# Using JMP® and R integration to Assess Inter-rater Reliability in Diagnosing Penetrating Abdominal Injuries from MDCT Radiological Imaging

Melvin Alexander, Operations Research Analyst, Social Security Administration


## ABSTRACT

Fleiss' kappa (in JMP's Attribute Gauge platform) using ordinal rating scales helped assess inter-rater agreement between independent radiologists who diagnosed patients with penetrating abdominal injuries.

One drawback of Fleiss' kappa is that it does not estimate inter-rater reliability well enough since it is limited to disagreement distributions between raters being treated equally. Typically, disagreements are not all alike and have different magnitudes with multiple raters.

When compared to Fleiss' kappa, Krippendorff's alpha better differentiates between rater disagreements for various sample sizes; and estimates judgments, with or without missing data, across multiple measurement scales (binary, nominal, ordinal, interval, and ratio) for multiple raters.

Currently, Krippendorff's alpha is not available in JMP, but is available in the R open-source statistical programming language. JMP connects to R via JSL to execute R commands and exchange data.

This presentation will demonstrate how JMP and R integration allows users to take advantage of the powerful capabilities in both tools. Combining JMP and R helps users gain more insight and get better analytic results. Results helped radiologists discern which imaging signs detected injuries from signs that needed improved detection training.


## INTRODUCTION

This paper describes how JMP and R integration (via the JMP Scripting language) was used to prospectively assess the degree of agreement independent radiologists had in detecting penetrating GI tract- Bowel injuries (a.k.a. PBI). GI tract-Bowel injuries have led to serious morbidity and mortalities in trauma patients with sustaining penetrating torso injuries. Triple contrast (oral, rectal, and intravenous) multi-detector computed tomography (MDCT) has been used in the past decade as the primary means of evaluating penetrating GI tract- Bowel injuries (PBI) resulting from gunshot or stab wounds. Contrast media increases visibility of internal abdominal structures in CT imaging. Few studies have reported with high accuracy that triple contrast CT predicts the need for surgical treatment of penetrating abdominal injuries.

## MATERIALS AND METHODS

The Institutional Review Board (IRB) of the University of Maryland Medical Center (UMMC) approved the prospective observational study used in this example with a waiver of informed consent. Sixteen signs have been cited in the medical literature as key signs indicating penetrating GI Tract-Bowel injuries (PBIs). Such signs include: peritoneal violation; bone and bullet fragments extending to the GI tract; visible wound track hemorrhaging; gas and inflammation near the colon, intestines, and other abdominal organs; leakage of contrast media, etc.

Working with radiologists, a 5-point ordinal scale was developed to spot signs indicating PBI. This was the first rating scale (part of six-sigma's measurement phase) developed for MDCT PBI imaging.

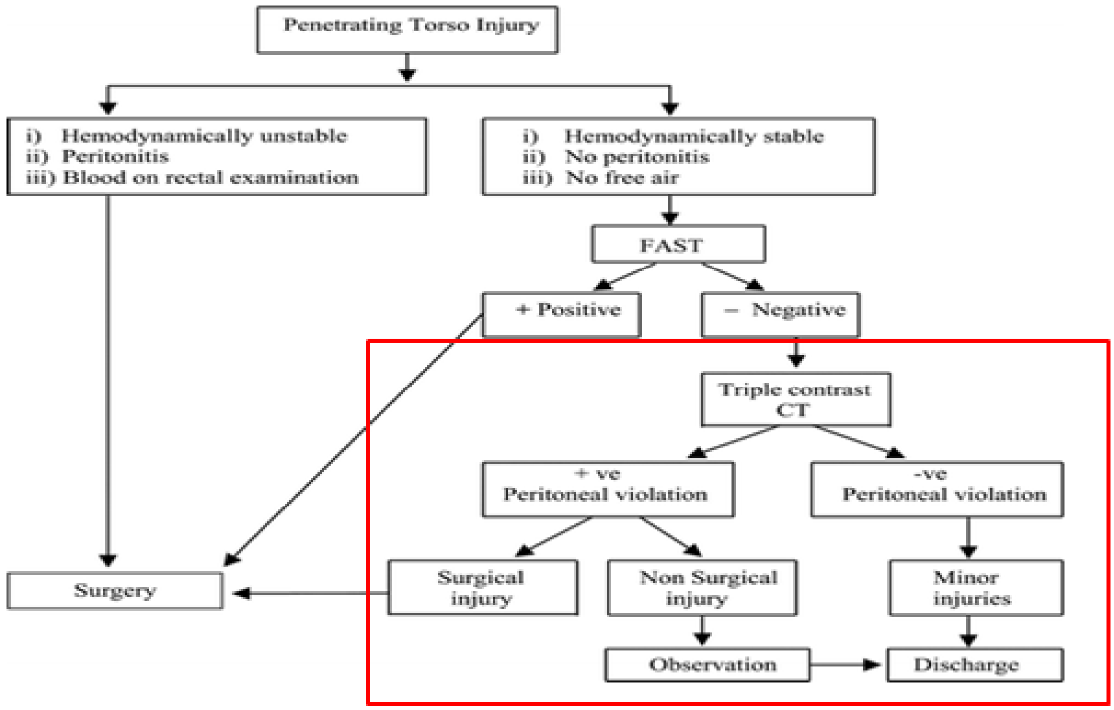
Each sign used a 5-point confidence scale (1-definitely absent, 2-possibly present but unlikely, 3-equivocal, 4-likely present, 5-definitely present).

The attending trauma surgeons and radiologists (reference standard) determined the presence or absence of GI tract injury intra-operatively (1=BI injury, 0 = No BI injury) after patient admission.

All radiologists scored a binary variable indicating whether the patient required an operative management specifically for GI tract injury or not.

Figure 1 [1, p. 783] depicts the **F**ocused **A**ssessment with **S**onography for **T**rauma (FAST) algorithm. FAST is the Rapid Sonographic, Ultrasound, MDCT Examination used by radiologists, surgeons, emergency physicians, and certain paramedics to screen for blood around the heart or abdominal organs after trauma. The FAST protocol was applied to detect the key signs indicating penetrating GI Tract-Bowel injuries (PBIs).

**FIGURE 1: TRIPLE-CONTRAST IMAGING ALGORITHM (FAST) FOR STABLE PATIENTS WITH PENETRATING TRAUMA (Reprinted with permission from the Radiological Society of North America)**




The FAST protocol in Figure 1 directs patient movement after they were admitted and imaged. No Peritonitis means that patients had no inflammed peritoneum (i.e., membrane lining the abdominal wall and internal tissues) nor any presence of free air or gas in the abdominal (peritoneal) cavity. The right side of the FAST path (inside the red rectangle) determines the patient treatment course options (surgery, observation, discharge) based on CT findings.

Table 1 lists the direct and indirect signs after the FAST protocol was used. A 17th overall CT diagnosis of GI injury (CToverall) was computed using the same 5-point confidence scale based on assessed combinations of the 16 signs.

