Biomarkers of exposure in food safety risk assessment

Rudolf Kaaks

German Cancer Research Center

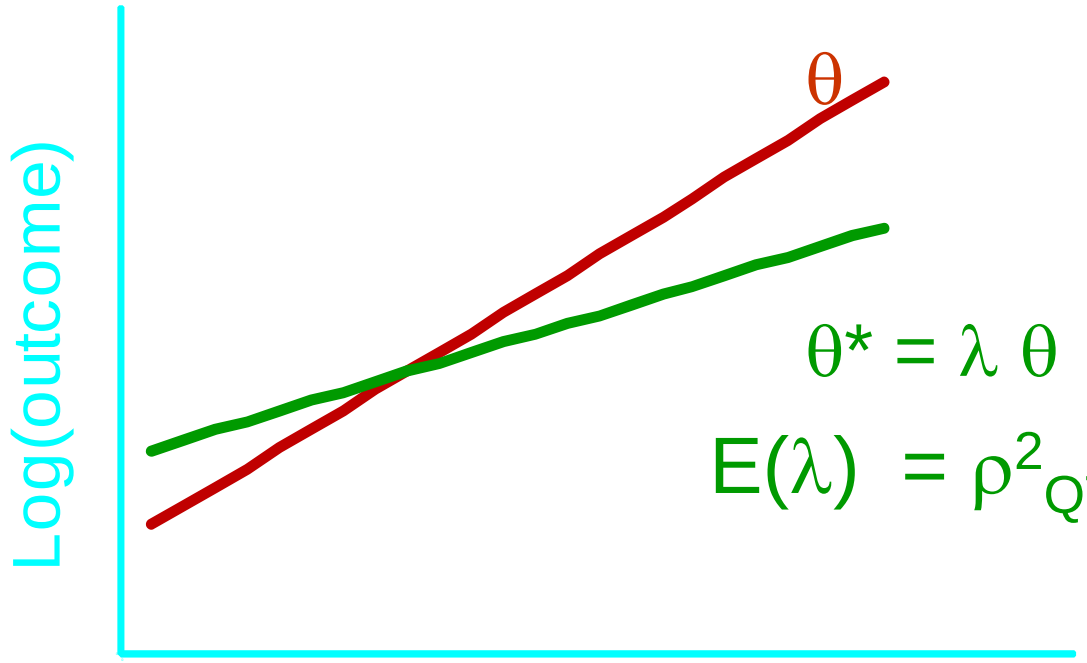
In nutritional epidemiology, dietary assessment errors (food frequency questionnaires) cause:

a. Attenuation & scaling biases in relative risk estimates

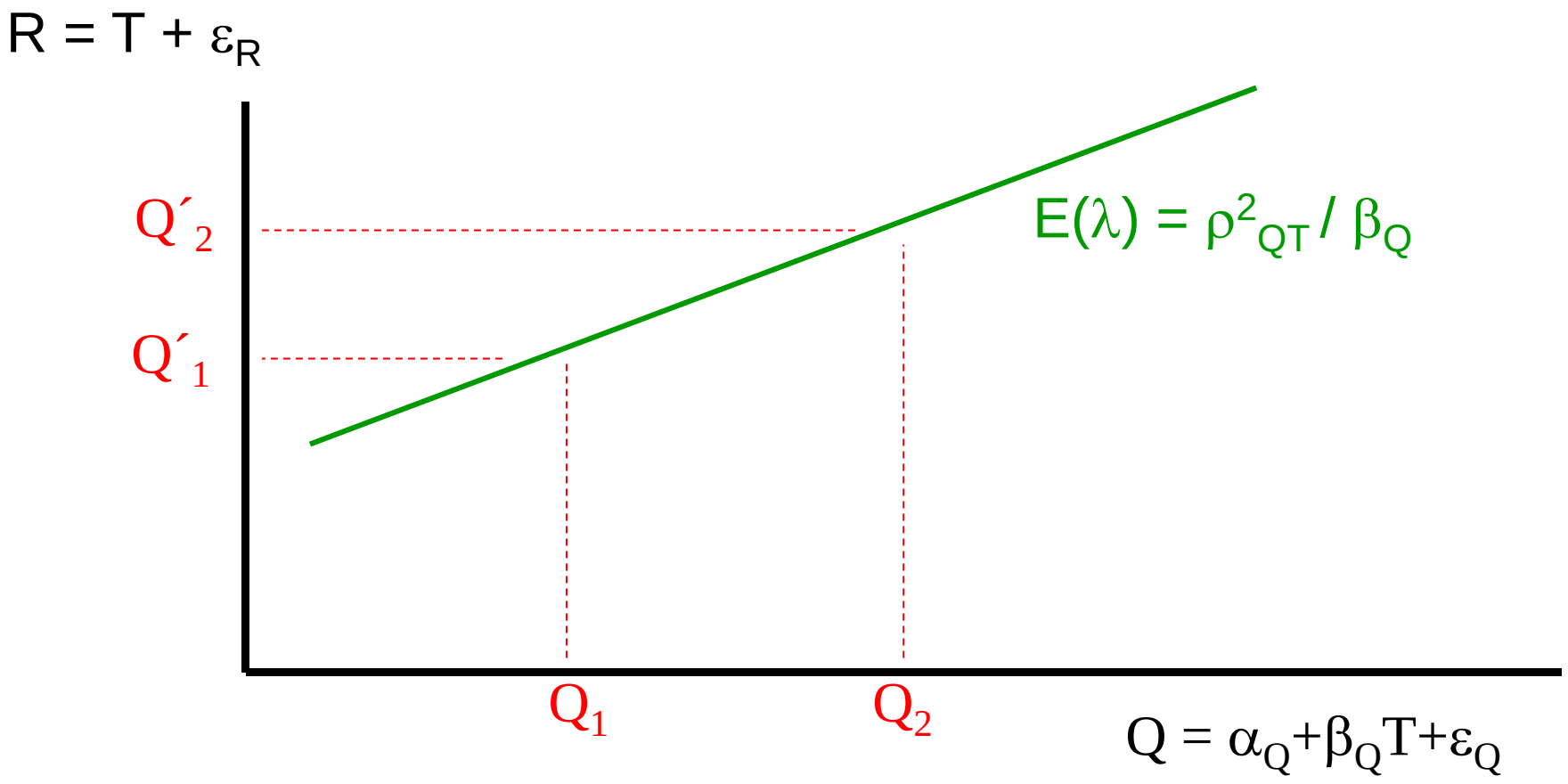
b. Biases in the estimated population distribution of intake levels

- Sub-studies are needed to assess accuracy of dietary exposure measurements, and to correct for biases due to error.

Regression bias, resulting from measurement errors:

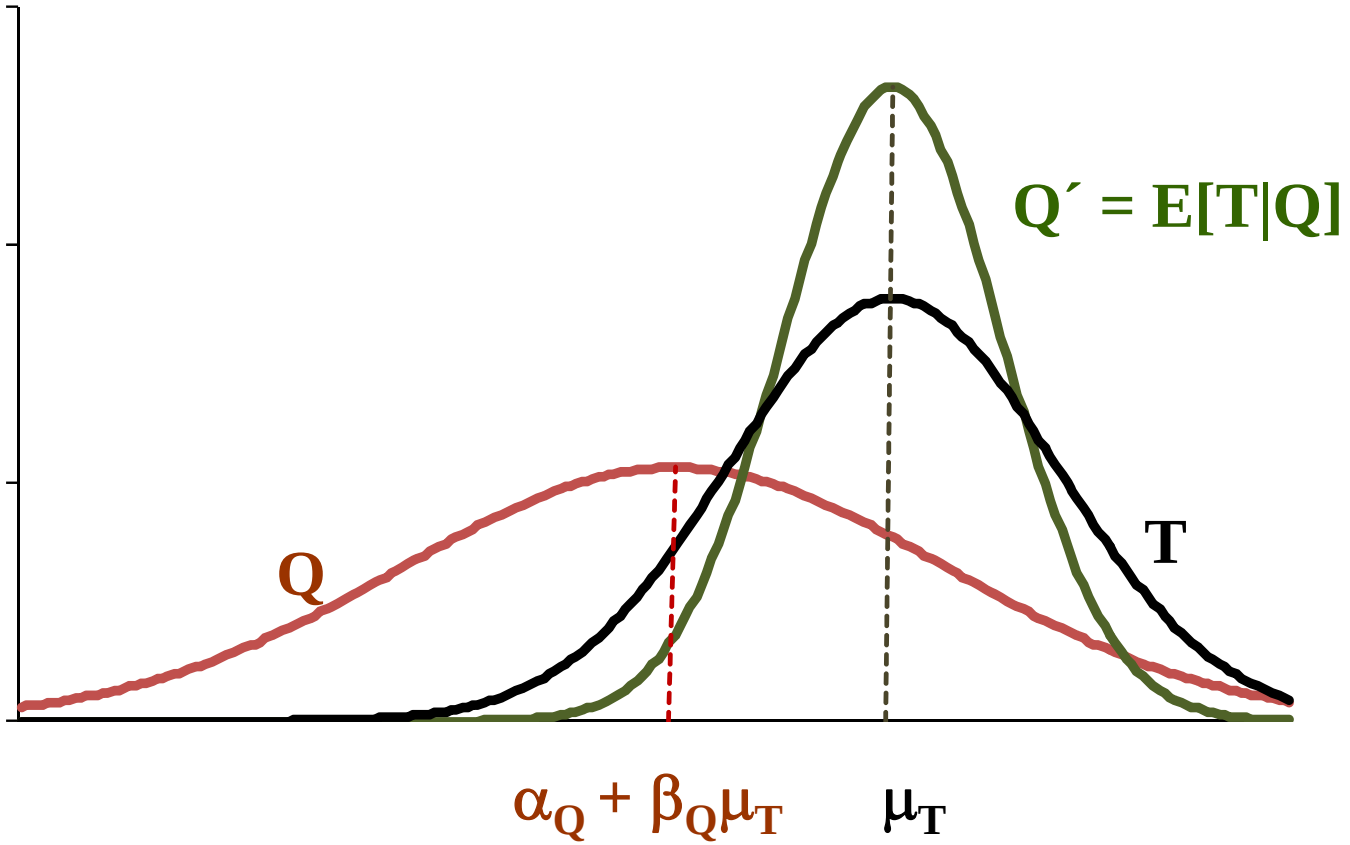


Estimation of λ_Q (calibration approach):



(Rosner et al, Stat Med 1989)

True, measured and predicted (calibrated) exposure distributions



Calibration and validation studies are designed to estimate, respectively:

- the calibration factor, λ
- the "validity coefficient", ρ_{QT}

Classical approach / model assumptions:

$$Q = \alpha_Q + \beta_Q T + \varepsilon_Q$$

$$R_i = T_i + \varepsilon_R \quad (i = 1, \ldots, p)$$

$$E(\varepsilon_Q | T) = E(\varepsilon_{Ri} | T) = 0$$

$$Cov(\varepsilon_Q, \varepsilon_{Ri}) = Cov(\varepsilon_{Ri}, \varepsilon_{Rj}) = 0$$

Estimation of ρ_{QT} (validation): classical approach

Comparison of Q with $R_1, \ldots, R_p$.

- computation of $\rho_{Q\bar{R}}$
- adjust for attenuation due to within-person variability in R.

- **<u>Problem / Question</u>**:
  Are ε_Q and ε_{Ri} , ε_{Ri} and ε_{Rj} really uncorrelated ?

Biochemical markers of diet

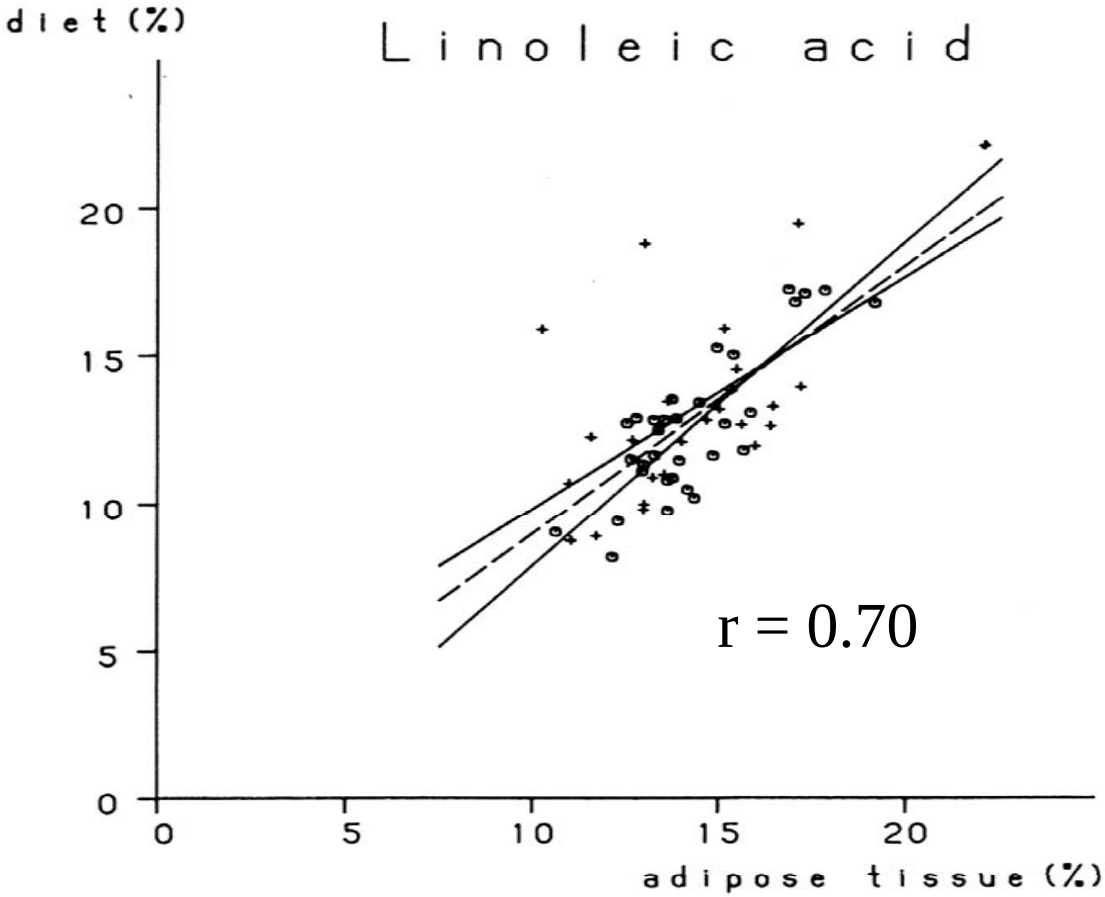
- for some aspects of diet may provide more accurate measurements than traditional assessment methods
- can be considered “objective” measurements; random errors can be assumed uncorrelated with those of Q and R
- can be statistically combined with measurements Q, R (e.g., in a structural equation model)
- may provide information about substances not included in food composition tables
- based either on concentration, or on recovery

