Probability

PSLS chapters 9 and 10

Part II: issues and examples

Randomness and probability

Probability theory is a branch of math based on axioms and logic.

Probability models are useful tools to describe uncertainty and randomness. It is important to use a model that is *appropriate* for the random events described. Probability models may be based on:

Properties of the event studied
Observed frequencies
Subjective / personal probability

Mendelian genetic theory predicts 50% males (XY) and 50% females (XX) among newborns. There are actually more male than female live births in the U.S. (~ 51.2% males each year, based on birth certificates).

The 2011 National Youth Risk Behavior Survey provides insight on the physical activity of US high school students. Physical activity was defined as any activity that increases heart rate. Here is the probability model obtain by asking, “During the past 7 days, on how many days were you physically active for a total of at least 60 minutes per day?"

Days	0	1	2	3	4	5	6	7
Probability	0.15	0.08	0.10	0.11	0.10	0.12	0.07	0.27

What is the probability of a randomly selected US high school student was physically active **at least 1 day** in the past 7 days?

A) 0.08 **B)** 0.23 **C)** 0.77 **D)** 0.85 **E)** 0.92

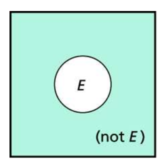
If X is the number of physically active days in a week, then the probability above is

A) $P(X < 1)$ **B)** $P(X \leq 1)$ **C)** $P(X = 1)$ **D)** $P(X \geq 1)$ **E)** $P(X > 1)$

Essential rules / equations

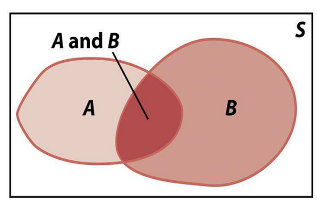
Probability of the **complement**:

$P(\text{not } E) = 1 - P(E)$



General addition rule for any two events A and B:

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$



Conditional probability of event B, given knowledge of event A: $P(B \mid A)$

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

Ways that events may relate

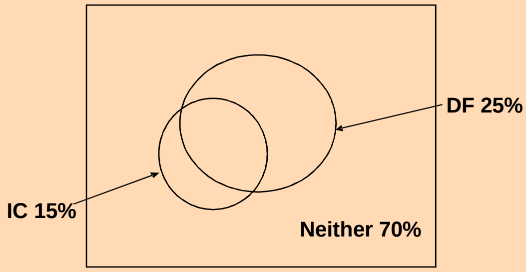
Two events are **disjoint (mutually exclusive)** if they never occur together on the same outcome; their joint probability is zero.

$P(A \text{ and } B) = 0 \iff$ A, B are mutually exclusive

Two events are **independent** if the knowledge that one event is true or has happened does not affect the probability of the other event.

$P(B \mid A) = P(B \mid \text{not } A) = P(B) \iff$ A, B are independent

In a large city, data on public swimming pools indicate that 15% of pools have inadequate levels of chlorine, 25% have dirty filters, and 70% have neither.



***P*(IC) = 0.15**
***P*(DF) = 0.25**
***P*(not IC *and* not DF) = 0.7**

	DF	not DF	Total
IC	10%	5%	**15%**
not IC	15%	70%	**85%**
Total	**25%**	**75%**	**100%**

If you find out that a public swimming pool in this city has dirty filters, then what is the chance that it has inadequate levels of chlorine too?

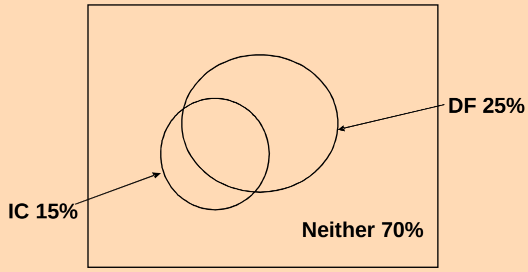
A) 0.10

B) 0.15

C) 0.25

D) 0.40

E) 0.80



	DF	not DF	Total
IC	10%	5%	15%
not IC	15%	70%	85%
Total	25%	75%	100%

A) DF, IC are independent.

