Published in final edited form as:

J Nurs Meas. 2021 August 01; 29(2): 213–226. doi:10.1891/JNM-D-19-00106.

Predictive ability of an illness severity measure: Implications for nursing research

Lisa C. Lindley, PhD, RN, FPCN [Associate Professor | Nightingale Endowed Faculty Fellow],

College of Nursing, University of Tennessee, Knoxville, Knoxville, Tennessee 37996

Christine A. Fortney, PhD RN [Assistant Professor],

Martha S. Pitzer Center for Women, Children and Youth, College of Nursing, The Ohio State University, Columbus, Ohio 43210

Melanie J. Cozad, PhD [Assistant Professor]

Department of Health Services Policy and Management, Center for Effectiveness Research in Orthopedics, University of South Carolina, Columbia, SC 29201

Abstract

Background and Purpose: Illness severity among children with life-limiting illnesses is measured with the pediatric complex chronic conditions (CCC) measure. Developed in 2000/2001, it was revised in 2014 to include infant-specific categories.

Methods: Discrimination, calibration, accuracy, and validation tests were used to examine the predictive performance of the measures.

Results: Among the 10,175 infants in the analysis, both measures poorly discriminated - palliative care consultation (C-statistics 0.6396 vs. C-statistics 0.5905) and any inpatient procedure (C-statistics 0.6101 vs. C-statistics 0.5160). The Hosmer-Lemeshow goodness-of-fit tests revealed good calibration for both measures. The original measure was more accurate in predicting end-of-life outcomes - palliative care consultation (Brier Score 0.3892 vs. 0.7787) and any inpatient procedures (Brier Score 0.3115 vs. 0.4738).

Conclusions: The revised measure did not perform any better than the original in predicting end-of-life outcomes among infants.

Keywords

complex chronic conditions; illness severity; pediatric; end-of-life; measurement testing; infant

The pediatric complex chronic conditions (CCC) measure has been the accepted method of measuring illness severity among children with life-limiting illnesses. The measure was originally developed by Feudtner and colleagues in 2000/2001 (Feudtner et al., 2000;

llindley@utk.edu .

Financial Disclosure: The authors have no financial relationships relevant to this article to disclose.

Conflict of Interest: Dr. Lindley declares she has no conflict of interest. Dr. Fortney declares she has no conflict of interest. Dr. Cozad declares she has no conflict of interest.

Feudtner et al., 2001) and is defined as medical conditions that would reasonably be expected to last at least 12 months and involve either several organ systems or 1 organ system requiring specialty pediatric care (Feudtner et al., 2000). The original CCC measure was created using 1980 to 1997 Washington State data from pediatric decedents under 19 years. The researchers used *International Classification of Disease, Ninth Revision (ICD-9)* codes most commonly associated with pediatric mortality to create nine categories: neurologic/ neuromuscular, cardiovascular, respiratory, renal, gastrointestinal, hematologic/ immunologic, metabolic, congenital/genetic defect, or malignancy. The measure was significantly revised in 2014 to include infant-specific categories (i.e., premature/neonatal, miscellaneous), along with new procedure codes (e.g., transplantation) and medical device codes under each category (Feudtner et al., 2014).

To date, there has been no head-to-head evaluation of the performance (e.g., discrimination, accuracy, calibration) of the two measures in predicting end-of-life outcomes. Prediction of end of life outcomes is especially important because it can help families with decision-making for their infant's care and goals of treatment. Overly optimistic predictions may delay referral to beneficial services, such as palliative care. Early referral to palliative care services can help the family navigate the long trajectory of illness that infants in the NICU sometimes experience, as well as provide seamless transition to hospice or end-of-life care if needed. Additionally, the performance of the two measures has not been validated using rigorous statistical methods. Testing the original versus the revised CCC measure's ability to predict end of life outcomes provides critical evidence for nurse researchers which enables them to select the most reliable and valid measure of illness severity for their pediatric research purposes. Therefore, the objective of this study was to compare the predictive performance of the original versus the revised CCC measure among infants.

