Blood test

Angie's friend was clearly not listening, and she seemed a thousand miles away. "What's wrong Melissa?"

So it seems she had a check-up a week ago and a routine blood tested indicated a rare form of leukemia. The doctor told her not to worry, that it was likely nothing 'cause these tests can be wrong, but that she'd have to go to the hospital on Friday for a series of more detailed tests.

"The doctor's probably right—it's likely nothing. Some bad reading. You *feel* okay, right?"
Yeah but this morning I went over to the medical library and read a some stuff about it, and they talk about that blood test right there, and it says it's 98% reliable!
"98% reliable? Jeez. What does that *mean* anyway"?
Well, that it's right 98% of the time!
Yeah?
What else!? What else could it mean!?
"Jeez. That's terrible. Hey I'm sure it'll be alright.
Baby, I'm so scared!
I'll come over tonight, okay?

Angie went home, but the more she tried to make sense of everything, the stranger it all seemed, and within an hour she was down at the library herself. Melissa had forgotten the name of the book, but she had given Angie the call number, so it was a cinch to find.

What Angie discovered is that the rare form that Melissa is talking about affects 1 in every 100,000 people. The disease produces a characteristic substance that shows up in a standard blood test, and—now here's the part she had to read real carefully—2% of the time, the blood test will give a false positive reading for that substance. *To be precise, of those people who don't have the disease, 2% will nevertheless test positive, and the remaining 98% will test negative*. That's where Melissa's 98% reliability came from. Among those who test positive, further tests of a more complicated nature can be done to detect the disease, and these are presumably what Melissa was having on Friday.

Something still bothered Angie—all those numbers—what did they really amount to? As soon as she got home, she took out a pad of paper and started playing with the figures. That led to even more confusion, so she started trying some simple diagrams. Well, things began to take shape, but the picture that emerged was a complete surprise—so much so, that it took her a full hour to be sure enough of her ground to pick up the phone.

"Melissa?"
Yeah? Oh hi Ange.
Pick any number between one and two thousand and one.
What?
Pick any number between one and two thousand and one.
Two thousand and one?
"Yep."
Isn't that the name of a movie?
"Yeah but that's something else. Look I'm serious. I've got a number written down here between one and two thousand and one. You've gotta try and guess what it is."
You gotta be kidding!
"Seriously—go ahead. Take a guess.
Ah, okay. Thirteen fifty six.
"Wrong! It's eight twenty seven."
Well that's dumb. How was I supposed to guess that?
You weren't. That's the whole point. You're not gonna believe this, but your chances of having leukemia are exactly the same as your chances were of guessing my number.
What? What are you talking about?
I went to the library. I read about your test, and I've done some figuring besides.
You mean that 98% accuracy stuff was wrong?
"Nope. That was correct."
I don't get it.
"I'll be right over."

On her way over to Melissa's, Angie tried to figure out the best way to present the situation so she'd see how it all worked. Melissa was a kind of a math phobe, so equations sure weren't going to fly. Gotta draw just the right picture.

Melissa had some tea ready, and as soon as it was poured, Angie took out her lined pad and began to make a table. Okay, think of a whole bunch of people who take the test, the test you took this morning. Some get positive and some get negative, right? *I guess.* Good. But also some have the disease and some don't. So that gives us 4 possible categories, according to whether you're positive or negative, and sick or well.

Now we're going to keep track of the numbers of each category with this table. For example, A is the number of people who are sick and test positive and C is the number of people who are sick and test negative. And then A+C is the total number of sick people. And the same for the well people. Okay?
Is this algebra?
Not really because in a moment I'm going to put real numbers in.

	Sick	**Well**	**Total**
POS	A	B	A+B
NEG	C	D	C+D
Total	A+C	B+D	

Here's what I found out at the library. First of all the disease is very rare. For every person who has the disease there are 100,000 who don't have it. So let's take a total of 100,001 people, one of whom is sick, and the others well. We'll put them in the totals row at the bottom.
So the Sick column has the sick person, the Well column has the well people, and the last column is the total.
Right!

	Sick	**Well**	**Total**
POS	A	B	A+B
NEG	C	D	C+D
Total	1	100,000	100,001

So how do we get the A, B, C and D?
You have to read more. You have to know the test statistics for each type of person, sick or well. For example, what you read in the library was that 2% of the well people will test positive (though they don't have the disease), while the remaining 98% will test negative. *That's where my 98% reliability came from.*
Right! Now how might we put that into the table?
Hmmm.
We're talking only about the 100,000 well people. How many of these test positive?
Well 2%. I guess that's 2000.
That's right. And the remaining 98,000 test negative.
So those are the B and the D.
Right! Hey I thought you didn't like math.
I don't!
Well you're sure getting it right now.