**TABLE 1: 16 SIGNS CITED IN THE MEDICAL LITERATURE AS INDICATIVE OF PRESENCE OF GI INJURY.**

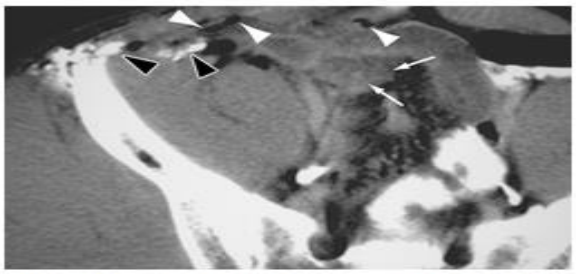
| JMP Column | Sign | Peritoneum – Membrane (thin tissues) that lines the abdomen and inner abdominal organs |
|---|---|---|
| Q1 | Peritoneal violation | Indirect sign of injury to the peritoneum |
| Q2 | Retroperitoneal violation | Indirect sign of injury to retro-peritoneal colon |
| Q3a | Free gas at wound track | Indirect sign of free intraperitoneal/retroperitoneal gas *adjacent to* the GI injury site |
| Q3b | Free gas away from wound track | Indirect sign of free intraperitoneal/retroperitoneal gas *remote to* the GI injury site |
| Q4 | Intramural gas | Direct sign of intramural air |
| Q5 | Leakage of luminal content | Direct sign bleeding into GI lumen (inside space of intestinal tubular structure) |
| Q6 | Leakage of oral/rectal contrast | Direct Sign of leakage of enteric contrast material |
| Q7 | Discontinuity of bowel wall | Direct sign of GI |
| Q8 | Bowel wall thickening | Direct sign of peritoneal thickening or enhancement |
| Q9 | Free fluid | Indirect sign of free intraperitoneal fluid |
| Q10 | Mesenteric hematoma | Indirect sign of blood leakage into intestinal organs |
| Q11 | Active mesenteric bleeding | Indirect sign of extensive hemorrhaging into intestinal organs |
| Q12 | Signs of peritonitis | Indirect sign of peritoneum inflammation |
| Q13 | Signs of solid organ injury | Indirect sign of injuries to other abdominal organs |
| Q14 | Intraluminal bleeding | Direct sign of bleeding into GI lumen |
| Q15 | Wound track extending up to bowel | Direct sign of visible penetrating wound track hemorrhage, air, and/or ballistic fragments that extended up to the GI wall |

CT images of 171 patients underwent MDCT imaging for surgery (77/171, 45.0%) or clinical follow-up (94/171, 55.0%) between October, 2011 and April, 2013.

Images were interpreted by three independent radiologists, (one attending radiologist and two secondary readers). Each radiologist interpreted each patient scan and recorded findings on dedicated worksheets. The radiologists were blind to one another's imaging, clinical data, and patient management outcomes.

Figure 2 is an example of a CT image showing PBI from [1, p. 779]. Signs shown are: Bone and bullet fragments (black arrowheads), ileum bleeding, and the small loop outline of the small bowel around the wound tract (Q15); free intraperitoneal air (Q3a, Q3b-white arrowheads); and free fluids (Q9- white arrows) indicating peritoneal violation (Q1).

**FIGURE 2: EXAMPLE OF PENETRATING BOWEL INJURY (PBI) (Reprinted with permission from the Radiological Society of North America)**



Transverse CT scan shows a wound track that extends to the bowel in a 33-year-old man shot in the right hemipelvis. Note bullet and bone fragments (black arrowheads) outlining the wound track that extends to a loop of small bowel (Q15). Free intraperitoneal air (Q3a, Q3b - white arrowheads) and fluid (Q9-arrows) are seen from peritoneal violation (Q1). An injury to the ileum was confirmed surgically.

**Penetrating Torso Trauma: Triple-Contrast Helical CT in Peritoneal Violation and Organ Injury—A Prospective Study in 200 Patients[1]**

K. Shanmuganathan, MD
Stuart E. Mirvis, MD
William C. Chiu, MD
Karen L. Killeen, MD[2]
Gerald J. F. Hogan, MD
Thomas M. Scalea, MD

**Index terms:**
Computed tomography (CT), helical, 78.12115
Computed tomography (CT), multi–detector row, 78.12112
Emergency radiology
Trauma

**Published online before print**
10.1148/radiol.2313030126
**Radiology 2004;** 231:775–784

Table 2 gives a sample of selected records of the JMP Data Table with the scores from each radiologist (Readernumber column) for each patient (No column) and specific signs (columns Q1-Ctoverall). Readernumber 1 was the attending radiologist.

**TABLE 2: SELECTED RECORDS OF THE DATA**

| | No | BI | Readernumber | Q1 | Q2 | Q3a | Q3b | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Q11 | Q12 | Q13 | Q14 | Q15 | Ctoverall | Numeric_BI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Y | 1 | 5 | 1 | 1 | 5 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 |
| 2 | 1 | Y | 2 | 5 | 1 | 1 | 5 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 1 |
| 3 | 1 | Y | 3 | 5 | 1 | 5 | 5 | 1 | 2 | 1 | 1 | 1 | 5 | 2 | 1 | 1 | 1 | 3 | 3 | 2 | 1 |
| 4 | 2 | N | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 5 | 2 | N | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 6 | 2 | N | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 7 | 3 | N | 1 | 5 | 5 | 5 | 1 | 1 | 1 | • | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 2 | 0 |
| 8 | 3 | N | 2 | 5 | 1 | 1 | 1 | 1 | 1 | • | 3 | 5 | 5 | 5 | 1 | 1 | 5 | 3 | 4 | 4 | 0 |
| 9 | 3 | N | 3 | 5 | 5 | 5 | 1 | 1 | 1 | • | 1 | 1 | 5 | 1 | 1 | 1 | 5 | 1 | 1 | 2 | 0 |
| 10 | 4 | N | 1 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 1 | 5 | 1 | 1 | 1 | 0 |
| 11 | 4 | N | 2 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 1 | 5 | 1 | 1 | 1 | 0 |
| 12 | 4 | N | 3 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 1 | 5 | 1 | 5 | 1 | 0 |
| 13 | 5 | N | 1 | 5 | 1 | 5 | 2 | 1 | 2 | • | 2 | 1 | 5 | 2 | 1 | 1 | 5 | 2 | 2 | 2 | 0 |
| 14 | 5 | N | 2 | 5 | 1 | 5 | 5 | 1 | 1 | • | 1 | 4 | 5 | 1 | 1 | 2 | 5 | 1 | 4 | 2 | 0 |
| 15 | 5 | N | 3 | 5 | 1 | 5 | 5 | 1 | 2 | • | 1 | 2 | 5 | 1 | 1 | 1 | 5 | 1 | 1 | 2 | 0 |
| 16 | 6 | Y | 1 | 1 | 5 | 5 | 5 | 1 | 1 | 5 | 5 | 5 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 1 |
| 17 | 6 | Y | 2 | 2 | 5 | 5 | 1 | 2 | 4 | 5 | 5 | 5 | 1 | 1 | 1 | 1 | 1 | 3 | 5 | 5 | 1 |
| 18 | 6 | Y | 3 | 1 | 5 | 5 | 5 | 1 | 1 | 1 | 1 | 5 | 5 | 1 | 1 | 1 | 1 | 1 | 5 | 4 | 1 |
| 19 | 7 | N | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 20 | 7 | N | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| 21 | 7 | N | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |

JMP's Attribute Gauge Platform [2] allowed comparisons: between raters; within each other; of raters vs. a standard; and across the scales. Kappa [3, 4] measured the degree of agreement by different raters beyond chance agreement over the rating scale. Krippendorff's alpha [5-14] is a coefficient of reliability that measures the extent of agreement among different raters. It is distinct among other similar measures of agreement (Cohen and Fleiss' kappas, Scott's pi, Intraclass Correlation Coefficient, Kendall's Coefficient of Concordance, etc.) in that it applies to any number of observers, and generalizes across scales of measurement (binary, nominal, ordinal, interval, ratio) with or without missing data.

Krippemdorff alpha provided a more generalized measure of inter-rater reliability than kappa. Kappa treats disagreement distributions equally among different raters. Krippendorff alpha takes the degree of disagreement among different raters into account more generally for many types of measurement scales. For example, suppose three ratings were assigned to cases, like (+1), neutral (0), and dislike (-1). The degree of difference between two raters of like (+1) vs. neutral (0) or dislike (-1) vs. neutral (0) would be taken into account the same way as paired ratings of like (+1) vs. dislike (-1) using kappa.