BACKGROUND AND CONCEPTUAL FRAMEWORK

Over the past two decades, the original CCC measure has been frequently utilized in pediatric nursing research. Although the measure can assist with identification of pediatric samples at end of life (Ananth et al., 2015; Feinstein et al., 2014; Hudson, et al., 2014; Golden & Nageswaran, 2012;), it is more commonly used by nursing researchers as a measure of illness severity and multimorbidity (Keele et al., 2013; Lindley & Keim-Malpass, 2017). For example, the original CCC measure is used within a quantitative research study as a proxy for illness severity. Its strength in accurately capturing illness severity among the pediatric population stems from the origin of its development using diagnostic codes among children who died in the hospital. Additionally, if a child has multiple CCCs, the original CCC measure commonly serves as a proxy for multimorbidity among pediatric patients. Because most measures of illness severity (e.g., Charlson Comorbidity Index) were developed using adult information and diagnoses, the original CCC measure developed using pediatric information is considered the best option for measuring illness severity among pediatric patients.

Among the infant population, the search for an adequate measure of illness severity and predictor of outcomes at end of life that could be used both clinically and for research purposes is ongoing. Several measures have been developed and used since the 1990s to

assess the severity of illness in infants in the NICU because markers of risk that are routinely available, such as birth weight, gestational age, and sex do not adequately capture the illness severity of neonatal patients. The Clinical Risk Index for Babies (CRIB) score and the Score for Neonatal Acute Physiology (SNAP) were validated as predictors of mortality and morbidity, but there were issues with these early measures (International Neonatal Network, 1993; Richardson et al., 1993). The CRIB scale was difficult to apply to infants born outside the hospital and the SNAP score required the collection of 34 data points, which was time-consuming (Richardson et al, 2001). Although these measures were revised and validated, they were still difficult to use in either clinical or research settings across all infants (Richardson et al., 2001). Furthermore, a systematic review of the utility of the revised SNAP by Morse, et al, found that it was useful to clinicians and researchers in evaluating outcomes at the population level, but the need for a precise measurement tool that can measure severity of illness and predict mortality and morbidity at the individual level is still needed (Morse et al., 2015). Thus, these scales are not sufficient to capture illness severity in studies of infants related to palliative and end-of-life care.

Given the need for an improved measure, the original CCC measure underwent a significant modification in 2014 (Feudtner et al., 2014). Feudtner and colleagues revised the original CCC system because ICD-10 had replaced the ICD-9 system, the original measure lacked health conditions originating in the neonatal period, and the original did not include information about medical technology utilization or organ transplantation, which would indicate CCC status. The revised CCC measure added two additional categories and revised/expanded original subcategories. The Premature and Neonatal category, as an example, was added and codes for medical devices and transplantation were included in most categories. The authors provided conversions tables from ICD-9-CM to ICD-10-CM, along with programming language in SAS and STATA to assist in creating the revised CCC measure.

Despite the availability of two measures, there has been a lack of rigorous, head-to-head comparison between the performance of the original and revised measures. Research has primarily focused on using the measure to identify children with a CCC. Feudtner and colleagues (2014) conducted a basic assessment of the two measures in creating the revised CCC measure. They compared the proportion of revised CCC categories to the original categories using the 2009 Kids' Inpatient Database -Healthcare Cost and Utilization Project (HCUP KID), and 2010 National Emergency Department Sample (NEDS) data files. They found that the revised measure identified more children as having a CCC than the original measure with exceptions in the malignancy and cardiovascular CCC categories. This was consistent with a more recent study using 2012 HCUP KID data, which tested the specificity, sensitivity, and positive predictive ability of the original versus the revised measures (Lindley et al., 2019). This study found that revised measure represented an improvement in identifying infants with a CCC compared to the original.

Only a single study was identified that examined the predictive ability of the CCC measures. From a recent investigation using 2006 to 2014 New York City hospital data, Hessel and colleagues reported the performance of the original pediatric CCC measure versus the adult Charlson/Deyo illness severity measure to predict mortality and length of hospital stay (Hessels et al., 2018). These researchers found that the original CCC measure performed

better at predicting both outcomes, compared to the Charlson/Deyo. Although knowledge about severity of illness measures has improved, very little is still known about how the measures perform in predicting end-of-life outcomes, especially among infants. In fact, the sparse literature suggests that even the two CCC measures may be different in their predictive capability. Therefore, the lack of valid and reliable evidence about the measures is a significant gap in nursing knowledge.

