Recommendations for Assessing the Validity of a Variable Crosswalk

By: Mitch Sevigny

Introduction

There exists many longitudinal research databases that collect multiple outcome measures. As research progresses, new instruments are developed to do a better job measuring these outcomes of interest. These new instruments are then used in replace of the traditional legacy measures within the longitudinal research database. Since those legacy measures are lost within the database, the subject's past data are also lost which effects researcher's ability to conduct longitudinal research projects. One solution for this issue is to convert the subject's traditional legacy measure into the new instrument measure using a variable crosswalk. Variable Crosswalks can be a valuable tool for researchers but issues can arise if the crosswalk is not reliable and does not properly convert the measures. Thus, it is essential for a variable crosswalk to be fully evaluated before it is implemented.

Despite the fact that a variable crosswalk converts an individual's measure from one instrument to another, it is important to evaluate the crosswalk at both the individual and group levels. At the individual level, the crosswalk must provide a converted legacy measure that is similar to that of the subject's actual outcome measure from the new instrument and it needs to be able to reduce the amount of uncertainty between these two measures. At the group level, the crosswalk converted legacy measure needs to possess a similar distribution to the actual outcome measure from the new instrument and be consistent across the different subpopulations. Here are some proposed recommendations to assess the validity of a variable crosswalk at both the individual and group level.

Solution

Individual Level

A valid variable crosswalk will be able to accurately convert an individual subject's legacy measure into the new instrument measure. This will be assessed by ensuring that the subject's crosswalk converted legacy measure is within a pre-determined range of their actual new outcome measure. The crosswalk conversion must also reduce the amount of uncertainty between the converted and actual outcome measures.

Reduction of Uncertainty

When converting the legacy measure into the new outcome measure, the variable crosswalk must be able to reduce the uncertainty by at least 50%. This can be calculated using the Reduction in Uncertainty (RiU) statistical index. RiU is a metric score that calculates the percentage of uncertainty reduction between two variables (Dorans, 2004). In order to reach 50% uncertainty reduction, the crosswalk converted legacy measure and the actual new outcome measure need to be highly correlated (r > 0.866).

Uncertainty Equation

Reduction in Uncertainty (RiU):	1 - CoA
Coefficient of Alienation[a] (CoA):	$\sqrt{1-r^2}$
Coefficient of Determination[b]:	r^2
Pearson Correlation[c]:	r

[a] : The proportion of variance in the independent variable that is not accounted for by the independent variable
[b] : Proportion of the variance in the dependent variable that is predictable form the independent variable
[c] : Strength of association between two continuous variables

Measure Comparison

An individual's crosswalk converted measure should be within a pre-determined or clinical range of their actual new outcome measure. A minimum of 75% of the subjects must have converted outcome measures within this pre-determined range of their actual outcome measure for the variable crosswalk to be valid. Suggested thresholds for the pre-determined range are half of a standard deviation (SD), the minimal clinically important difference (MCID), or some other clinically significant value supported by the literature. These thresholds can differ depending on the use and design of the variable crosswalk.

Difference Equation

$$\left|X_{(converted)} - X_{(actual)}\right| \leq \Delta$$

Δ : Threshold for pre-determined range

Group Level

A valid variable crosswalk will also be able to accurately convert the legacy measure to the new instrument measure so that the converted measure still represents the overall sample. The distribution of this converted legacy measure needs to be similar to that of the actual measure from the new instrument with a high percentage of overlap. The crosswalk converted measure must also remain consistent across the sample subpopulations like gender and race.

Population Invariance

The variable crosswalk should be unique and similar when converting measures across the sample's subpopulations. There shouldn't be any inconsistencies with the converted outcome measures when looking between different demographic and injury characteristics. One way of assessing this similarity between subpopulations is with population invariance. Population invariance can be calculated using the standardized mean difference between the two subpopulations divided by the pooled SD. The population invariance for the crosswalk convert measure and the actual outcome measure must have similar direction and magnitude within each subpopulation.

Subpopulation Invariance

$$\frac{Mean_{(a)} - Mean_{(b)}}{SD_{(pooled)}}$$

Statistical Moments

In order for the distribution of the crosswalk converted measure to be similar to the actual outcome measure, both need to have complementary statistical moments. Cooper (1989) provided guidelines for evaluating the first four statistical moments (mean, SD, Skewness, and Kurtosis) between two measures. The means between the converted and actual outcome measures should be within one SD. The SDs between the two measures should be within one unit. The 95% confidence interval (CI) for each measure's skewness must overlap. Similarly, the kurtosis 95% CI for both measures needs to overlap (Byers, 2004).

Effect Sizes and Percentage of Overlap

Calculating the percentage of overlap between two distributions is quite difficult, but one way to measure it is by using effect size. Effect size is a statistical concept used for measuring the strength of the relationship between two variables. Cohen's D is a common effect size method that is calculated using the mean difference between two measures, divided by the pooled SD. Smaller effect sizes indicate a higher percentage of overlap between the two measures (Cohen, 1988). The Cohen's D effect size will computed between the converted and actual outcome measures.