Krippendorff alpha would treat the like (+1) vs. dislike (-1) more severely than like (+1) vs. neutral (0) or dislike (-1) vs. neutral (0) classification by two or more raters. This is partly due to the fact that kappa treats such classifications strictly as nominal. On the other hand, Krippendorff alpha discerns the measurement classifications as ordinal, especially when multiple raters have different magnitudes of disagreements or there are incomplete data. Different formulas are used to compute alphas, depending upon the nominal, ordinal, interval, ratio (even polar and circular) metrics for the data.

Since JMP did not compute this statistic, the kripp.alpha function in the lpSolve and irr packages of R [11] was used to obtain the metric using JMP and R integration (see Appendix 1). The bootalpha function [12] produced 95% confidence intervals of krippendorff alphas from 1000 bootstrapped resamples for each sign (Table 3).

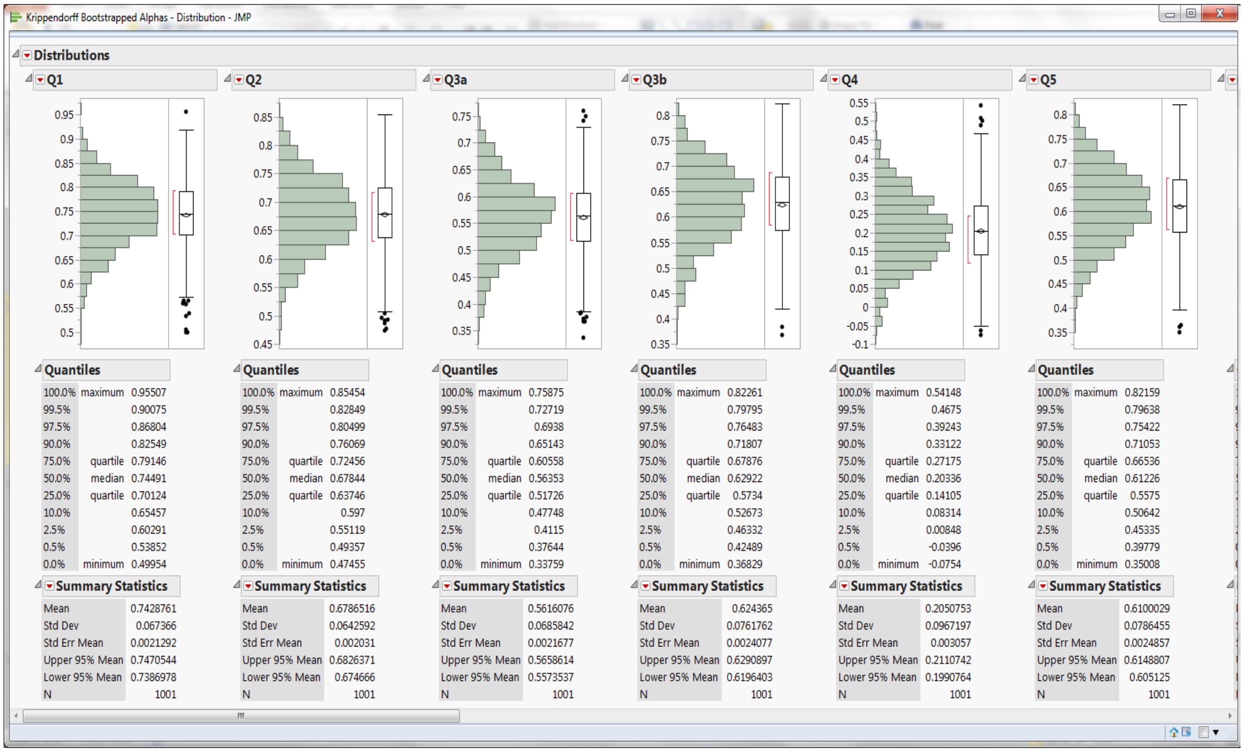
**TABLE 3: SELECTED SIGNS OF 1001 KRIPPENDORFF'S ALPHAS (FIRST ROW IS ARE THE KRIPPENDORFF ALPHA VALUES FOR EACH SIGN OF THE ORIGINAL, SOURCE DATA)**

Krippendorff Bootstrapped Alphas - JMP

| | Q1 | Q2 | Q3a | Q3b | Q4 | Q5 | Q6 | Q7 | Q8 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.7503407205 | 0.6863852424 | 0.5696314041 | 0.6282646652 | 0.2139495116 | 0.6155022449 | 0.7821915078 | 0.4424166103 | 0.477530414 |
| 2 | 0.7210360746 | 0.7859200824 | 0.5046050838 | 0.6056230326 | 0.1947785763 | 0.7059581525 | 0.8079280009 | 0.4604220466 | 0.389775604 |
| 3 | 0.5720361086 | 0.6482426264 | 0.514438717 | 0.6588734576 | 0.1879729034 | 0.5646124566 | 0.7947567094 | 0.1381087392 | 0.538654539 |
| 4 | 0.7506966294 | 0.6023449735 | 0.6158203202 | 0.6521212777 | 0.2025975456 | 0.5054917158 | 0.5698358853 | 0.3346991954 | 0.438355545 |
| 5 | 0.7715459011 | 0.6272030508 | 0.6195849086 | 0.6861634784 | 0.1673904465 | 0.5048999755 | 0.8252894444 | 0.5993458945 | 0.528086890 |
| 6 | 0.6280803851 | 0.5600545567 | 0.6716101968 | 0.6711610727 | 0.3420326486 | 0.539202459 | 0.5018901591 | 0.4802878074 | 0.72450041 |
| 7 | 0.7732603879 | 0.6761144729 | 0.5274906005 | 0.7066767284 | 0.1077347 | 0.7335028986 | 0.7750818821 | 0.3739184774 | 0.419341013 |
| 8 | 0.7338822678 | 0.6876724506 | 0.5193174203 | 0.6429510506 | 0.1451848429 | 0.5748235679 | 0.8108623581 | 0.4449492196 | 0.364786389 |
| 9 | 0.6769203931 | 0.6227643765 | 0.7497746122 | 0.6337941785 | 0.2846360342 | 0.457483872 | 0.8227957925 | 0.3241417287 | 0.361501689 |
| 10 | 0.7235056939 | 0.768018476 | 0.6630945567 | 0.7272999747 | 0.2305584332 | 0.5848968593 | 0.7870634607 | 0.3672754197 | 0.429156673 |
| 11 | 0.7605945339 | 0.6635798092 | 0.5727918574 | 0.6622085244 | 0.2106267704 | 0.6614732251 | 0.8006302235 | 0.4187850251 | 0.307332009 |
| 12 | 0.7708749225 | 0.682339275 | 0.5350938842 | 0.5976630835 | 0.1321558467 | 0.5854656389 | 0.7604468865 | 0.3719499532 | 0.589711358 |
| 13 | 0.6372039997 | 0.6426098698 | 0.5718887318 | 0.6707687996 | 0.1458339579 | 0.5999384779 | 0.7190363121 | 0.4021871538 | 0.545018387 |
| 14 | 0.7640478029 | 0.6141551123 | 0.5221268047 | 0.7026533601 | 0.1616578682 | 0.7105800187 | 0.9514525139 | 0.5125407724 | 0.40452465 |
| 15 | 0.7600176898 | 0.6494131183 | 0.6084140732 | 0.7979545441 | 0.180615877 | 0.6792807285 | 0.8240211237 | 0.4751102005 | 0.606648207 |
| 16 | 0.5807408431 | 0.7718705223 | 0.6321302269 | 0.5778995766 | 0.2933670118 | 0.6802872078 | 0.7666736916 | 0.3665353209 | 0.487058037 |
| 17 | 0.7805272207 | 0.6730492708 | 0.6605950713 | 0.5793462048 | 0.3540457742 | 0.6750115674 | 0.7197741931 | 0.4351805292 | 0.497322298 |
| 18 | 0.6507940291 | 0.7230874432 | 0.5974847323 | 0.444380781 | 0.2207898559 | 0.6709899714 | 0.6418147685 | 0.5783670043 | 0.409557003 |
| 19 | 0.7419538518 | 0.7715424138 | 0.5384472799 | 0.6425875489 | 0.2165720164 | 0.6810537925 | 0.7496650449 | 0.5140910758 | 0.375117370 |
| 20 | 0.7616445982 | 0.7040156087 | 0.4802663329 | 0.7130578605 | 0.2429206431 | 0.7159862525 | 0.7384735901 | 0.5340342135 | 0.508478506 |
| 21 | 0.6912089883 | 0.5453255004 | 0.4940759801 | 0.600797481 | 0.124466469 | 0.6323482586 | 0.7156257938 | 0.5265215436 | 0.453879359 |

Figure 3 shows the bootstrapped resampling distribution of Krippendorff alphas for selected signs.