Understanding the performance of the pediatric CCC measures is important and timely in advancing the nursing science of pediatric end-of-life care (Berry et al., 2015). Knowledge about the predictive performance of the two measures is critical to advance the nursing science of end-of-life care for infants, which has been identified as an important national health care goal (Institute of Medicine, 2001). Given the increasing health complexity of children at end of life, improving knowledge about classifying children is critical to advancing the science. Approximately 16,000 infants in the United States will die annually from their health condition; while others will go on to develop CCCs that require ongoing intensive medical care (Decourcey et al., 2018; Murphy et al., 2017a;2017b). Ensuring that these children are included in nursing research across the lifespan is increasingly important to external research funders such as National Institute of Health (NIH). Children are defined by the NIH as individuals under 18 years old, however nurse researchers must include appropriate groups of children for the study design while also protecting this vulnerable group. Furthermore, the study results must be generalizable to the population under study. Although the inclusion of different and/or additional diagnoses in the revised CCC measure may suggest that it is an improved predictor, a systematic evaluation with rigorous analytical methods will guide researchers in understanding why one measure might be better than another and in what manner. This work will ultimately enable nurse researchers to select the most appropriate measure of illness severity for infants.

METHODS

Study Design & Data Source

This study used a non-experimental design with data from the 2012 Healthcare Cost and Utilization Project (HCUP), Kids' Inpatient Database (KID) (Agency for Healthcare Research and Quality (AHRQ), 2012). HCUP KID is the largest, all-payer pediatric inpatient care database in the United States (US). It has been produced every three years since 1997 through 2012, which is the most current year available. It yields national estimates of hospital inpatient stays for patients younger than 21 years of age. HCUP KID data set includes inpatient records from 4179 acute care hospitals in 44 states in a calendar year. The data set is comprehensive with up to 25 International Classification of Diseases, Ninth Revision, Clinical Modification (*ICD-9-CM*) diagnosis codes, demographic data, procedure codes, and payment information. The files are available through AHRQ (<https://www.hcup-us.ahrq.gov/kidoverview.jsp>), which also maintains data quality. This study was approved by the Institutional Review Board of the University of Tennessee, Knoxville.

Study Population

The population of interest was infants admitted to US hospitals. Infants were defined as less than 1 year (Centers for Disease Control and Prevention, 2018). The focus was on infants because the revised CCC measure predominantly included infant-specific modifications (Feudtner et al., 2014). Infants were included if they had a discharge disposition of death in the hospital occurring from January 1 to December 21, 2012. The use of pediatric decedents is consistent with the Feudtner et al. (2000; 2001) studies; which used data from decedents to create the CCC classification measure. Additionally, infants have the highest mortality of any pediatric age group (Murphy et al., 2017b). All infants, who had an external cause of injury or poisoning *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* code were excluded because accident and injury hospital admissions are typically unrelated to a health condition. The 2012 KID data files used ICD-9-CM codes because hospital conversion to ICD-10-CM codes was not federally-mandated until 2015. We also excluded infants with any missing observations or invalid ICD-9-CM codes from the analysis. The final sample size after apply the inclusion and exclusion criteria was 10,175 infants.

Analysis

Predictive Models.—Using pediatric end-of-life literature and clinical knowledge, two models were created to test the performance of the CCC measures to predict end-of-life outcomes (Lindley & Newnam, 2017; Lindley & Fortney, 2019) [See prior publications by Lindley et al. (2019) for further clarification regarding how the CCC classification systems were used and what they include]:

Model 1: Model 1 included the original CCC measure. A measure of the original CCC was operationalized as whether an infant had a neurologic/neuromuscular, cardiovascular, respiratory, renal, gastrointestinal, hematologic/immunologic, metabolic, other congenital or genetic defect (chromosomal abnormalities, bone/joint abnormalities, diaphragm/abdominal abnormalities, other abnormalities), or malignancy health condition (Feudtner et al., 2000; 2001). The measure was tested against two end-of-life outcomes: palliative care consultation and inpatient procedures. Palliative care consultation was defined as having a palliative care consult during the last hospital stay prior to death (yes, no) and any inpatient procedures during the last hospital stay prior to death (yes, no). The following covariates were included in the model: gender (male, female), race/ethnicity (Caucasian, African American, Hispanic, Other), household income (<\$39,000, \$39,000-\$47,999, \$48,000-\$62,999, or >\$63,000), insurance type (Medicaid, commercial, other), region of residence (South, Midwest, West, Northeast), and urban/non-urban.

Model 2: The revised CCC measure was included in Model 2. The revised CCC measure was operationalized as whether an infant had a neurologic/neuromuscular, cardiovascular, respiratory, renal, gastrointestinal, hematologic/immunologic, metabolic, other congenital or genetic defect (chromosomal abnormalities, bone/joint abnormalities, diaphragm/abdominal abnormalities, other abnormalities), malignancy, premature/neonatal, or miscellaneous condition (Feudtner et al., 2014). The outcomes and covariates from Model 1 were included in Model 2.

