Bayesian Statistics (2)

Classical vs Bayesian

Frequentist (Classical)

- Population parameters are fixed, but unknown constants.
- Probabilities are interpreted as long-term relative frequency.
- Statistical procedures are judged by how well they perform in the long-term over an infinite number of hypothetical repetitions.

Bayesian

- Population parameters are random variables.
- Probability statements about parameters must be interpreted as "degree of belief".
- We revise our “beliefs” about parameters (prior probabilities) after getting the observed data. This gives our posterior distribution.

Since the parameters are variables, we can make probability statements about them, posterior to the data. This contrasts with the conventional approach where inference probabilities are based on all possible data sets that could have occurred for the fixed parameter value.

Bayes Factor

Bayes Factor (B):

.

$$B = \frac{P(D/M_1)}{P(D/M_2)}$$

D: data observed

M_1*: model 1*

M_2*: model 2*

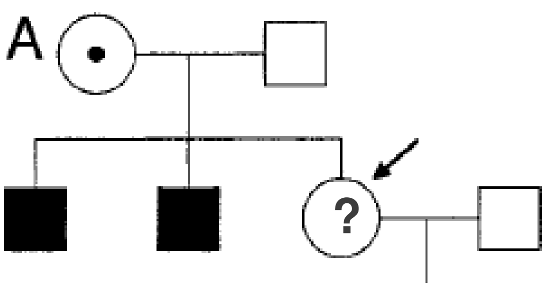
If *B > 1*, then the data (*D*) has made us believe that model 1 is more probable than model 2, and vice-versa.

Bayes in Genetics

Only Pedigree

The two brothers of the consultant have the Kennedy disease [X-linked spinal and bulbar muscular atrophy], which is caused by a mutation in the androgen receptor (AR) gene. Therefore, the mother is an obligate carrier of a disease allele.

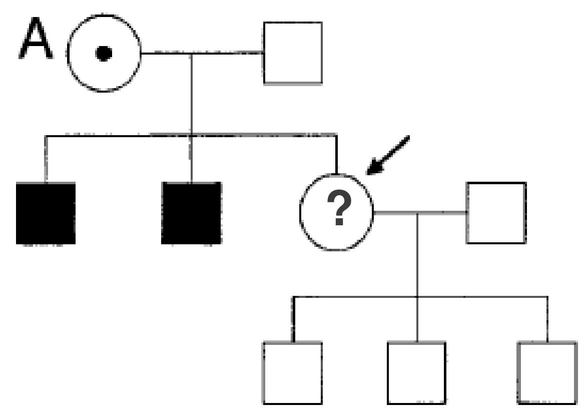
Image source:
Sumner et al. Neurology, 2002



A priori, the consultant has a 50% probability of being also a carrier.

Bayes in Genetics

But we have more information. The consultant has three children, and all of them are healthy. Intuitively, we know that this changes the probability that the consultant is a carrier.



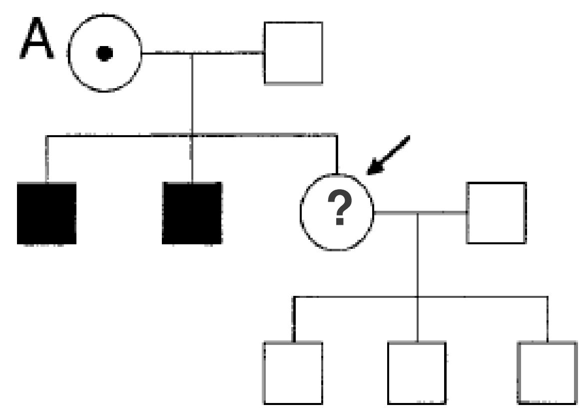
But which is the probability now?

Data source:
Ogino et al. JMD, 2004

Bayes in Genetics

The conditional probability of having 3 healthy sons, given that the consultant is non-carrier is 1 (they must be healthy). But if the consultant would be carrier, the conditional probability would be only 1/8 = P(healthy son | carrier mother)^3 = $(1/2)^3$.
The joint probability is the product of the rows above.