Markers based on concentration

- concentrations of specific molecules in biological fluids, specific tissues or cells, lipoproteins, cell membranes, DNA, or specific proteins
- no time dimension
- variable quantitative relationships with dietary intake levels (interference of other, background factors)

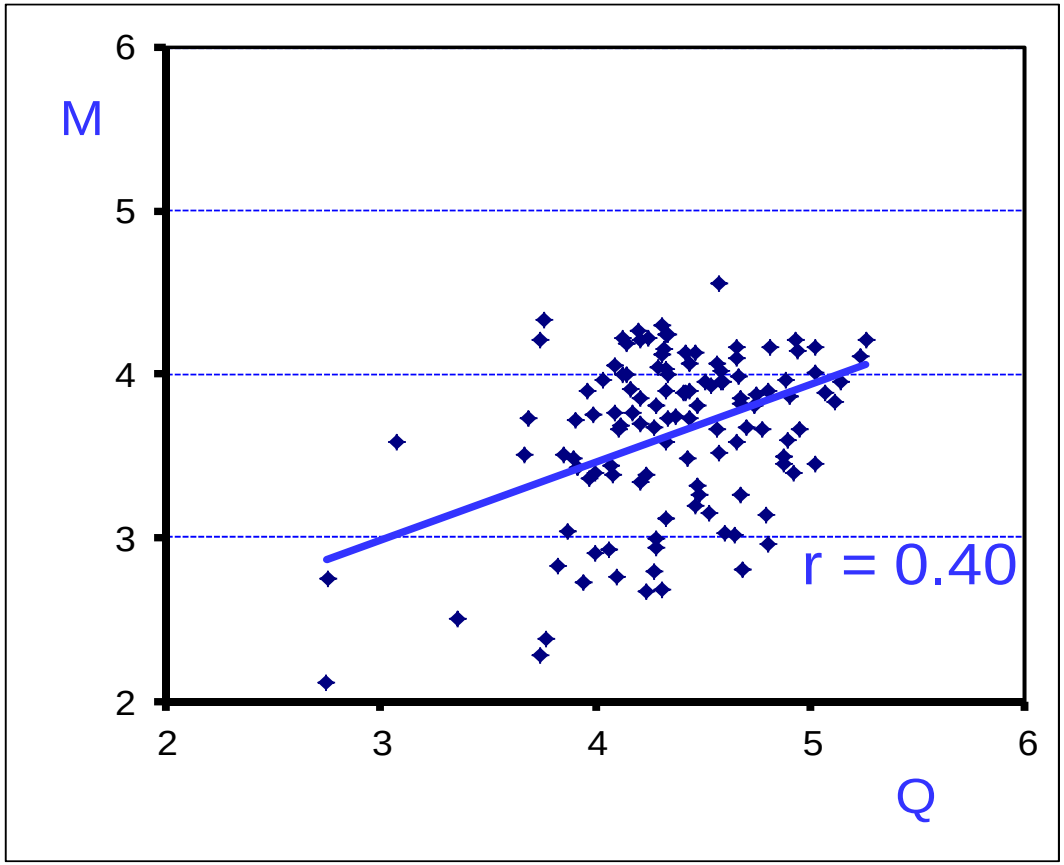
Kaaks et al., Public Health Nutr 2002

Linoleic acid in adipose tissue biopsies in relation to dietary intake



van Staveren, et al. 1986. Am J Epidem, 131, 987-94.

