

Paolo Vineis

Validation of biomarkers

Markers need to be validated

What is validation?

- **To achieve an accurate estimate of the association between any marker and disease, in epidemiology we need reliable and valid measurements of exposure, of covariates (potential confounders and effect modifiers), and of outcomes.**

Levels of validation:

**Intra-individual variation
Inter-individual variation
and confounding**

**Intra-laboratory variation
Inter-laboratory variation**

Validity (vs a standard) and predictive value

Time relationships

Dose-response

Ability to predict outcome

Validation and relevance: some examples

Inter-centre variation (and potential confounding) for an intermediate marker (plasma DNA amount in EPIC)

Univariate and multivariate analysis: plasma DNA amount (logarithm transformation, dependent variable), by center, age, gender and time between blood drawing and diagnosis (for cases only).

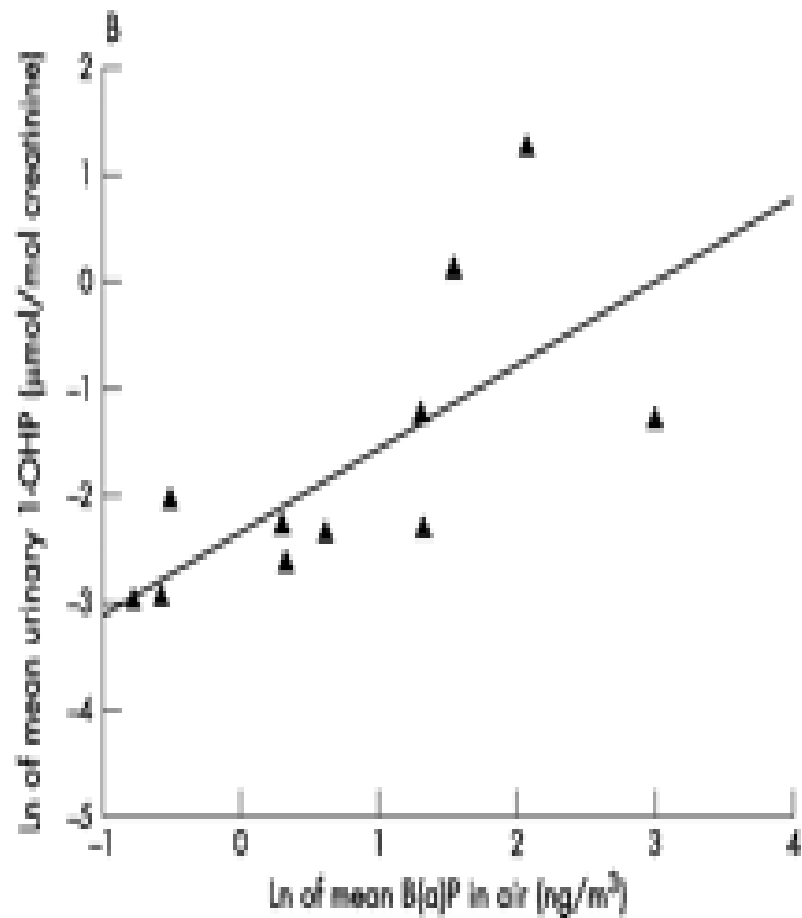
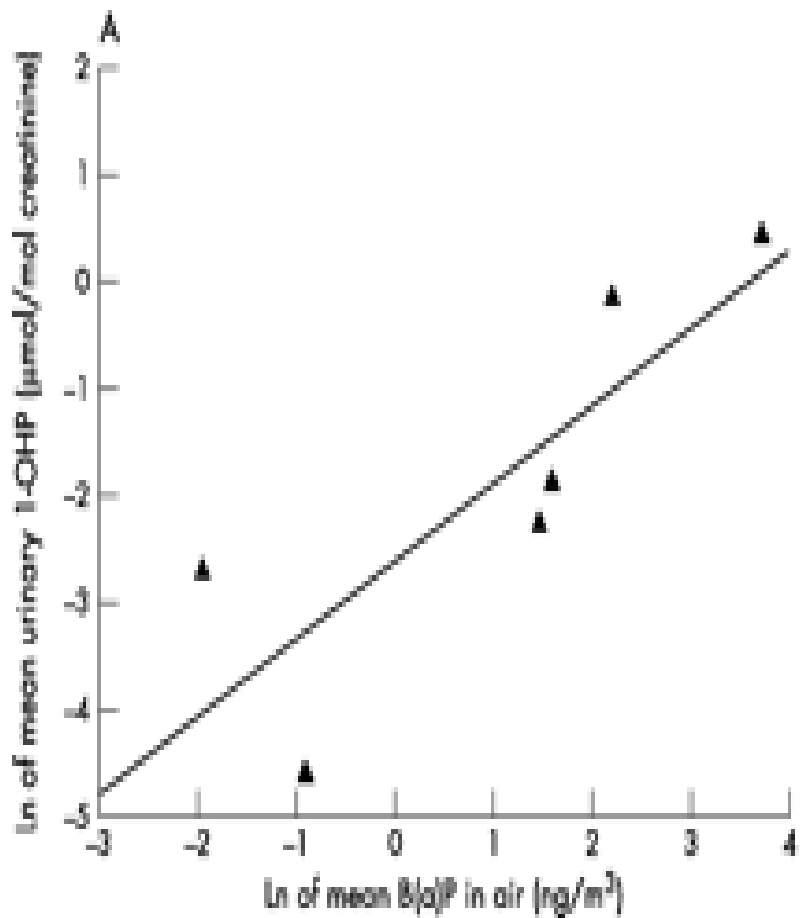
Univariate analysis:

Variable	F-value	DF	p-value
Controls only (N=778)			
Center	11.23	22	<0.0001
Age	5.21	1 (a)	0.023
Gender	0.52	1	0.47

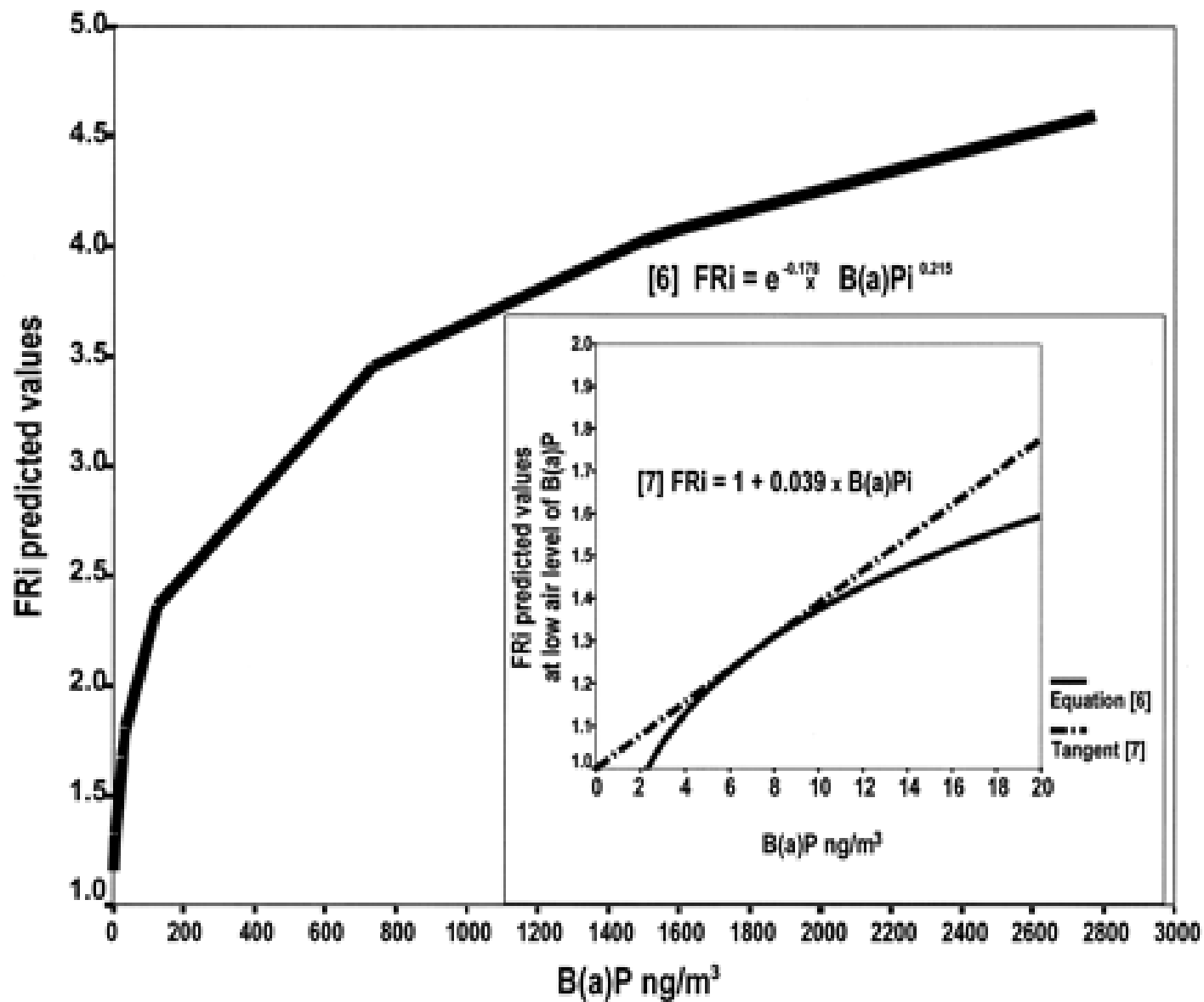
Cases and controls (N=1185):