	Carrier	Non-carrier
Prior prob.	1/2	1/2
Cond. 3 sons	1/8	1
Joint prob.	1/16	1/2



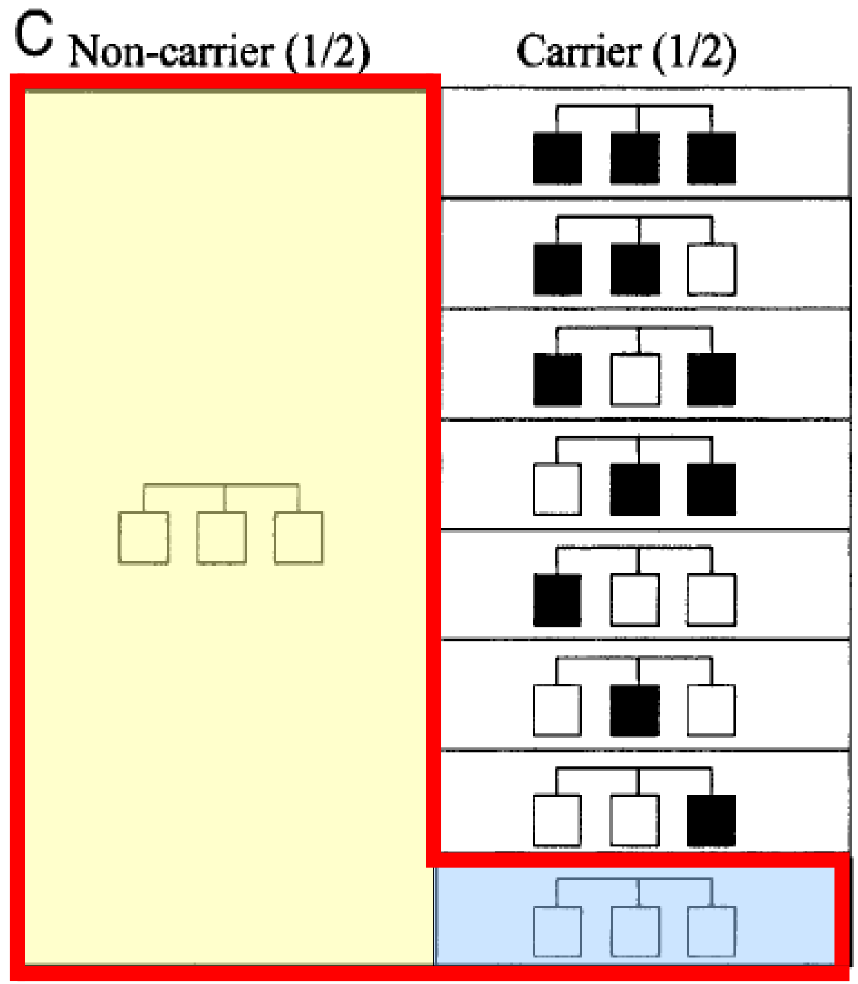
Bayes in Genetics

The posterior probability for each hypothesis is the joint probability of that hypothesis divided by the total joint probability = 1/16 + 1/2 = 9/16. Taking into account the prior family history and the information that she has three unaffected sons, the probability that the consultant is a carrier is 1/9 and the probability that she is a non-carrier is 8/9.

	Carrier	**Non-carrier**
Prior prob.	1/2	1/2
Cond. 3 sons	1/8	1
Joint prob.	1/16	1/2
Posterior prob.	(1/16) / (9/16) =**1/9**	(1/2) / (9/16) = **8/9**

Bayes in Genetics

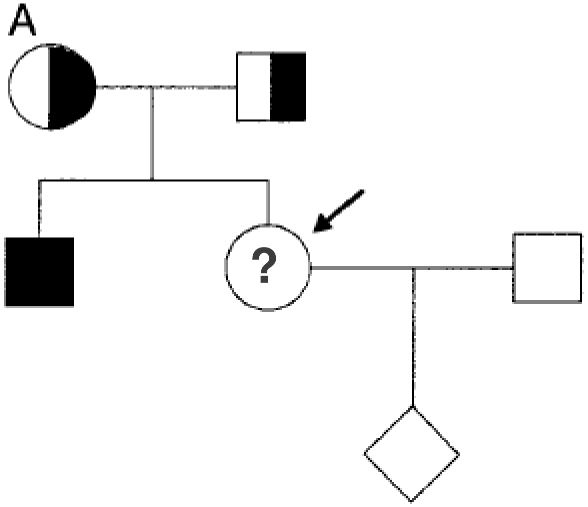
	Carrier	Non-carrier
Prior prob.	1/2	1/2
Cond. 3 sons	1/8	1
Joint prob.	1/16	1/2
Posterior prob.	**1/9**	**8/9**



Bayes in Genetics

Pedigree and genetic testing

CF is caused by mutations in the cystic fibrosis transmembrane conductance regulator gene (CFTR). It generates, among others, scar tissue in the lungs and cysts in the pancreas. The consultant is pregnant and wants to know if she is a carrier.