**FIGURE 3: DISTRIBUTION OF 1001 BOOTSTRAPPED KRIPPENDORFF'S ALPHAS FOR SELECTED SIGNS**




The remaining script in Appendix 1 vertically concatenated the 95% Krippendorff alpha confidence interval matrices of all signs and formed a final data table.Table 4 lists the resulting Data Table of the observed Krippendorff alpha values and the 95% confidence intervals for each sign from the original data in Table 2.

**TABLE 4: KRIPPENDORFF'S ALPHA DATA TABLE WITH 95% CONFIDENCE INTERVALS COMPUTED FROM 1000 BOOTSTRAPPED RESAMPLES**

Krippendorff Alpha and 95%CI - JMP

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

Krippendorff Alpha and 95%CI
Stack Alphas and 95% CIs and plc

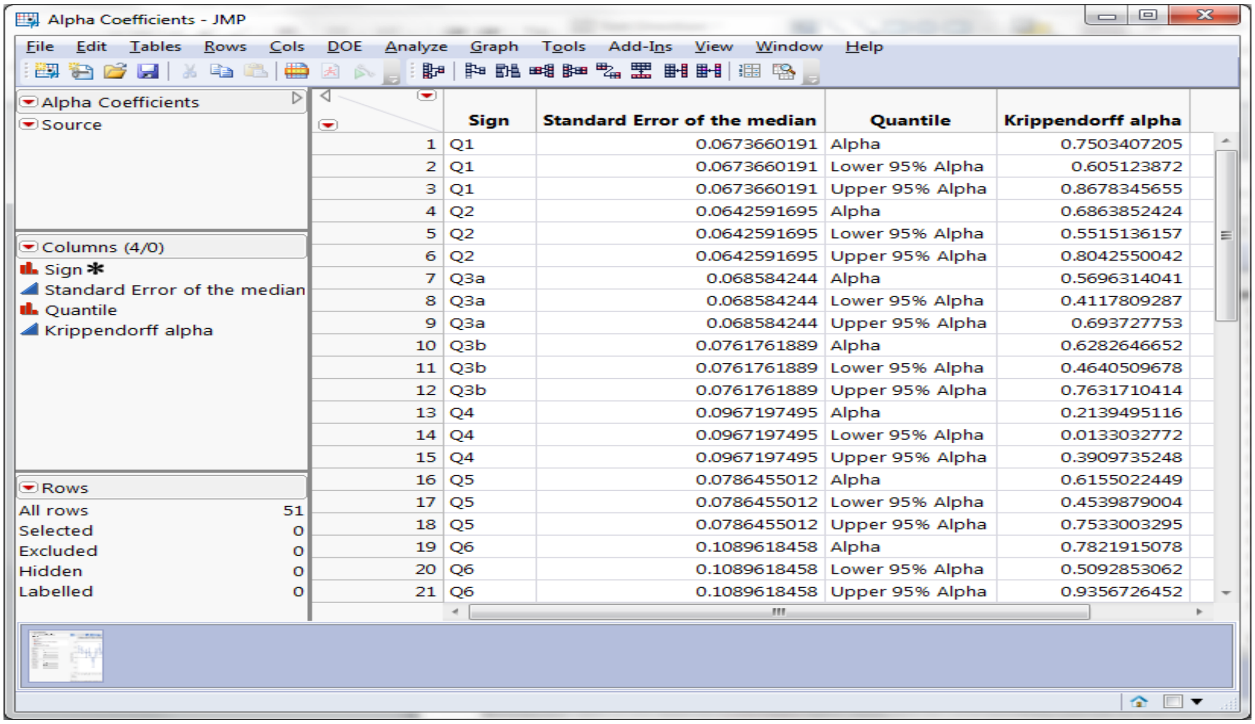
Columns (5/0)
Sign
Alpha
Standard Error of the median
Lower 95% Alpha
Upper 95% Alpha

Rows
All rows 17
Selected 0
Excluded 0
Hidden 0
Labelled 0

| | Sign | Alpha | Standard Error of the median | Lower 95% Alpha | Upper 95% Alpha |
|---|---|---|---|---|---|
| 1 | Q1 | 0.7503407205 | 0.0673660191 | 0.605123872 | 0.8678345655 |
| 2 | Q2 | 0.6863852424 | 0.0642591695 | 0.5515136157 | 0.8042550042 |
| 3 | Q3a | 0.5696314041 | 0.068584244 | 0.4117809287 | 0.693727753 |
| 4 | Q3b | 0.6282646652 | 0.0761761889 | 0.4640509678 | 0.7631710414 |
| 5 | Q4 | 0.2139495116 | 0.0967197495 | 0.0133032772 | 0.3909735248 |
| 6 | Q5 | 0.6155022449 | 0.0786455012 | 0.4539879004 | 0.7533003295 |
| 7 | Q6 | 0.7821915078 | 0.1089618458 | 0.5092853062 | 0.9356726452 |
| 8 | Q7 | 0.4424166103 | 0.0939232452 | 0.237861742 | 0.6028726855 |
| 9 | Q8 | 0.4775304141 | 0.0769103174 | 0.3276420937 | 0.6268635406 |
| 10 | Q9 | 0.6790435774 | 0.0637603594 | 0.5384322913 | 0.7895230257 |
| 11 | Q10 | 0.4134978718 | 0.0818086061 | 0.2397866828 | 0.5604381094 |
| 12 | Q11 | 0.4477476322 | 0.1623261924 | 0.0834909273 | 0.7251071094 |
| 13 | Q12 | -0.019908968 | 0.0650933494 | -0.038277248 | -0.002173913 |
| 14 | Q13 | 0.8223807001 | 0.047156654 | 0.720483304 | 0.8992787812 |
| 15 | Q14 | 0.1285898652 | 0.1073809241 | -0.088778552 | 0.3285515847 |
| 16 | Q15 | 0.5571417835 | 0.0715141086 | 0.4083534376 | 0.6852612036 |
| 17 | Ctoverall | 0.792743252 | 0.0391764202 | 0.6985028447 | 0.8509944042 |

Table 5 shows the stacked columns of Table 4 that was used to prepare Figure 4.