Frequencies, means, and standard deviations were used to describe the characteristics of the infants in the study. Multivariate logistic regressions with robust standard errors were conducted to predict the association between the CCC measures and end-of-life outcomes. Separate regressions were conducted for each model and for each outcome. Adjusted odds ratios (aOR) and associated 95% confidence intervals (CI) were reported. All statistical analyses were performed using Stata 14.0 (StataCorp, 2015).

Predictive Performance.—To assess the predictive performance of the CCC measures, discrimination, calibration, accuracy, and validation tests were conducted using the predictive models:

Discrimination. Discrimination refers to the ability of the CCC measures to discriminate between infants with and without an end-of-life outcome. For this study, discrimination was measured using the C-statistic, which is equivalent to the area under the receiver operating characteristics curve (AUC) for dichotomous outcomes (Ikeda et al., 2002; Harrell et al., 1996). The C-statistic ranges from zero to one with a value of one representing perfect prediction and a value of 0.5 representing chance prediction (Alba et al., 2017). A value between 0.7 and 0.8 is considered to demonstrate acceptable predictive performance, while a value greater than 0.8 demonstrates excellent discriminative performance (Alba et al., 2017). The 95% confidence intervals were computed and differences in the C-statistic between the models were tested using the DeLong et al. method (DeLong et al., 1988). The percentage change in the C-statistic was also computed.

Calibration. The agreement between observed and predicted outcomes using the CCC measures is a measure of calibration. Poorly calibrated models will under- or over-estimate the outcome of interest (Steyerberg et al., 2010). Calibration was assessed in this study using the Hosmer-Lemeshow goodness-of-fit statistic, which is reported as a χ^2 value and p value. A small p -value means the model is a poor fit ($p < 0.05$).

Accuracy. The Brier score measures accuracy in predicting outcomes that incorporates features of discrimination and calibration as a single measure. It was calculated for each CCC measure. The Brier score, which is appropriate for categorical outcomes, ranges from zero to one (Smith et al., 2015). A lower score indicates less prediction error and more accurate performance.

Validation. To internally validate the performance of the CCC measures, bootstrapping simulation (1000 iterations) to predict end-of-life outcomes was utilized (Chu et al., 2010; Cooke et al., 2011; Matheny et al., 2005; Stukenborg, 2011). This approach was selected rather than split sample internal validation because it provides more accurate, unbiased estimates of performance in samples less than 20,000 participants (Harrell, 2018). No external validation was conducted because rigorous internal validation such as bootstrapping account for externalities (Harrell, 2018).

RESULTS

Table 1 contains the summary data of the study participants. Among the infants in the overall sample, slightly more than 10% had a palliative care consults, while more than half of infants had a procedure during their last hospital stay prior to death. Less than half of the infants were girls and the most common race was Caucasian (36.71%) and least common was Hispanic (16.26%). Infants were most commonly from households with a median income of less than \$39,000/annual. A majority of infants were Medicaid beneficiaries. Infants most commonly resided in the south (39.46%) and least commonly in the Northeast (16.09%). More than 10% of infants and their families resided in non-urban areas. In comparing infant characteristics by original versus revised CCC measures, more original CCC infants received a palliative care consultation (15.25% vs. 10.95%) and had any inpatient procedure (78.84% vs. 54.37%). Revised CCC infants were more frequently African American (23.30% vs. 17.59%). Original CCC infants were more often Medicaid beneficiaries, compared to revised CCC infants (54.53% vs. 50.93%).

Discrimination

Table 2 presents parameter estimates and discrimination fit indices for the predictive models assessing the relationship between each of the CCC measures and end-of-life outcomes. The original and revised CCC measures were positively and significantly associated with a palliative care consultation, respectively (OR=2.40, 95%CI:2.10–2.74 vs. OR=2.01, 95%CI:1.59–2.55). The association between CCC measures and any inpatient procedures varied. The original CCC measure was positively related to inpatient procedures (OR=5.88, 95%CI:5.36–6.43); whereas the revised CCC measure was negatively associated to inpatient procedures (OR=0.84, 95%CI:0.74–0.94). The C-statistics indicated that both models poorly discriminated between infants with and without a palliative care consultation (C-statistics 0.6396 vs. C-statistics 0.5905) and those who did or did not receive any type of inpatient procedure (C-statistics 0.6101 vs. C-statistics 0.5160).

Calibration

The Hosmer-Lemeshow goodness-of-fit tests revealed good calibration ($p>0.05$) for both CCC models (Table 3). Although the revised CCC model performed better than the original, both models showed significant agreement between observed and predicted outcomes using the CCC measure as a measure of calibration.