Relationship between plasma vitamin C (M) and questionnaire assessments (Q) of vitamin C in a Swedish validation study (N=122).



Callmer E et al., *J Intern Med* 1993; 233:53-57.

Use of markers based on concentration in a Structural Equation Model (a)

$$Q = \alpha_Q + \beta_Q T + \varepsilon_Q$$

$$R = T + \varepsilon_R$$

$$M_1 = \alpha_M + \beta_M T + \varepsilon_M$$

$$M_2 = \alpha_M + \beta_M T + \varepsilon_M$$

Assumptions:

$$Cov(\varepsilon_Q, \varepsilon_{Mj}) = Cov(\varepsilon_R, \varepsilon_{Mj}) = 0$$

$$E(\varepsilon_Q|T) = E(\varepsilon_R|T) = E(\varepsilon_{Mi}|T) = 0$$

Use of markers based on concentration in a Structural Equation Model (b)

	Q	**R**	**M_1**	**M_2**
Q	$\beta^2_Q \sigma^2_T \sigma^2_{\varepsilon Q}$			
R	$\sigma^2_T + \sigma^2_{\varepsilon R,\varepsilon Q}$	$\sigma^2_T + \sigma^2_{\varepsilon R}$		
M_1	$\beta_Q\beta_M \sigma^2_T$	$\beta_M \sigma^2_T$	$\beta^2_M \sigma^2_T + \sigma^2_{\varepsilon M1}$	
M_2	$\beta_Q\beta_M \sigma^2_T$	$\beta_M \sigma^2_T$	$\beta^2_M \sigma^2_T + \sigma^2_{\varepsilon M1,\varepsilon M2}$	$\beta^2_M \sigma^2_T + \sigma^2_{\varepsilon M2}$

8 unknown parameters: σ^2_T, β_Q, β_M, $\sigma^2_{\varepsilon Q}$, $\sigma^2_{\varepsilon R}$, $\sigma^2_{\varepsilon M}$, $\sigma^2_{\varepsilon R,\varepsilon Q}$, $\sigma^2_{\varepsilon M1,\varepsilon M2}$

Markers based on concentration can have many endogenous and exogenous determinants:

- absorption
- distribution over body compartments or tissues
- metabolism
- endogenous synthesis
- excretion

$\Rightarrow$ possible sources of correlation between errors of replicate measurements