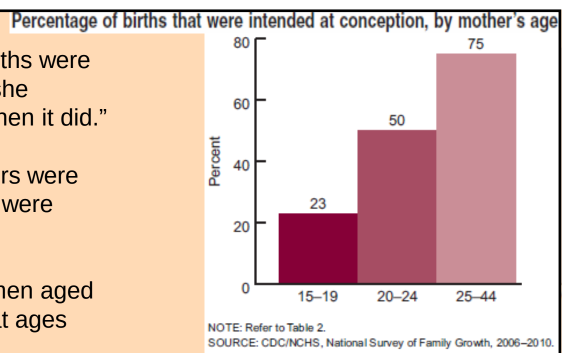
B) DF, IC are not independent.

C) DF, IC may or may not be independent.

“During 2006–2010, 63% of all births were intended by the mother—that is, she wanted the pregnancy to occur when it did.”

“Only 23% of births to teen mothers were intended in 2006–2010, and 77% were unintended.”

“Among births to young adult women aged 20–24, 50% were intended, and at ages 25–44, 75% were intended.”



Express the values cited here in terms of probability for the population of births during the 2006-2010 study period.

0.63 =

0.23 =

0.77 =

0.50 =

0.75 =

A) *P*(Intended)

B) *P*(Teen)

C) *P*(Intended and Teen)

D) *P*(Intended | Teen)

E) *P*(Teen | Intended)

Explaining conditional probabilities

***P*(*B* | *A*)**, the conditional probability of event *B,* given the knowledge of event *A,* can be computed from other probabilities representing "out of all *A* outcomes, how often *B* would also occur":

$$\boxed{P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}}$$

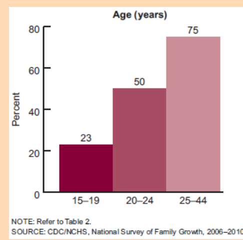
$P(\text{Intended} \mid \text{Teen})$

"Only 23% of births to teen mothers were intended."

$$= \frac{\#\,\text{births to teen moms and intended}}{\#\,\text{births to teen moms}}$$

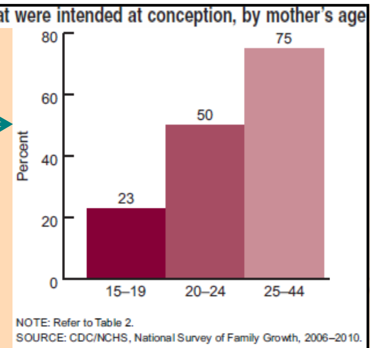
$$= \frac{\#\,\text{births to teen moms and intended} \,/\, \text{total}\,\#\,\text{births}}{\#\,\text{births to teen moms} \,/\, \text{total}\,\#\,\text{births}}$$

$$= \frac{P(\text{Teen and Intended})}{P(\text{Teen})}$$



"During 2006–2010, 63% of all births were intended by the mother… Only 23% of births to teen mothers were intended... Among births to young adult women aged 20–24, 50% were intended, and at ages 25–44, 75% were intended."

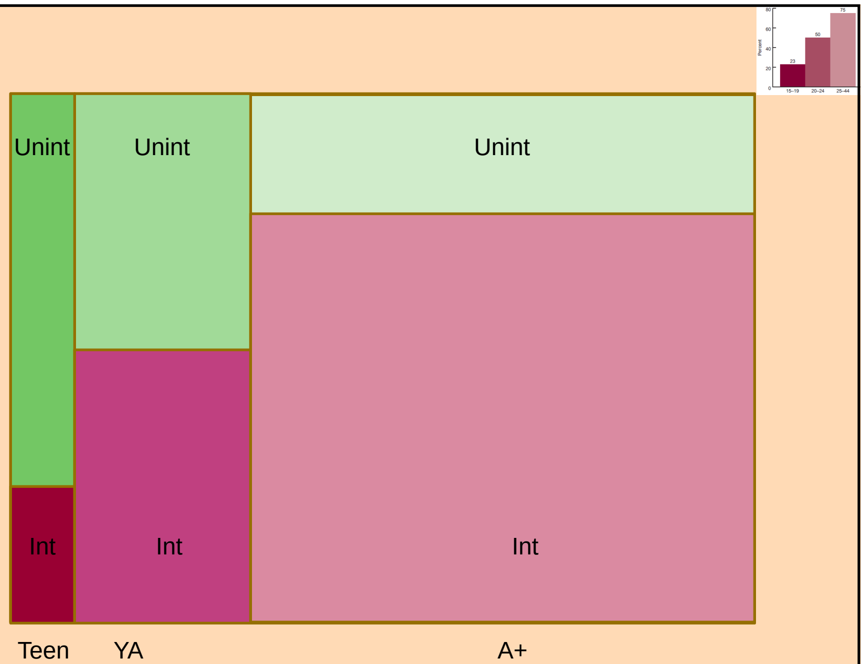
Percentage of births that were intended at conception, by mother's age



