LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Use of biomarkers in research

Isabel dos Santos Silva

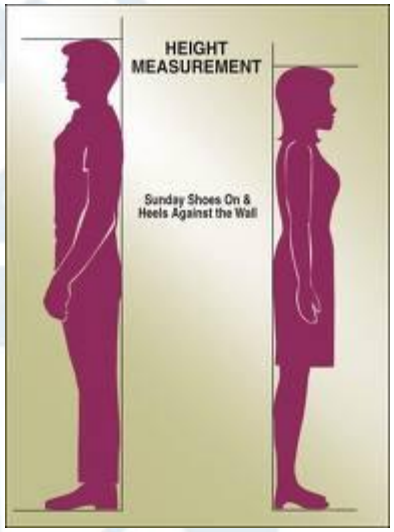
LSHTM, 4 April 2013

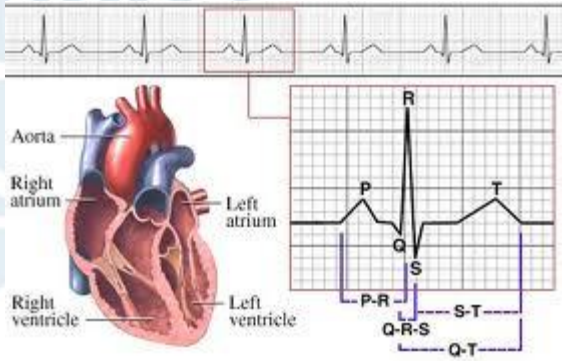
What is a biomarker?

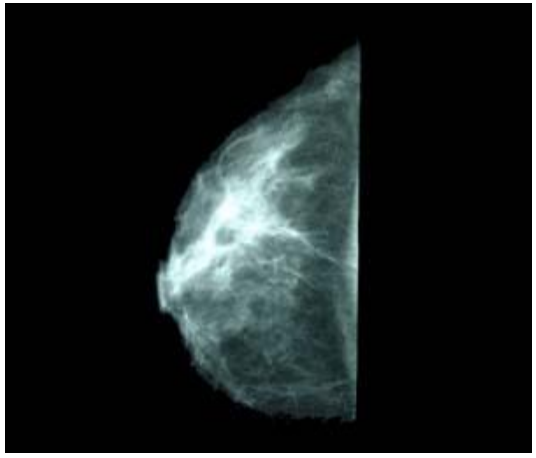
- A ***biological marker*** (abbreviated to ***biomarker***) is a substance or structure measured in body tissues, fluids or body products.
  - Endobiotic (if normally present in the body)
  - Xenobiotic (if foreign to the body)

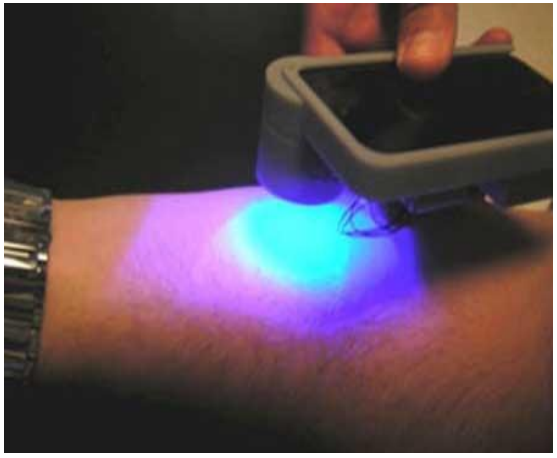
"Molecular epidemiology"

Examples of biomarkers









LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Examples of disease associated biomarkers

Disease	Biomarker
Coronary heart disease	LDL cholesterol
Coronary heart disease	Blood pressure
Diabetes	Fasting glucose, A1C, Insulin resistance
Colorectal cancer	Polyps
Prostate cancer	PSA
Osteoporosis	Bone mineral density
Dementia	Mild cognitive impairment

This is a rapidly advancing field…

Measuring the immeasurable…

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

The discovery of the double helix structure of the DNA, 1953

James Watson & Francis Crick

Nobel Prize, 1962

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

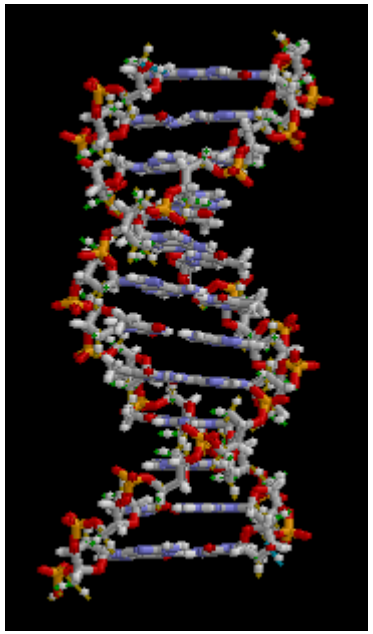
A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a



The Human Genome Project

In 2003 - just 50 years later – the sequence of the entire human genome was completed.

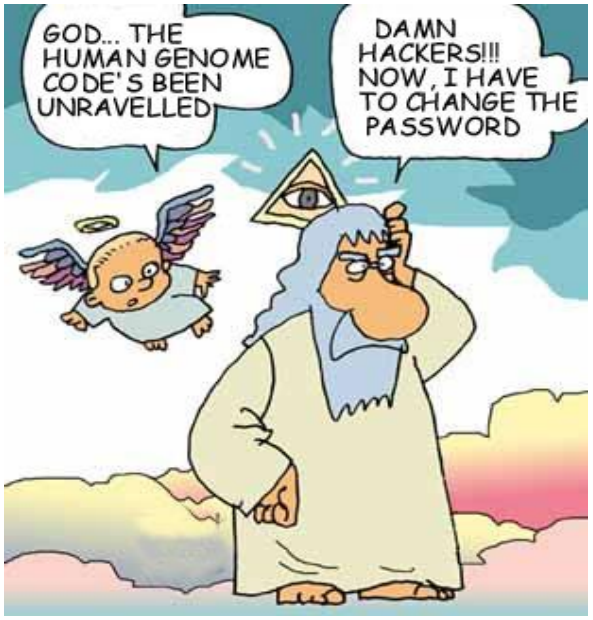
Nature – 21 October 2004
(431 :931-945)

articles

Finishing the euchromatic sequence of the human genome

International Human Genome Sequencing Consortium*

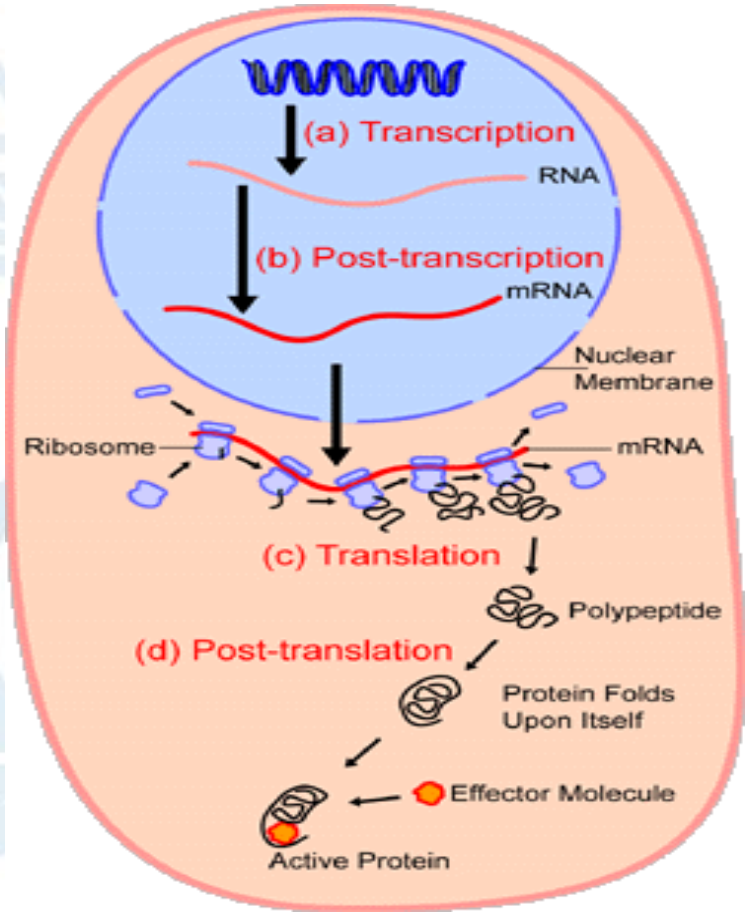
* A list of authors and their affiliations appears in the Supplementary Information




MY, THEY KNOW HOW TO FIND AN EXCUSE FOR A PARTY!
June 2000 Draft genome finished
Feb 2001 Draft genome published
April 2003 Genome 'finished'
Oct 2004 'Finished' genome published
May 2006 Last chromosome published...

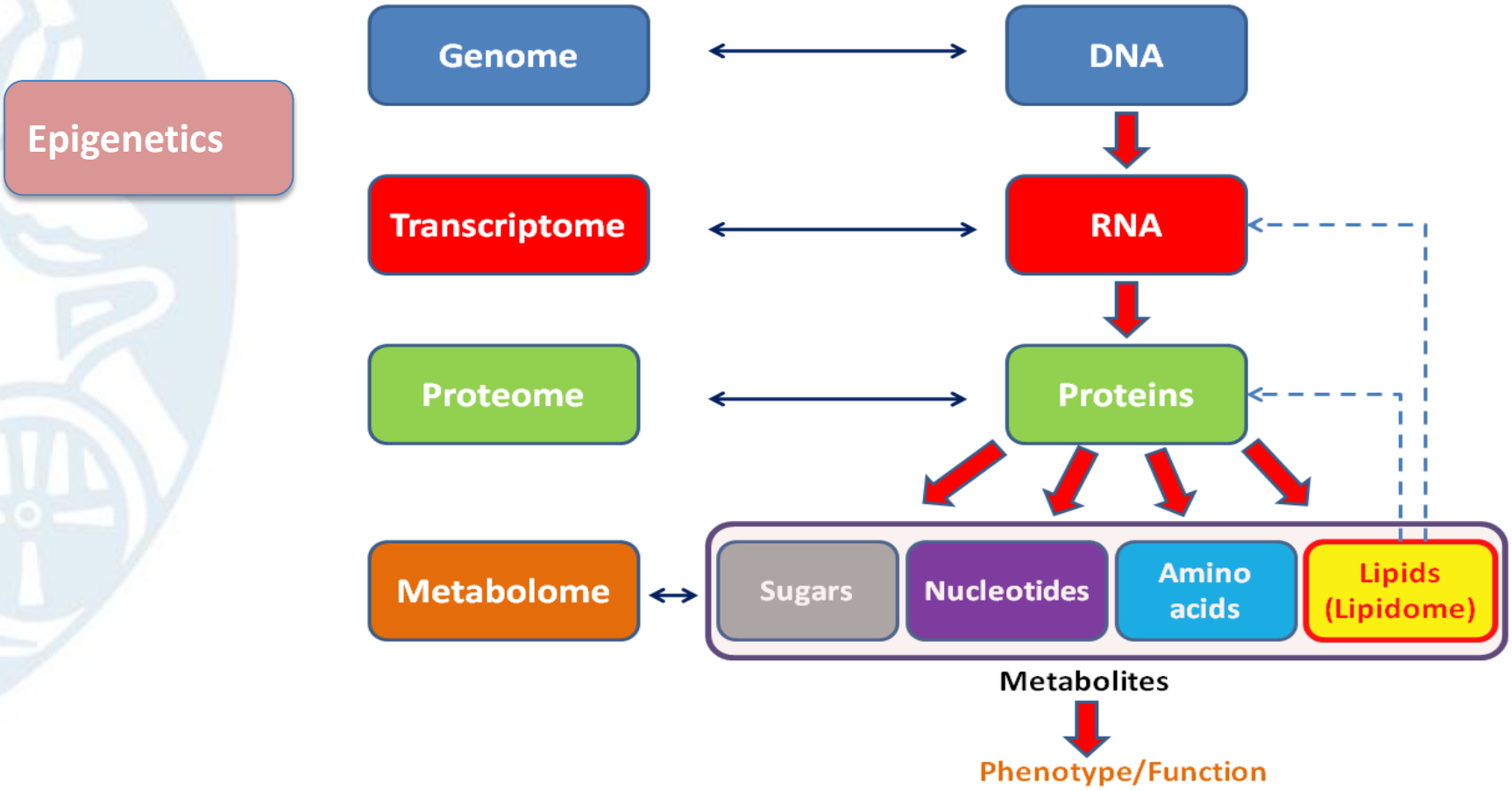
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

The shape of things to come…



From DNA to proteins…

The shape of things to come…



Published in final edited form as:
Adv Chronic Kidney Dis. 2010 November ; 17(6): 469–479. doi:10.1053/j.ackd.2010.09.002.

