

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Association for Academic Surgery

Inter- and intra-rater reliability of a grading system for congenital diaphragmatic hernia defect size



Chelsea E. Hunter, BS,^a Zoe M. Saenz, BS,^a Daisy Nunez, MS,^a
Lava Timsina, MPH, PhD,^b and Brian W. Gray, MD^{a,*}

^a Division of Pediatric Surgery, Department of Surgery, Indiana University School of Medicine, Indianapolis, Indiana

^b Department of Surgery, Indiana University School of Medicine, Center for Outcomes Research in Surgery, Indianapolis, Indiana

ARTICLE INFO

Article history:

Received 2 March 2018

Received in revised form
17 May 2018

Accepted 13 July 2018

Available online xxx

Keywords:

Congenital diaphragmatic hernia
CDH
Grading system
Reliability

ABSTRACT

Background: The Congenital Diaphragmatic Hernia Study Group (CDHSG) registry is a multi-institutional tool to track outcomes of patients with CDH. The CDHSG asks surgeons to categorize diaphragmatic defect sizes as type A-D based on published guidelines. The reported size of the defect has been correlated with patient outcomes, but the reliability of this system has never been studied. Our goal was to evaluate the inter- and intra-rater reliability of the CDHSG grading system.

Materials and methods: Forty-six operative notes from CDH patients that underwent surgical repair at a single institution were collected and cropped to include only the information necessary to grade the hernia defect based on the CDHSG guidelines. The defects were graded by nine pediatric surgeons on two separate occasions (18 wk apart). Inter-rater reliability was calculated using a Cohen's kappa (κ). Intra-rater reliability was calculated using an intraclass correlation coefficient.

Results: Inter-rater reliability was minimal to weak ($\kappa_{\text{round1}} = 0.395$, $\kappa_{\text{round2}} = 0.424$). Agreement ranged from 19.57% ($\kappa = -0.0745$) to 82.61% ($\kappa = 0.7543$). Inter-rater agreement was similar despite operative findings and outcomes: survival yes/no ($\kappa = 0.3690$, $\kappa = 0.3518$), need for ECMO yes/no ($\kappa = 0.3323$, $\kappa = 0.3362$), patch repair yes/no ($\kappa = 0.2050$, $\kappa = 0.1916$), and liver up/down ($\kappa = 0.2941$, $\kappa = 0.4404$).

Intra-rater reliability was good to excellent (intraclass correlation coefficient = 0.88, 95% CI [0.83-0.92]). Agreement with oneself ranged from 71.74% to 93.48%.

Conclusions: The demonstrated weak inter-rater reliability of the current CDHSG grading system shows the need for improvement in how the grading system is applied by surgeons when reporting CDH defect size.

© 2018 Elsevier Inc. All rights reserved.

* Corresponding author. Division of Pediatric Surgery, Department of Surgery, Riley Hospital for Children, Indiana University School of Medicine, 705 Riley Hospital Drive, RI2500, Indianapolis, IN 46202. Tel.: +1 317 274 5706; fax: +1 317 274 4491.

E-mail address: graybw@iupui.edu (B.W. Gray).

0022-4804/\$ – see front matter © 2018 Elsevier Inc. All rights reserved.

<https://doi.org/10.1016/j.jss.2018.07.045>

Introduction

Congenital diaphragmatic hernia (CDH) is a condition that is associated with an approximate 80% overall survival, and size of the defect has been found to be one of many important prognostic indicators.¹ Owing to the low incidence and wide-ranging severity of this disease, a unified database is vital to study overall disease burden, outcomes, and new treatment modalities. The Congenital Diaphragmatic Hernia Study Group (CDHSG) registry is a multi-institutional tool that collects data on CDH patients from participating centers worldwide.

After showing that patients with larger defects had worse survival in a 2007 report from the registry, the CDHSG updated the data collection form and started gathering more specific information regarding hernia defect size.² A grading system was created that divides the size of the hernia defect into four types (A, B, C, and D) based on intraoperative findings. When institutions enter their CDH patients into the registry, hernia defect size is reported for those patients who underwent surgical repair. This is accomplished by interpreting the information that the surgeon provides in the operative note, which can include naming an exact defect type (A-D), reporting the diaphragmatic defect as a percentage, or purely qualitative descriptions.