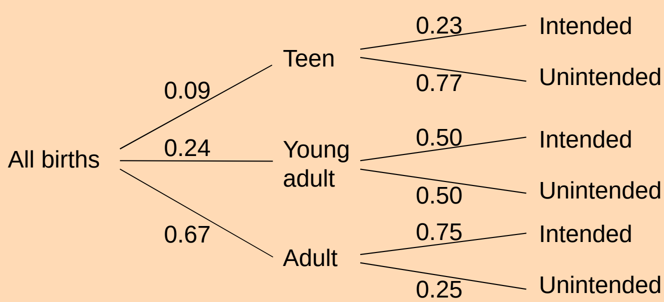
We also know that during those years in the USA, about 9% of all births were to teen mothers (ages 15–19), 24% were to young adult mothers (ages 20–24), and the remaining roughly 67% were to adult mothers (ages 25–44).

What is the probability that a birth in 2010 was to a teen mother and intended?

A) 0.02
B) 0.06
C) 0.09
D) 0.23
E) 0.63



What is the probability that a birth in 2010 was intended?



Intended and Unintended Births in the United States: 1982–2010

Confusion of the inverse: $P(B|A) \neq P(A|B)$!!!

Confusing $P(B|A)$ with $P(A|B)$ is a lot less likely to happen if you translate the probability notation into a full sentence.

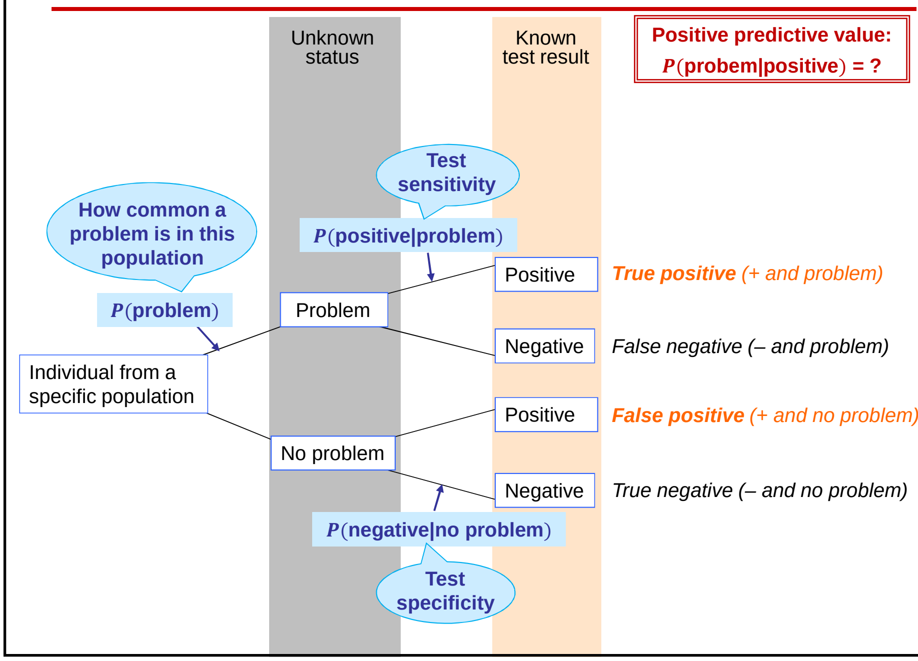
A woman in her 40s considers getting a mammogram to screen for breast cancer. She is told that if she has breast cancer, there is 80% chance that the mammogram will find it (positive result).

P(positive|disease): test sensitivity

But what really matters is this: If she receives a positive diagnosis, what is the chance that she actually has breast cancer?

