5.2.A Abdominal Tenderness to Palpation

Yen et al[1] compared abdominal exam findings suggestive of appendicitis, such as tenderness to palpation, between pediatric emergency physicians and pediatric surgical residents.

Assume that the emergency physician and the surgeon each examine the same 10 patients for right lower quadrant tenderness with the following results:

Emergency Physician	Surgeon		
	Tender	**Not Tender**	**Total**
Tender	3	2	**5**
Not Tender	2	3	**5**
Total	**5**	**5**	**10**

a) Note that the observed agreement is 3 + 3 = 6/10 = 60%. Calculate kappa.

Now, assume that the emergency physician and the surgeon both find a higher prevalence of right lower quadrant tenderness, but still have 60% observed agreement:

Emergency Physician	Surgeon		
	Tender	**Not Tender**	**Total**
Tender	5	2	**7**
Not Tender	2	1	**3**
Total	**7**	**3**	**10**

b) Calculate kappa.

c) Compare the values of kappa for the tables in part (a) and part (b). The observed agreement was 60% in both cases, why is kappa different?

Now, assume that the surgeon has a higher threshold than the emergency physician for calling tenderness. This is a source of systematic disagreement.[1] Results follow:

Emergency Physician	Surgeon		
	Tender	**Not Tender**	**Total**
Tender	3	4	**7**
Not Tender	0	3	**3**
Total	**3**	**7**	**10**

[1] In fact, in the Yen et al. study, abdominal tenderness was reported much more frequently by the emergency department residents (73.5%) and attendings (72.1%) than by the surgical residents (43.5%).

d) Note that the observed agreement is still 6/10 or 60% and calculate kappa.

e) If you answered (a), (b) and (d) correctly, you found that the highest value of kappa occurred in (d) when disagreements were unbalanced. Why?

1. Yen K, Karpas A, Pinkerton HJ, Gorelick MH. Interexaminer reliability in physical examination of pediatric patients with abdominal pain. Arch Pediatr Adolesc Med. 2005;159(4):373-6.