

Letters

RESEARCH LETTER

R Package for Pediatric Complex Chronic Condition Classification

Identification of children with complex chronic conditions (CCCs) is necessary to improve health care delivery and perform clinical research, because this patient population uses significant inpatient and outpatient medical resources.¹ The original CCC classification was published in 2000.² A second version was published in 2014 to reflect additions to the *International Classification of Diseases* system and the US adoption of the *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*.³ The CCC classification is widely used in research (currently cited in more than 100 peer-reviewed journal publications). However, the current approach to assigning the CCC categories in health care-related data sets is limited by proprietary software and computational inefficiency. SAS and Stata software to assign CCC categories were published as appendices to the 2014 update,³ but not all investigators have access to these statistical packages. In addition, increasingly large data sets are available to investigators. Although the data processing capability of individual computers continues to improve, the SAS and Stata software can take significant time to run on data sets with millions of observations. The objective of this project was to develop computationally efficient software to generate the CCC categories using R, a free, open-source statistical environment.⁴ We then compared the SAS, Stata, and R software with respect to accuracy and speed of classification on a typical desktop system.

Methods | We developed the pccc R package based on the 2014 version 2 CCC system.³ To maximize computational efficiency, we leveraged the ability to call C++ from within R using the Rcpp package.⁵ We used standard software engineering practices, including distributed version control, issue tracking, and unit testing. We tested the pccc package using the same Healthcare Cost and Utilization Project data sets from the Agency for Healthcare Research and Quality used to develop the 2014 software (2009 Kids' Inpatient Database [KID] and 2010 Nationwide Emergency Department Sample [NEDS]).⁶ On the same desktop system (i7 dual-core, 16-GB RAM), we classified each record using the SAS, Stata, and R software and compared the results. We tested the accuracy (percentage correctly classified) of the R software using SAS as the criterion standard. To test the relative speed of the 3 implementations, we compared processing time (in minutes) for the 3 407 146-record KID data set and the 28 584 301-record NEDS data set. The latest release of the R package is available on the Comprehensive R Archive Network (<https://cran.r-project.org/web/packages/pccc/index.html>), and the developmental version is on GitHub (<https://github.com/CUD2V/pccc>). Institutional

review board approval was not required for this study using publicly available data sets.

Table 1. Issues Revealed by the Unit Testing Process

Type of Issue by Specific Code Affected	Resolution
Duplicates ^a	
Neuromuscular	
343	Duplicate deleted
G253	Duplicate deleted
Technology dependence	
T84498A	Duplicate deleted
T86890	Duplicate deleted
T86891	Duplicate deleted
T86899	Duplicate deleted
Transplantation	
T86890	Duplicate deleted
T86891	Duplicate deleted
T86899	Duplicate deleted
Deletions and additions ^b	
Neuromuscular	
331	Added
3311	Dropped from Stata only
3318	Dropped from Stata only
35921	Added
35922	Added
35923	Added
35929	Added
9782	Dropped from Stata only
E750	Dropped, matched by E75
E751	Dropped, matched by E75
E752	Dropped, matched by E75
E754	Dropped, matched by E75
G3189	Dropped, matched by G318
G3289	Added in Stata only
G4735	Added
G800	Added
G804	Added
G808	Added
Q851	Added in SAS only
Cardiovascular	
4160	Added
Q219	Added in SAS only
Q258	Added in SAS only
Q259	Added in SAS only
Q268	Added
T82121A	Added in Stata only; also flags technology dependence

(continued)

Table 1. Issues Revealed by the Unit Testing Process (continued)

Type of Issue by Specific Code Affected	Resolution
Hematologic/immunologic	
D869	Dropped, already matched by D86
Metabolic	
D841	Deleted
Respiratory	
4160	Added, previously only in Stata
51630	Added
51637	Added
Errors ^c	
Respiratory	
9620	Changed to J9620 in SAS only
G4753	Changed to G4735 in SAS only
Metabolic	
2359	Changed to 2539
Substring errors ^d	
Neuromuscular	
359	Now uses exact matching
3592	Now uses exact matching
G80	Now uses exact matching
Respiratory	
5163	Now uses exact matching
Cardiovascular	
416	Now uses exact matching; previously in Stata only
Metabolic	
624	Now uses exact matching
Shift in categorizations ^e	
Cardiovascular and respiratory	
143	Now only in cardiovascular category
Hematologic/immunologic and metabolic	
D84	Now only in heme/immunologic category
Metabolic	
E75	Now in neuromuscular category

Abbreviation: CCCs, complex chronic conditions.

^a Includes duplicate codes that were present in the SAS CCC version 2 software.

^b Includes codes that should (or should not) have been classified as CCCs.

^c Includes erroneous codes due to, for example, typos and keystroke errors.

^d Includes issues with codes where matching on a substring led to erroneous inclusion of more specific codes that did not correspond to a CCC.

