The following data were obtained to estimate the number of births with spinal bifida. There were three sources of data. For example, there are 13 people who are in Source 1, but not in Sources 2 and 3.

		Source 1*				
		Yes		No		
		Source 2*		Source 2		
		Yes	No	Yes	No	x
Source 3*	Yes	0	3	1	8	
	No	8	13	3	0	$N_{obs} = 36$

Question 1

Fit a log linear model that assumes all the data sources are independent of each other. What is your estimate for the total number of births with spinal bifida?

Using the model assuming independence between sources, the total number of births with spina bifida is estimated to be 55 people.

Question 2

Now fit a model that assumes that there is dependence between sources 1 and 2 and sources 2 and 3. What is your estimate for the total number of births with spinal bifida based on this model?

Using the model assuming interaction between sources 1 and 2 and sources 2 and 3, the total number of births with spina bifida is estimated to be 71 people.

Question 3

Assuming the model from the previous question is the most correct model, what does this tell you about source dependence in this setting? Does it exist? If so, what kind is there?

Assuming that the model accounting for interaction between sources is the best model (which one can see from the output that it is not), the results would imply that the sources are not independent of each other.

The interaction term between source 1 and 2 is positive, which indicates positive source dependence. Infants with spina bifida who show up in source 1 are more likely to show up in source 2, resulting in underestimation.

The interaction term between source 2 and 3 is negative, which indicates negative source dependence. Sources 2 and 3 must have a high degree of mutual exclusive groups of infants with spina bifida, resulting in overestimation.

Also, the estimate using interaction terms is larger than the estimate assuming independence, which indicates positive source dependence and an underestimation of the true population size.

Question 4

What is capture recapture used for? What are alternative methods to capture recapture?

Capture recapture is a sampling approach to estimate an unknown population size by using two or more samples from that population. Alternative methods to estimating prevalence include LQAS, surveys, and multiplier methods.