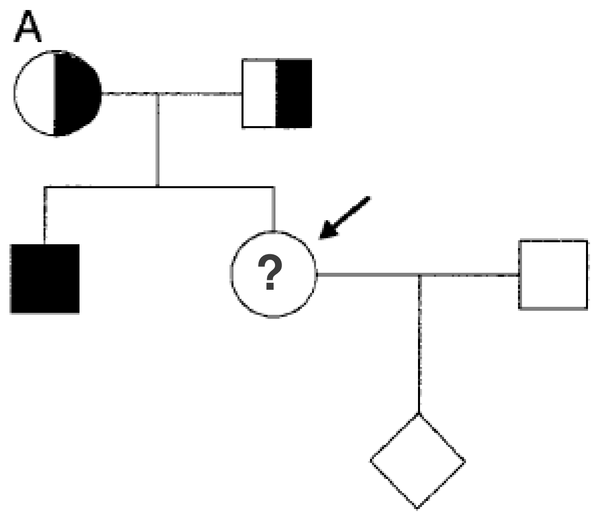
Data source:
Ogino et al. JMD, 2004

A priori, the consultant has a 75% probability of being also a carrier.

Bayes in Genetics

She undergoes a genetic testing. We typically analyze 25 potential mutations that cover 90% of all CF mutations. She is negative for all of them. What is her risk now?

~1%	~ 5%	~10%	~15%	~30%	~60%	~90%
☐	☐	☐	☐	☐	☐	☐



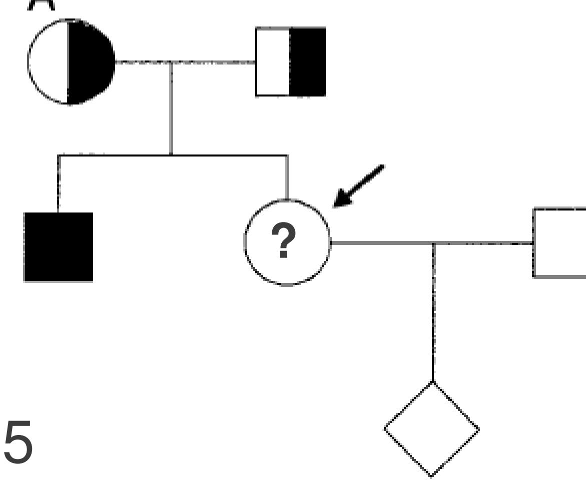
Bayes in Genetics

She undergoes a genetic testing. We typically analyze 25 potential mutations that cover 90% of all CF mutations. She is negative for all of them. What is her risk now?

We calculate conditional probability of having a negative genetic testing given that she is a carrier, and given that she is non-carrier.

	Carrier	**Non-carrier**
Prior prob.	2/3	1/3
Cond. neg test	1/10	1
Joint prob.	1/15	1/3

A

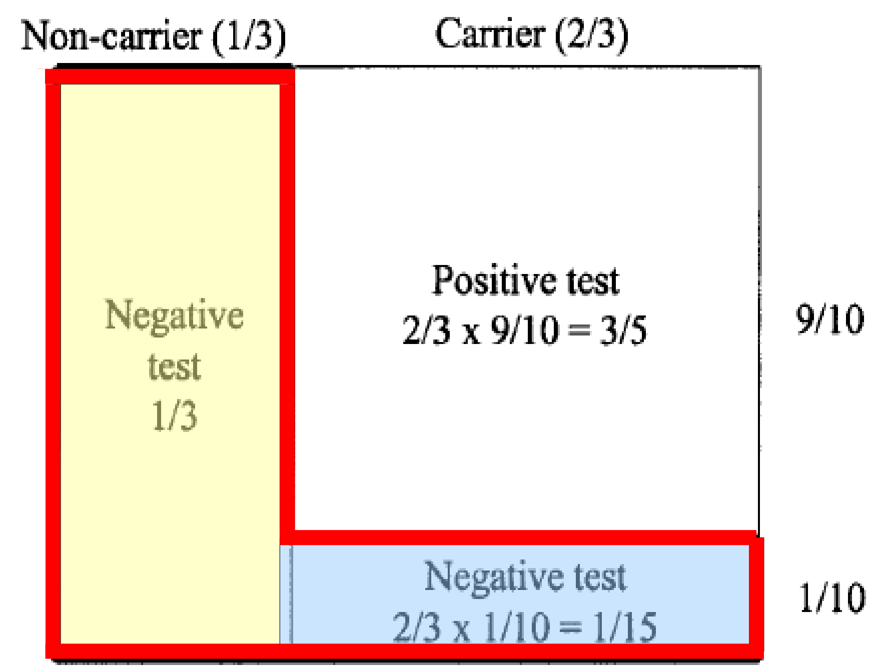


The total joint probability is 1/15 + 1/3 = 6/15

Bayes in Genetics

The posterior probability of being a carrier is:
(1/15) / (6/15) = 1/6 (~17%)