Accuracy

The Brier scores showed that the original CCC measure was more accurate in predicting end-of-life outcomes, compared to the revised CCC measure (Table 4). For both palliative care consultation (Brier Score 0.3892 vs. 0.7787) and any inpatient procedures (Brier Score 0.3115 vs. 0.4738), the model with the original CCC measure had the lowest Brier score.

Validation

The bootstrap validation was repeated using the 1000 iterations to internally validate the best performing model. The statistical performance originally obtained between the CCC models

was essentially identical to the results obtained by repeating the validation process on the bootstrap resampling with 1000 samples (no results listed).

DISCUSSION

As one of the first studies to rigorously evaluate the predictive performance of measures important in nursing research, the analysis revealed that the revised CCC measure did not perform any better than the original measure in predicting end-of-life outcomes among infants. Based on prior work that demonstrated the revised measure was better at identifying infants with a CCC compared to the original, it was expected that the revised measure would also perform better than the original measure at predicting outcomes. One possible explanation for this finding may relate to the timing of data collection in the HCUP KID files. The KIDs diagnostic data were collected at discharge. However, hospitalized infants at end of life have rapidly changing health conditions. For example, many hospitals have palliative care consults triggered on an inpatient diagnosis or change in status. Those changing diagnoses may not be the final diagnosis reported at death. Thus, the nature of the HCUP KID data might have introduced measurement bias because of when ICD-9 codes were recorded. In fact, it is possible that because the revised and original CCC measures were based on diagnoses at discharge, this might explain why both measures performed poorly in their predictive ability of inpatient-related end-of-life outcomes. Future research might explore prospective designs that capture diagnoses in real-time and as they change. Although a study retrospectively analyzing this clinical data would generally have smaller sample size and limited generalizability, it might provide much-needed additional testing of these measures.

An alternative explanation for the poor performance of the revised CCC measure may relate to a difference in CCC conceptual definitions. Previous work comparing the revised and original measures found the revised measure identified more infants, but the composition of the infants identified was different (Lindley et al., 2019). These differences were attributed to the different constructs being measured by the CCC measures. The original CCC measured infants with life-limiting illness, while the additions made to the revised CCC measured medical complexity. Life-limiting illnesses are those health conditions that are likely to lead to death. Since medical complexity is defined as multisystem diseases with a need for significant health care services, along with a marked functional impairment and/or technology dependence for activities of daily living (Cohen et al., 2011), it could be that using the revised CCC captures infants who already had significant technology dependence and transplants. This may mean that infants identified by the revised CCC are less likely to need any inpatient procedures. This may explain why within the discrimination results, the revised CCC was negatively associated with any inpatient procedures. The focus of the original CCC on life-limiting illness also may explain why it was more accurate in predicting a palliative care consult and any inpatient procedure. Having a technology dependence or a transplant, while meeting the definition of complex chronic condition, may be viewed as less life-threatening. Thus, it may explain why the revised CCC was less accurate than the original in predicting important end-of-life outcomes.

Limitations

There were limitations to this study affecting its generalizability. Because the study sample included US infants admitted to children's hospitals, it limits generalizability of the study findings to other pediatric decedents such as toddlers or adolescents. Additionally, the findings may not be generalizable to other health care institutions. Another limitation, as noted above, was that the HCUP KID data files only included discharge diagnoses and these are often different from admission diagnoses due to rapidly changing infant conditions during their inpatient stay. The diagnosis timing may introduce measurement error. Despite these limitations, HCUP KID is the largest publicly-available all-payer pediatric inpatient database in the United States. These files allow for a large, national sample size and improve external validity. The results provide valuable insight into the performance of two classifications measures for pediatric nursing research in predicting patient outcomes.

Relevance to Nursing Practice, Education, or Research

For nurse researchers, the findings of this study suggest that there is no clear advantage to using the revised CCC measure over the original as a predictor of end-of-life outcomes. As a consequence, researchers should consider other factors when deciding which CCC measure to use. First, nurse researchers should select the measures that conceptually fits with their research question. For example, the original CCC measure captures illness severity as a construct of life-limiting health conditions; whereas the revised measure is better aligned with the construct of medical complexity. Therefore, researchers whose studies focus on predicting life-limiting health conditions may want to use the original because its conceptual origin fits better with those types of research questions. Additionally, ease of use might influence measure selection. The original CCC measure includes the ICD9 codes for the 9 CCC categories, which are relatively easy to identify in a database (e.g., claims data, electronic medical records), but programming language has not been included. Conversely, the revised CCC measure includes a significant number of ICD9/10 codes and procedure codes. The authors also published programming language in SAS, SPSS, and Stata to assist researchers. Therefore, conceptualization of measure with study design and ease of use might be additional factors to consider in the measurement selection process.