	F-value	DF	p-value
Center	16.6	23	<0.0001
Age	1.56	1 (a)	0.21
All deaths and tumours	2.3	6	0.03

Urinary 1-OHP vs. external measurements



Bulky DNA adducts and dose-response relationship



Comments

- The fact that adducts and other markers are related to exposure does not imply that they are a better measure
- Biomarkers can increase biological plausibility of associations
- They can be useful for example if it is possible to show that intra-individual variability is lower with the marker than with external exposure measurements
- They can address issues such as saturation of enzymes at high levels of exposure (dose-response, risk assessment)
- DNA adducts are an integrated marker (over several sources of exposure) that expresses also individual susceptibility (eg for DNA repair), and can be predictive of cancer onset

Genotyping

Comparison of four genotyping methods at the Cambridge laboratory. The standard is represented by a panel evaluation of all results (courtesy of A Dunning).

<u>Method</u>	<u>Sensitivity</u>	<u>%</u>	<u>Specificity</u>	<u>%</u>
ASO	836/864	97	753/836	90
Taqman	826/864	96	812/826	98
RsaI digest	125/173	72	103/125	82
Invader	62/92	67	45/62	73

According to estimates, the common genotyping method Taqman has 96% sensitivity and 98% specificity, thus allowing little error in classification.

On the contrary, sensitivity in environmental exposure assessment is quite often lower than 70% and specificity even lower.

Genotype is stable, measured accurately (sens, spec=90-100%), frequency of alleles is high

Environmental exposures are changing (life-course events), often measured inaccurately, frequency may be too low

In addition, genetic polymorphisms are investigated with high-throughput technologies that allow researchers to investigate thousands of SNP at a time: with the usual p-values this originates a large number of false positives (see Bayesian strategy proposed by Colhoun et al, Lancet 2003 361: 865-872)

False positives with genetic research vs false positives+negatives with environmental research?

Examples of validation for nutritional biomarkers

**QUALITY CONTROL OF A BIOMARKER:
MEASUREMENT ERROR**

**MEASUREMENT ERROR IS CLASSIFIED AS PREANALYTICAL
(BIOLOGICAL and SAMPLING ERROR) OR ANALYTICAL
(LABORATORY) ERROR**

**LABORATORY ERROR FOCUSES ON METHOD, INSTRUMENT,
REAGENT OR MATRIX EFFECTS**

PREANALYTICAL ERROR:

- individual genetic, environmental, behavioural and health status-related variability (including smoking status, weight and weight loss, physical exercise)

Example of genetic source of variation: FOLATE AND MTHFR

Health status-related: retinol or ascorbate and trauma, several biomarkers and inflammation

- sampling error: within subject variation due to hourly, daily, weekly, monthly ... changes

EXAMPLE OF QUALITY CONTROL PROGRAM: NATIONAL CHOLESTEROL EDUCATION PROGRAM (US)