**TABLE 5: STACKED COLUMN OF KRIPPENDORFF'S ALPHAS (COEFFICIENT AND 95% CONFIDENCE INTERVAL VALUES)**



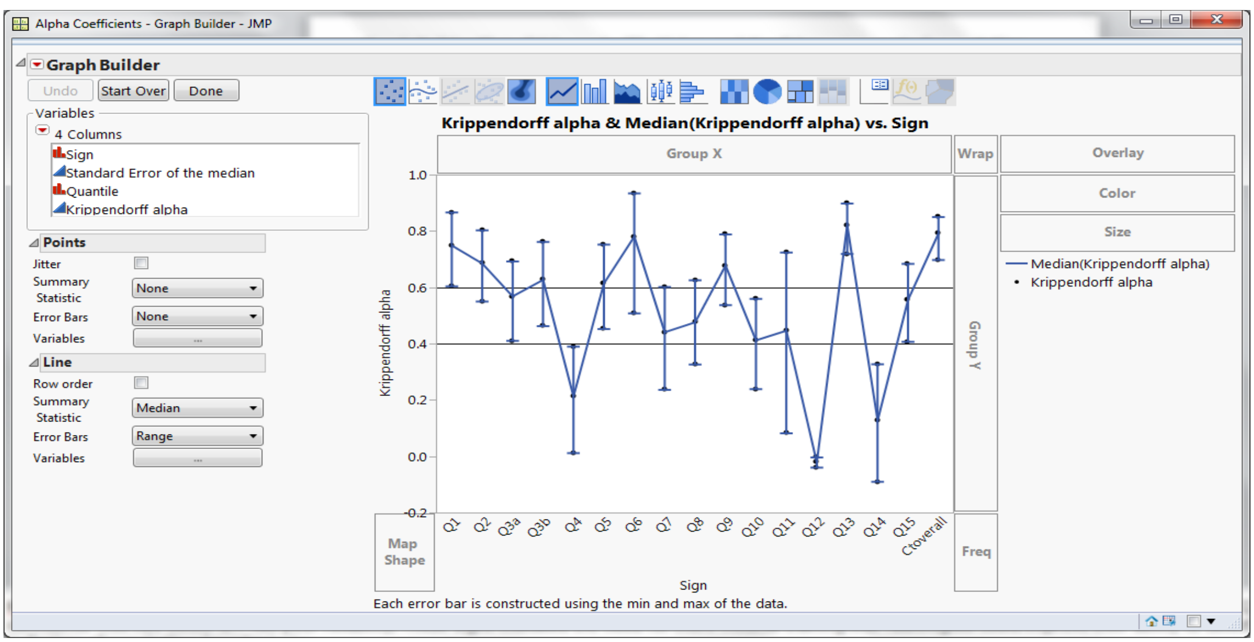
Alpha Coefficients - JMP

| | Sign | Standard Error of the median | Quantile | Krippendorff alpha |
|---|---|---|---|---|
| 1 | Q1 | 0.0673660191 | Alpha | 0.7503407205 |
| 2 | Q1 | 0.0673660191 | Lower 95% Alpha | 0.605123872 |
| 3 | Q1 | 0.0673660191 | Upper 95% Alpha | 0.8678345655 |
| 4 | Q2 | 0.0642591695 | Alpha | 0.6863852424 |
| 5 | Q2 | 0.0642591695 | Lower 95% Alpha | 0.5515136157 |
| 6 | Q2 | 0.0642591695 | Upper 95% Alpha | 0.8042550042 |
| 7 | Q3a | 0.068584244 | Alpha | 0.5696314041 |
| 8 | Q3a | 0.068584244 | Lower 95% Alpha | 0.4117809287 |
| 9 | Q3a | 0.068584244 | Upper 95% Alpha | 0.693727753 |
| 10 | Q3b | 0.0761761889 | Alpha | 0.6282646652 |
| 11 | Q3b | 0.0761761889 | Lower 95% Alpha | 0.4640509678 |
| 12 | Q3b | 0.0761761889 | Upper 95% Alpha | 0.7631710414 |
| 13 | Q4 | 0.0967197495 | Alpha | 0.2139495116 |
| 14 | Q4 | 0.0967197495 | Lower 95% Alpha | 0.0133032772 |
| 15 | Q4 | 0.0967197495 | Upper 95% Alpha | 0.3909735248 |
| 16 | Q5 | 0.0786455012 | Alpha | 0.6155022449 |
| 17 | Q5 | 0.0786455012 | Lower 95% Alpha | 0.4539879004 |
| 18 | Q5 | 0.0786455012 | Upper 95% Alpha | 0.7533003295 |
| 19 | Q6 | 0.1089618458 | Alpha | 0.7821915078 |
| 20 | Q6 | 0.1089618458 | Lower 95% Alpha | 0.5092853062 |
| 21 | Q6 | 0.1089618458 | Upper 95% Alpha | 0.9356726452 |



Figure 4 is a Graph-Builder plot of the Krippendorff alpha 95% Confidence Intervals. Alpha values above 0.8 (Q13 and Ctoverall) indicated good reliability in detection between raters. Three Signs (Q4, Q12, and Q14) with confidence intervals below 0.4 had poor reliability. These signs indicated the need for more medical training so that radiologists can improve their ability to detect them as indicators of penetrating bowel injuries.

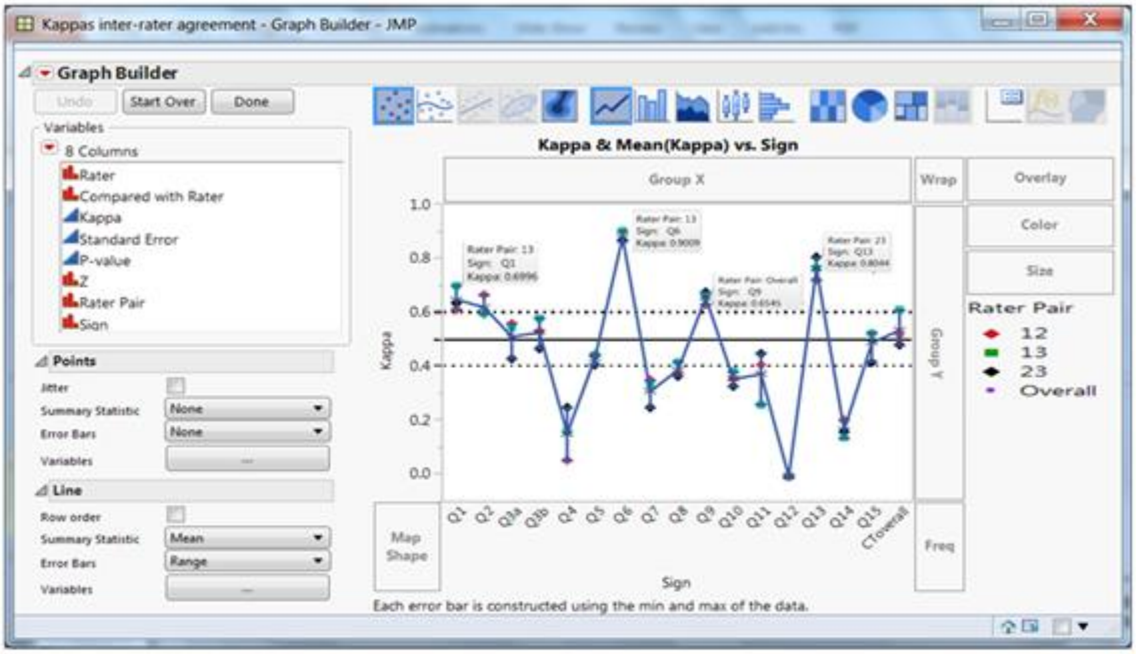
**FIGURE 4: GRAPH-BUILDER PLOT OF KRIPPENDORFF'S ALPHA AND CONFIDENCE INTERVALS FOR EACH SIGN**





The Kappa values above 0.6 (Q1,Q2,Q6, Q9, and Q13) in Figure 5 indicate good agreement between raters. Five Signs (Q3a, Q3b, Q5, Q15, and Ctoverall) had moderate kappa values between 0.4-0.6. Four signs (Q7, Q8, Q10, and Q11) had fair agreement with kappa values between 0.2-0.4. Signs Q4, Q12, and Q14 had poor agreement. Like the Krippendorff alpha, the fair-to-poor kappa values (below 0.2) suggested the need for increased medical training in detecting these signs in order to improve the measurement system capability of the rating scales.