	Sick	**Well**	**Total**
POS	A	2000	A+B
NEG	C	98000	C+D
Total	1	100,000	100,001

Well what about A and C. If you're sick what happens?
It turns out that if you're sick the test will always catch you. Everyone who has the disease tests positive.
So A is 1 and C is zero.
Exactly, and the table's done.

	Sick	**Well**	**Total**
POS	1	2000	2001
NEG	0	98000	98000
Total	1	100,000	100,001

So I'm kind of imagining the people standing in those four middle squares. There one person on the first square, 2000 on the next, 98000 on the one below that, and no one on the last one.
Right.
So where am I?

Okay. The wonderful thing about this table is that it will answer all your questions. For example, let's try the question you are anxious about. If you test positive, what is the probability that you're sick?
Okay. How does that work?
Well, we take all the people who test positive. Where are they?
That's the first row.
Right. There are 2001 of them in our population. How many test positive?
Just the one?
That's right—just the one. So what are the chances of being sick if you test positive?
One in 2001?
You got it!
Wow.

	Sick	**Well**	**Total**
POS	1	2000	2001
NEG	0	98000	98000
Total	1	100,000	100,001

Problems

1. There is a rare genetic disorder which will cause illness to the carrier in middle age. There is a treatment which can be administered to children at age 10 which will greatly reduce the severity of the illness, but the treatment is difficult, so it is important to try to identify those children who carry the bad gene. Recently a test, which is 99% accurate, has been developed for the presence of the gene—more precisely, a child who carries the gene will always test positive, but a child who does not carry the gene will test negative only 99% of the time (and will test positive 1% of the time). The question is this: suppose a child takes the test and tests positive. What is the probability he carries the gene? By the way, it is estimated that 1 child in every 400,000 is a carrier.

2. A virus has been discovered which affects two percent of the individuals in a certain population. A test for the presence of the virus is 95% accurate, that is, someone who has the virus will always test positive, but someone who does not have the virus will test negative only 95% of the time (and will test positive 5% of the time). Suppose an individual takes the test and tests positive. What is the probability that he has the virus.

3. A cab was involved in a hit and run accident at night. There was a single witness who was able to identify only that the colour of the cab was blue. Now there are only two cab companies in town, one uses green cars and the other uses blue cars. Also, 85% of cab accidents in the city involve green cabs and 15% involve blue cabs. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness could correctly identify each one of the two colours 80% of the time and would fail 20% of the time. Given this data, what is the probability that the cab involved in the accident was blue?
Well, here's a simple argument. The witness is correct 80% of the time, and he said it was blue so the probability that the cab was blue is 80%! What else need be considered? Nothing, right?

4. A survey of all students in a local high school showed that "girls were twice as likely to be smokers as boys." But when the results of the survey were reported in the student paper, it was claimed that among the students at the school, "smokers were twice as likely to be girls than boys." Provide a convincing argument to show that these two statements are different.

5. On one cigarette package it is stated that "smokers are three times as likely to develop lung cancer as non-smokers." On another cigarette package it is stated that "lung cancer patients are three times as likely to be smokers as non-smokers." Find a simple condition (which can be easily stated and which anyone would understand) under which these two statements would be equivalent.

6. At a recent murder trial, the jury was given the following information. The victim was walking home late at night, was accosted at a dark corner a few blocks from his home, robbed, and after putting up a bit of a struggle, was murdered. Witness say that the assailant, of indeterminate build and wearing a mask, fled from the scene. A DNA analysis was done of skin cells found underneath the victim's fingernails. A few days later a genetic match was found to a man who lived within a 1 km radius of the crime and had a prison record for a violent crime. The prosecutor called a forensic scientist as a witness who testified that two random individuals chosen from the population would have such a DNA match with probability one in ten thousand. The prosecutor summed up the case with the statement, "the defendant is 99.99% guilty." What do you think of this conclusion?

6. Consider the following "advertisement" for bicycle helmets.

80 cyclists died this year in Canada in car/bicycle collisions.
Only 12 of these were wearing helmets.

Discuss this ad. What do you think its purpose is? What conclusion does it want you to draw? Does the data it presents justify this conclusion? Present what you think would be a better ad, manufacturing additional data if you need to.