The Use of Targeted Biomarkers for Chronic Kidney Disease

Prasad Devarajan, M.D.
Nephrology and Hypertension, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, MLC 7022, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, Ph: 513-636-4531; FAX: 513-636-7407

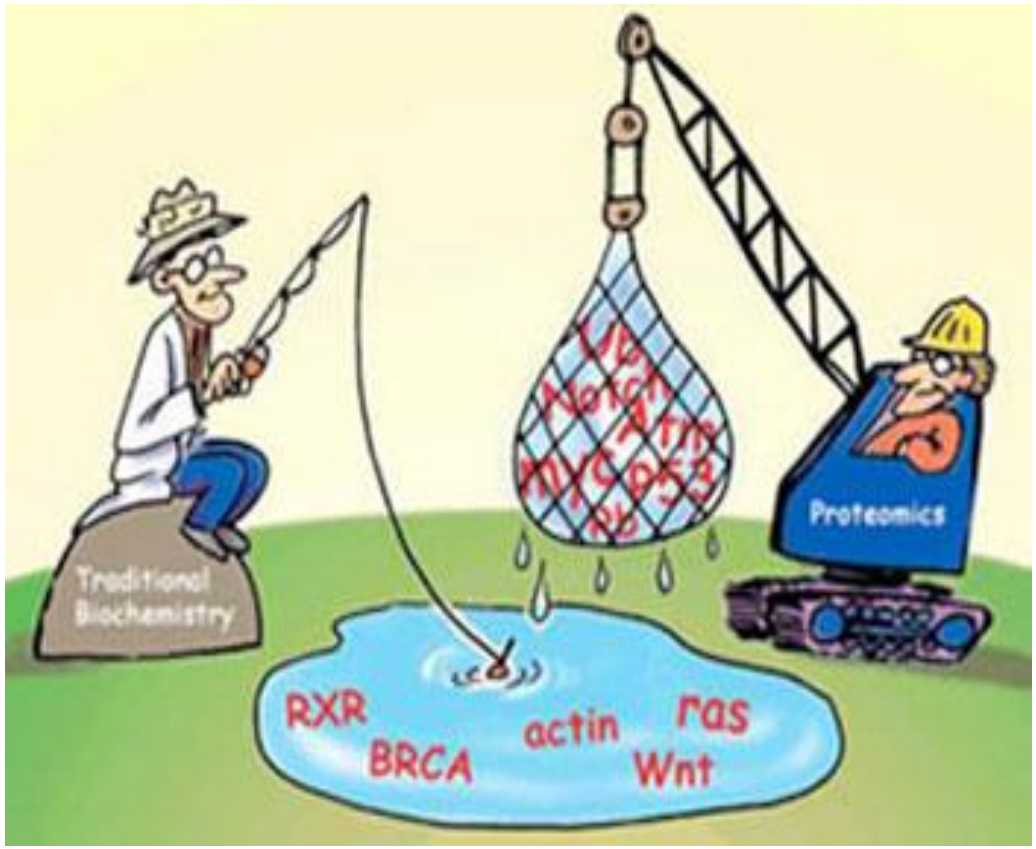
Prasad Devarajan: prasad.devarajan@cchmc.org

Abstract

There is a paucity of sensitive and specific biomarkers for the early prediction of chronic kidney disease (CKD) progression. The recent application of innovative technologies such as functional genomics, proteomics, and biofluid profiling has uncovered a number of new candidates that are emerging as predictive biomarkers of CKD. The most promising among these include urinary proteins such as neutrophil gelatinase-associated lipocalin (NGAL), kidney injury molecule-1 (KIM-1), and liver-type fatty acid binding protein (L-FABP). In addition, an improved

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

The shape of things to come…
very large-scale!



Belief that biomarkers have favourable characteristics relative to other methods:

- Objective
- Individualised (target to each individual at relevant times
- More specific and sensitive

Biomarkers in Social Sciences

Aims:

(i) To validate or complement other measurements (e.g. self-reports)

(ii) To analyse interactions between social and biological factors

(iii) To examine pathways and establish causation

Traditional approach

Poverty and ageing

Distal exposure

(e.g. SE deprivation of area of residence, educational level, income)

??????

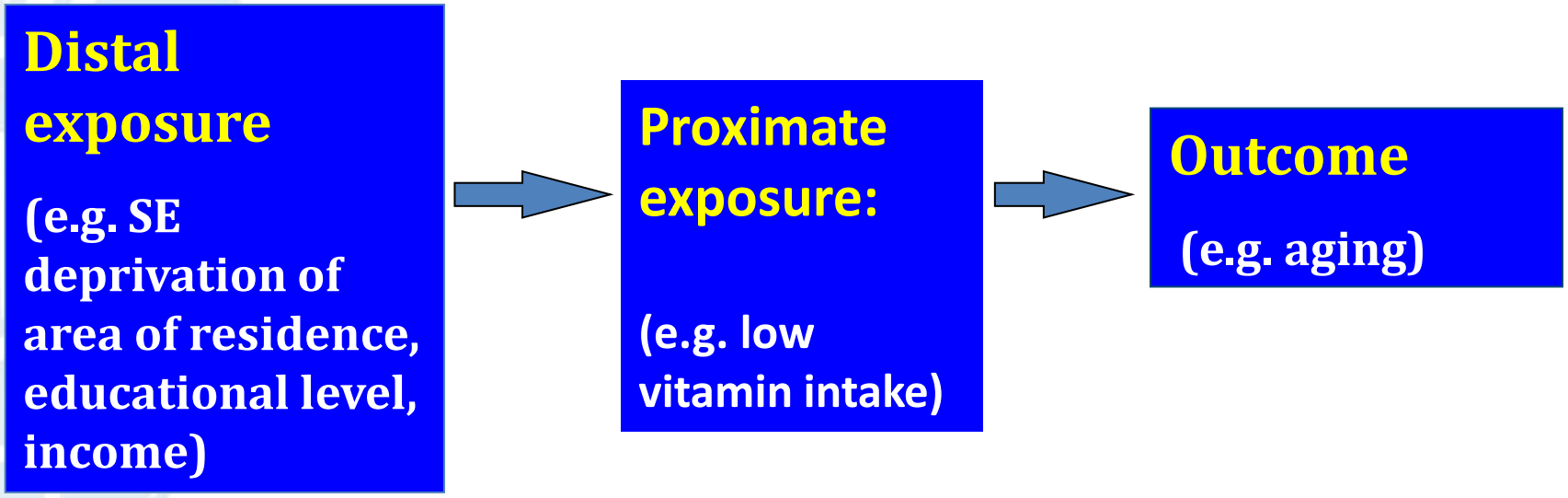
Outcome

(e.g. aging)

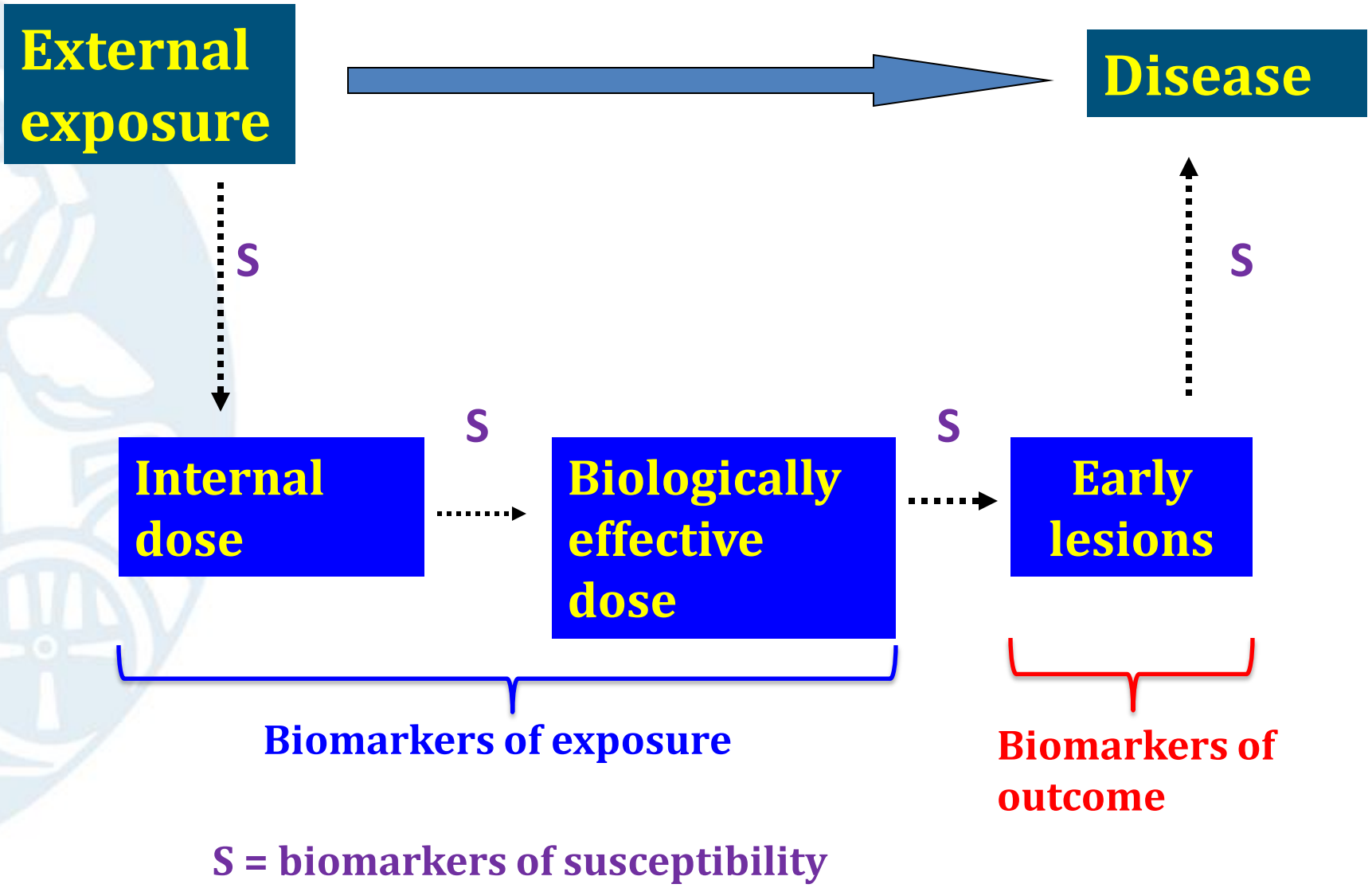
"Black box approach"

"Biomarker" approach

Poverty and ageing

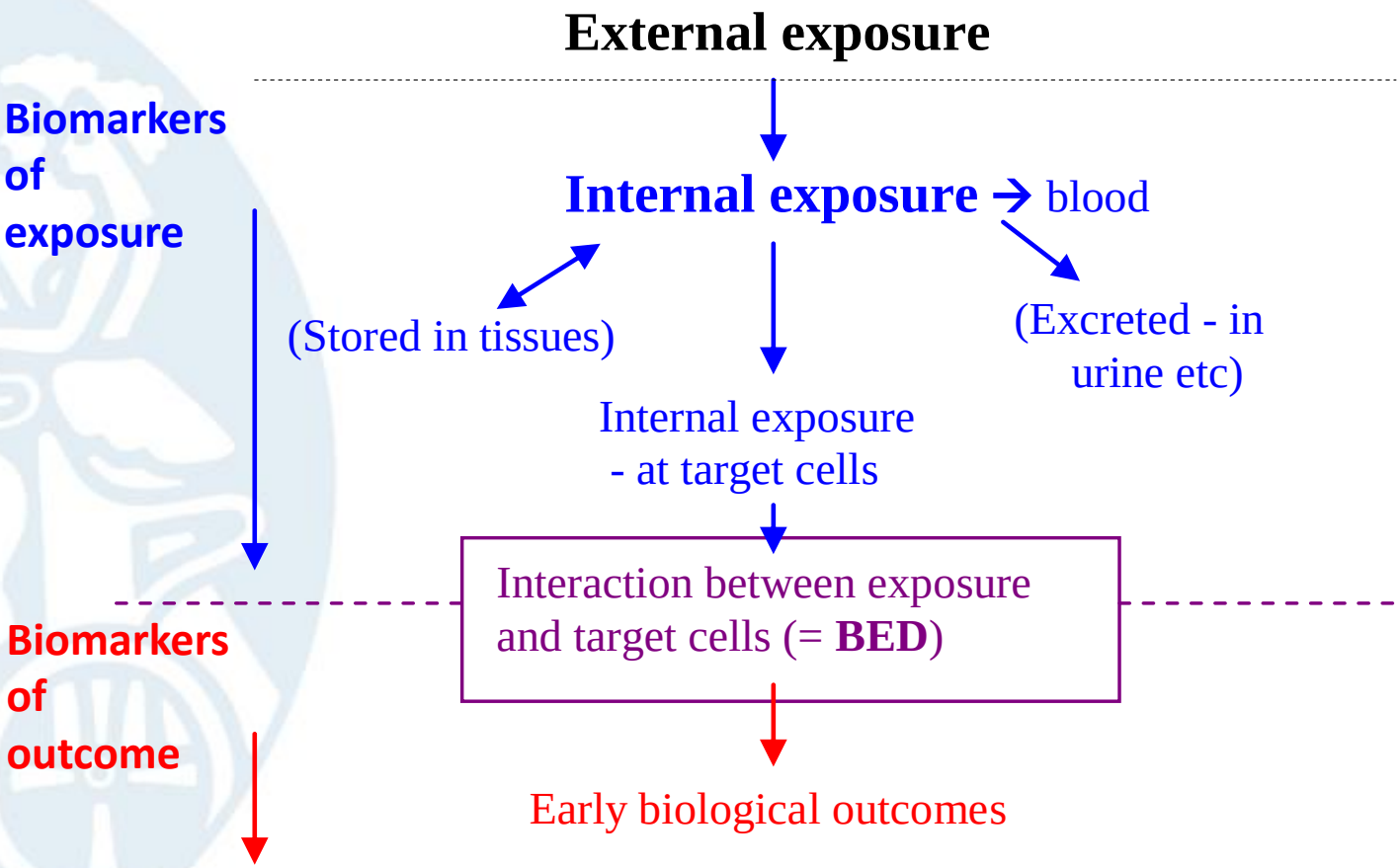


Examination of of pathways


External exposure
Disease
S
S
S
S
Internal dose
Biologically effective dose
Early lesions
Biomarkers of exposure
Biomarkers of outcome
S = biomarkers of susceptibility

Uses of biomarkers in research:

- Improve measurement of exposure
- Measure differences between individuals' susceptibility to the effects of exposure or progression to disease
- Provide measures of early outcome


External exposure
Biomarkers of exposure
Internal exposure → blood
(Stored in tissues)
(Excreted - in urine etc)
Internal exposure - at target cells
Interaction between exposure and target cells (= BED)
Biomarkers of outcome
Early biological outcomes