P(disease|positive): positive predictive value, PPV

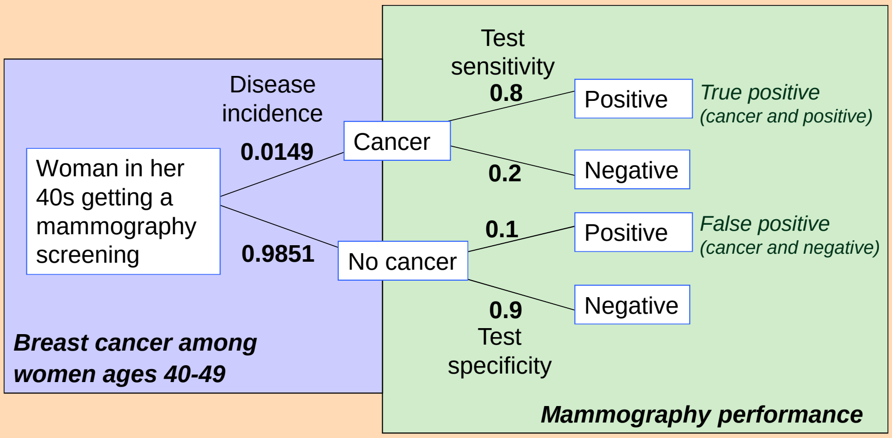
Diagnostic and screening tests

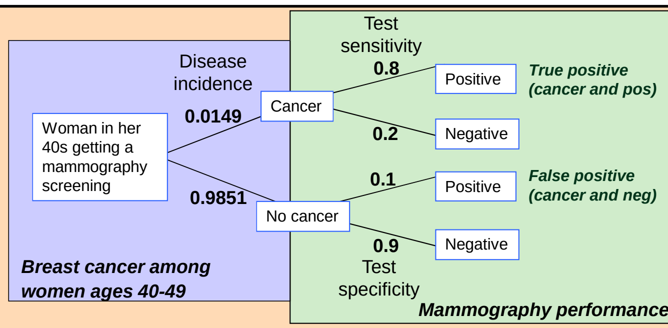


-Breast cancer occurs in roughly 1.49% of women in their 40s.

-If a woman has breast cancer, there is 80% chance that the mammogram will find it (be positive). This is the test sensitivity.

-If a woman does not have breast cancer, there is 90% chance that the mammogram will come up negative (positive result). This is the test specificity.





If a woman in her 40s screened for breast cancer receives a positive test result, what is the probability that she actually has breast cancer?

Positive predictive value: $P(\text{disease}|\text{positive})$

$$\frac{\#\text{ true positives}}{\#\text{ positives, true or false}}$$

$$P(\text{cancer} \mid \text{test+}) = \frac{P(\text{cancer and test+})}{P(\text{test+})} = \frac{P(\text{true +})}{P(\text{true +}) + P(\text{false +})}$$

$$= \frac{0.0149 * 0.8}{0.0149 * 0.8 + 0.9851 * 0.1} = \frac{0.01192}{0.11043} \approx 0.108\ (11\%)$$

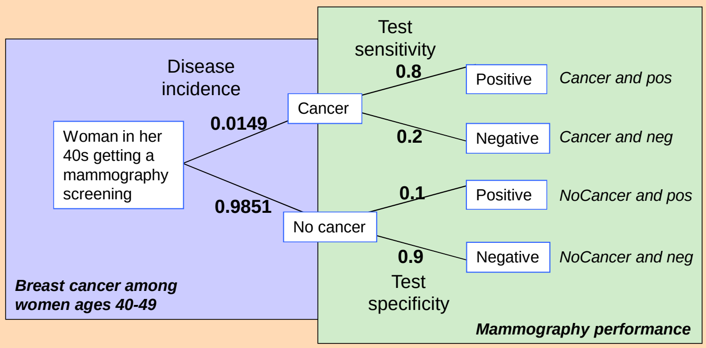
	Positive test	Negative test
Cancer 15	True + 15*.8 = 12	
No cancer 985	False + 985*.1 ≈ 99	

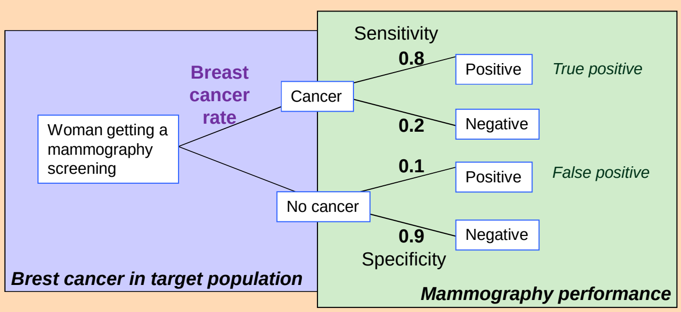
Consider 1,000 women

Studies show that thinking in terms of frequencies can help.