Goals:

- 1. Attain analytical accuracy and precision (<3% cv)**
- 2. Identify individual determinants of cholesterol variation (lifestyle factors)**
- 3. Identify clinical determinants of variation (metabolic states, illness)**
- 4. Other sampling sources (fasting status, posture, serum vs. plasma)**

The NCEP guidelines have proven adequate to ensure 90% correct classification

OVERALL MEASURE OF ERROR IS THE COEFFICIENT OF VARIATION:

SD/MEAN x 100%

(SD in repeated measurements)

IDEALLY CALCULATED FOR SAMPLES AT THE BOTTOM, MIDDLE AND TOP OF THE REFERENCE CONCENTRATION RANGE DETERMINED IN HEALTHY SUBJECTS

OTHER EXAMPLES OF QUALITY CONTROL:

- Gunter et al (1996), international round-robin for folate involving 20 labs: CV of 27% for serum folate, and 36% for whole blood folate, with substantial intermethod variation**
- Pfeiffer et al (1999), interlaboratory comparison of homocysteine in plasma samples (14 labs): CV=9% among labs, and 6% within labs**

TWO MAJOR APPROACHES TO REDUCING MEASUREMENT ERROR ARE:

- 1. TO BLIND THE ANALYST TO THE CASE-CONTROL STATUS OF SPECIMENS**
- 2. TO ELIMINATE SYSTEMATIC DIFFERENCES IN THE WAY CASE AND CONTROL SPECIMENS ARE HANDLED**

Examples of nutritional biomarker reference sample and certified quality control

sources H. Michels Blanck et al **Laboratory Issues: Use of Nutritional Biomarkers** J. Nutr. 133:888S-894S, 2003

- National Institute of Standards and Technology Standard Reference Materials (U.S.A.)
- National Institute of Biological Standards and Controls (U.S.A.)
- World Health Organization, Blood Safety and Clinical Technology
- International Federation of Clinical Chemistry, Scientific Division
- Centers for Disease Control and Prevention, Division of Laboratory Sciences
- Northwest Lipid Research Laboratories, University of Washington, Seattle, WA
- Solomons Park Research Laboratories, Kirkland, WA
- Commercial companies (primary standards)
- Proficiency testing programs
- National reference material institutions

General considerations in choosing a nutritional biomarker for an epidemiologic research study

- 1. Timing relative to dietary exposure: recent intake versus usual intake, acute versus chronic exposure.**
- 2. Type of measurement: direct measure (static indicator) versus functional assay.**
- 3. Would a dietary assessment method such as a food-frequency questionnaire or a 24-h recall provide adequate dietary information precluding the need for biomarker assessment?**
- 4. Has within-person (intra-) and between-person (inter-) variance been documented for the biomarker measurement method of interest? If yes, is the between-person variance larger than the within-person variance? If not, it will be difficult to assess associations without an extremely large sample size.**

EXAMPLE

In the measurement of oxidative damage to DNA, routine phenol-based DNA purification procedure can increase 8-hydroxydeoxyguanosine levels 20-fold in samples that are exposed to air following removal of the phenol. Such gross contamination would seriously bias an epidemiological study if subsets (batches) coming from different subgroups in the study population (e.g. exposed vs. unexposed) undergo different technical procedures that result in different levels of error.

MEASUREMENT ERROR: BETWEEN LABS

Correlation coefficients (r) for the measurement of estrone by different laboratories and resulting observed relative risks given true relative risks of 1.5, 2.0 and 2.5 (from Hankinson et al, 1994).

Laboratory	r	True relative risks		
		RR _t =1.5	RR _t =2.0	RR _t =2.5

		Observed relative risks		
Lab 1	0.12	1.1	1.1	1.1
Lab 2	0.82	1.4	1.8	2.1
Lab 3	0.57	1.3	1.5	1.7
Lab 4	0.90	1.4	1.9	2.3

Observed RR = exp (ln RR_t * r)

INTRA-INDIVIDUAL VARIATION OVER TIME:

IMPACT OF SAMPLE DEGRADATION (CASES AND CONTROLS) AND EFFECT OF DISEASE ON METABOLISM (CASES)

Mean retinol levels in the blood of cancer cases and controls (Wald et al, 1986). Concentrations are in ug/l.

