5.4.A. Agreement on Elements of History in Chest Pain Patients

Cruz(1) et al studied the agreement between research assistants (RAs) and emergency physicians (MDs) on the presence or absence of certain symptom characteristics in patients presenting to the emergency department with chest pain. The following table shows responses to the question "Was the quality of the chest pain crushing?"

Note to non-clinicians: "crushing" chest pain suggests a possible myocardial infarction (heart attack), which is something emergency physicians always worry about in people with chest pain.

"Crushing" Pain?	MD Recorded Yes	MD Recorded No	
RA recorded Yes	117	6	123
RA recorded No	18	2	20
Total	135	8	143

a) What is the observed percent agreement?

(117 + 2)/143 = 83.2%

b) What is expected agreement based on the marginals?

Expected Cell Counts Based on Marginals

	MD Recorded Yes	MD Recorded No	
RA recorded Yes	116.1	6.9	**123**
RA recorded No	18.9	1.1	**20**
Total	**135**	**8**	143

Expected Agreement = (116.1 + 1.1)/143 = 82.0%

c) What is Kappa?

Kappa: (Actual – Expected)/(Perfect – Expected) = (83.2 – 82.0)/(100 – 82.0) = 0.07

d) What does it mean when we say the disagreements were "not balanced" in this study?

It means the disagreements tended to be in a particular direction, so numbers on one side of the diagonal were significantly higher than on the other side.

Of the 24 disagreements, there were 18 in which only the MD thought the pain was "crushing," and 6 in which only the RA did.

There is a simple statistical test for unbalanced disagreement. In this case, the test asks: given that there were 24 disagreements, if the probability of each type of disagreement were 0.5 (i.e., if the probability of being in the upper right and lower left cells of the 2×2 table were the same), what would be the chances of observing an 18:6 or greater imbalance (in either direction)? This is also the probability of obtaining ≥ 18 or ≤ 6 heads on 24 coin tosses.

For Stata users you can use the binomial probability test:

```
. bitesti 24 6 0.5
Pr(k <= 6 or k >= 18) = 0.022656  (two-sided test)
```

e) Why do you think there was imbalance in the direction observed in this study?

The direction of imbalance suggests that the MDs had a lower threshold for considering chest pain crushing, perhaps because their clinical experience made them more worried about missing a possible heart attack.

1. Cruz CO, Meshberg EB, Shofer FS, McCusker CM, Chang AM, Hollander JE. Interrater reliability and accuracy of clinicians and trained research assistants performing prospective data collection in emergency department patients with potential acute coronary syndrome. Ann Emerg Med. 2009;54(1):1-7.