PPV ≈ 12 / [99+12]

≈ 0.108 (11%)





From National Cancer Institute:

Rate of breast cancer in women

ages 40 to 49:
1 out of 67 (**0.0149**)

$$PPV = \frac{0.0149 * 0.8}{0.0149 * 0.8 + 0.9851 * 0.1} \approx 11\%$$

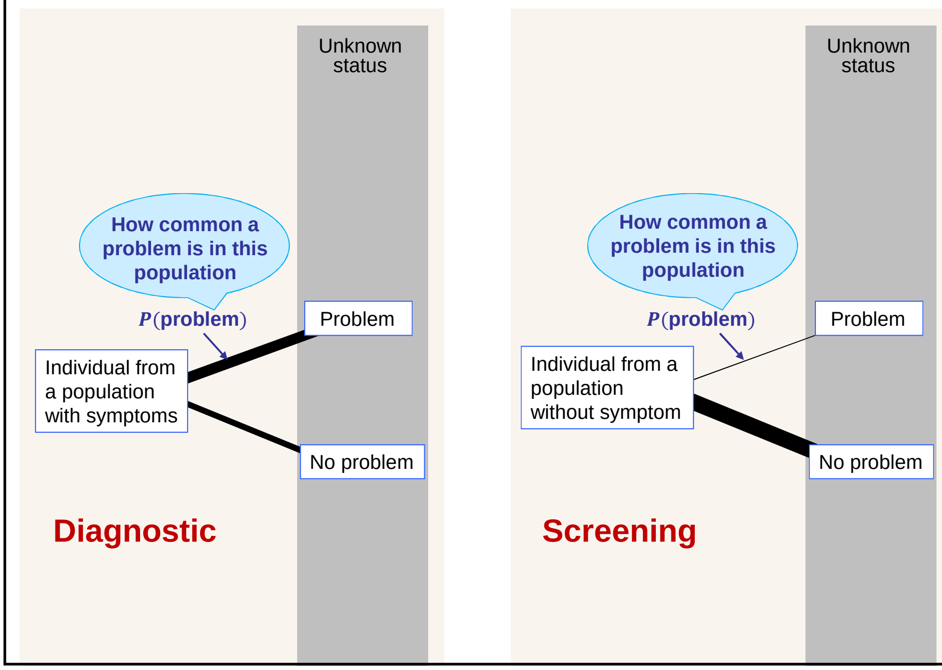
ages 50 to 59:
1 out of 35 (**0.0286**)

$$PPV = \frac{0.0286 * 0.8}{0.0286 * 0.8 + 0.9714 * 0.1} \approx 19\%$$

ages 60 to 69:
1 out of 28 (**0.0357**)

$$PPV = \frac{0.0357 * 0.8}{0.0357 * 0.8 + 0.9643 * 0.1} \approx 23\%$$

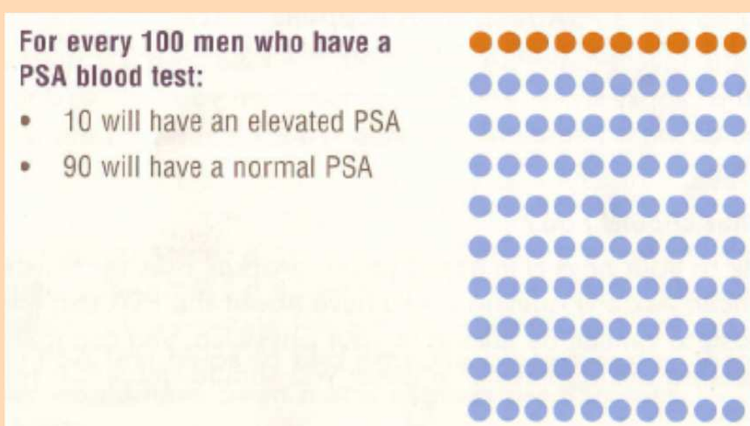
Diagnostic versus screening tests



The PSA test is used as a screening test for prostate cancer among mature men.

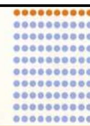
"Should you get a PSA test?"

clinical practice guidelines for members



What is the rate of prostate cancer ("PC") among mature men?
P(PC)= ?