	Time between blood collection and cancer onset					
	Less than 1 year		1-2 years		3+ years	
	N	mean	N	mean	N	mean
Cases	66	641	45	650	116	694
Controls	132	722 (a)	90	701 (b)	232	633

(a) $p < 0.001$

(b) $p < 0.01$

INTRA-INDIVIDUAL VARIATION : SAMPLING ISSUES ACCORDING TO BIOMARKER TYPE

	Intraindividual variation over time	Biological sample variation
<i>Internal dose</i>		
Hormones	Yes (diurnal variation)	No
Water-soluble nutrients	Yes (short half-life)	No
Organochlorine	No (long half-life)	No
<i>Biologically effective dose</i>		
White blood cells	Yes (half life wks-mo.)	No
Urothelial cells	Yes (half-life months)	Yes
<i>Early effects</i>		
Chromosome aberrations	Stable	?

SAMPLING ISSUES ACCORDING TO BIOMARKER TYPE

	Intraindivual variation	Biological sample variation
<i>Intermediate markers</i>		
Cervical dysplasia	Yes	Yes
Colon hyperproliferation	Yes	Yes
<i>Genetic susceptibility</i>		
Genotype	No	No
Non-inducible phenotype	No	No
Inducible phenotype	Yes	No

It has been estimated that 20% of the variability of the rectal mucosa proliferation index (measured by nuclear antigen immunohistochemistry) is due to inter-individual variation, 30% to the site of biopsy within the subject, and 50% is due to inclusion of crypts (i.e., micro-anatomic location) within a biopsy. In other words, as much as 80% of variation is related to sampling.

Further biomarker validation:

**Comparison between FFQ and biochemical
measurements**

Reciprocal correlation among different biomarkers

EPIC: Spearman correlation coefficients (a) among carotenoids, and (b) between carotenoids and food items (R^2)(Al-Delhaimi et al, 2004)

	Lutein	Lycopene	B_cryptoxanthin	Zeaxanthin	A_carotene	B_carotene
Lutein	1	0.39 (<.0001)	0.36 (<.0001)	0.73 (<.0001)	0.22 (<.0001)	0.34 (<.0001)
Lycopene	0.39 (<.0001)	1	0.19 (<.0001)	0.26 (<.0001)	0.11 (<.0001)	0.30 (<.0001)
B_cryptoxanthin	0.36 (<.0001)	0.19 (<.0001)	1	0.48 (<.0001)	0.11 (<.0001)	0.25 (<.0001)
Zeaxanthin	0.73 (<.0001)	0.26 (<.0001)	0.48 (<.0001)	1	0.13 (<.0001)	0.24 (<.0001)
A_carotene	0.22 (<.0001)	0.11 (<.0001)	0.11 (<.0001)	0.13 (<.0001)	1	0.71 (<.0001)
B_carotene	0.34 (<.0001)	0.30 (<.0001)	0.25 (<.0001)	0.24 (<.0001)	0.71 (<.0001)	1

	dietary method	LUTEIN	ZEAXANTHIN	B_CRYPTOXANTHIN	LYCOPENE
Fruits & vegetables	24HDR	38.6% (0.0009)	66.0% (<0.0001)	27.9% (0.007)	40.4% (0.0006)
	FFQ	NS	16.5% (0.044)	19.8% (0.026)	23.4% (0.014)
Vegetables	24HDR	21.8% (0.019)	54.2 (<0.0001)	NS	17.6% (0.038)
	FFQ	NS	NS	NS	NS
Leafy vegetables	24HDR	NS	38.7% (0.0009)	NS	NS
	FFQ	NS	22.5% (0.017)	NS	NS
Root vegetables	24HDR	NS	16.5 (0.044)	NS	NS
	FFQ	NS	NS	NS	NS
Cabbages	24HDR	NS	NS	16.3% (0.045)	NS
	FFQ	NS	NS	NS	NS

Comments:

- different carotenoids are strongly correlated with each other
- good correlations between questionnaire estimates and biomarkers at the regional level (not at individual level!)