Knowledge of the CCC classification measures is also relevant to the ethics of nursing research. Given that children under 2 years of age are often excluded from pediatric nursing research studies because of difficulties with diagnosis, prognostication, and measurement, improving knowledge about measuring illness severity among infants, and which measures performs best as a predictor of end-of-life outcomes is critical to closing gaps and advancing the science. This assurance would meet the moral requirement illuminated in the Belmont Report that states that "there should be fair procedures and outcomes in the selection of research participants." (Belmont Report, 1979). Patterns of injustice can arise from social, racial, sexual, and cultural biases that are engrained in our society, even if a researcher has selected individuals in a seemingly fair manner (Belmont Report, 1979; Diekema, 2009). Because of their age, infants must be protected from the risks of medical research, however they should not be unfairly excluded from participating in nursing research that could be beneficial to them or future infants (Diekema, 2009).

In summary, the study findings demonstrated that the revised CCC measure did not perform any better than the original measure in predicting end-of-life outcomes among infants. Both measures poorly discriminated between infants with and without a palliative care consultation and those who did or did not receive any type of inpatient procedure. The revised CCC measure performed slightly better than the original in terms of calibration. However, the original CCC measure was more accurate in predicting end-of-life outcomes, compared to the revised CCC measure. As nurse researchers consider which version of the CCC measure to include as a predictor in their outcome models, they might consider the ease of programming and the conceptual fit (life-limiting vs. medical complexity) in their decision making because the revised CCC measure did not offer superior predictive ability compared to the original measure.

Funding Source:

This publication was made possible by Grant Number R01NR017848 from the National Institute of Nursing Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Nursing Research or National Institutes of Health.

References

- Agency for Healthcare Research and Quality (2012). 2012 Healthcare Cost and Utilization Project (HCUP), Kids' Inpatient Database (KID) [Data set]. <https://www.hcup-us.ahrq.gov/kidoverview.jsp>
- Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, & Guyatt G (2017). Discrimination and calibration of clinical prediction models. *Journal of the American Medical Association*, 318(14), 1377–1384. 10.1001/jama.2017.12126 [PubMed: 29049590]
- Anath P, Melvin P, Feudtner C, Wolfe J, & Berry JG (2015). Hospital use in the last year of life for children with life-threatening complex chronic conditions. *Pediatrics*, 136(5), 938–946. 10.1542/peds.2015-0260 [PubMed: 26438707]
- Berry JG, Hall M, Cohen E, O'Neill M, & Feudtner C (2015). Ways to identify children with medical complexity and the importance of why. *Journal of Pediatrics*, 167(2):229–237. 10.1016/j.jpeds.2015.04.068
- Centers for Disease Control and Prevention. (2019). Child Development-Infants. <https://www.cdc.gov/ncbddd/childdevelopment/positiveparenting/infants.html>.
- Cohen E, Kuo DZ, Berry Agrawai R, Bhagat JG, Simon SK, D. T, & Srivastava R (2011). Children with medical complexity: an emerging population for clinical and research initiatives. *Pediatrics*, 127(3), 529–538. 10.1542/peds.2010-0910 [PubMed: 21339266]
- Cooke CR, Joo MJ, Anderson SM, Udris EM, Johnson E, & Au DH (2011). The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. *BMC Health Services Research*, 11(37), epub. 10.1186/1472-6963-11-37
- Chu Y, Ng Y, & Wu S (2010). Comparison of different comorbidity measures for use with administrative data in predicting short- and long-term mortality. *BMC Health Services Research*, 10 (140), epub. 10.1186/1472-6963-10-140
- Decourcey DD, Silverman M, Oladunjoye A, Balkin EM, & Wolfe J (2018). Advance Care Planning and Parent-Reported End-of-Life Outcomes in Children, Adolescents, and Young Adults with Complex Chronic Conditions. *Journal of Pediatrics*, 193, 196–203.e2. 10.1097/CCM.0000000000003472
- DeLong ER, DeLong DM, & Clarke-Pearson DL (1988). Comparing the areas under two or more correlated receiver operating characteristics curves: A nonparametric approach. *Biometrics*, 44, 837–845. [PubMed: 3203132]
- Diekema DS (2009). Ethical issues in research involving infants. *Seminars in Perinatology*, 33, 364–371. 10.1053/j.semperi.2009.07.003 [PubMed: 19914520]