**FIGURE 5: GRAPH-BUILDER PLOT OF PAIRED RATERS AND OVERALL KAPPA VALUES FOR EACH SIGN**




Next, I joined the Fleiss overall kappa statistics with the Krippendorff alpha data table as shown in Figure 6 to form Table 6 and Figure 7.

**FIGURE 6: OPENED FLEISS KAPPA AND KRIPPENDORFF ALPHA DATA TABLES WITH SCRIPT TO JOIN THEM**

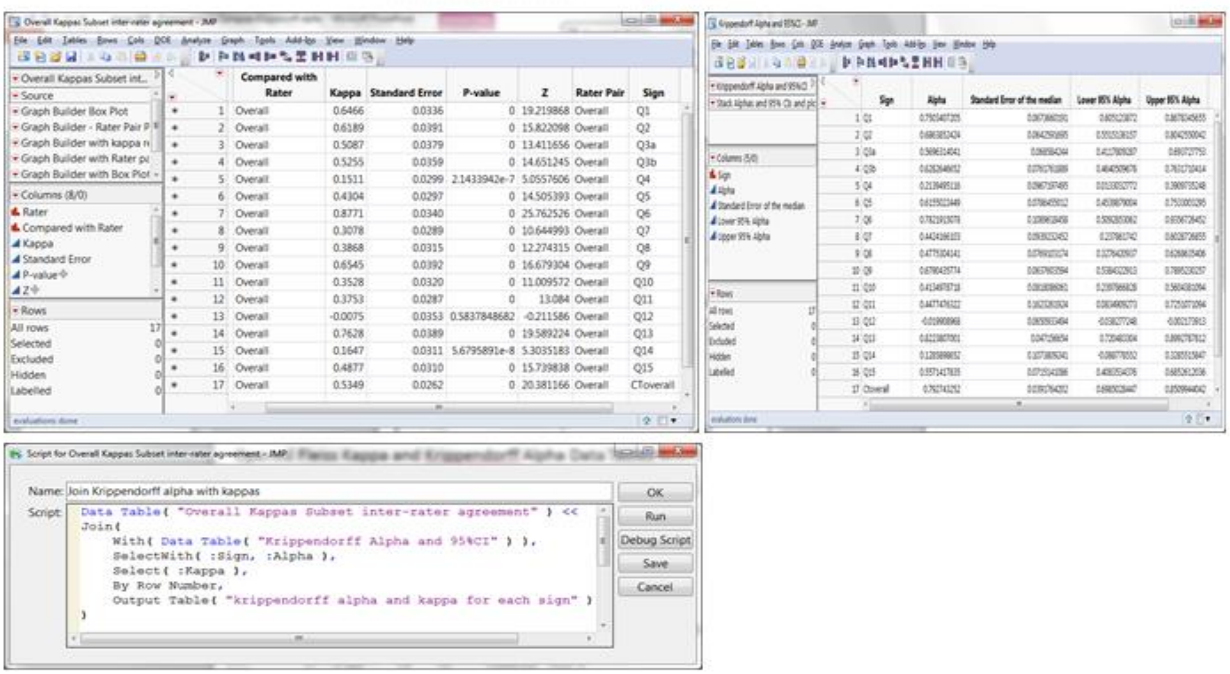
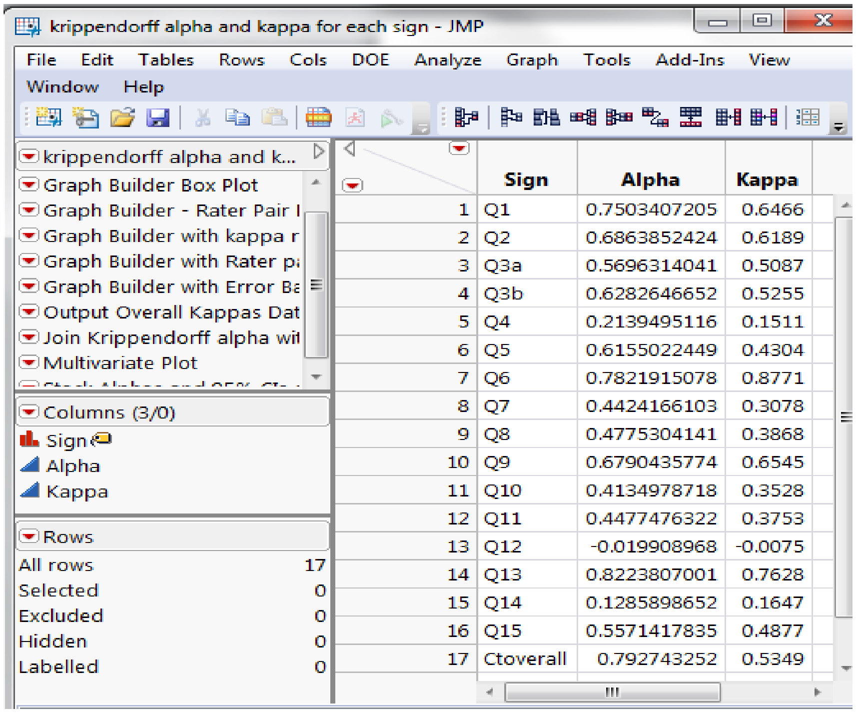


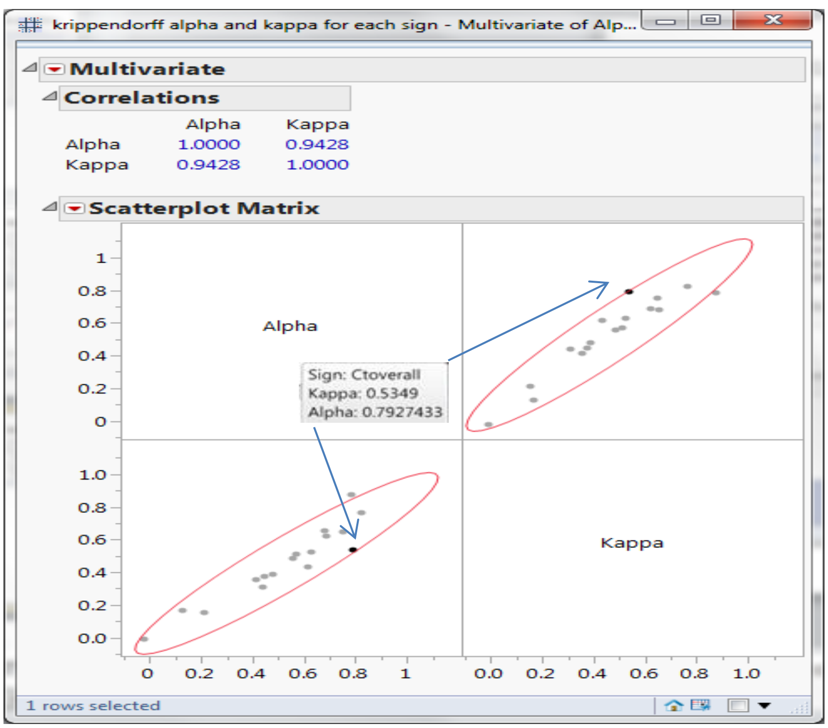
**TABLE 6: KRIPPENDORFF ALPHA AND OVERALL KAPPA DATA TABLE**



| | Sign | Alpha | Kappa |
|---|---|---|---|
| 1 | Q1 | 0.7503407205 | 0.6466 |
| 2 | Q2 | 0.6863852424 | 0.6189 |
| 3 | Q3a | 0.5696314041 | 0.5087 |
| 4 | Q3b | 0.6282646652 | 0.5255 |
| 5 | Q4 | 0.2139495116 | 0.1511 |
| 6 | Q5 | 0.6155022449 | 0.4304 |
| 7 | Q6 | 0.7821915078 | 0.8771 |
| 8 | Q7 | 0.4424166103 | 0.3078 |
| 9 | Q8 | 0.4775304141 | 0.3868 |
| 10 | Q9 | 0.6790435774 | 0.6545 |
| 11 | Q10 | 0.4134978718 | 0.3528 |
| 12 | Q11 | 0.4477476322 | 0.3753 |
| 13 | Q12 | -0.019908968 | -0.0075 |
| 14 | Q13 | 0.8223807001 | 0.7628 |
| 15 | Q14 | 0.1285898652 | 0.1647 |
| 16 | Q15 | 0.5571417835 | 0.4877 |
| 17 | Ctoverall | 0.792743252 | 0.5349 |

The Scatterplot Matrix density ellipses of Figure 7 showed the strong positive correlation (0.9428) between the overall Fleiss' kappas and Krippendorff alphas across all signs.