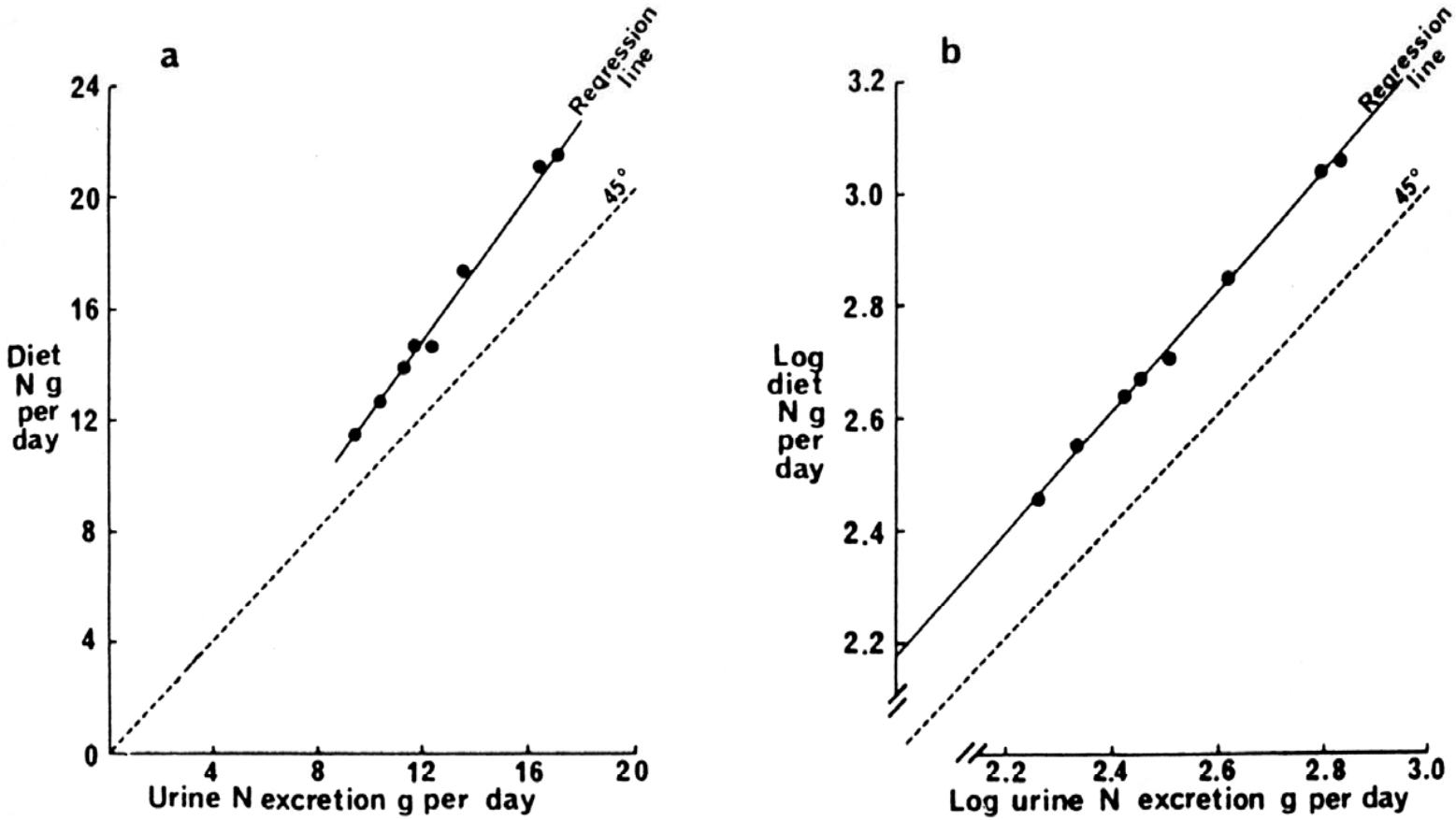
Kaaks et al., Public Health Nutr 2002

Markers based on recovery

- based on the balance between intake and output
- contain a time dimension
- translatable into absolute intake

Kaaks et al., Public Health Nutr 2002

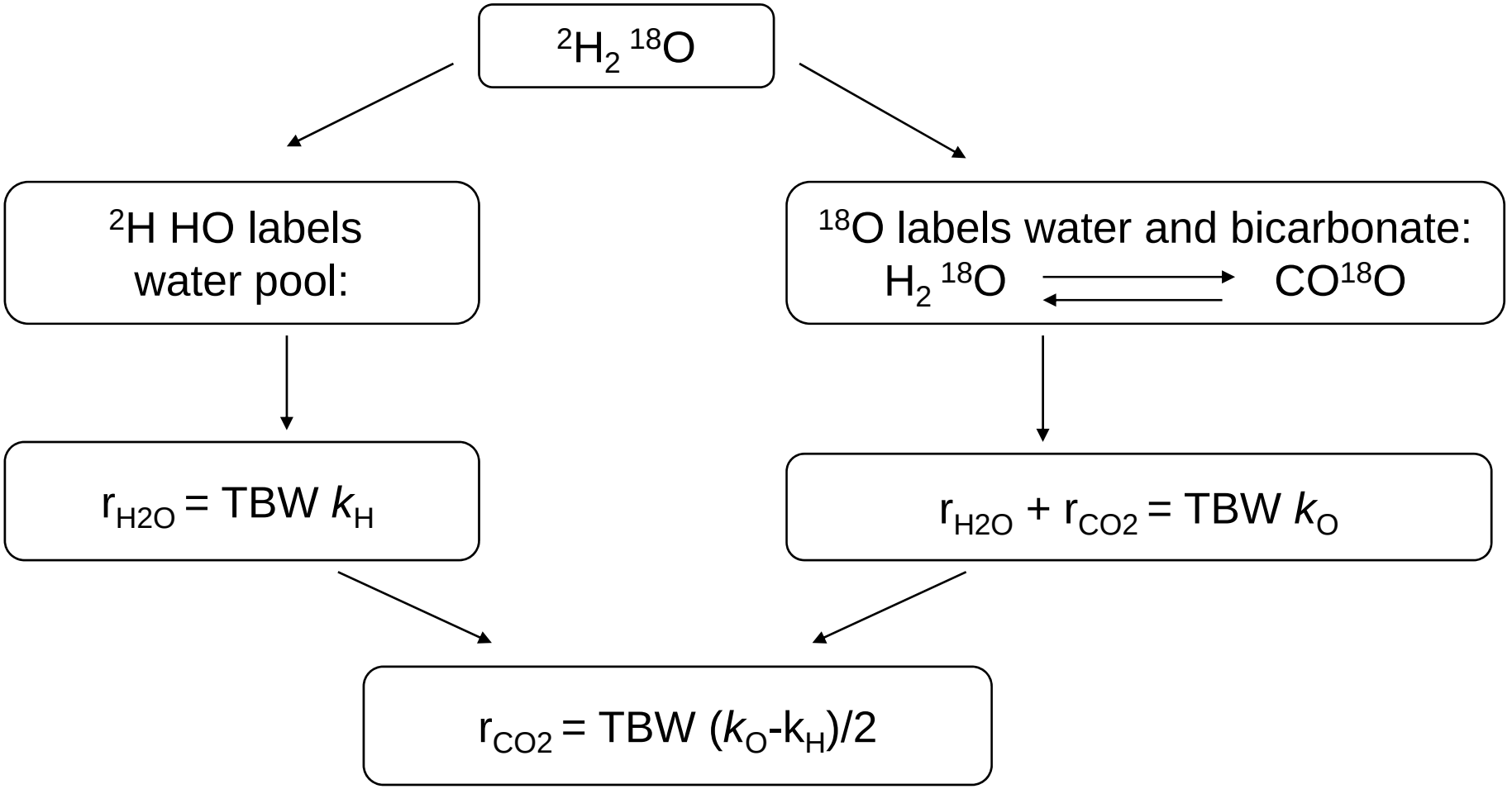
Markers based on recovery: Urinary nitrogen