Biomarkers of internal dose

Capture external exposure, absorption and sometimes metabolism by the body:

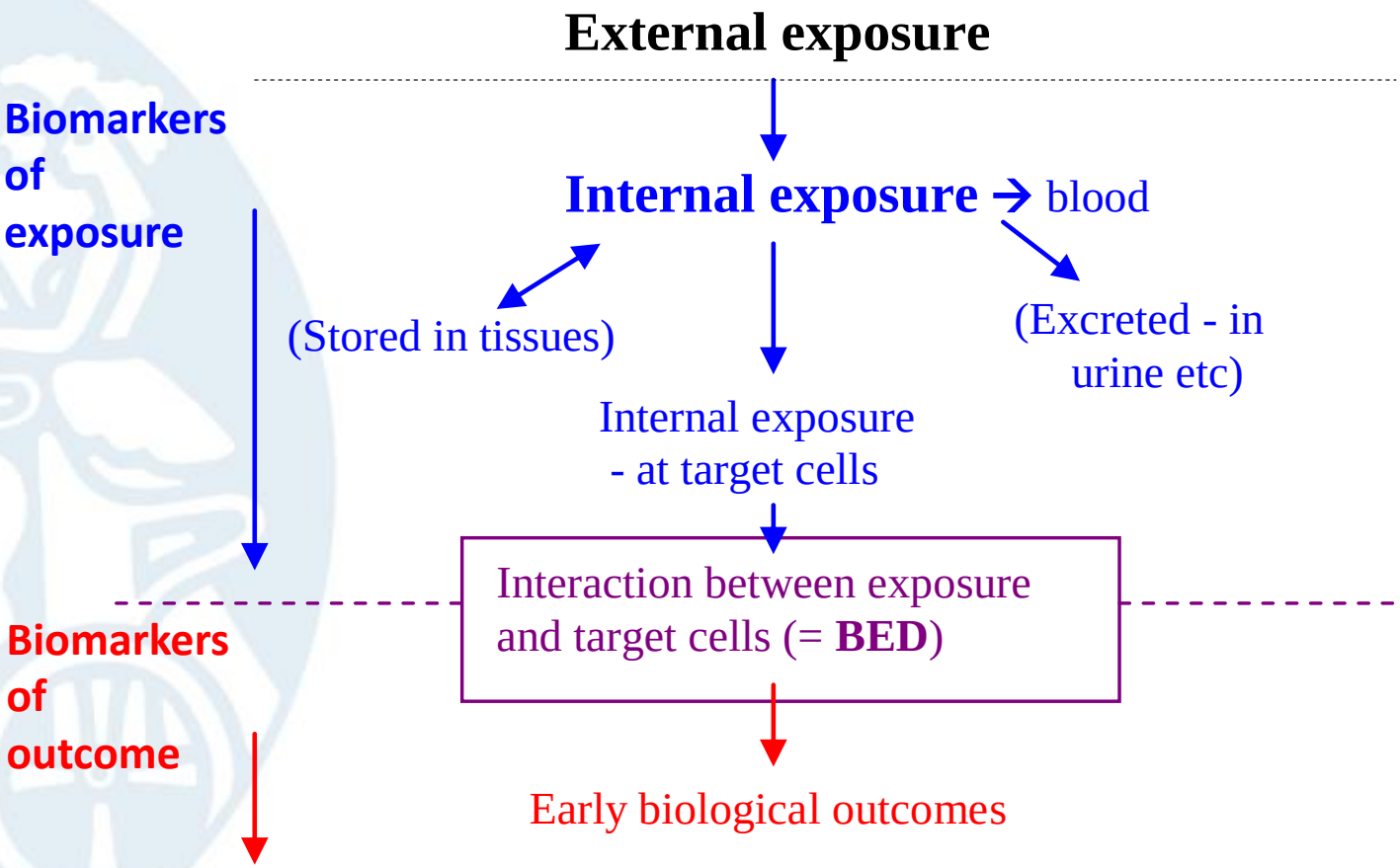
- ***Chemical exposures:*** *e.g.* levels of Vitamin C in the blood or arsenic in the hair

- ***Metabolites of exposures:*** *e.g.* levels of cotinine (a metabolite of tobacco) in the urine

Biomarkers of internal dose

Validation studies:

- E.g. levels of cotinine in the urine measured in a sub-sample to validate questionnaire data on current smoking habits


External exposure
Biomarkers of exposure
Internal exposure → blood
(Stored in tissues)
(Excreted - in urine etc)
Internal exposure - at target cells
Interaction between exposure and target cells (= BED)
Biomarkers of outcome
Early biological outcomes

Biomarkers of biologically effective dose

The interaction between the exposure and target cells represents the bridging stage between exposure and outcome – the **biologically effective dose (BED)**

Carcinogen adducts

Chemical carcinogens are **genotoxic** – they can damage DNA:

- Carcinogens are metabolised inside the body, where they are either *inactivated* or *activated*
- The activated carcinogen can bond with DNA or proteins in the target cells and in other tissues
- Carcinogen-DNA or carcinogen-protein products are called **adducts**

Chemical carcinogens

- The adduct distorts the DNA structure. If it is not repaired before the cell divides, part of the genetic code may be misread, resulting in a genetic mutation.

- In most cells the affected DNA bases do not have a coding function - the mutations are 'silent'.

- Some mutations may lead to changes in structure or function. These changes may result in cell death, or in development of cancers.

Carcinogenic adducts

Carcinogen adduct levels measure the ***biologically effective dose*** **(BED)** of the carcinogen.

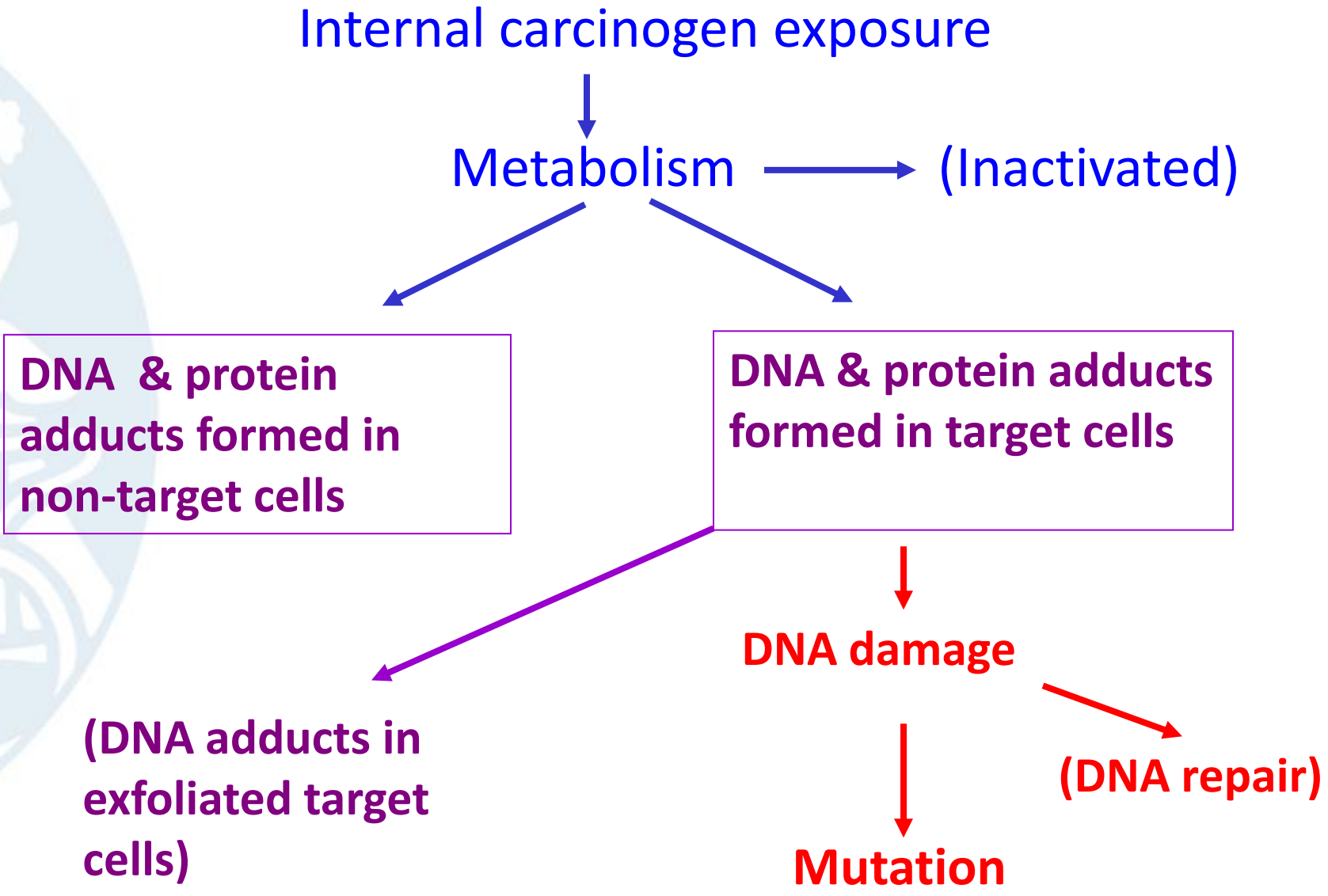
The BED represents the net effect of external exposure to the carcinogen as well as multiple internal metabolic steps:

- absorption by the body
- metabolic activation
- DNA repair rates
- target cell turnover

Carcinogenic adducts

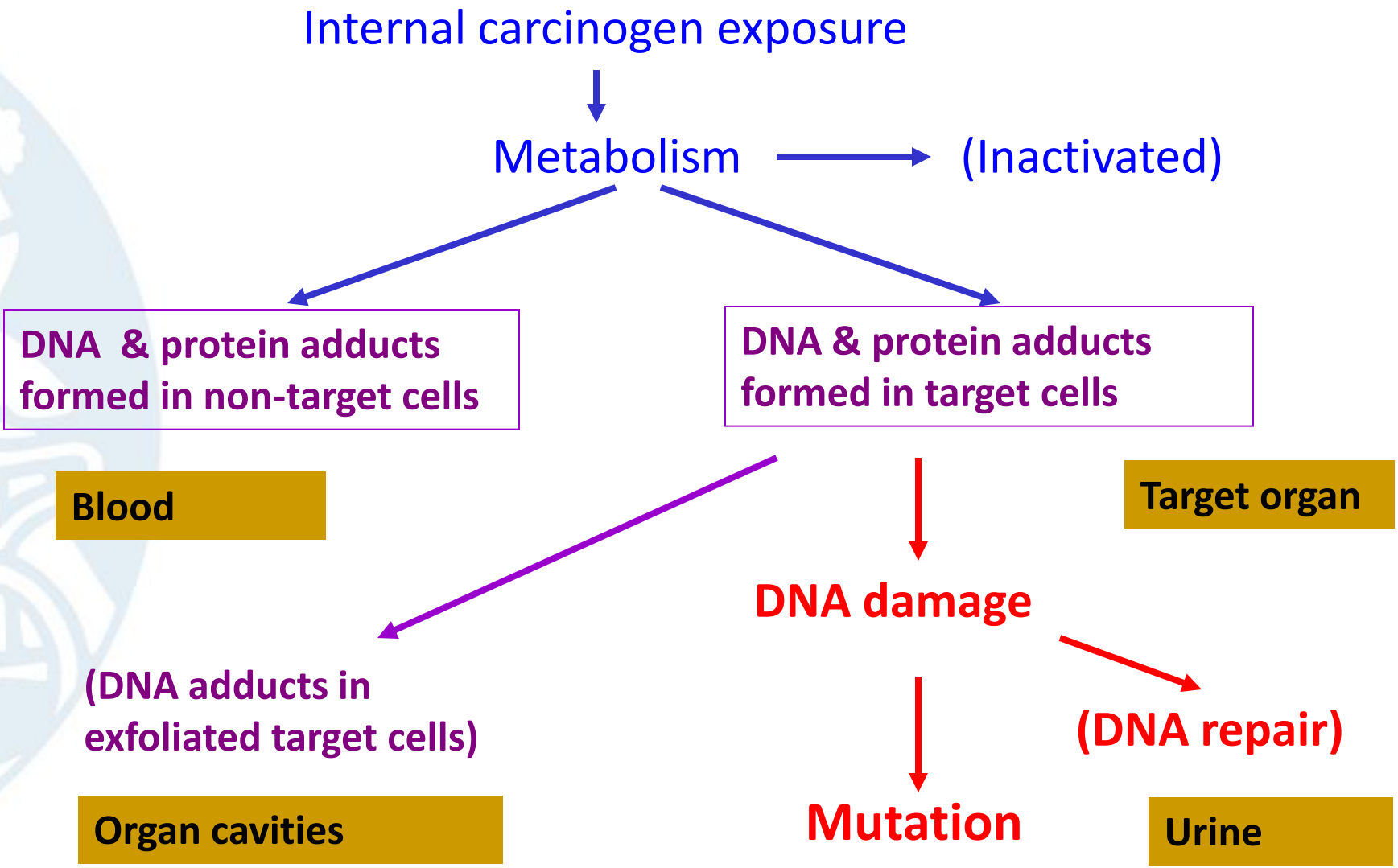
If a group of individuals receive exactly the same external exposure would you expect them to have exactly the same biologically effective dose of that exposure?

e.g. if they all smoke 20 of the same brand of cigarettes a day for the same number of years would they all have the same levels of nicotine-DNA adducts?