	Carrier	Non-carrier
Prior prob.	2/3	1/3
Cond. neg test	1/10*	1*
Joint prob.	1/15	1/3
Posterior prob.	**1/6**	**5/6**

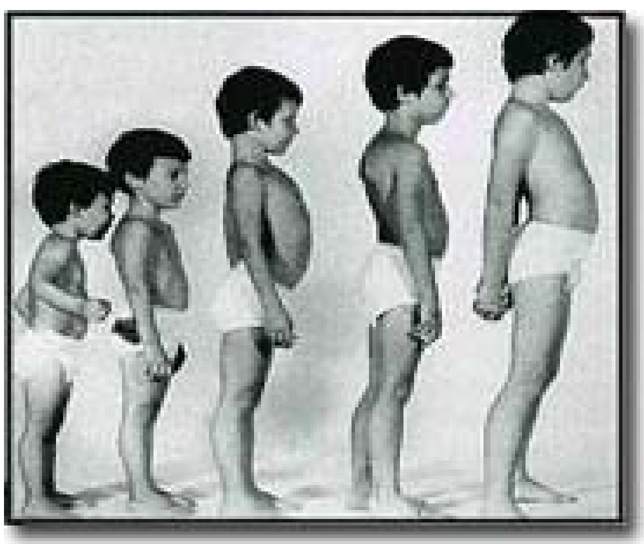


* to simplify we consider test specificity and sensitivity 100%

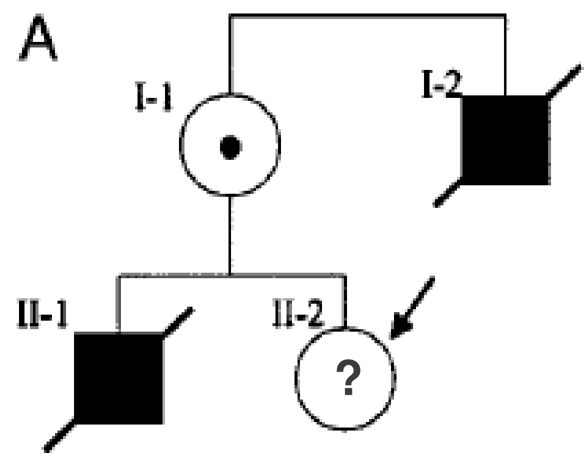
Bayes in Genetics

Pedigree and molecular testing

Duchenne's Muscular Distrophy (DMD) [X-linked recessive degenerative disease caused by mutations in the DMD gene]. Affects only boys. The consultant wants to know if she is a carrier.



Data source:
Ogino et al. JMD, 2004



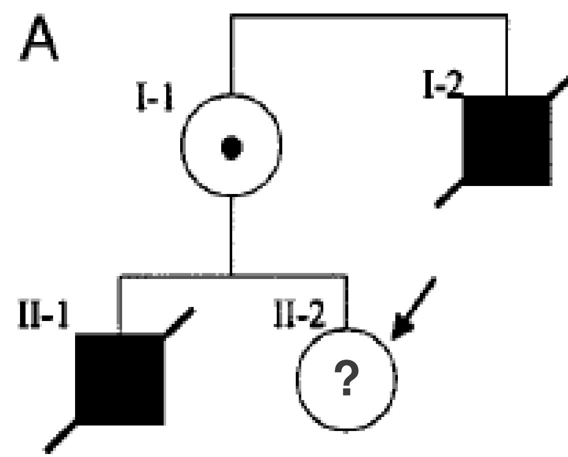
A priori, the consultant has a 50% probability of being also a carrier.

Bayes in Genetics

Pedigree and molecular testing

DMD carriers can be detected thanks to an elevated level of creatine phosphokinase (CPK) in serum, which occurs in 67% of DMD carrier women (sensitivity 67%). 5% of non-carrier women have an abnormal serum CPK (specificity 95%). The CPK test is negative. What is the risk now?