Examples of markers

1-carbon metabolism

**Complex of folate, B vitamins, including B12, choline
and methionine**

- **Modest dietary inadequacies of some of these compounds can cause severe diseases:**
 - **inadequate folate in pregnancy leads to neural tube defects**
 - **inadequate intake of folate, B-6 or B-12 raises serum levels of homocysteine, highly associated with CHD**
 - **inadequate intake of folate and methionine is implicated in carcinogenesis**

These effects occur at levels far from those of classical deficiency (megaloblastic anemia)

Folate and Cancer

- **Methyl-deplete diet, low folate associated with a number of cancers in epidemiologic studies:**
colorectal, esophageal, gastric, cervical, and breast

Methyl-deplete diet and status:

Low folate and methionine and high alcohol

Low vitamins B12 and B6

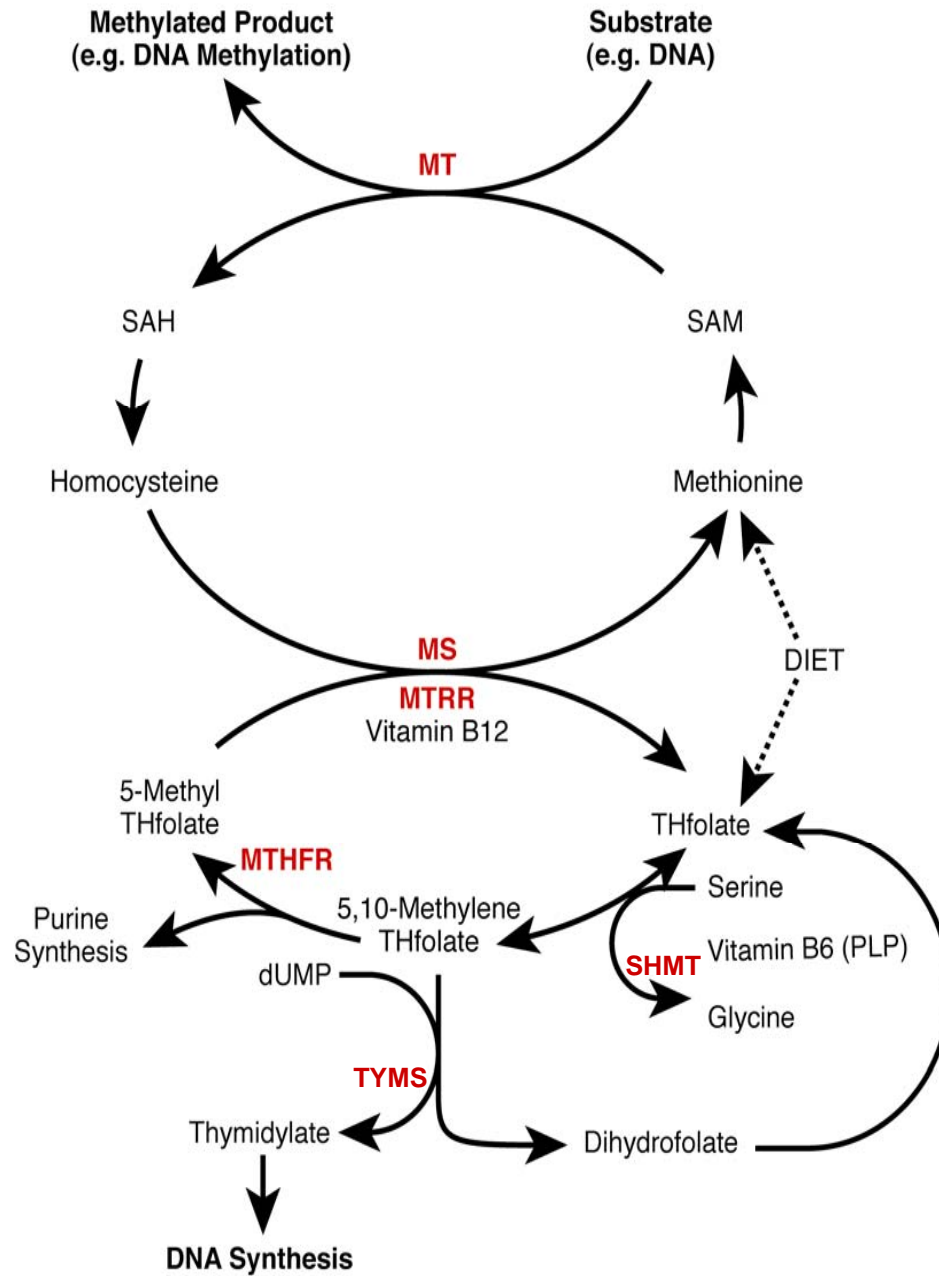
Cigarette smoke

- **Polymorphisms in genes involved in folate metabolism associated with cancer**