^e Includes codes that were misclassified or erroneously included in ≥ 2 CCC categories.

Results | Unit testing of the new pccc package revealed several different types of issues present in the 2014 SAS and Stata software (Table 1). We collaborated with the authors of the 2000 (C.F.) and 2014 (J.A.F., C.F., and D.D.) CCC systems to resolve those issues. Subsequently, the R package and the updated SAS and Stata software yielded identical patient CCC categorizations when run on each row of patient data in the KID and NEDS data sets. Processing the same data, the R package was comparable to SAS and significantly more efficient than Stata (Table 2). The updated SAS and Stata

Table 2. Processing Time by Software Type

Software	KID (N = 3 407 146)	NEDS (N = 28 584 301)
R	4 min 48 s	18 min 21 s
SAS	3 min 1 s	14 min 57 s
Stata	22 min 45 s	69 min 11 s

Abbreviations: KID, Kids' Inpatient Database; NEDS, Nationwide Emergency Department Sample.

software packages are available at https://feudtnerlab.research.chop.edu/ccc_version_2.php.

Discussion | The free and open-source pccc R package provides accurate, efficient, and reproducible pediatric CCC categorization for large files of administrative records. The ability of R to call C++ directly can improve computational efficiency and is an advantage for package developers. Software development practices, including unit testing, can identify errors before code release. Code in the pccc package was developed collaboratively and that process, including issue tracking, is publicly visible in the GitHub repository. Suggestions or improvements can be submitted through GitHub's pull request mechanism.

James A. Feinstein, MD, MPH

Seth Russell, MS

Peter E. DeWitt, PhD

Chris Feudtner, MD, PhD, MPH

Dingwei Dai, PhD

Tellen D. Bennett, MD, MS

Author Affiliations: Adult and Child Consortium for Health Outcomes Research and Delivery Science, Children's Hospital Colorado, University of Colorado School of Medicine, Aurora (Feinstein, Bennett); Division of General Pediatrics, University of Colorado School of Medicine, Aurora (Feinstein); Data Science to Patient Value, University of Colorado Anschutz Medical Campus, Aurora (Russell, Bennett); Neptune and Company Inc, Lakewood, Colorado (DeWitt); Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania (Feudtner, Dai); Section of Pediatric Critical Care, University of Colorado School of Medicine, Aurora (Bennett).

Accepted for Publication: January 17, 2018.

Corresponding Author: James A. Feinstein, MD, MPH, Adult and Child Consortium for Health Outcomes Research and Delivery Science, Children's Hospital Colorado, University of Colorado School of Medicine, 13199 E Montview Blvd, Ste 300, Room 312A, Aurora, CO 80045 (james.feinstein@ucdenver.edu).

Published Online: April 23, 2018. doi:10.1001/jamapediatrics.2018.0256

Author Contributions: Drs Feinstein and Bennett had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Feinstein, Russell, DeWitt, Feudtner, Bennett.

Acquisition, analysis, or interpretation of data: Feinstein, Russell, Dai, Bennett.

Drafting of the manuscript: Feinstein, Russell, DeWitt.

Critical revision of the manuscript for important intellectual content: Feinstein, Russell, Feudtner, Dai, Bennett.

Statistical analysis: Feinstein, Russell, Dai, Bennett.

Obtained funding: Bennett.

Administrative, technical, or material support: Russell, DeWitt, Dai, Bennett.

Study supervision: Bennett.

Conflict of Interest Disclosures: None reported.

Funding/Support: This work was supported by grants K23HD091295 (Dr Feinstein) and K23HD074620 (Dr Bennett) from the Eunice Kennedy Shriver

National Institute for Child Health and Human Development and by the Data Science to Patient Value Initiative, University of Colorado Anschutz Medical Campus.

Role of the Funder/Sponsor: The sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

1. Cohen E, Berry JG, Camacho X, Anderson G, Wodchis W, Guttmann A. Patterns and costs of health care use of children with medical complexity. *Pediatrics*. 2012; 130(6):e1463-e1470.
2. Feudtner C, Hays RM, Haynes G, Geyer JR, Neff JM, Koepsell TD. Deaths attributed to pediatric complex chronic conditions. *Pediatrics*. 2001;107(6):E99.
3. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2. *BMC Pediatr*. 2014;14:199.
4. The R Foundation for Statistical Computing. The R Project for Statistical Computing. <https://www.r-project.org/>. 2017. Accessed January 10, 2018.
5. Rcpp.org. Rcpp for Seamless R and C++ Integration. <http://www.rcpp.org/>. Accessed January 10, 2018.
6. Healthcare Cost and Utilization Project. Overview of Nationwide Emergency Department Sample (NEDS). <http://www.hcup-us.ahrq.gov/nedsoverview.jsp>. December 2017. Accessed January 10, 2018.