- Feinstein JA, Feudtner C, & Kempe A (2014). Adverse drug event-related emergency department visits associated with complex chronic conditions. *Pediatrics*, 133(6), e1575–e1585. 10.1542/peds.2013-3060 [PubMed: 24843054]
- Feudtner C, Christakis DA, & Connell FA (2000). Pediatric death attributable to complex chronic conditions: a population-based study of Washington state, 1980–1997. *Pediatrics*, 106(1Pt 2), 205–209. [PubMed: 10888693]
- Feudtner C, Hays R, Haynes G, Geyer J, Neff J, & Koepsell T (2001). Deaths attributed to pediatric complex chronic conditions: National trends and implications for supportive care services. *Pediatrics*, 107(6), e99–e103. 10.1542/peds.107.6.e99. [PubMed: 11389297]
- Feudtner C, Feinstein JA, Zhong W, Hall M, & Dai D (2014). Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex technology dependence and transplantation. *BMC Pediatrics*, 14, 199–205. 10.1186/1471-2431-14-199 [PubMed: 25102958]
- Golden SL, & Nageswaran S (2012). Caregivers voices: Coordinating care for children with complex chronic conditions. *Clinical Pediatrics*, 51(8), 723–729. 10.1177/000992281244592 [PubMed: 22563062]
- Harrell FE (2018). Split-Sample Model Validation. <http://www.fharrell.com/post/split-val/>
- Harrell FE, Lee KL, & Mark DB (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics & Medicine*, 15, 361–387.
- Hessels AJ, Liu J, Cohen B, Shang J, & Larson EL (2018). Severity of illness measures for pediatric patients. *Journal of Healthcare Quality*, 40(5), e77–e89. 10.1097/JHQ.0000000000000135
- Hudson SM, Mueller M, Hester WH, Magwood GS, Newman SD, & Laken MA (2014). At-risk characteristics for hospital admissions and ED visits. *Journal of Specialist in Pediatric Nursing*, 19(2), 183–193. 10.1111/jspn.12068
- Ikeda M, Ishigaki T, & Tamauchi K (2002). Relationship between Brier score and area under the binormal ROC curve. *Computer Methods and Programming in Biomedicine*, 67, 187–194.
- International Neonatal Network. (1993). The CRIB (Clinical Risk Index for Babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet*, 342(8865), 193–198. [PubMed: 8100927]
- Institute of Medicine. (2003). When children die: Improving palliative and end-of-life care for children and their families. National Academy Press.
- Keele L, Keenan HT, Sheetz J, & Bratton SL (2013). Differences in characteristics of dying children who receive and do not receive palliative care. *Pediatrics*, 132, 72–78. 10.1542/peds.2013-0470 [PubMed: 23753086]
- Lindley LC, Cozad MJ, & Fortney CA (2019). Pediatric complex chronic conditions: Evaluating two versions of the classification system. *Western Journal of Nursing Research*, Epub. 10.1177/0193945919867266
- Lindley LC, & Fortney CA (2019). Pediatric complex chronic conditions: Does the classification system work for infants? *American Journal of Hospice and Palliative Medicine*, epub. 10.1177/1049909119838985
- Lindley LC, & Keim-Malpass J (2017). Quality of pediatric hospice care for children with and without multiple complex chronic conditions. *International Journal of Palliative Nursing*, 23(5), 112–119. 10.12968/ijpn.2017.23.5.230
- Lindley LC, & Newnam KM (2017). Hospice use for infants with life-threatening health conditions, 2007 to 2010. *Journal of Pediatric Health Care*, 31(1), 96–103. 10.1016/j.pedhc.2016.04.015 [PubMed: 27245660]
- Matheny ME, Ohno-Machado L & Resnic FS (2005). Discrimination and calibration of mortality risk prediction models in interventional cardiology. *Journal of Biomedical Informatics*, 38, 367–375. 10.1016/j.jbi.2005.02.007 [PubMed: 16198996]
- Morse S, Groer M, Shelton MM, Maguire D, Ashmeade T (2015). A Systematic Review: The Utility of the Revised Version of the Score for Neonatal Acute Physiology Among Critically Ill Neonates. *Journal of Perinatal and Neonatal Nursing*, 29(4): 315–344. 10.1097/JPN.0000000000000135 [PubMed: 26505848]

