

May 22nd, 4:30 PM - 6:00 PM

Using bootstrap to compare the validity of PRO measures in discriminating among CKD patients

Nina Deng

University of Massachusetts Medical School, nina.deng@umassmed.edu

John E. Ware Jr.

University of Massachusetts Medical School, john.ware@umassmed.edu

Follow this and additional works at: http://escholarship.umassmed.edu/cts_retreat

 Part of the [Statistics and Probability Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#).

Nina Deng and John E. Ware Jr., "Using bootstrap to compare the validity of PRO measures in discriminating among CKD patients" (May 22, 2012). *UMass Center for Clinical and Translational Science Research Retreat*. Paper 15.
http://escholarship.umassmed.edu/cts_retreat/2012/posters/15

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in UMass Center for Clinical and Translational Science Research Retreat by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

USING BOOTSTRAP TO COMPARE THE VALIDITY OF PRO MEASURES IN DISCRIMINATING AMONG CKD PATIENTS

Nina Deng, John E. Ware Jr.

University of Massachusetts Medical School, Department of Quantitative Health Sciences

Contact Information: nina.deng@massmed.edu; AC7-053, 55 Lake Ave. North, Worcester, MA, 01655

BACKGROUND: Patient-reported outcome (PRO) research requires valid and sensitive measures. Relative Validity (RV) offers an objective way to compare the validity of different PRO measures in discriminating groups of patients or occasions.

There is no significance test associated with RV. We applied bootstrap to estimate the confidence interval (CI) of RV to better interpret the differences in RV.

METHODS: The CKD-specific legacy (KDQOL Burden, Symptom, and Effect), generic health scales (SF-12), and Kidney Disease Impact Scale (KDIS) were administered to 453 CKD patients. ANOVA-based RV coefficients were computed to compare how well each scale discriminated between three clinically-defined severity groups (Dialysis > Stage 3-5 > Transplant). Bootstrap was used to construct CI to determine whether the differences in RV were significant in comparisons between each scale and the best legacy standard- KDQOL Burden. Factors of sample size, number of bootstrap replications, bootstrap method were varied to investigate their impacts.

RESULTS: In comparison with KDQOL Burden (RV=1), using 95% CI, differences were non-significant for KDIS (RV=1.13), KDQOL Effect (RV=.99), SF-12 RP (RV=.77) and PF (RV=.70). SF-12 PCS (RV=.60) was at borderline. The other measures were significantly poorer in discriminating the patients.

Sample size played a substantial role. 300 patients for 3 groups greatly reduced the standard errors compared to 100 patients. A larger sample size greatly increased the power of detecting the differences.

The number of replications did not have consequential influence. The types of BCa and percentile intervals were preferred as all bootstrap distributions were skewed. The magnitude of chosen standard measure's F-statistics appeared to have a noticeable impact on CI too.

CONCLUSIONS: Bootstrapping appears to be valuable in comparing the validity of PRO measures from a statistical perspective. The significance test of RV was affected by the sample size, magnitude of RV, and F-statistic of standard measure.