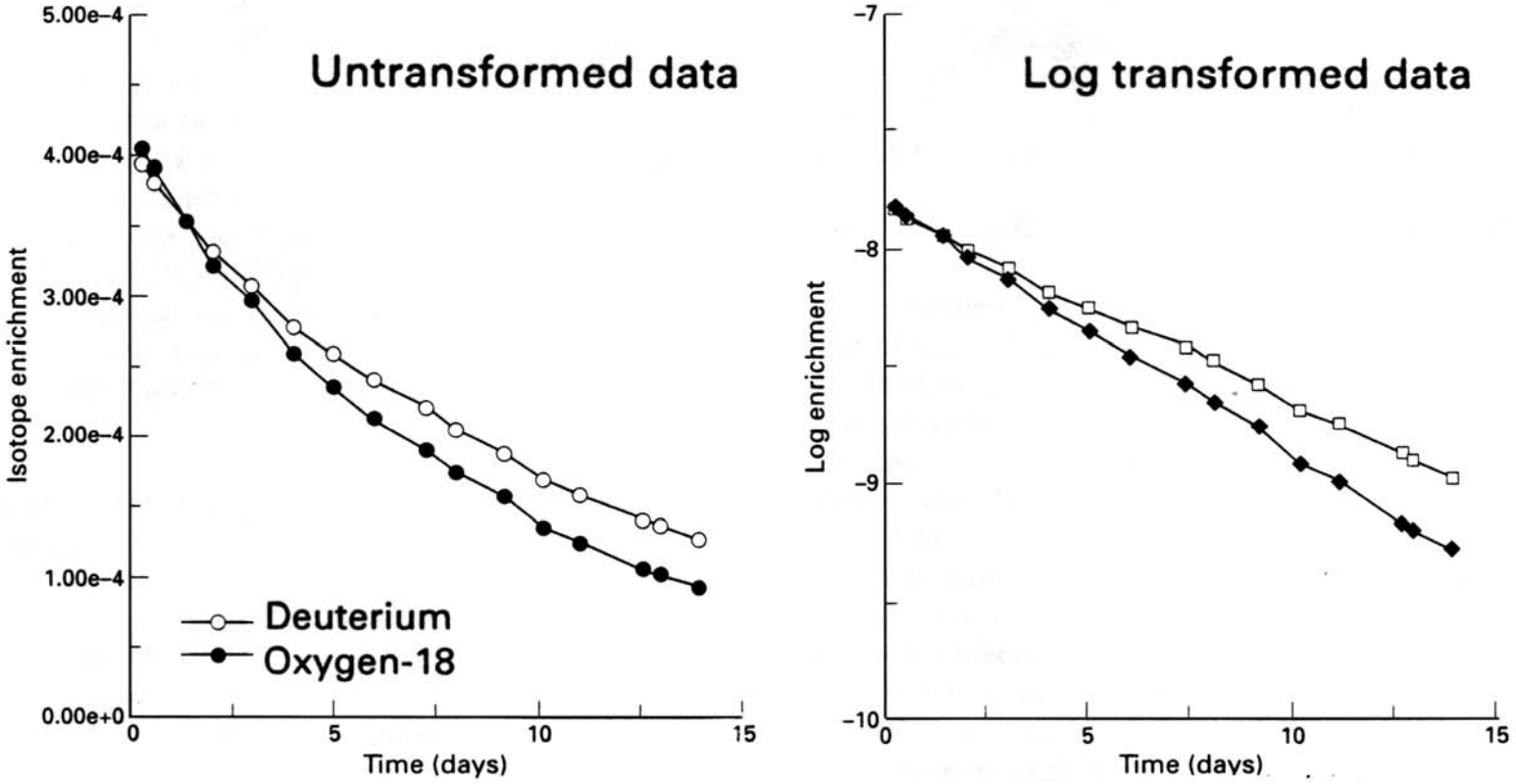
Relationship of 28 days dietary N intake and 20-28 days urine N