	Carrier	Non-carrier
Prior prob.	0.5	0.5
Cond. neg test	0.33	0.95
Joint prob.	0.165	0.475



Bayes in Genetics

The posterior probability of being a carrier if the test was negative is:
0.165 / (0.165 + 0.475) = 0.26

The posterior probability of being a carrier if the test was positive is:
0.335 / (0.335 + 0.025) = 0.93

	Carrier	**Non-carrier**
Prior prob.	0.5	0.5
Cond. neg test	0.33	0.95
Joint prob.	0.165	0.475
Posterior prob.	0.26	0.74

	Carrier	**Non-carrier**
Prior prob.	0.5	0.5
Cond. pos test	0.67	0.05
Joint prob.	0.335	0.025
Posterior prob.	0.93	0.07

Bayes in Genetics

Factors that affect Bayesian risk assessment:

- **Sensitivity and Specificity** of a molecular/genetic test

- **Population data**: frequency of a SNP (i.e. Hardy-Weinberg equilibrium) and/or of a disease in a particular population, frequency coverage of a genetic test for that population, ...

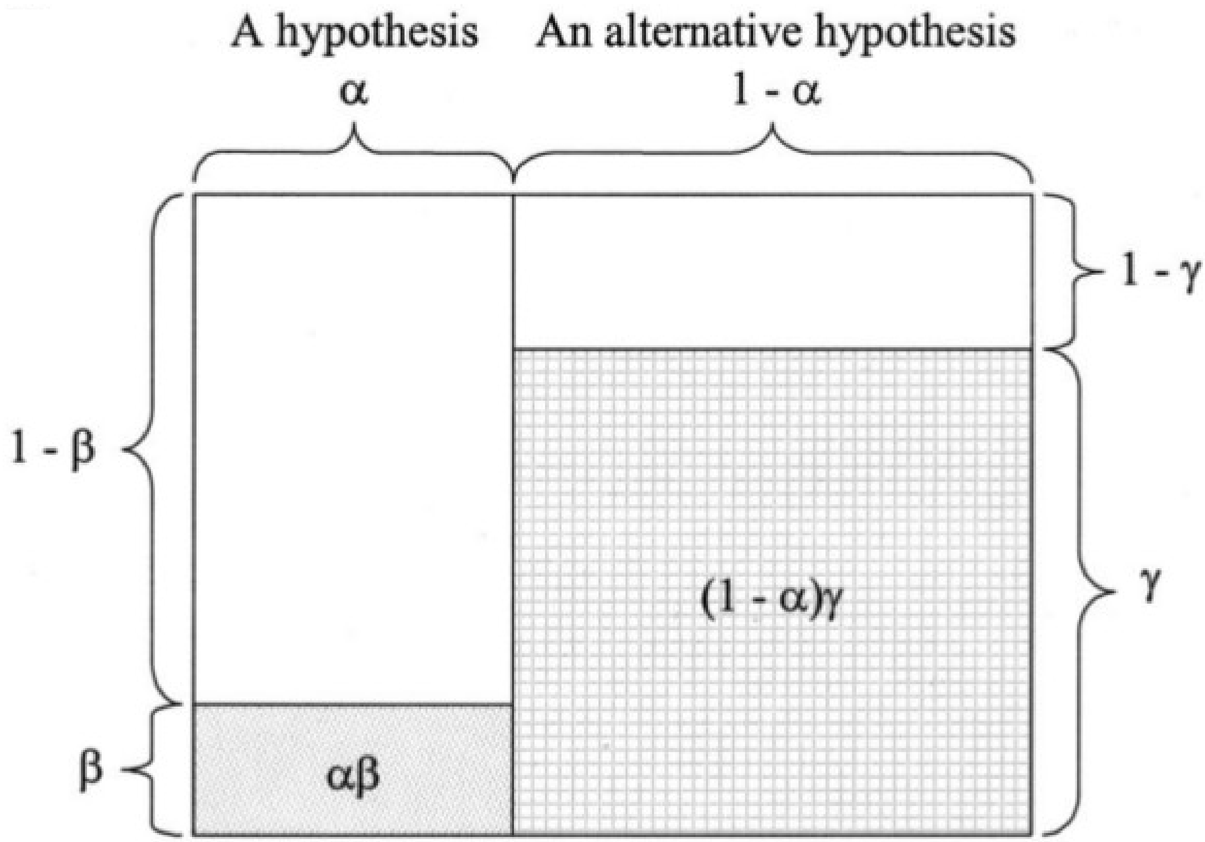
- **Family data**: disease status and/or genotyping of a family member

- **Multifactoriality**: polygenic diseases are much more difficult to analyze

- **Non-Mendelian inheritance**: e.g. mitochondrial inheritance, genetic imprinting, ...