For every 100 men who have a PSA blood test:
- 10 will have an elevated PSA
- 90 will have a normal PSA

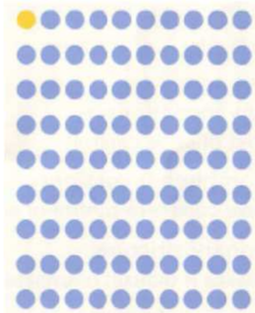


Of the 10 with elevated PSAs:
- 3 will have prostate cancer (true positive)
- 7 will not have prostate cancer (false positive)

How "sensitive" is the PSA test? *P*(positive | PC)= ?

Of the 90 with normal PSAs:
- 1 will have prostate cancer (false negative)
- 89 will not have prostate cancer (true negative)

How "specific" is the PSA test? *P*(negative | noPC)= ?



If a man gets an elevated PSA test (positive), what is the probability that he has prostate cancer?
PPV = *P*(PC | positive)= ?

A) 0-20%, **B)** 20-40%, **C)** 40-60%, **D)** 60-80%, **E)** 80-100%

KAISER PERMANENTE

"Should you get a PSA test?"

What is a PSA test?
The prostate-specific antigen (PSA) test is a blood test used to help identify men who **may** have prostate cancer. PSA is a substance from the prostate; the PSA blood test tells how much PSA is in your blood. All men have PSA in the blood, but a high PSA level may be a sign of prostate problems (such as an enlarged prostate). Older age may also cause a higher PSA level. Prostate cancer is another cause of a high PSA. Therefore, **a high PSA *may or may not* be a sign of cancer.**

Why should I get a PSA test?
Getting a PSA test routinely may help identify aggressive cancers. If you have cancer, a routine PSA test can help tell whether the cancer is growing fast. Catching a prostate cancer early may make it easier to treat. If the cancer spreads outside of the prostate there are fewer treatment options. **The PSA test helps to find prostate cancers earlier than they would have been found without the PSA test.**

Why shouldn't I get a PSA test?
Prostate cancer usually grows very slowly. Even if a man has prostate cancer, it may never affect his health. **Most men with prostate cancer die of other causes.** Many men never even know they have prostate cancer. They die of other causes before the prostate cancer grows enough to cause health problems. Treatment for prostate cancer may result in bad side effects. **A man who has prostate cancer may be worse off if he gets treated than if he did not get treatment.** The side effects from the treatment may cause harm when the cancer may not have. Some men think the treatment side effects are worse than having the cancer.

If I do get a PSA test, what happens next?
If you have a high PSA test result, the next step is a biopsy. If the biopsy shows you have cancer then you will need to decide what treatment is best for you.

What should I do?
Talk to your physician about your individual risks for prostate cancer. Ask any questions you have about the PSA test and prostate cancer. By talking to your physician, you can make a shared decision about whether you should get a PSA test.

Why can't my physician decide what's best for me?
Whether or not you should get a PSA test depends on how **you** balance the pros and cons. Would you feel better knowing you have prostate cancer? Would you feel better not knowing? Some things to think about: What happens if your PSA is elevated? What happens if you do have cancer? What difference will it make for you to know?

If you have prostate cancer, you must make a tough decision of whether or not to treat it. Most treatments can have bad side effects. It is not known whether treatment or no treatment results in better quality of life. It is not known whether treatment prolongs life. Here are some of the things that can happen after treatment:

Possible Outcomes of Treatment

	Radiation	Surgery
Improved survival	Unknown	Unknown
Death from treatment	2 in 1,000	1 in 200 younger men 1–3 in 100 older men
Impotence (difficulty with erection)	40 in 100	30*–90 in 100 *nerve-sparing surgery
Any incontinence (loss of urine control)	60 in 100	32 in 100
Complete incontinence (lose complete control of urine)	1 in 100	7 in 100
Urinary stricture (makes it difficult to urinate)	5 in 100	12–20 in 100
Any rectal injury (discomfort/trouble with bowel movements)	11 in 100	30 in 100

It is up to you and your physician to decide whether you should have a PSA test. The correct decision is an informed decision. Talk to your physician and decide what factors are important to you. Here are some pros and cons to consider:

Possible advantages to having a PSA test:
- It may give reassurance if it's normal.
- It can find many cancers earlier than is possible by a digital rectal exam.
- Treatment at early stages **may** help men live longer and avoid cancer complications.

Possible disadvantages to having a PSA test:
- It may miss cancer and give a false reassurance that there is no cancer.
- It may lead to a biopsy and anxiety when a man has no cancer.
- Treatment at early stages **may not** help men live longer, and treatment has risks of side effects.

clinical practice guidelines for members