Since this data has been added to the registry, studies have shown that defect size correlates with outcomes, such as morbidity, mortality, and hospital length of stay.^{3,4} The ability to conduct meaningful clinical research regarding the implications of the size of a CDH depends heavily on the reliability of the registry data, particularly the reported grade of the hernia defect. However, the reliability of this grading system has not been evaluated. The goal of this study was to determine the inter-rater and intra-rater reliability of the current CDHSG grading system.

Materials and methods

Study design

This study was approved by the Indiana University Institutional Review Board (#1705606018). Operative notes were collected from all patients that underwent surgical repair of a unilateral CDH at Riley Hospital for Children at Indiana University Health between 2010 and 2016. These 46 operative notes were cropped to include only the information necessary to grade the hernia defect based on the CDHSG grading system guidelines. Descriptions of the size and location of the defect, description of the repair, and size of the patch, when applicable, was retained. Any information that may bias the surgeon when grading the defect size (such as ECMO status and viscera found in the chest) was removed from the operative notes. Each note was cut and copied to fit on a single page and then placed in a random order. In addition to the cropped operative notes, the surgeons were provided with the standard CDHSG diagram that depicts the four grades of diaphragmatic hernias. The diagram included a description of each defect type, using percentages (Table 1). The surgeons were provided with no formal training in the use or application of the grading system. Each hernia was initially graded by nine pediatric surgeons. The exercise was then repeated 18 wk later with the order of the operative notes

Table 1 – Description of hernia grades.

Hernia grade	% of hemidiaphragm present	% of chest wall involvement
A	>90%	<10%
B	50-75%	<50%
C	<50%	>50%
D	<10%	>90%

rearranged. The packets were otherwise not changed in any way. All surgeons were part of the same group, but they have varied training pedigrees. Each surgeon reported their number of years in practice and estimated the number of CDH repairs that they have performed. In addition, intraoperative findings (liver up versus down, ECMO status, repair type) and patient outcomes (length of stay, survival to discharge or transfer) were recorded.

Statistical analysis

Inter-rater reliability between multiple raters was assessed using Cohen's kappa generalized for more than two rating outcomes (A, B, C, and D) by nine raters.^{5,6} Kappa analysis for inter-rater reliability adjusts for the percentage of concordant and discordant agreements that arise from chance alone between the raters. We used $-\text{kap}$ in Stata/SE 14.2⁷ to compute unweighted Kappa's statistics.⁸

Unweighted kappa scores were used to more accurately evaluate the true reliability of the grading system. For use in the clinical setting, a stricter interpretation of kappa has been proposed and recently applied in clinical research.^{9,10} The following cutoffs were used to interpret κ : ≤ 0.20 —no agreement, 0.21-0.39—minimal agreement, 0.40-0.59—weak agreement, 0.60-0.79—moderate agreement, 0.80-0.90—strong agreement, and >0.90 —almost perfect.

We also analyzed intraclass correlation coefficient (ICC) using a multilevel model with random effects.^{11,12} Because two ratings were taken for each patient and there exist multiple ratings per patient, we fit a three-level mixed-effect model with the following random effects: a random intercept and random slope on follow-up time at the patient level, and a random intercept at the rater's level. This model assumed that the residuals were independent with constant variance (homoskedastic). We then estimated ICC for this three-level nested model: ICC for the patient, the correlation between CDH ratings in the same patient, and ICC for the raters within a patient, the correlation between CDH ratings in the same rater and patient. We also reported 95% CI and P-value at 0.05 level of significance. The following cutoffs were used to interpret ICC¹¹: <0.50 —poor reliability, 0.50-0.75—moderate reliability, 0.75-0.90—good reliability, and >0.90 —excellent reliability.

Results

Demographics

Patient characteristics and outcomes of the 46 CDH patients in our study group are displayed in Table 2.

Download English Version:

<https://daneshyari.com/en/article/8835158>

Download Persian Version:

<https://daneshyari.com/article/8835158>

[Daneshyari.com](https://daneshyari.com)