Bayes in Genetics

Hypothesis	A particular hypothesis	An alternative hypothesis
Prior probability	α	$1 - \alpha$
Conditional probability (Probability of particular information)	β	γ
Joint probability	$\alpha\beta$	$(1 - \alpha)\gamma$
Posterior probability	$\alpha\beta/[\alpha\beta + (1 - \alpha)\gamma]$	$(1 - \alpha)\gamma/[\alpha\beta + (1 - \alpha)\gamma]$



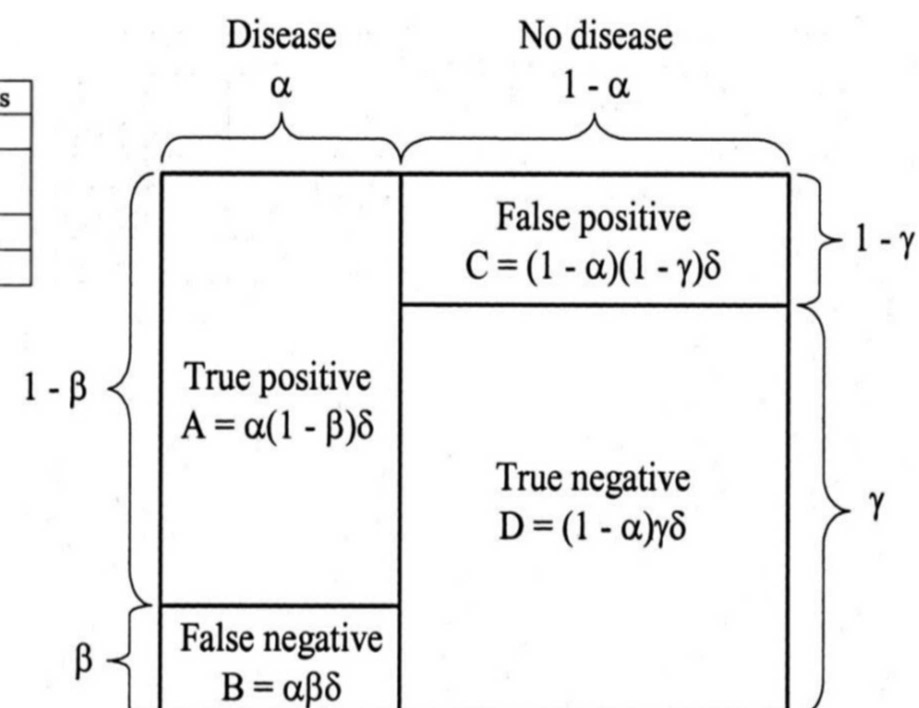
Bayes and contingency

	Disease	No disease	Total
A positive test result	A	C	A + C
A negative test result	B	D	B + D
Total	A + B	C + D	A + B + C + D

Hypothesis	Disease	No disease
Prior probability	(A + B) / (A + B + C + D)	(C + D) / (A + B + C + D)
Conditional probability of a negative result	B / (A + B)	D / (C + D)
Joint probability	B / (A + B + C + D)	D / (A + B + C + D)
Posterior probability	B / (B + D)	D / (B + D)

Hypothesis	A particular hypothesis	An alternative hypothesis
Prior probability	α	$1 - \alpha$
Conditional probability (as odds ratio)	1	γ/β
Joint probability	α	$(1 - \alpha)\gamma/\beta$
Posterior probability	$\alpha\beta/[\alpha\beta + (1 - \alpha)\gamma]$	$(1 - \alpha)\gamma/[\alpha\beta + (1 - \alpha)\gamma]$

$$\delta = A + B + C + D$$



Bayes in Genetics

Pedigree, genetic testing, test accuracy, population data

Further examples (based on the CF pedigree):

- Calculate the posterior probability of being a carrier for the consultant, knowing that the sensitivity of the genetic test was 99% and the specificity 95%.

- Calculate the posterior probability of being a carrier of the consultant's husband. He is an Ashkenazi Jew without CF family history (in that case his prior carrier risk is ~1/25) and tests negative for all 25 mutations (they cover 97% of all CF mutations in the Ashkenazi population).

- How does the posterior risk change for the mother and for the father, knowing that the newborn is diseased by CF?