Internal carcinogen exposure
Metabolism
(Inactivated)
DNA & protein adducts formed in non-target cells
DNA & protein adducts formed in target cells
DNA damage
(DNA adducts in exfoliated target cells)
(DNA repair)
Mutation

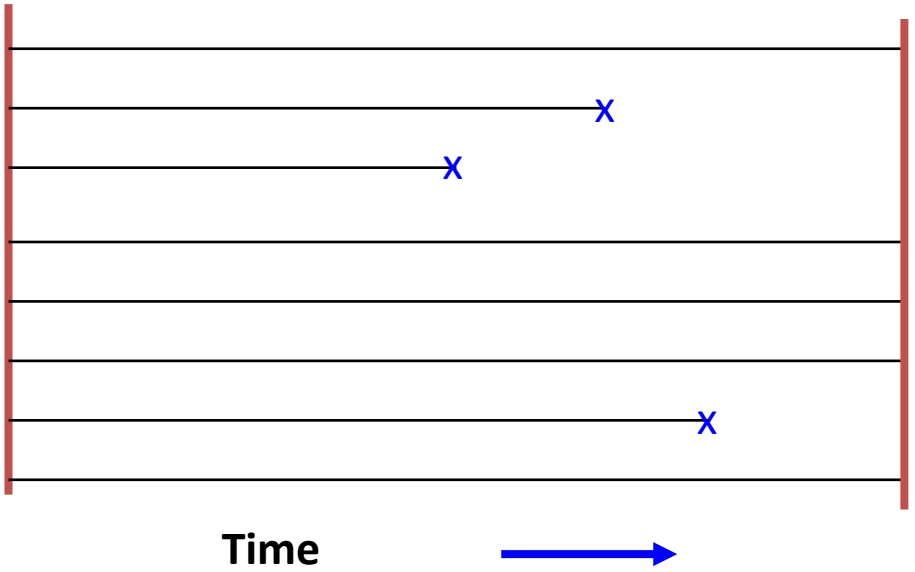
Where to measure carcinogen adduct levels

- Target issues (ideal but usually invasive)
- Exfoliated cells
- Surrogate tissues or fluids
  - body fluids such as urine
  - peripheral blood:
    - Protein adducts are formed when the carcinogen binds to blood proteins such as haemoglobin or albumin. This method assumes that protein adduct formation parallels DNA adduct formation in target tissue.


Internal carcinogen exposure
Metabolism
(Inactivated)
DNA & protein adducts formed in non-target cells
DNA & protein adducts formed in target cells
Blood
Target organ
DNA damage
(DNA adducts in exfoliated target cells)
(DNA repair)
Organ cavities
Mutation
Urine

Studies nested within the cohort

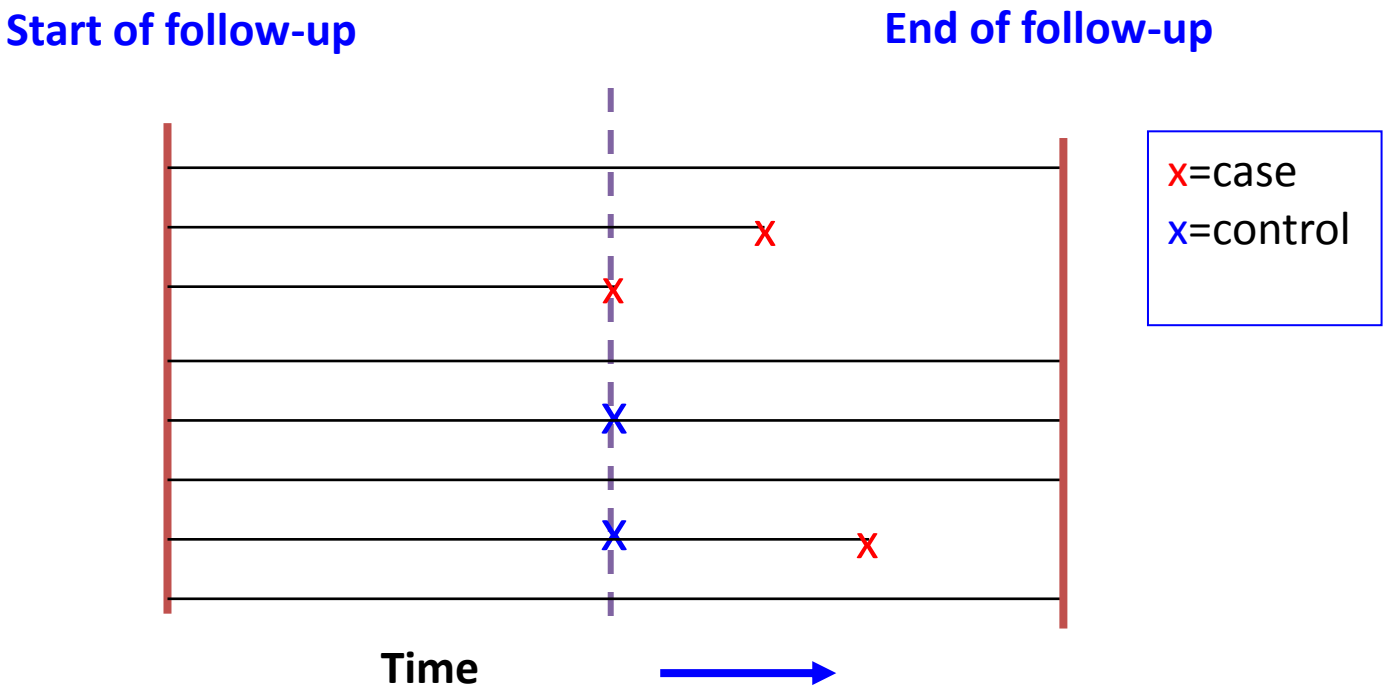
Each line represents one subject
x = case



Studies nested within the cohort

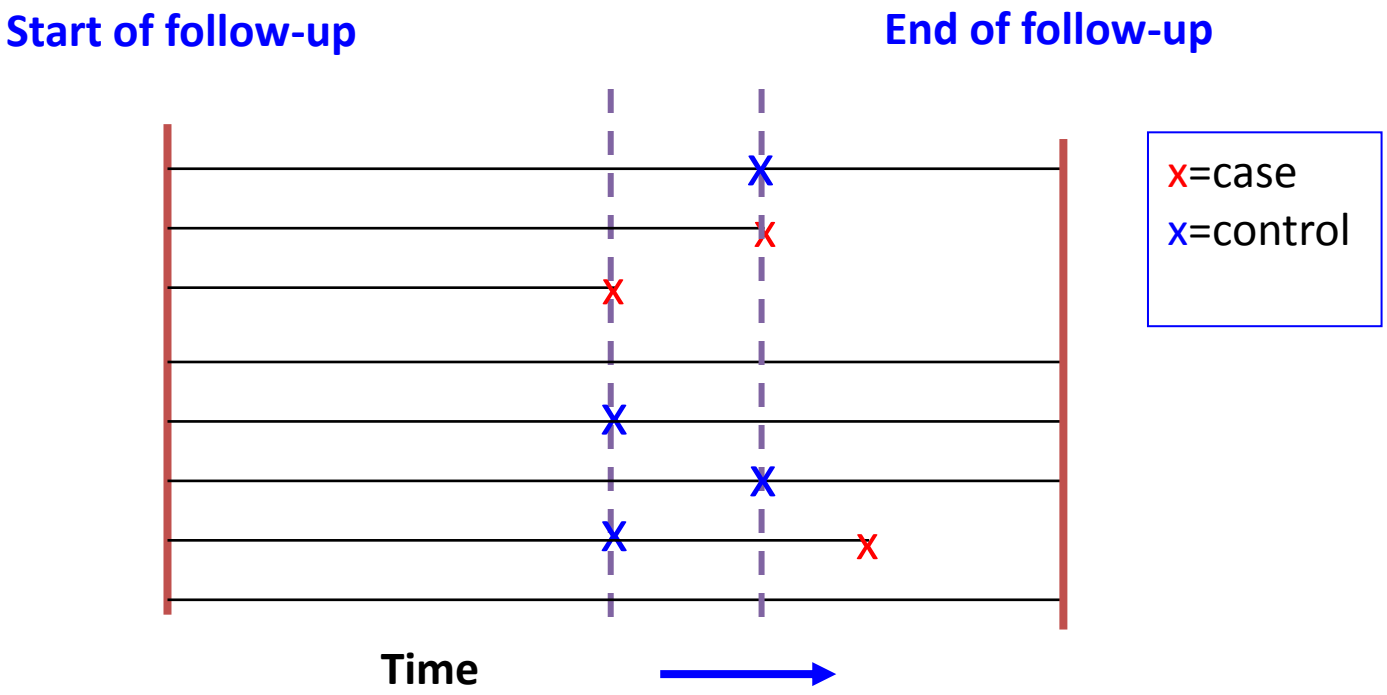
For each case, one or more controls are selected among those who are still at risk (disease-free) at the time of diagnosis of the case.

So, controls are matched to the cases on time



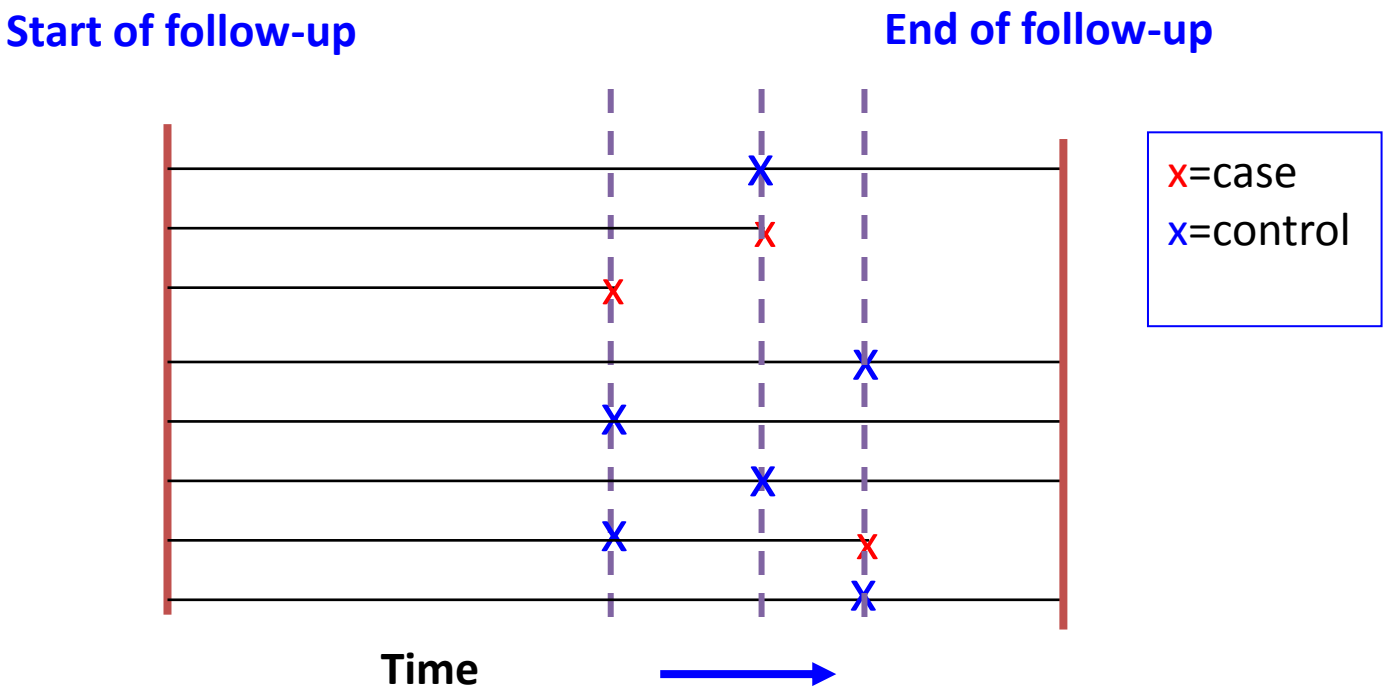
Studies nested within the cohort

Controls matched to the cases on time – risk sets



Studies nested within the cohort

Controls matched to the cases on time – risk sets



Nested case-control design

- Advantages:
  - Less costly (e.g. only need to conduct lab assays for the cases and controls rather than the whole cohort)
  - No need to collect exposure data for cases and controls beyond the time of follow-up of the case
  - Statistically-efficient

- Disadvantages:
  - Requires selection of controls as cases arise (with eventual collection of additional exposure data)
  - Impossible to examine data using multiple time scales (i.e. analyses will have to use the time scale on which controls were matched to the cases)
  - Controls selected for a particular outcome cannot be used to examine other outcomes

Example:

- Aflatoxin - metabolic product of fungi of the *Aspergillus spp.* that grow on maize and peanut crops
- Risk factor for liver cancer
- Aflatoxin forms adducts with DNA in liver cells, which correlates with a mutation in codon 249 of the p53 gene
- Descriptive studies conducted to determine the prevalence of aflatoxin-DNA adducts in liver biopsy samples from individuals with liver cancer

Example:

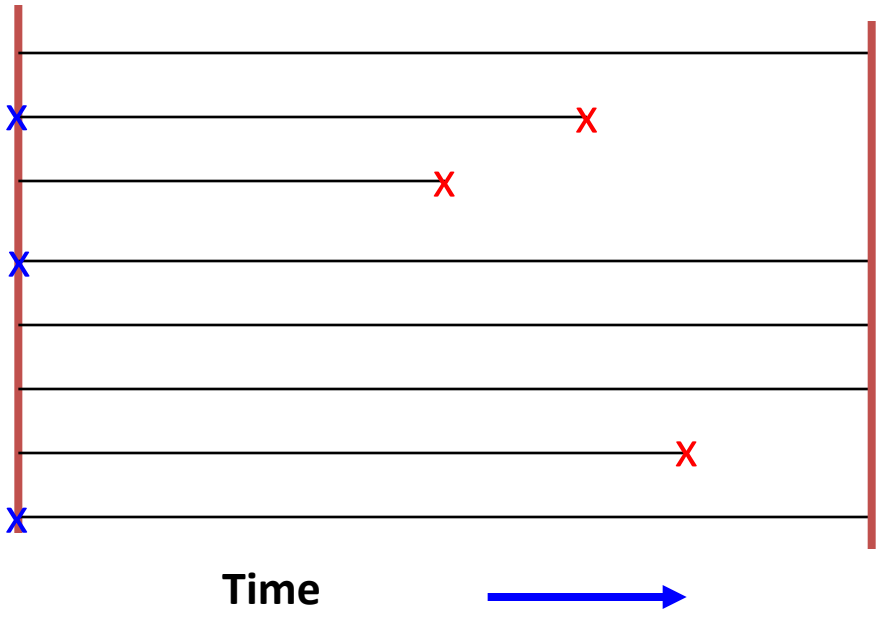
- Urine samples from 18,244 men in Shanghai.
- ~70,000 person-years of follow up

  55 cases of liver cancer ascertained.
- The presence of aflatoxin adduct was compared in the stored urine samples from the individuals with liver cancer and from 267 individuals without liver cancer, randomly selected from the cohort.
- Positive association between the presence of urinary aflatoxin adduct and risk of liver cancer.