Bingham et al, Am J Clin Nutr 1985

Markers based on recovery: The doubly labeled water method



The doubly labeled water method



Isotope disappearance curves from a typical adult subject

Murgatroyd et al., Int J Obes 1993

Use of markers based on recovery in a Structural Equation Model (1) :

$$Q = \alpha_Q + \beta_Q T + \varepsilon_Q$$

$$R = T + \varepsilon_R \qquad \text{(alternatively: } R = \alpha_R + \beta_R T + \varepsilon_R\text{)}$$

$$M_1 = T + \varepsilon_M$$

$$M_2 = T + \varepsilon_M$$

Assumptions:

$$\mathrm{Cov}(\varepsilon_Q, \varepsilon_{Mi}) = \mathrm{Cov}(\varepsilon_R, \varepsilon_{Mi}) = 0$$

$$\mathrm{Cov}(\varepsilon_{Mi}, \varepsilon_{Mj}) = 0$$

$$E(\varepsilon_Q|T) = E(\varepsilon_R|T) = E(\varepsilon_{Mi}|T) = 0$$

Use of markers based on recovery in a Structural Equation Model (b)

	Q	R	M_1	M_2
Q	$\beta^2_Q \sigma^2_T \sigma^2_{\varepsilon Q}$			
R	$\beta_Q \sigma^2_T + \sigma^2_{\varepsilon Q, \varepsilon R}$	$\sigma^2_T + \sigma^2_{\varepsilon R}$		
M_1	$\beta_Q \sigma^2_T$	σ^2_T	$\sigma^2_T + \sigma^2_{\varepsilon M}$	
M_2	$\beta_Q \sigma^2_T$	σ^2_T	σ^2_T	$\sigma^2_T + \sigma^2_{\varepsilon M}$

6 unknown parameters: σ^2_T, β_Q, $\sigma^2_{\varepsilon Q}$, $\sigma^2_{\varepsilon R}$, $\sigma^2_{\varepsilon M}$, $\sigma^2_{\varepsilon R, \varepsilon Q}$

Conclusions (i)

- Markers based on concentration can provide only a correlate of diet, and hence must be combined with other measurements that can provide a reference scale.

- Problem: Random errors of replicate marker measurements based on concentration are likely to be correlated

Conclusions (ii)

- Validation models identifiable when using replicate recovery-based marker measurements as reference.
- Markers based on recovery are ideal (prove a reference scale, and have uncorrelated random errors on replication) but only few are available (N, K, DLW).
- Other study designs have potential problems:
  - Q *vs.* R_1 *vs.* R_2: $Cov(\varepsilon_Q,\varepsilon_{Ri}) \neq 0$, $Cov(\varepsilon_{R1},\varepsilon_{R2}) \neq 0$
  - Q *vs.* R *vs.* M: $Cov(\varepsilon_Q,\varepsilon_R) \neq 0$
  - Q *vs.* M_1 *vs.* M_2 (concentration): $Cov(\varepsilon_M,\varepsilon_M) \neq 0$
  - Q *vs.* M *vs.* P (instrumental variable)

Thank you for your attention

Conclusions (i)

- Validation:
  - Is the estimation of parameters in a postulated measurement error model;
  - Applies to sets of measurements obtained in a specific context; does not apply to methods.