**FIGURE 7: MULTIVARIATE CORRELATIONS BETWEEN THE KRIPPENDORFF ALPHA AND OVERALL FLEISS' KAPPA STATISTICS**



## CONCLUSION

This presentation showed how JMP's JSL with R integration computed a Krippendorff's alpha inter-rater reliability statistic that complements Fleiss' kappa in JMP. Although Krippendorff alpha is not as widely used and is more computationally complex than Fleiss' kappa, it has gained more acceptance by researchers in content analysis and measurement reproducibility studies as a robust, flexible, inter-coder reliability metric. Combining JMP's dynamic interactivity with R's unique functionalities and packages gives users the ability to develop custom applications that can be implemented in JMP.

## REFERENCES


[1]. Shanmuganathan, K., Mirvis, S.E., Chiu, W.C., Killeen, K.L., Hogan, G.J.F., Scalea, T.M. (2004), "Penetrating Torso Trauma: Triple-Contrast Helical CT in Peritoneal Violation and Organ Injury—A Prospective Study in 200 Patients," *Radiology*, 231, 775-784.

[2]. SAS Institute Inc. (2014). "Attribute Gauge Charts," *JMP® 11 Quality and Process Methods*. Cary, NC: SAS Institute Inc., pp. 193-205.

[3]. De Mast, J and van Wieringen, W. (2004). "Measurement System Analysis for Bounded Ordinal Data," *Quality and Reliability Engineering International*, 20, 383-395.

[4]. Fleiss, J.L., Levin, B., and Paik, M.C. (2003), *Statistical Methods for Rates and Proportions (3rd ed.).* New York: Wiley.

[5]. Wikipedia, "Krippendorff's Alpha," http://en.wikipedia.org/wiki/Krippendorff's_alpha.

[6]. Hayes, A. F., & Krippendorff, K. (2007). "Answering the Call for a Standard Reliability Measure for Coding Data," *Communication Methods and Measures, 1*, 77-89.

[7]. Kang, N. et al. (1993), "A SAS® Macro for Calculating Intercoder Agreement in Content Analysis," *Journal of Advertising*, 23, 17-28.

[8]. Krippendorff, K. (2013). *Content Analysis: An Introduction to its Methodology (3rd ed.)*. Thousand Oaks, CA: Sage.

[9]. Krippendorff, K. (2007), "Computing Krippendorff's Alpha-Relaibility," http://www.asc.upenn.edu/usr/krippendorff/mwebreliability5.pdf.

[10]. Gwet, K. (2011), "On the Krippendorff's Alpha Coefficient," http://www.agreestat.com/research_papers/onkrippendorffalpha_old.pdf.

[11]. Gamer, M. et al., (2012) "Various Coefficients of Interrater Reliability and Agreement," http://cran.r-project.org/web/packages/irr/irr.pdf.

[12]. Gruszcynski, M. (2013). "R Convenience functions to perform bootstrap resampling on intercoder reliability data for Krippendorff's statistics," https://github.com/MikeGruz/kripp.boot/blob/master/R/kripp.boot.R.

[13]. Hayes, A.F. (2011), "KALPHA SAS Macro," http://www.afhayes.com/public/kalpha.sas.

[14]. Artstein, R. and Poesio, M. (2008). "Inter-Coder Agreement for Computational Linguistics," *Computational Linguistics*, 34, 555-596.


## ACKNOWLEDGMENTS


I thank Nitima Saksobhavivat, MD and K. Shanmuganathan, MD of the University of Maryland's Medical Center for providing the impetus to this paper; Robin Hughes Moran, Jeff Perkinson, Tom Donnelly, Kathy Walker, Holly Connors, Lori Harris, John Sall, and the JMP Division of SAS® for their contributions and support; and Lucia Ward-Alexander for her review and editorial assistance.


## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:


Melvin Alexander
Social Security Administration
6401 Security Blvd.; East High Rise Building (5-A-10)
Baltimore, MD 21235
Phone: (410) 966-2155
Fax: (410) 966-4337
E-mail: Melvin.Alexander@ssa.gov

## DISCLAIMER

The views expressed in this presentation are the author’s and do not represent the views of the Social Security Administration, the University of Maryland Medical Center, or SAS Institute, Inc.