(Quian et al., 1994)

Case-cohort studies (risk ratio)

Random sample of the cohort
(sub-cohort)



All cases diagnosed
during follow-up

Case-cohort design

- **Advantages:**
  - Less costly (e.g. only need to conduct lab assays for the cases and the sub-cohort)
  - The sub-cohort can be identified at the start of the follow-up
  - New cases do not require identification of new controls
  - The same sub-cohort can be used to examine multiple outcomes
  - The sub-cohort can be used to estimates incidence/prevalence
  - Possible to examine multiple time scales

- **Disadvantages:**
  - Requires constant up-date of exposure data for the whole sub-cohort

Belief that biomarkers have several strengths:

- Objective
- Individualised (target to each individual at relevant times)
- More specific and sensitive

Limitations of exposure biomarkers:

1. **Not always objective**

- Tend to be automated (hence not affected by knowledge of the exposed/unexposed, or case-control, status of subjects

- But subjectivity may be introduced if subjects have to collaborate in the collection of specimens

Limitations of exposure biomarkers:

2. The wrong biomarker may be chosen for the exposure of interest:

- it may not be specific to a single external exposure

- the correct component(s) of a complex external exposure may not have been chosen (e.g. vit. A)

- biomarkers in surrogate tissues may not reflect events at the target site

-it may measure early outcome rather than exposure

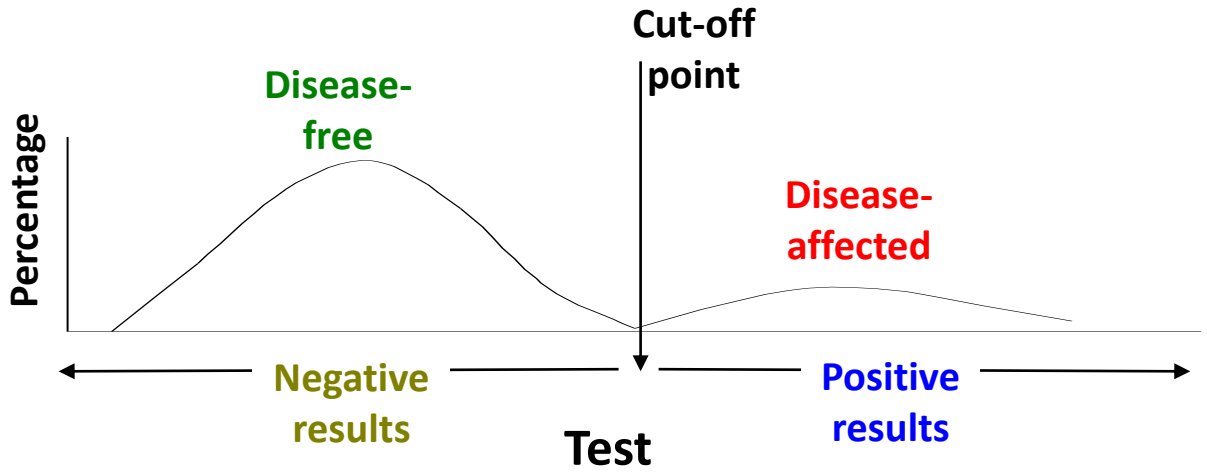
Limitations of exposure biomarkers:

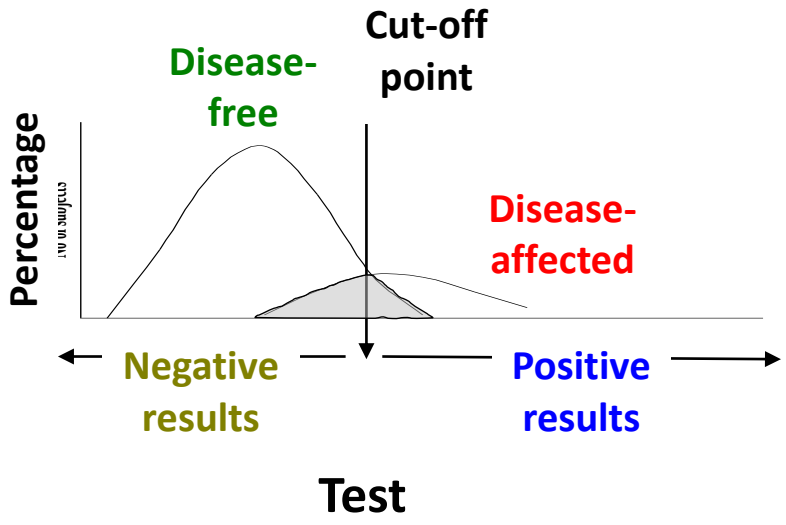
2. The biomarker assay may have poor validity:

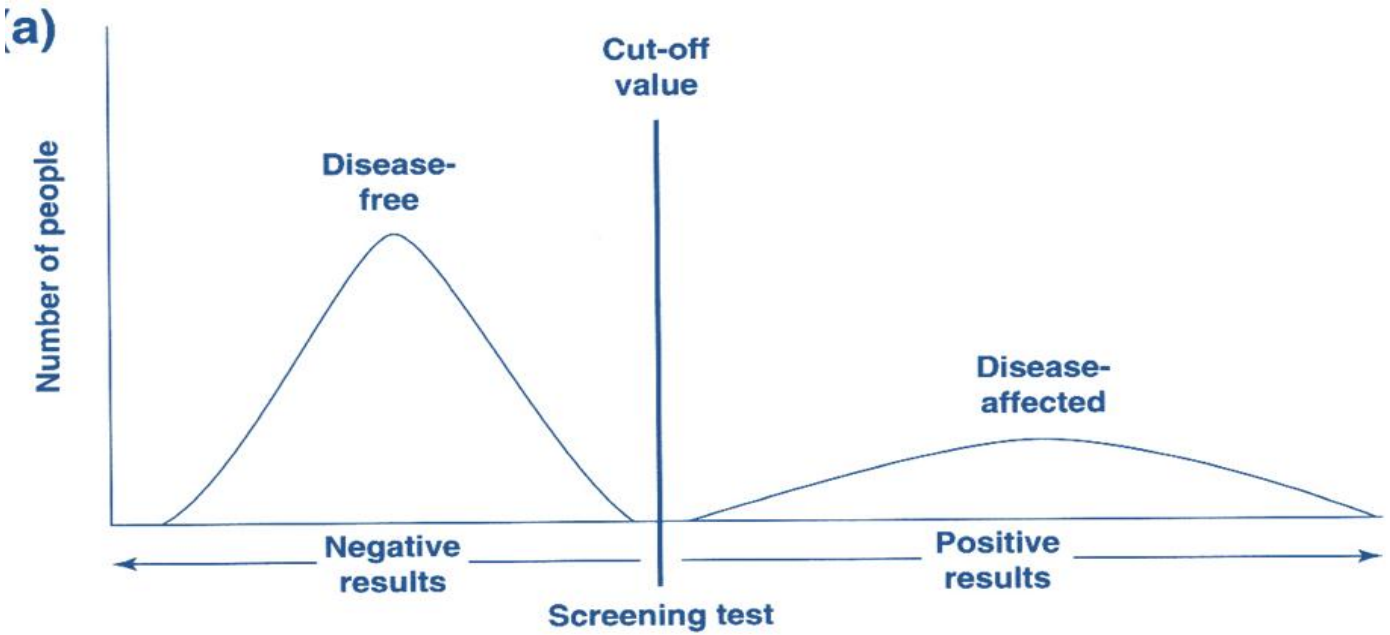
- it is important to assess how well the test (biomarker assay) measures the true value of the exposure (the biomarker).

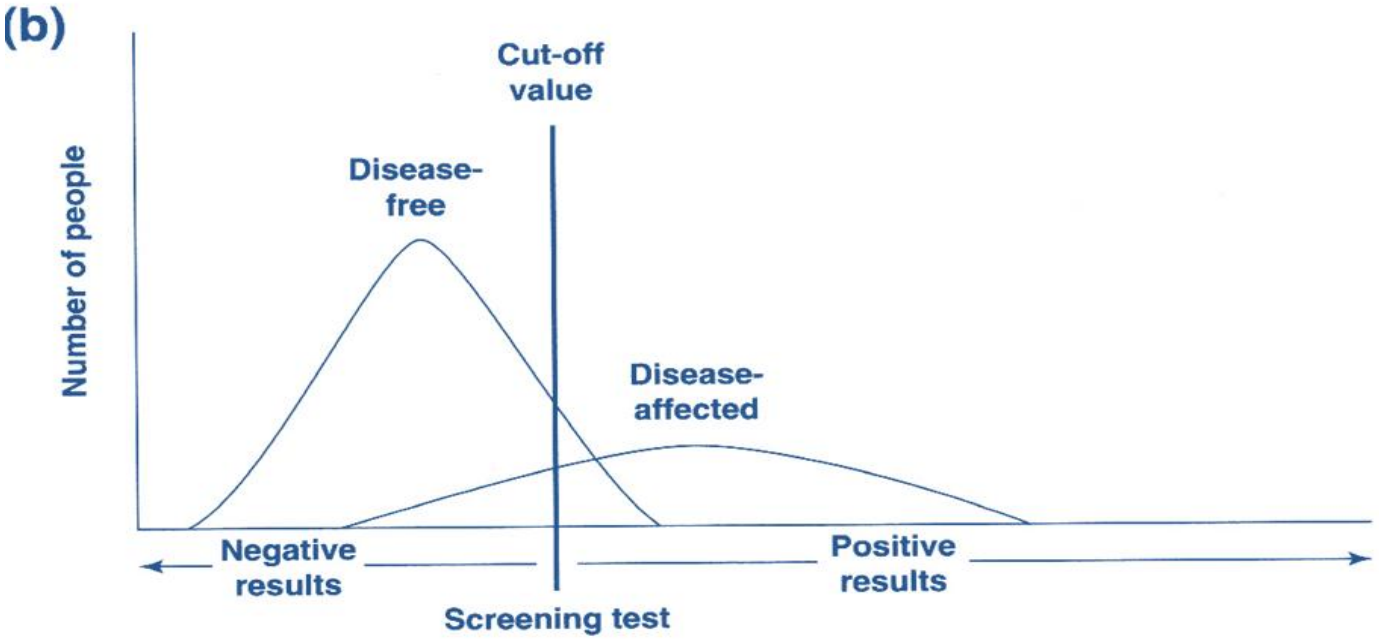
» Validity (sensitivity and specificity)

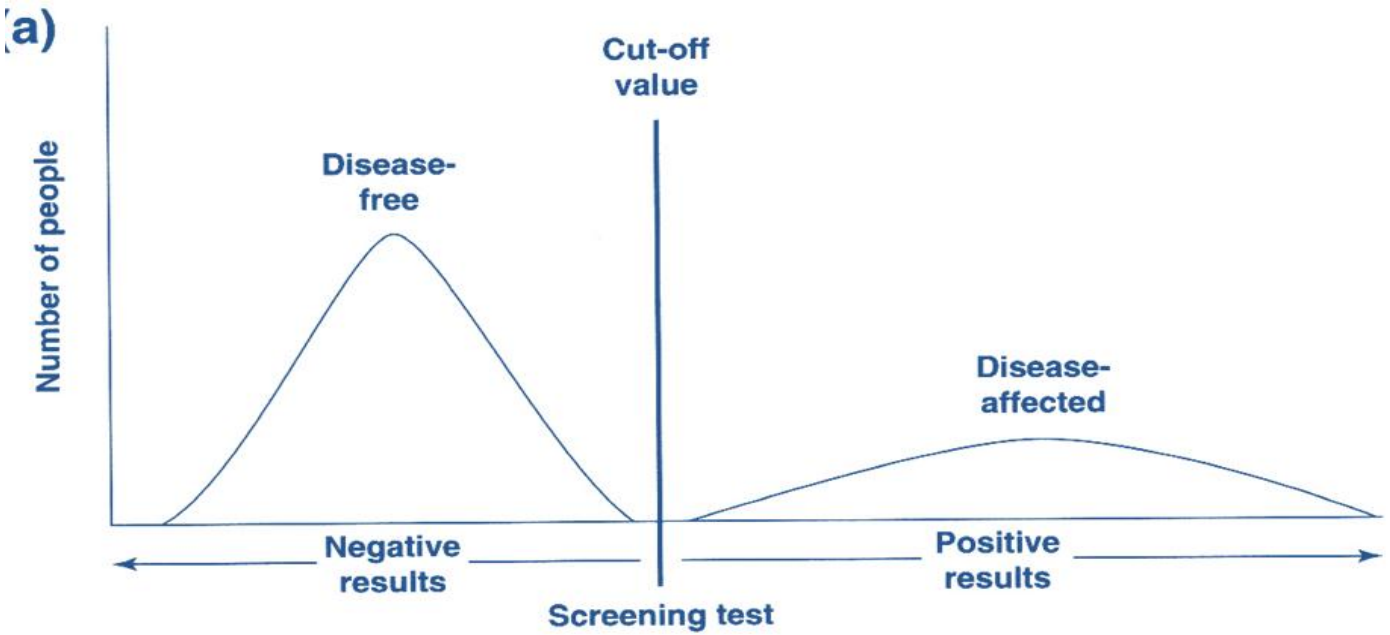
» Reliability (Kappa statistics, inter-coefficient of variation)