Cohen's D Effect Size

$$\frac{Mean_{(converted)} - Mean_{(actual)}}{SD_{(pooled)}}$$

Conclusion

A variable crosswalk is a valuable tool for researchers. However, research cannot progress utilizing unreliable or inaccurate data and tools. That's why evaluating a variable crosswalk at the individual and group level is important and essential. This recommended procedure for assessing the validity of a variable crosswalk will allow researchers to confidently conduct studies using replaced outcome measures within longitudinal research databases.

Example: TBI-QoL Depression T-Score and PHQ-9

An example for how to validate a variable crosswalk can be done using two measures of depression, the Patient Health Questionnaire (PHQ-9) and the TBI-QoL Depression T-score (TBI-QoL). The TBI-QoL is a patient-reported outcome measure of depression that was suggested to replace the traditional PHQ-9 score in the Traumatic Brain Injury Model Systems (TBIMS). A crosswalk was created to convert a subject's PHQ-9 score into the TBI-QoL score. Using 1381 subjects from the TBIMS, this variable crosswalk was evaluated using the recommended methods.

PHQ-9*	TBI-QoL		PHQ-9*	TBI-QoL
9	X_0		23	X_{14}
10	X_1		24	X_{15}
11	X_2		25	X_{16}
12	X_3		26	X_{17}
13	X_4		27	X_{18}
14	X_5		28	X_{19}
15	X_6		29	X_{20}
16	X_7		30	X_{21}
17	X_8		31	X_{22}
18	X_9		32	X_{23}
19	X_{10}		33	X_{24}
20	X_{11}		34	X_{25}
21	X_{12}		35	X_{26}
22	X_{13}		36	X_{27}

*: PHQ-9 scores +9

Individual Level

Reduction of Uncertainty

The crosswalk was evaluated at the individual level by calculating the reduction in uncertainty and comparing the scores. For the uncertainty, the correlation between the subject's actual TBI-QoL score and the crosswalk converted TBI-QoL score was calculated. There was a moderate correlation found between the two scores but this correlation was less than 0.866. Thus, the RiU was also below the recommendation of 50% reduction in uncertainty.

Measure Comparison

Each subject's crosswalk converted TBI-QoL score was compared to their actual TBI-QoL score using a threshold of half a SD. Less than 75% of the subjects were found to have their crosswalk converted scores within this threshold. This did not meet the suggested criteria of 75% of the subjects.

Group Level

Population Invariance

The TBI-QoL variable crosswalk was then evaluated at the group level by calculating the population invariance, comparing the statistical methods, and reviewing the effect size. The population invariance was evaluated across four different subpopulations; gender (males vs females), race (white vs non-white), ethnicity (Hispanic vs not Hispanic), and the subject's Glasgow Coma Scale (GCS) score (mild vs not mild). The direction of the invariance was the same between the crosswalk converted TBI-QoL score and the actual TBI-QoL score for all four subpopulations. The magnitude of the invariance was similar between the two score for the subject's GCS. The crosswalk converted score was slightly higher for the other three subpopulations. Overall, these population invariances met the recommended criteria.

Population Invariance

	Gender	Race	Ethnicity	GCS
Actual TBI-QoL	X_1	Y_1	Z_1	W_1
Converted TBI-QoL	X_2	Y_2	Z_2	W_2

Statistical Moments

The four statistical moments were compared between the actual and crosswalk converted TBI-QoL scores. The mean difference between the two scores was less than one SD and the difference in SDs was less than the recommendation of one unit. Both the skewness and kurtosis 95% CIs overlapped between the two scores. All four moments between the two scores met the recommended criteria.

	Actual TBI-QoL	Converted TBI-QoL
Mean	$\overline{X}$	$\overline{Y}$
Standard Deviation	SD_X	SD_Y
Skewness 95% CI	(LB_X, UB_X)	(LB_Y, UB_Y)
Kurtosis 95% CI	(LB_X, UB_X)	(LB_Y, UB_Y)

Effect Sizes and Percentage of Overlap

The Cohen's D effect size between the actual TBI-QoL score and crosswalk converted TBI-QoL score was small (<0.2). This small effect size translated to a large percentage of overlap between the two score distributions. This percentage of overlap met the recommended criteria.

Conclusion

The TBI-QoL variable crosswalk did not meet the recommended criteria at the individual level for either the reduction in uncertainty or the score comparison. However, this crosswalk did meet all the recommended criteria at the group level. Based on the suggested method for assessing a variable crosswalk, the TBI-QoL crosswalk could be used to convert the PHQ-9 for a large study sample but caution would be advised at the individual level.

References

Dorans NJ. (2004). Equating, concordance, and expectation. Applied Psychological Measurement, 28, 227–246.

McNemar, Q. (1969). Psychological statistics (4th ed.). New York: John Wiley.

(Cooper, 1989)*

Byers KL. (2004). Testing the accuracy of linking healthcare data across the continuum of care (Doctoral dissertation). Retrieved from http://etd.fcla.edu/UF/UFE0005961/byers_k.pdf

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

*: Article referenced in Byers dissertation