- Murphy SL, Mathews TJ, Martin JA, Minkovitz CS, & Strobino DM (2017a). Annual Summary of Vital Statistics 2013–2014. *Pediatrics*, 139(6), e20163239. 10.1542/peds.2016.3239 [PubMed: 28814547]
- Murphy SL, Xu JQ, Kochanek KD, Curtin SC, & Arias E (2017b). Deaths: Final data for 2015. *National Vital Statistics Report*. 66(6), 1–75.
- Richardson DK, Gray JE, McCormick MC, Workman K, & Goldman DA (1993) Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. *Pediatrics*, 91(3), 617–23. [PubMed: 8441569]
- Richardson DK, Corcoran JD, Escobar GJ, Lee SK (2001). SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *Journal of Pediatrics*, 138(1), 92–100. 10.1067/mpd.2001.109608
- Skuenborg GJ (2011). Hospital mortality risk adjustment for heart failure patients using present on admission diagnoses: Improved classification and calibration. *Medical Care*, 49(8), 744–751. 10.1097/MLR.0b013e31821a9812 [PubMed: 21577166]
- Smith ADAC, Tilling K, Lawlor DA, & Nelson SM (2015). External validation and calibration of IVF predict: A national prospective cohort study of 130,960 in vitro fertilization cycles. *PLOS One*, epub. 10.1371/journal.pone.0121357
- StataCorp. (2015). *Stata Statistical Software: Release 14*. StataCorp LP.
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, & Kattan MW (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, 21(1), 128–138. 10.197/EDE.0b013e3181c30fb2 [PubMed: 20010215]
- U.S. Department of Health & Human Services. (1979). Belmont Report. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>

Table 1.

Characteristics of Study Participants

Variables	Overall % (N=10,175)	Original CCC Measure % (n=4,196)	Revised CCC Measure % (n=8,807)
Palliative Care Consultation	10.26%	15.25%	10.95%
Any Inpatient Procedures	54.94%	78.84%	54.37%
Female	44.46%	46.12%	44.83%
Race/Ethnicity			
Caucasian	36.71%	40.20%	36.26%
African American	23.19%	17.59%	23.30%
Hispanic	16.26%	17.76%	16.60%
Other	23.84%	24.45%	23.84%
Household Income (median)			
<\$9,000/annual	35.76%	35.13%	35.84%
\$9,000–\$47,999/annual	25.43%	26.24%	25.20%
\$48,000–\$62,999/annual	22.40%	22.16%	22.43%
>\$63,000/annual	16.41%	16.47%	16.54%
Insurance Type			
Medicaid	51.13%	54.53%	50.93%
Commercial	35.81%	35.44%	36.40%
Other	13.06%	10.03%	12.67%
Region of Residence			
South	39.46%	40.44%	39.29%
Midwest	23.11%	22.28%	22.80%
West	21.35%	22.47%	21.19%
Northeast	16.09%	14.80%	16.73%
Non -Urban	14.00%	15.49%	13.64%

Table 2.

Discrimination of CCC Measures

Outcomes	Model 1 Original CCC aOR (95% CI)	C-Statistic (95% CI)	Model 2 Revised CCC aOR (95% CI)	C-Statistic (95% CI)	%
Palliative Care Consultation	2.40(2.10–2.74)	0.6396(0.62–0.66)	2.01(1.59–2.55)	0.5905(0.57–0.61)	–0.0768
Any Inpatient Procedures	5.88(5.36–6.43)	0.6101(0.59–0.63)	0.84(0.74–0.94)	0.5160(0.50–0.53)	–0.1982

Note: aOR= adjusted odds ratio

Note: Controlled for gender, race/ethnicity, household income, insurance type, region of residence, and urban/non-urban.

Table 3.

Calibration of CCC Measures

Outcome	Model 1	Original CCC	HL χ^2	HL (p)	Model 2 Revised	CCC HL	HL χ^2	HL (p)
Palliative Care Consultation	2.85			0.42	2.36			0.50
Any Inpatient Procedures	6.94			0.07	1.95			0.58

Note: HL= Hosmer-Lemshow Statistic

Table 4.

Accuracy of CCC Measures

Outcomes	Model 1 Original CCC Classification Brier Score	Model 2 Revised CCC Classification Brier Score
Palliative Care Consultation	0.3892	0.7787
Any Inpatient Procedures	0.3115	0.4738

Note: Controlled for gender, race/ethnicity, household income, insurance type, region of residence, urban/non-urban.