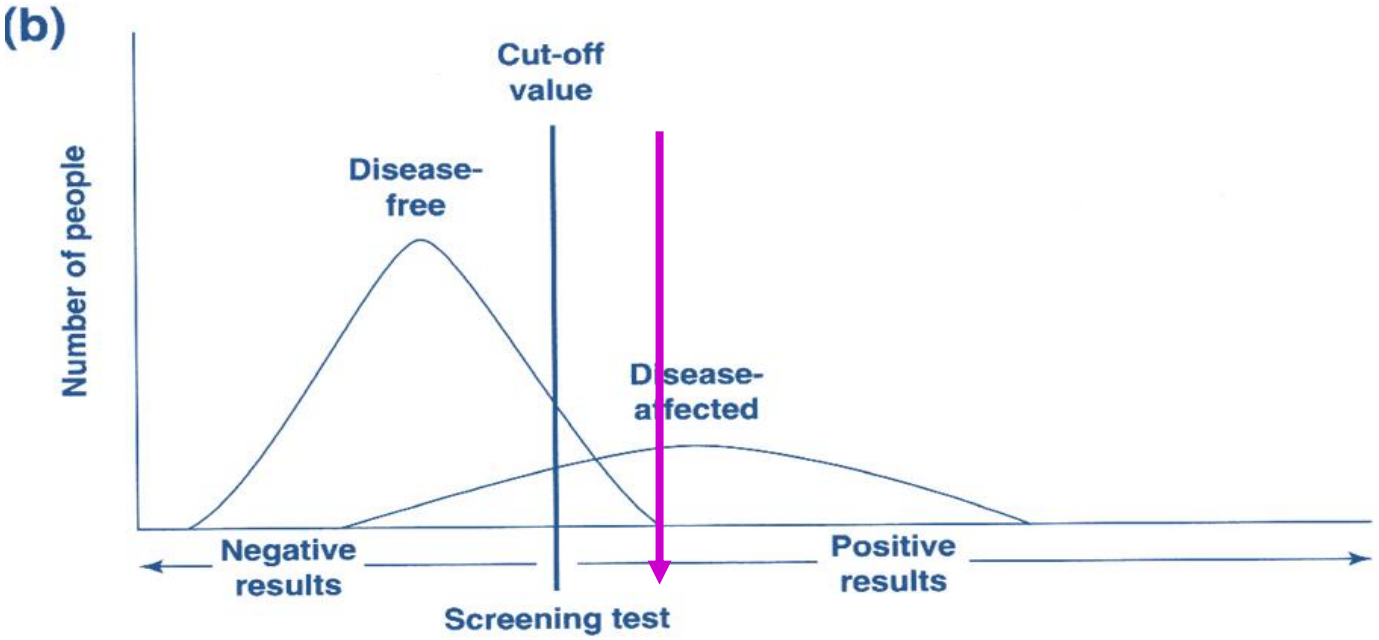

Cut-off point
Disease-free
Disease-affected
Percentage
Negative results
Positive results
Test

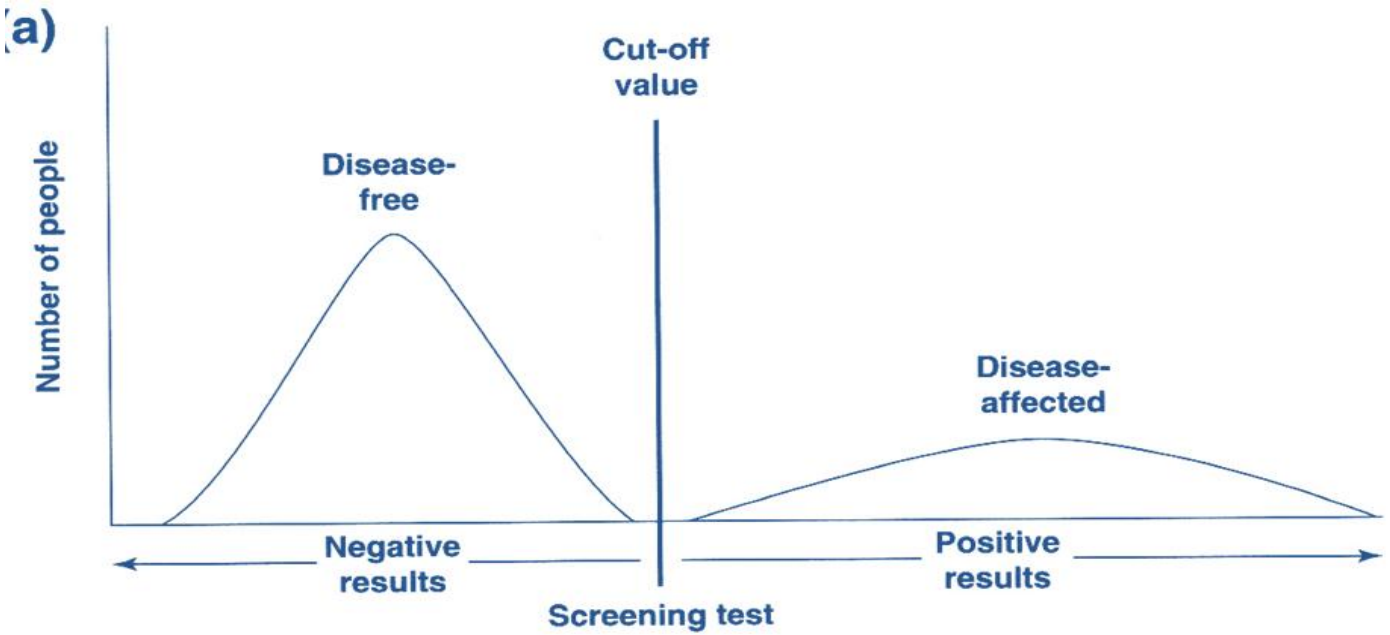

Cut-off point
Disease-free
Disease-affected
Percentage
Negative results
Positive results
Test

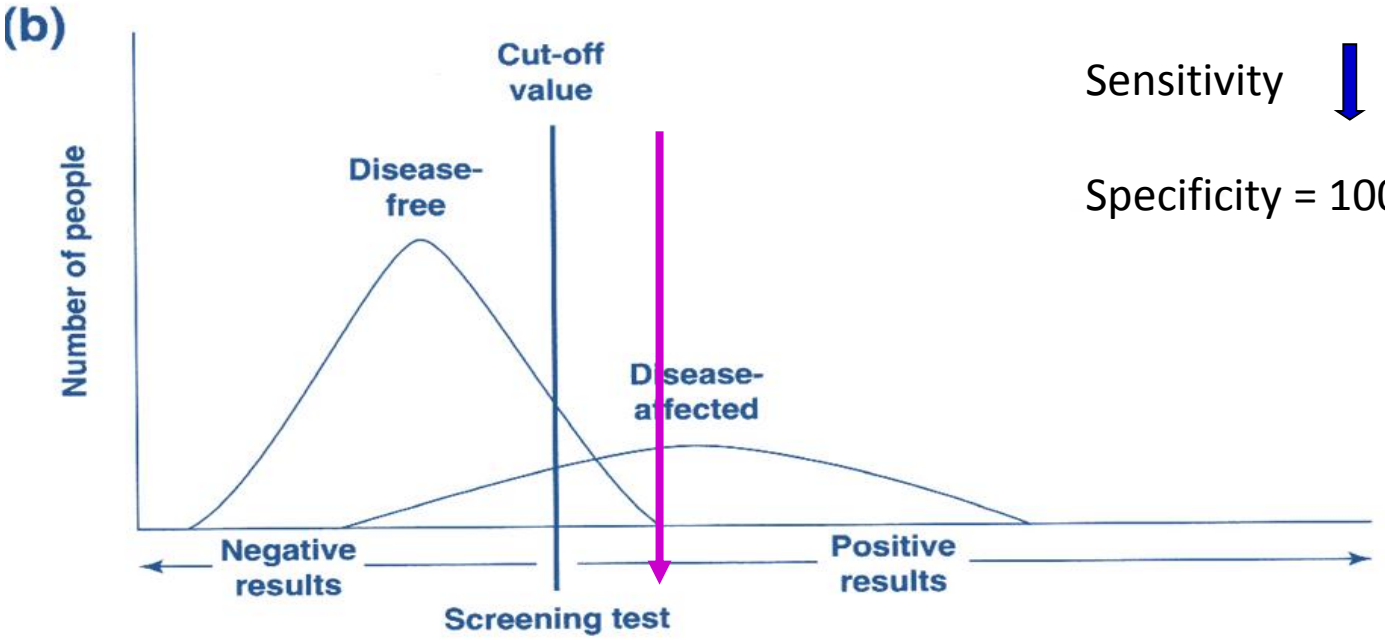

(a)
Cut-off value
Disease-free
Number of people
Disease-affected
Negative results
Positive results
Screening test


(b)
Cut-off value
Disease-free
Number of people
Disease-affected
Negative results
Positive results
Screening test


(a)
Cut-off value
Disease-free
Number of people
Disease-affected
Negative results
Positive results
Screening test

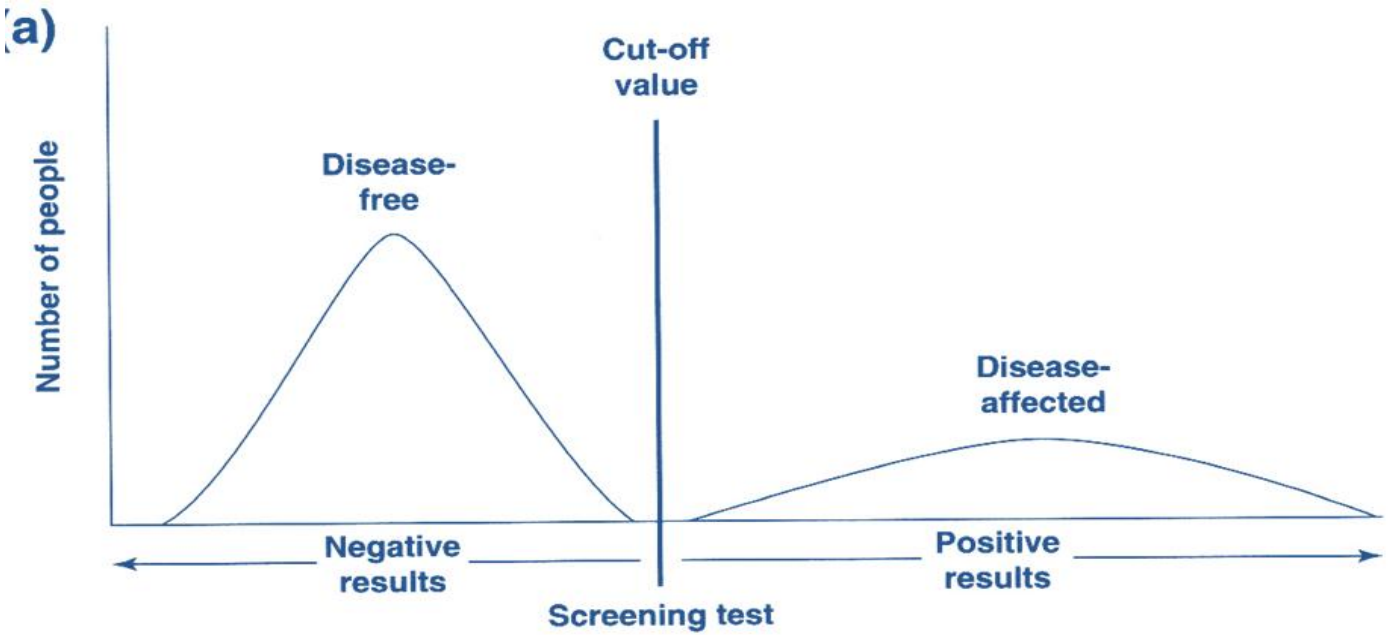

(b)
Cut-off value
Disease-free
Number of people
Disease-affected
Negative results
Positive results
Screening test

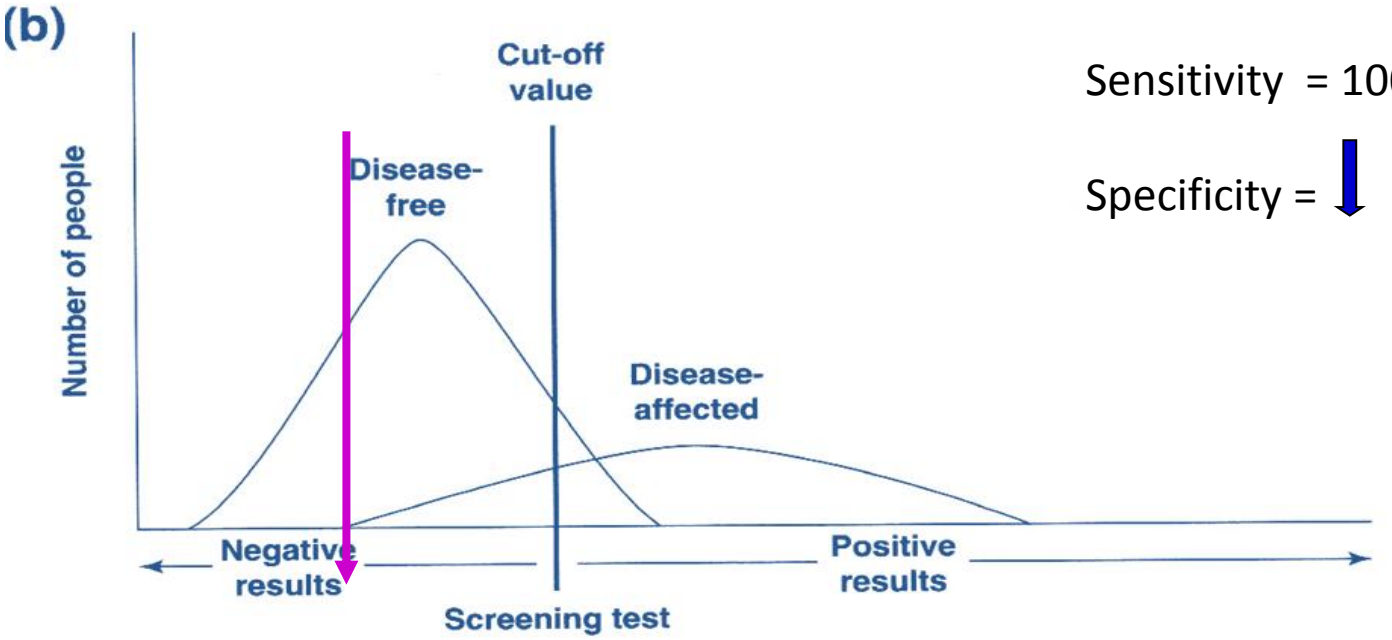




Sensitivity ↓

Specificity = 100%





Sensitivity = 100%

Specificity = ↓

TABLE 1 *Biased odds ratios according to sensitivity and specificity with OR = 2 and $P_o = 0.1$*

