharged to a location other than home. These findings may be important when devising approaches to improve

CPC reliability and when interpreting and powering studies using CPC scores as an outcome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Ms. Leary has received research support from the Laerdal Foundation and the American Heart Association, as well as in-kind support from Laerdal Medical and Physio-Control. Dr. Gaieski provides consulting for CR Bard and Stryker Medical and has received honoraria from CR Bard. Dr. Abella has received research funding from the National Institutes of Health, the Patient-Centered Outcomes Research Institute, American Heart Association and CR Bard, in-kind research support from Laerdal Medical Corporation, and honoraria from CR Bard. Dr. Perman receives research support from the National Institutes of Health.

References

1. Moulaert VR, Verbunt JA, van Heugten CM, Wade DT. Cognitive impairments in survivors of out-of-hospital cardiac arrest: a systematic review. Resuscitation. 2009; 80:297–305. [PubMed: 19117659]
2. Perkins GD, Jacobs IG, Nadkarni VM, et al. Cardiac Arrest and Cardiopulmonary Resuscitation Outcome Reports: Update of the Utstein Resuscitation Registry Templates for Out-of-Hospital Cardiac Arrest: A Statement for Healthcare Professionals From a Task Force of the International Liaison Committee on Resuscitation (American Heart Association, European Resuscitation Council, Australian and New Zealand Council on Resuscitation, Heart and Stroke Foundation of Canada, InterAmerican Heart Foundation, Resuscitation Council of Southern Africa, Resuscitation Council of Asia); and the American Heart Association Emergency Cardiovascular Care Committee and the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation. Resuscitation. 2015; 96:328–40. [PubMed: 25438254]
3. Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet. 1975; 1:480–4. [PubMed: 46957]
4. Hypothermia after Cardiac Arrest Study Group. Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. N Engl J Med. 2002; 346:549–56. [PubMed: 11856793]
5. Bernard SA, Gray TW, Buist MD, et al. Treatment of comatose survivors of out-of-hospital cardiac arrest with induced hypothermia. N Engl J Med. 2002; 346:557–63. [PubMed: 11856794]
6. Nielsen N, Wetterslev J, Cronberg T, et al. Targeted temperature management at 33 degrees C versus 36 degrees C after cardiac arrest. N Engl J Med. 2013; 369:2197–206. [PubMed: 24237006]
7. Mak M, Moulaert VR, Pijls RW, Verbunt JA. Measuring outcome after cardiac arrest: construct validity of Cerebral Performance Category. Resuscitation. 2015; 100:6–10. [PubMed: 26744101]
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–74. [PubMed: 843571]
10. Ajam K, Gold LS, Beck SS, Damon S, Phelps R, Rea TD. Reliability of the Cerebral Performance Category to classify neurological status among survivors of ventricular fibrillation arrest: a cohort study. Scand J Trauma Resusc Emerg Med. 2011; 19:38. http://dx.doi.org/10.1186/1757-7241-19-38. [PubMed: 21672267]
11. Hsu JW, Madsen CD, Callaham ML. Quality-of-life and formal functional testing of survivors of out-of-hospital cardiac arrest correlates poorly with traditional neurologic outcome scales. Ann Emerg Med. 1996; 28:597–605. [PubMed: 8953946]
12. Stiell IG, Nesbitt LP, Nichol G, et al. Comparison of the Cerebral Performance Category score and the Health Utilities Index for survivors of cardiac arrest. Ann Emerg Med. 2009; 53:241–8. [PubMed: 18450329]
13. Sabedra AR, Kristan J, Raina K, et al. Neurocognitive outcomes following successful resuscitation from cardiac arrest. Resuscitation. 2015; 90:67–72. [PubMed: 25737082]

14. Tjepkema-Cloostermans MC, van Meulen FB, Meinsma G, van Putten MJ. A Cerebral Recovery Index (CRI) for early prognosis in patients after cardiac arrest. Crit Care. 2013; 17:R252. [PubMed: 24148747]

15. Raina KD, Callaway C, Rittenberger JC, Holm MB. Neurological and functional status following cardiac arrest: method and tool utility. Resuscitation. 2008; 79:249–56. [PubMed: 18692288]