### APPENDIX 1: JSL SCRIPT TO COMPUTE KRIPPENDORFF’S ALPHA USING JMP AND R INTEGRATION

```
/***************************************************************
  This JMP and R Integration example computes Krippendorff's alpha
  reliability coefficient that measures agreement among
  different raters or measuring instruments drawing distinctions
  among unstructured phenomena or assign computable values to
  items.

  Alpha emerged from content analysis whenever two or more
  methods of generating data are applied to the same subjects.
  Alpha answers the question of how much the resulting data
  can be trusted to represent something real.

  The R implementation uses the lpSolve and irr library packages
  with the kripp.alpha function with ordinal measures.

  Visit http://en.wikipedia.org/wiki/Krippendorff's_alpha
  for more information.

  Author: Melvin Alexander
  Version: 1, 12Feb14 – Initial coding
  Updates:

**************************************************************/

R Init() ;
//Open Data Table with Response variable (BI) and Signs (Explanatory variables)
//pbi 11 for Mel July_02_2013 contains all data
dt = Open("G:\Nitima\Surgical Subset of pbi 11 for Mel July_02_2013.jmp",
        Select Columns("No", "Readernumber", "BI", "Q1","Q2", "Q3a", "Q3b", "Q4",
"Q5", "Q6", "Q7", "Q8", "Q9", "Q10", "Q11", "Q12", "Q13", "Q14", "Q15", "Ctoverall" )
) <<   New Column( "Numeric_BI",
             Nominal,
             Format( "Best", 12 ),
             Formula( :BI == "Y" ));

// Form column vectors of signs (Q1-Ctoverall)
col=Column("Q1");
q1mtx = col<<GetAsMatrix;
dt=Current Data Table() ;
q2mtx=Column("Q2")<<GetAsMatrix;
dt=Current Data Table();

< Code Removed for other column vectors >
Dt=Current Data Table();
qoallmtx=Column("Ctoverall")<<GetAsMatrix;

//Send JMP vectors to R variables and compute Krippendorff alphas
R Send(q1mtx);
R Submit( "\[
# load lpSolve and irr packages
library (lpSolve)
library (irr)
```

```
Q1<-matrix(q1mtx,nrow=3)
Q1alpha <- kripp.alpha(Q1,"ordinal")
# Mel Alexander's adaptation of Mike Gruszczynski's kripp.boot.R code
# Reference: https://github.com/MikeGruz/kripp.boot/blob/master/R/kripp.boot.R
# bootalpha generates bootstrap resamples of krippendorff alphas several times
(default iter=100)

bootalpha <- function (x, raters='rows', probs=c(0.025,0.975), iter=100,
method=c('nominal','ordinal','interval','ratio'))
               {
                 alphas <- numeric(iter)

                 for( i in 1:iter)
                    {
                      alphas[i] <- kripp.alpha(x[,sample(ncol(x),
                                                    size=ncol(x),
                                                    replace=TRUE)],
                                                    method=method)$value
                     }

       kripp.ci <- quantile(alphas, probs=probs, na.rm=TRUE)
       boot.stats <- list(mean.alpha=mean(alphas, na.rm=TRUE),
                                       upper=kripp.ci[2],
                                       lower=kripp.ci[1],
                                       alphas=alphas,
                                       raters=nrow(x),
                                       iter=iter,
                                       probs=probs,
                                       sealpha = sqrt(var(alphas)/length(alphas)),
                                       size=ncol(x))
                    class(boot.stats) <- 'bootalpha'
                    return(boot.stats)
                 }
Q1boot <- bootalpha(Q1, iter=1000,  method='ordinal')
#Q1boot
# combine the original alpha (Q1alpha$value) with the bootstrapped alphas
(Q1boot$alphas)
Q1allalphas <- matrix(c(Q1alpha$value, Q1boot$alphas))
#Q1allalphas

#hist(allalphas)
# Get the standard error of the median
Q1.median <- sapply(Q1allalphas,median)
Q1semed <- sqrt(var(Q1.median))

# join the original alpha (Q1alpha$value) with the standard error of the median
(Q1semed),
# 95% ci alpha values (Q1boot$lower, Q1boot$upper)
Q195ci <- matrix(c(Q1alpha$value, Q1semed, Q1boot$lower, Q1boot$upper), nrow=1)
Q195ci
]\" );

R Send(q2mtx);
R Submit( "\[
#library (irr)
Q2<-matrix(q2mtx,nrow=3)
Q2alpha <- kripp.alpha(Q2,"ordinal")
Q2boot <- bootalpha(Q2, iter=1000,  method='ordinal')
#Q2boot
# combine the original alpha (Q2alpha$value) with the bootstrapped alphas
(Q2boot$alphas)
Q2allalphas <- matrix(c(Q2alpha$value, Q2boot$alphas))
#Qallalphas
```

```
#hist(Q2allalphas)
# Get the standard error of the median
Q2.median <- sapply(Q2allalphas,median)
Q2semed <- sqrt(var(Q2.median))
# join the original alpha (Q2alpha$value) with the standard error of the median
(Q2semed)
# 95% ci alpha values (Q2boot$lower, Q2boot$upper)
Q295ci <- matrix(c(Q2alpha$value, Q2semed, Q2boot$lower, Q2boot$upper), nrow=1)
Q295ci
]\" );
```

**< Code removed for other R Send and Submit commands >**

```
/*Collect R krippendorff bootstrapped alphas as JMP variables*/
Q1allalpha = R Get (Q1allalphas);
Q2allalpha = R Get (Q2allalphas);
Q3aallalpha = R Get (Q3aallalphas);
Q3ballalpha = R Get (Q3ballalphas);
Q4allalpha = R Get (Q4allalphas);
Q5allalpha = R Get (Q5allalphas);
Q6allalpha = R Get (Q6allalphas);
Q7allalpha = R Get (Q7allalphas);
Q8allalpha = R Get (Q8allalphas);
Q9allalpha = R Get (Q9allalphas);
Q10allalpha = R Get (Q10allalphas);
Q11allalpha = R Get (Q11allalphas);
Q12allalpha = R Get (Q12allalphas);
Q13allalpha = R Get (Q13allalphas);
Q14allalpha = R Get (Q14allalphas);
Q15allalpha = R Get (Q15allalphas);
Qoallallalpha = R Get (Qoallallalphas);
//Collect R krippendorff alphas as JMP variables
Q1alpha95 = R Get (Q195ci) ;
Q2alpha95 = R Get (Q295ci) ;
Q3aalpha95 = R Get (Q3a95ci) ;
Q3balpha95 = R Get (Q3b95ci) ;
Q4alpha95 = R Get (Q495ci) ;
Q5alpha95 = R Get (Q595ci) ;
Q6alpha95 = R Get (Q695ci) ;
Q7alpha95 = R Get (Q795ci) ;
Q8alpha95 = R Get (Q895ci) ;
Q9alpha95 = R Get (Q995ci) ;
Q10alpha95 = R Get (Q1095ci) ;
Q11alpha95 = R Get (Q1195ci) ;
Q12alpha95 = R Get (Q1295ci) ;
Q13alpha95 = R Get (Q1395ci) ;
Q14alpha95 = R Get (Q1495ci) ;
Q15alpha95 = R Get (Q1595ci) ;
Qoallalpha95 = R Get (Qoall95ci) ;

/*Send krippendorff bootstrapped alphas and cis from R to JMP matrices and data tables
*/
allalpha=Concat(Q1allalpha,Q2allalpha,Q3aallalpha,Q3ballalpha,
Q4allalpha,Q5allalpha,Q6allalpha,Q7allalpha,Q8allalpha,
Q9allalpha,Q10allalpha, Q11allalpha,Q12allalpha,Q13allalpha,
Q14allalpha,Q15allalpha,Qoallallalpha );
// Vertically Concatenate the alphas and 95%ci alpha matrices
alpha95ci = VConcat(Q1alpha95, Q2alpha95,Q3aalpha95, Q3balpha95, Q4alpha95, Q5alpha95,
Q6alpha95, Q7alpha95, Q8alpha95, Q9alpha95, Q10alpha95, Q11alpha95,
Q12alpha95, Q13alpha95, Q14alpha95, Q15alpha95, Qoallalpha95 );

dtallalpha = As Table(allalpha)<< Set Name("Krippendorff Bootstrapped Alphas");
```

```
Data Table ("Krippendorff Bootstrapped Alphas");
col = Col1 << Set Name("Q1");
col = Col2 << Set Name("Q2") ;
col = Col3 << Set Name("Q3a");
col = Col4 << Set Name("Q3b");
col = Col5 << Set Name("Q4");
col = Col6 << Set Name("Q5") ;
col = Col7 << Set Name("Q6");
col = Col8 << Set Name("Q7");
col = Col9 << Set Name("Q8");
col = Col10 << Set Name("Q9") ;
col = Col11 << Set Name("Q10");
col = Col12 << Set Name("Q11");
col = Col13 << Set Name("Q12");
col = Col14 << Set Name("Q13") ;
col = Col15 << Set Name("Q14");
col = Col16 << Set Name("Q15");
col = Col17 << Set Name("Ctoverall");

dt95cialpha = As Table(alpha95ci)<< Set Name("Krippendorff Alpha and 95%CI");
col = Col1 << Set Name("Alpha");
col = Col2 << Set Name("Standard Error of the median") ;
col = Col3 << Set Name("Lower 95% Alpha ");
col = Col4 << Set Name("Upper 95% Alpha ");

// Add Sign column to Data Table and move it to first
dt95cialpha = Data Table ("Krippendorff Alpha and 95%CI") << New Column("Sign",
Character, Values({"Q1", "Q2", "Q3a", "Q3b", "Q4",
       "Q5","Q6","Q7","Q8","Q9","Q10","Q11","Q12","Q13","Q14","Q15","Ctoverall"}))  ;

dt95cialpha = Data Table("Krippendorff Alpha and 95%CI")
              << Get Selected Columns ("Sign")
              << Move Selected Columns({"Sign"}, To First);

// Open data table of kappas
dtkappa = Open("G:\Nitima\Kappas inter-rater agreement.jmp");

R Term( );
```