		Specificity					
		0.60	0.70	0.80	0.90	0.95	1.00
Sensitivity	0.60	1,069	1,115	1,188	1,337	1,504	1,918
	0.70	1,105	1,154	1,231	1,388	1,556	1,938
	0.80	1,141	1,192	1,273	1,436	1,603	1,957
	0.90	1,179	1,231	1,315	1,482	1,648	1,978
	1.00	1,217	1,270	1,357	1,526	1,690	2,000

Limitations of exposure biomarkers:

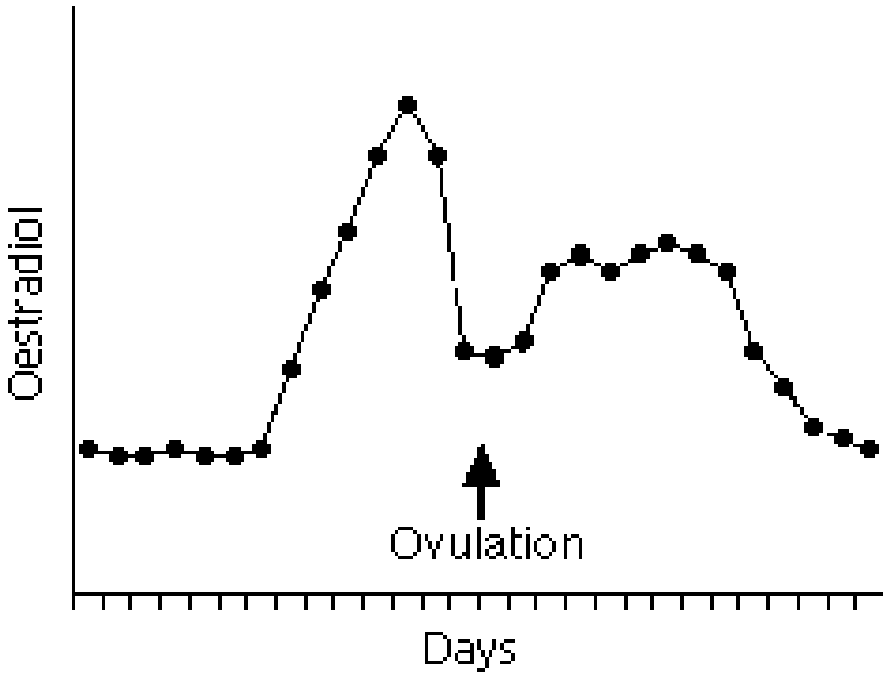
3. Biomarker may not reflect measurement "at the right time"

- Should reflect exposure at a time relevant aetiologically to the disease (often unknown)

- Internal exposure may vary *within an individual* due to physiological processes, etc. - short-term physiological variability

Collection of specimens at multiple time points may be required to capture long-term exposure patterns

Variation in oestradiol levels during the menstrual cycle



The biomarker may not measure exposure at the relevant time:

Biomarker	Average life span
Haemoglobin-adducts	3 months
Albumin-adducts	20-25 days
Cellular DNA	hours-years

Limitations of exposure biomarkers:

3. Biomarker may not reflect measurement "at the right time"

- Should reflect exposure at a time relevant aetiologically to the disease (often unknown)

- Internal exposure may vary *within an individual* due to physiological processes, etc. - short-term physiological variability

Collection of specimens at multiple time points may be required to capture long-term exposure patterns

Limitations of exposure biomarkers:

4. Biomarker may not reflect measurement "at the right site"

- Requires knowledge of the kinetics of absorption, distribution, storage, metabolism and elimination from the body

- **Crucial to establish the correct site and time for sampling**

Limitations of exposure biomarkers:

5. Feasibility and costs

- Nature of specimen collection (e.g. acceptability, safety, adversary effects)

- Complexity of storage and processing logistics

- Costs

Biomarkers of exposure: quality control concerns

- Minimise the effect of storage time by matching controls to cases on time of collection and duration of follow-up
- Account for other relevant variables such as number of freeze-thaw cycles
- All samples should be send to the same lab (high-quality one)
- Blind lab technicians to case-control status of the samples, exposure data and study hypothesis (avoid information bias)

Biomarkers of exposure: quality control concerns

- Use a common standardised protocol to perform the laboratory measurements (to minimises inter-assay variability).
- Samples should be analysed in case-control pairs to minimize inter-batch variability.
- Duplicate samples (with laboratory staff not being aware they are duplicates) should be randomly inserted within each batch to allow estimation of intra- and inter-batch variability.
- A random selection of samples should be measured using a "gold standard" method (or sent to a reference laboratory, if there is one) to check the validity of the laboratory assays.

Limitations of exposure biomarkers

In summary:

- exposure biomarkers have limitations
- external exposure vs. exposure biomarkers?

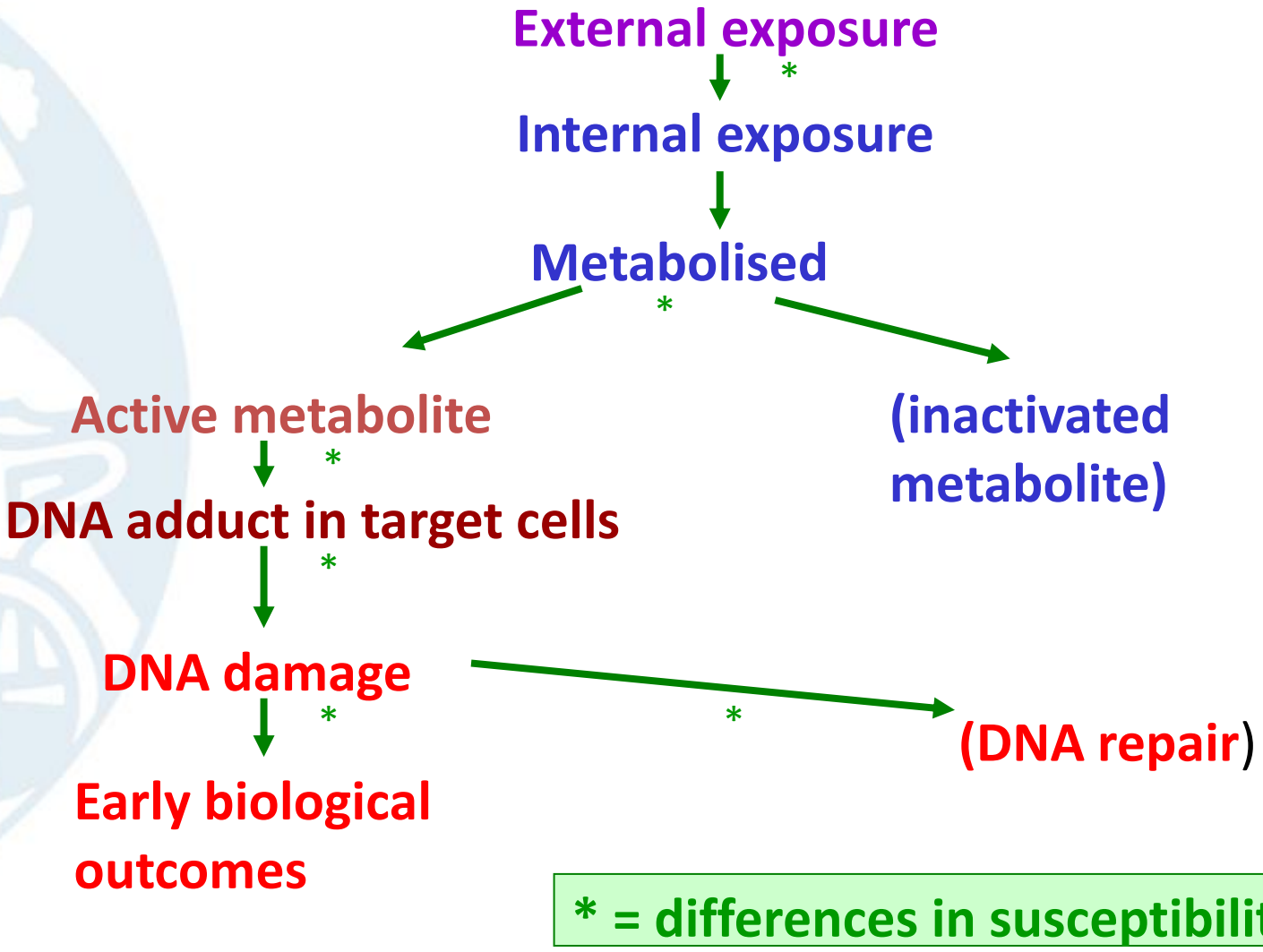
→ they measure different things!

Biomarkers to measure susceptibility

Individuals may have different *susceptibility* to the effect of external exposures

E.g. individuals may differ in the way they:

- absorb and metabolise the external exposure
- repair DNA
- respond to mutational events


External exposure
*
Internal exposure
Metabolised
*
Active metabolite
*
DNA adduct in target cells
*
DNA damage
*
Early biological outcomes
*
(DNA repair)
(inactivated metabolite)
* = differences in susceptibility

Biomarkers to measure susceptibility

When an internal exposure is metabolised, it may be:

- ***activated:*** the metabolite can interact with target cells
- ***inactivated:*** the metabolite can be excreted

The metabolic activity of many activating and inactivating enzymes is determined genetically

So, differences in activity can be measured using either **phenotypic** or **genotypic** biomarkers

Biomarkers to measure susceptibility

Example:

N-acetyltransferase-2 (NAT2) is an enzyme that inactivates several carcinogens, including compounds present in cigarette smoke.

NAT2 activity can be determined using:

- ***Phenotypic* tests** - e.g., drug clearance tests (by administering a substance that is metabolised by NAT2 to see how fast the metabolised substance is excreted from the body)

- ***Genotypic* tests** – to assess the structure of the NAT2 genes that code for the enzyme

Fast vs. slow acetylators

Biomarkers to measure susceptibility

- Population-based case control study of history of smoking and acetylator type as risk factors for renal cell (kidney) cancer (RCC).
- Smoking history was obtained at interview
- Acetylator type was determined by characterising the NAT2 alleles

(Semenza et al., 2001)

Biomarkers to measure susceptibility

Smoking status	Odds ratio for RRC	(n=374)
Non Smokers	**1.0**	
Smokers	**2.2 (1.3 - 3.7)**	

Is the effect modified by acetylators status?

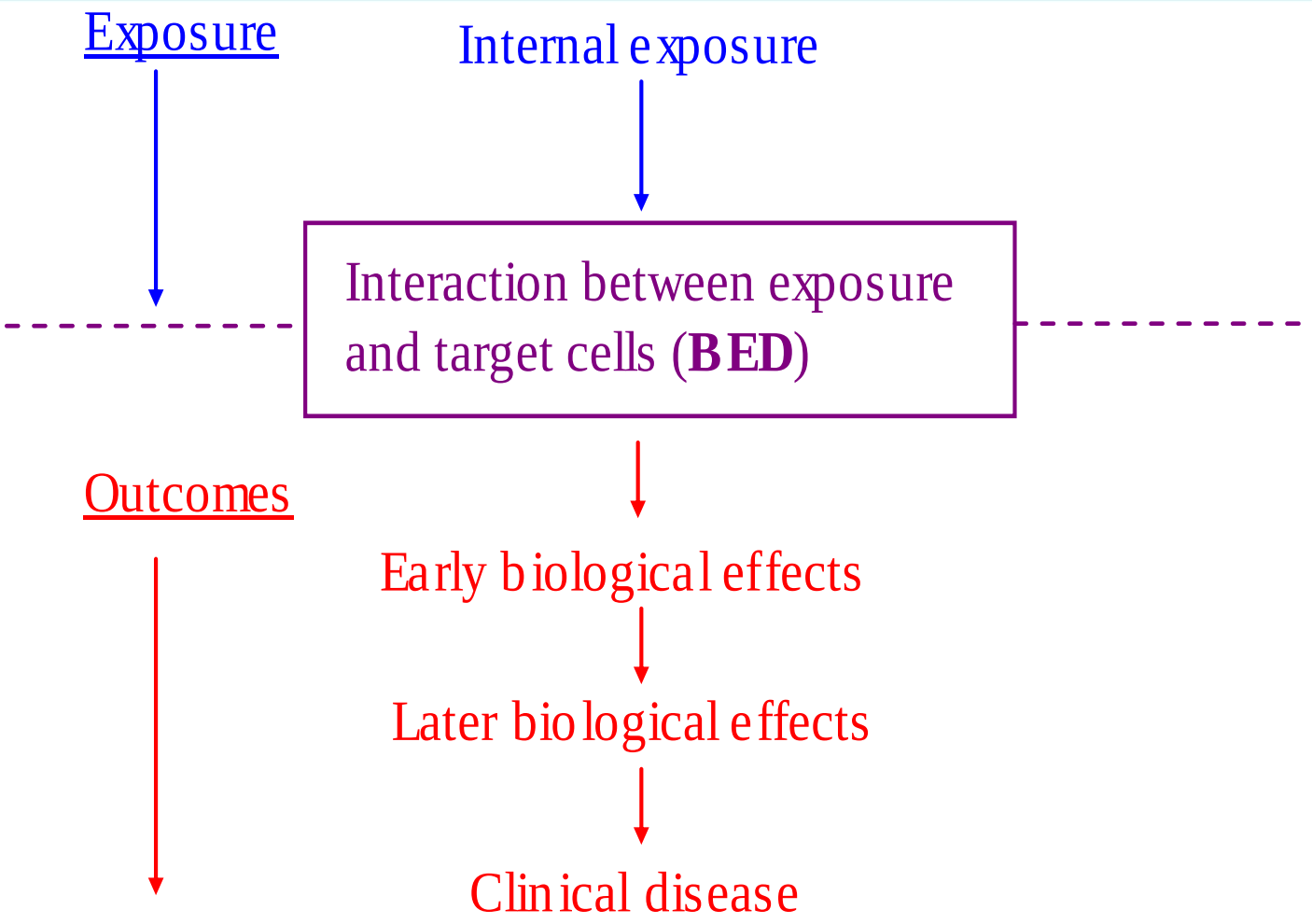
Biomarkers to measure susceptibility

OR for effect of smoking on RCC

Smoking Status	Slow acetylators (n=195)	Fast acetylators (n=179)
Non Smokers	1.0	1.0
Smokers	3.2 (1.7 - 6.1)	1.4 (0.7 - 2.9)

P for interaction = 0.04

Is there evidence evidence of effect modification?