16. Rittenberger JC, Raina K, Holm MB, Kim YJ, Callaway CW. Association between Cerebral Performance Category, Modified Rankin Scale, and discharge disposition after cardiac arrest. Resuscitation. 2011; 82:1036–40. [PubMed: 21524837]

17. Balouris SA, Raina KD, Rittenberger JC, Callaway CW, Rogers JC, Holm MB. Development and validation of the Cerebral Performance Categories-Extended (CPC-E). Resuscitation. 2015; 94:98–105. [PubMed: 26025569]

18. Nichol G, Guffey D, Stiell IG, et al. Post-discharge outcomes after resuscitation from out-of-hospital cardiac arrest: a ROC PRIMED substudy. Resuscitation. 2015; 93:74–81. [PubMed: 26025570]

19. Hsu CH, Li J, Cinousis MJ, et al. Cerebral performance category at hospital discharge predicts long-term survival of cardiac arrest survivors receiving targeted temperature management. Crit Care Med. 2014; 42:2575–81. [PubMed: 25072759]

20. Phelps R, Dumas F, Maynard C, Silver J, Rea T. Cerebral Performance Category and long-term prognosis following out-of-hospital cardiac arrest. Crit Care Med. 2013; 41:1252–7. [PubMed: 23388519]

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Cerebral performance category (CPC) score scale.[a]

CPC 1	Good cerebral performance: conscious, alert, able to work, might have mild neurologic or psychologic deficit.
CPC 2	Moderate cerebral disability: conscious, sufficient cerebral function for independent activities of daily life. Able to work in sheltered environment.
CPC 3	Severe cerebral disability: conscious, dependent on others for daily support because of impaired brain function. Ranges from ambulatory state to severe dementia or paralysis.
CPC 4	Coma or vegetative state: any degree of coma without the presence of all brain death criteria. Unawareness, even if appears awake (vegetative state) without interaction with environment; may have spontaneous eye opening and sleep/awake cycles. Cerebral unresponsiveness.
CPC 5	Brain death: apnea, areflexia, EEG silence, etc.

[a]From Safar P. Resuscitation after brain ischemia. In: Grenvik A, Safar P, editors. Brain failure and resuscitation. New York: Churchill Livingstone; 1981; p. 155–84.

Table 2

Table of CPC Assessments.

a. CPC Agreement (5 point scale)

		Registry CPC score			
		1	**2**	**3**	**4**
Database CPC score	1	71	6	1	0
	2	10	30	1	0
	3	8	10	19	2
	4	0	0	1	12

b. CPC Agreement (dichotomized)

		Registry CPC score	
		1–2	**3–4**
Database CPC score	1–2	117	2
	3–4	18	34

Table 3

Patient demographics stratified by concordance of CPC score.

	All patients (*n* = 171)	Concordant CPC score (*n* = 132)	Discordant CPC score (*n* = 39)	*p*-Value
Age (years)	55.6 ± 16.8	54.9 ± 15.5	58.2 ± 20.3	0.352
Male	106 (62.0)	81 (61.4)	25 (64.1)	0.757
Initial rhythm				
Asystole	25 (15.0)	16 (12.3)	9 (24.3)	0.113
PEA	68 (40.7)	52 (40.0)	16 (43.2)	
VF/VT	74 (44.3)	62 (47.7)	12 (32.4)	
Out-of-Hospital Arrest	106 (62.0)	86 (65.2)	20 (51.3)	0.117
Witnessed Arrest	137 (82.0)	107 (83.6)	30 (76.9)	0.348
Cardiac Etiology	111 (66.1)	92 (71.3)	19 (48.7)	0.009
Duration (min)	15 (7, 25)	14.5 (7, 25)	16 (9, 24)	0.461
Treatment with TTM	133 (77.8)	106 (80.3)	27 (69.2)	0.144
Discharged Home	79 (46.2)	71 (53.8)	8 (20.5)	<0.001
Median CPC score from in-house database	2 (IQR: 1, 3)	1 (1, 2)	3 (2, 3)	<0.001
Median CPC score from registry database	1 (IQR: 1, 2)	1 (1, 2)	2 (1, 2)	0.922