1 Exercise 1.7.23, Simpson's paradox

A researcher wants to determine the relative efficacies of two drugs. The results (differentiated between men and women were as follows.

women	drug I	drug II
succes	200	10
failure	1800	190

men	drug I	drug II
succes	19	1000
failure	1	1000

We are now faced with the question which drug is better. Here are two possible answers:
-Drug I was given to 2020 people, of whom 219 were cured. Drug II was given to 2200 people, of whom 1010 were cured. Therefore, drug II is much better.
-Amongst women, the success rate of drug I is 1/10, and for drug II the success rate is 1/20. Amongst men, these rates are 19/20 and 1/2 respectively. In both cases, that is, for both men and women, drug I wins, and is therefore better.

1.1 First question

Which of the two answers do you believe?
I believe neither answer is truly correct, as they both are incomplete. The correct answer would follow a bit of both answers explained above. The probability that a randomly chosen person is cured with either drug is dependant on the drugs efficiency per gender, and the probability that the chosen person is a man or a woman:

P(success)=P(isMale)*MaleSuccessRatio + P(isFemale)*FemaleSuccessRatio

Let us calculate this for drug I and II:

drug I: P(success)=$\frac{20}{2020}\frac{19}{20} + \frac{2000}{2020}\frac{1}{10} = \frac{219}{2020}$

drug II: P(success)=$\frac{2000}{2200}\frac{1}{2} + \frac{200}{2200}\frac{1}{20} = \frac{1010}{2200}$

So now we clearly see that drug II is the better drug. This lets us suspect that this answer is equivalent to answer (1). We can see this is true because P(isMale)*MaleSuccesRatio equals the number of cured males divided by the total number of males. The same is true for P(isFemale)*FemaleSuccesRatio; this equals the number of cured females divided by the total number of females. Now, those two put together should then return the number of cured persons divided by the total number of people in the test. Or, in other words, the overall success rate of the drug. That is exactly what answer (1) was all about.

1.2 Second question

Can you explain the paradox?
We saw answer (1) was the correct one, and the reason that answer (2) also seemed correct was that that answer actually was part of a way to get to answer (1), as we saw in the above. The only thing missing in (2) was that one should also include the probabilities whether a randomly chosen person is female or male. And in this case, the probability that such a person is male was so small that it almost nullified the high MaleSuccessRatio for drug I, which is the main reason why answer (2) failed to choose for the right drug.