Exposure
Internal exposure
Interaction between exposure and target cells (**BED**)
Outcomes
Early biological effects
Later biological effects
Clinical disease

Biomarkers to measure early outcomes

- Early outcomes are usually events on the causal pathway which lie between the *biologically effective dose* (BED) and clinical disease.

- Can be subdivided into *early* and *later biological effects*, depending on how close they are to the clinical outcome.

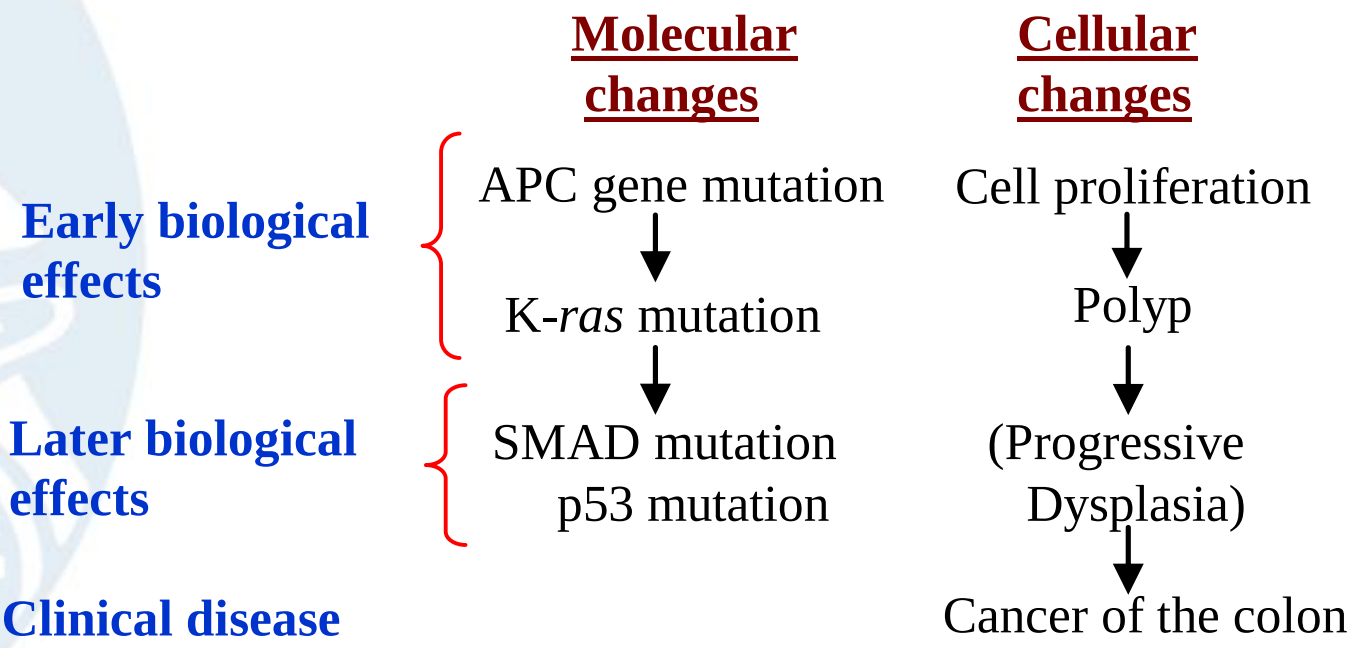
Biomarkers to measure early outcomes

Biological effects preceding cancer of the colon:

- ***Molecular* changes**: **early** changes include mutations which inactivate the tumour suppressor gene APC and cause dysregulation of the K-*ras* oncogene. **Later** changes include mutations which inactivate the SMAD and p53 genes.

- ***Cellular* changes**: **early** changes include proliferation of the cells lining the colon to form a polyp. **Later** changes include progressive dysplasia, and progress to cancerous cells.

Biomarkers to measure early outcomes



Biomarkers to measure early outcomes

- Early and late biological effects can be identified by carrying out **natural history studies**.

- Once these early outcome biomarkers have been identified, they can be considered as outcomes in epidemiological studies.

Biomarkers to measure early outcomes

Early outcome biomarkers can be used to obtain earlier endpoints in studies. This may be useful for:

- Screening for pre-clinical disease
- Analytical epidemiological studies
- Monitoring variations in population health risk

Biomarkers to measure early outcomes

What is the advantage of using early outcomes in analytical studies?

- Individuals can be followed up for a shorter period of time
- The sample size of the study can be smaller

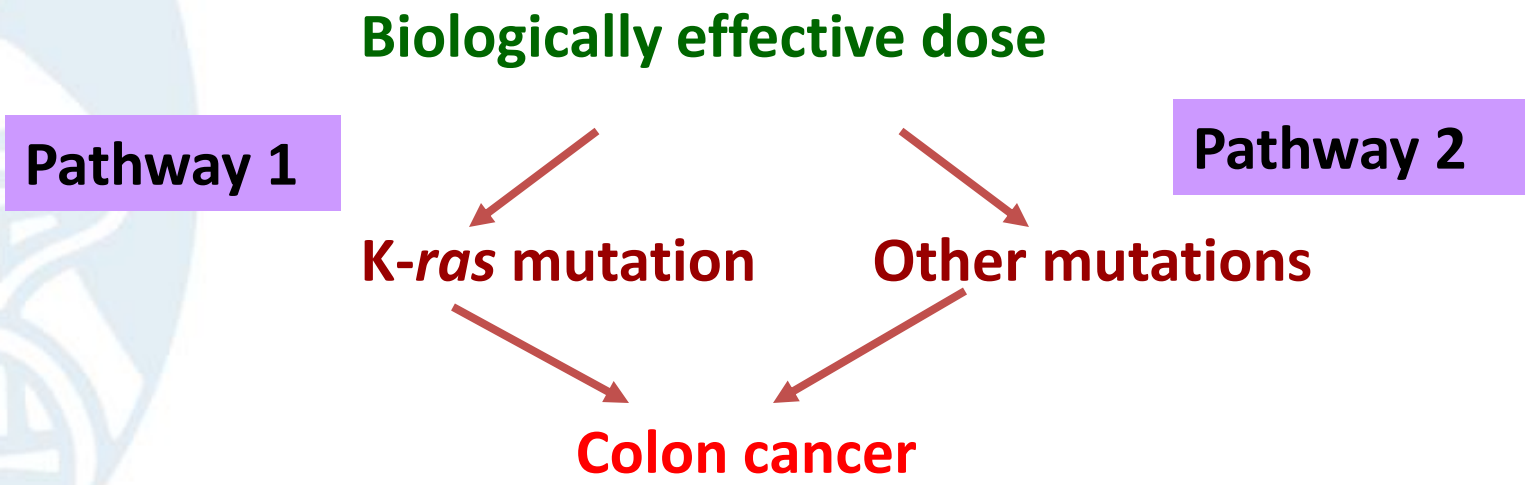
Biomarkers to measure early outcomes

Use of an ***early outcome biomarker*** may be limited if:

- it is not a necessary cause of the clinical disease
- it is not a sufficient cause of the clinical disease
- it is an early outcome for more than one clinical disease
- it is intermittently produced or unequally distributed throughout the tissue tested
- its assay has imperfect validity or reliability

Biomarkers to measure early outcomes

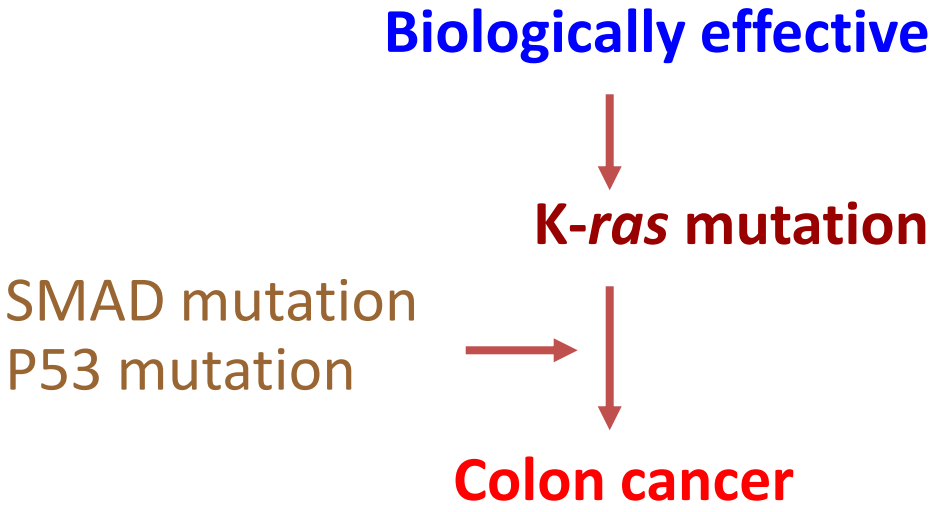
An early outcome biomarker may not be a necessary cause of the disease - there may be more than one causal pathway for the clinical outcome.



In this case, a positive K-*ras* biomarker will *underestimate* subsequent colon cancer.

Biomarkers to measure early outcomes

An early outcome biomarker may not be a <u>sufficient cause</u> of the disease that follows it.

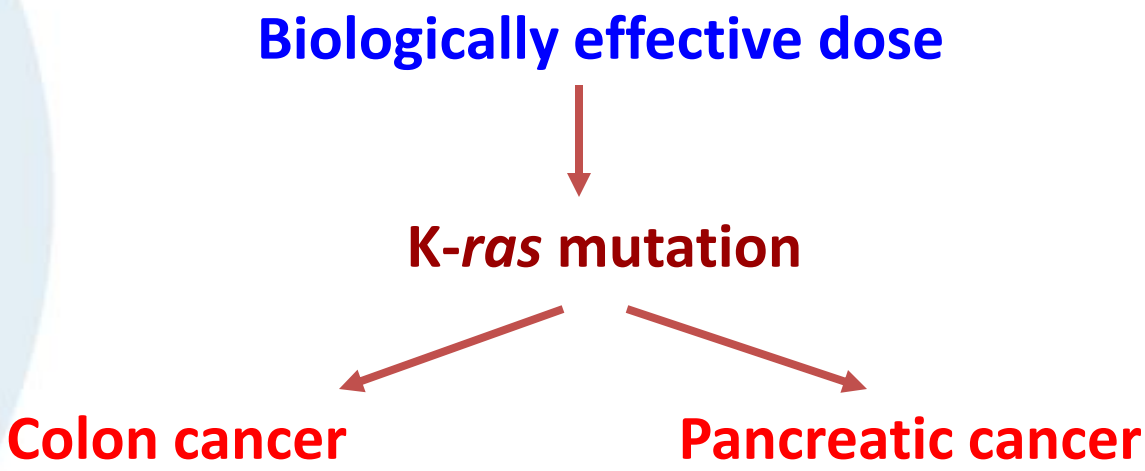


The K-*ras* mutation may not inevitably lead to colon cancer – subsequent mutations may also be needed.

Thus, a positive K-*ras* biomarker may *overestimate* subsequent colon cancer.

Biomarkers to measure early outcomes

The biomarker may not be specific to the disease of interest.



A positive K-*ras* biomarker may *overestimate* subsequent colon cancer

Biomarkers to measure early outcomes

Do understanding of disease mechanisms matter?

Biomarkers can also be used to clarify the intermediate stages in disease pathogenesis

- But:

  - understanding disease mechanisms is not essential for disease prevention

Do understand of disease mechanisms matter?

However, knowledge of biological mechanisms may:

- strengthen the evidence for a causal relationship between an exposure and a disease
- identify biomarkers for use in epidemiological research
- identify targets for preventative or therapeutic interventions
- identify aetiologically distinct disease sub-types

Introduction to the use of biomarkers in research: summary

- What are biomarkers?
- Use of biomarkers to measure exposure
- Use of biomarkers to measure susceptibility
- Use of biomarkers to measure early outcomes
- Use of biomarkers to clarify disease mechanisms
- Limitations of biomarkers