MTHFR (677TT protective for colorectal cancer with low status ↑ risk)

MTRR; MTR (MS); TYMS (TS); CSB

(Sharp and Little, AJE, March 1, 2004)



Hazard Ratio (HR) for Pancreatic Cancer of Dietary Methyl-Group Nutrients

Nutrient	Quintile 1	Quintile 3	Quintile 5	P-trend
Folate, mg/d	≤ 280	311-338	>373	
HR	1.00	0.59 (0.36-0.96)	0.52 (0.31-0.87)	0.05
B6, mg/d	≤ 2.09	2.34-2.54	> 2.81	
HR ^a	1.00	1.25 (0.76-2.05)	1.32 (0.75-2.30)	0.28
B12, $\mu\text{g/d}$	≤ 7.57	9.27-11.08	> 13.68	
HR ^a	1.00	1.21 (0.77-1.90)	0.88 (0.53-1.48)	0.91
Methionine, g/d	≤ 1.7	1.9-2.1	> 2.3	
HR ^a	1.00	1.30 (0.81-2.09)	1.00 (0.60-1.68)	0.88

Stolzenberg-Solomon,
AJE, 2001

Oxidative stress: disturbance in the pro-oxidant/antioxidant status in favour of the former

Reactive oxygen species attack various substrates including lipids, nucleic acids and proteins:

oxidatively modified LDL have been implicated in cardiovascular disease

oxidatively modified DNA in carcinogenesis

oxidatively modified protein in cataract

Biomarkers of exposure for antioxidant nutrients:

Beta-carotene and other carotenoids

Alpha-tocopherol/vitamin E

Vitamin C

Selenium

Biomarkers of oxidative stress status:

1. Biomarkers of lipid peroxidation:

breath hydrocarbons

LDL resistance to oxidation

F₂ isoprostanes

- 2. Biomarkers of DNA oxidation:**
 - 8-OHdG (see below; artifacts)**
 - COMET assay**

- 3. Biomarkers of protein oxidation:**
 - protein carbonyls**

**INTERVENTION STUDY ON FRUIT AND
VEGETABLE INTAKE AND LEVELS OF 8-
OHdG IN URINE
(THOMPSON ET AL, CARCINOGENESIS 20:
2261, 1999)**

BACKGROUND:

**A NUMBER OF INVESTIGATIONS HAVE
CAST DOUBT ON THE HYPOTHESIS THAT
ANTIOXIDANTS ACCOUNT FOR THE
PROTECTIVE EFFECT OF FRUIT AND
VEGETABLES**

**DNA OXIDATION WAS NOT MODIFIED BY
ANTIOXIDANT ADMINISTRATION
(PRIEME ET AL, 1997; COLLINS ET AL,
1998; VAN POPPEL ET AL, 1995)**

**THOMPSON ET AL, CARCINOGENESIS 20: 2261, 1999:
INTERVENTION BASED ON A “CUISINE” TAILORED TO
INCREASE THE INTAKE OF FRUIT AND VEGETABLES**

	PRE-INTERVENTION	POST-INTERVENTION	
SERVINGS/DAY	4.7	12.0	+155%
8-OHdG LYMPHOCYTES	8.6+-1.6 se	5.8+-1.0 se	-32%
8-OHdG URINE median	48.3+-18.1 se 27.1	20.8+-1.9 se 18.9	-57%

Interpretation

- in spite of large inter- and intra-individual variability, some intermediate markers such as DNA adducts and 8-OHdG are clearly influenced by dietary habits
- fruit and vegetables are likely to protect from cancer through an “effect modification” mechanism (e.g. by inducing metabolizing enzymes or DNA repair)

THE END