Nutrition Epidemiology
Choosing and evaluating indicators of nutritional status

Laura Caulfield, PhD
Professor
Center for Human Nutrition
Indicators of nutritional status

• How do I decide what to measure?
• Did the researchers measure the right indicator of nutritional status?
• There are so many indicators….how do I choose?!
• How good are the tests I am using?
Factors to consider in choosing an indicator of nutritional status

What should you know?

1. How many indicators do you have to choose from?

“Vitamin A”

BIOMARKERS
- plasma or serum retinol concentration
- Breast milk retinol concentration

NUTRIENT DEPENDENT FUNCTIONS
- relative dose response
- RBP response test
- isotope dilution techniques

TRANSPORT
- retinol binding protein (RBP)

CLINICAL SYMPTOMS
- night blindness history
- clinical indicators (X2, X3, X1B)

FUNCTIONAL TESTS
- vision restoration time
- rapid dark adaptation test (RDAT)
- conjunctival impression cytology (CIC, ICT)
Factors to consider in choosing an indicator of nutritional status

2. Purpose: What do you want to know, or what do you want to use it for?

a. Aspect of nutritional status measured
   - Body stores
   - Body pools (metabolically active stores)
   - Concentration in target tissue of interest
   - Transport
   - Nutrient dependent function (observational, experimental)
   - Excretion – urine or fecal (metabolite, regulation, recovery/balance)
   - Dietary intake
New terminology for dietary biomarkers

• Recovery: when metabolic balance governs intake and excretion such that they are highly correlated
  – Urinary nitrogen for protein
  – Urinary potassium and sodium
  – Doubly labelled water for energy expenditure

• Predictive: like recovery but it is not complete, but a dose-response relation is known
  – Urinary sucrose
  – Urinary fructose
New terminology for dietary biomarkers

- **Concentration**
  - Concentration biomarkers are traditional indicators of nutrient status, they are more variably associated with dietary intake

- **Replacement**
  - These are biomarkers for intake used instead of dietary measures, or when dietary measure does not exist
    - Pesticides, aflatoxin
    - Phytoestrogens, alkylresorcinols
    - Sodium/salt
New terminology for dietary biomarkers

• Metabonomics
  – Refers to a wide range of downstream metabolites of foods or dietary patterns
  – Metabolites result from diet, metabolism, microbiome influences and genetics
  – Exploratory/discovery
  – Applications primarily to cancer; this is where the research is moving quickly
Factors to consider in choosing an indicator of nutritional status

2. Purpose: What do you want to know, or what do you want to use it for?

b. Level of nutriture of interest
   - Deficiency (metabolic or clinical)
   - Variation among healthy individuals achievable through dietary means
   - Variation among healthy individuals achievable with supplements
   - Toxicity
Vitamin A indicators across the range of intakes/status
Vitamin A indicators across the range of intakes/status: relate to previous slide
Stages of Iron Deficiency and Development of Anemia

Illustrates a variety of iron status indicators and shows that anemia is a severe condition of iron deficiency.
Clinical studies relating dietary change and indicator changes: Vitamin C

FIG 3. Plasma ascorbic acid during periods of varied ascorbic acid intake for 11 adult males.
Clinical studies relating dietary change and indicator changes: Vitamin C

FIG 4. Red blood cell ascorbic acid during periods of varied ascorbic acid intake for 11 adult males.
Clinical studies relating dietary change and indicator changes: Vitamin C

FIG 2. Mean ± SD of plasma and leukocyte ascorbic acid for 11 and 8 adult males, respectively, who received varied ascorbate intakes (straight lines are drawn arbitrarily between leukocyte AA points and are not meant to imply a linear change in leukocyte AA levels between the measured times).
Clinical studies relating dietary change and indicator changes: Vitamin C

FIG 5. Urinary ascorbic acid (mean ± SD) during periods of varied ascorbic acid intake for 11 adult males.
Example of feeding studies: urinary 1-methylhistidinone for each volunteer by diet type (amount of red meat over 14 days).

2. Purpose: What do you want to know, or what do you want to use it for?

c. Epidemiologic purpose
   - Screening individuals/diagnosis
   - Prevalence estimation
   - Monitoring changes
   - Evaluating change
   - Etiologic investigation
   - Calibration or validation of dietary measures

Examples:
   a. Change in nutriture
   b. Compliance with treatment
   c. Causal pathway
Factors to consider in choosing an indicator of nutritional status

2. Purpose: What do you want to know, or what do you want to use it for?
   d. Time frame
      – Prospective (very few exist)
      – Concurrent
      – Retrospective (short, medium, long term)
      – Mixture of concurrent and retrospective

What is the relevant exposure period for your study?
Factors to consider in choosing an indicator of nutritional status

3. What does variation in the indicator imply with respect to…?

   a. Functional outcomes of interest

Functional outcomes are outcomes that are quantifiable as GOOD or BAD in terms that are widely accepted, and fall into 3 categories:

<table>
<thead>
<tr>
<th>Performance:</th>
<th>physical work performance, cognitive or behavioral function, fecundity/fertility, able to contribute to society</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health:</td>
<td>morbidity</td>
</tr>
<tr>
<td>Survival:</td>
<td>mortality</td>
</tr>
</tbody>
</table>
BMI became important because of its association with mortality.
FIG 2. Incidence of NIDDM according to BMI in a population of US women 30–64 y of age. (Adapted from reference 10.)
Figure 3. The prevalence of diabetes according to relative weight and body fat distribution in 20,325 women. Data from National Survey of Women in TOPS Club, Inc, in 1969.
From: Waist Circumference and All-Cause Mortality in a Large US Cohort


Figure Legend:

All-cause mortality by waist circumference in the Cancer Prevention Study II Nutrition Cohort, 1997-2006. Models were adjusted for age, race, educational level, marital status, smoking status, alcohol use, height, and physical activity. Models for women were also adjusted for hormone therapy.
FIG 3. Relative risk of nonfatal myocardial infarction and fatal coronary heart disease (combined), according to categories of BMI in a cohort of US women who were 30–55 y of age in 1976 and were followed for 8 y. (Adapted from reference 11.)

FIG 4. Relative risk of hypertension in a population of US women followed for 4 y. Women with BMI 23–25.9 kg/m² are the reference. Relative risks are adjusted for age and alcohol intake. (Adapted from reference 19.)
ASSOCIATION OF GIRTH MEASUREMENTS WITH DISEASE

PERCENTAGE WITH HYPERTENSION

WAIST TO HIP RATIO

SEVERELY OBESE

MODERATELY OBSESE

NON-OBSESE

Figure 2. The prevalence of hypertension according to relative weight and body fat distribution in 20,380 men aged 40 to 59. Data from National Survey of Men in TOPS Club, Inc., in 1969.
Fig. 2. Relation between body mass index and percentage of men (n=199) taking time off work through illness in the month before the survey (Bangladesh, 1993)

Functional outcomes are...

i. Usually assessed with methods from clinical epidemiology (e.g., sensitivity /specificity analysis; risk estimates)

ii. Useful for estimating the health consequences/economic costs associated with the nutritional problem or outcomes of interest

iii. Useful for influencing policy-makers
Factors to consider in choosing an indicator of nutritional status

4. What does variation in the indicator imply with respect to…?

c. Actual change resulting from treatment/intervention

   This is the responsiveness of the indicator

   \[
   \text{Responsiveness} = \frac{\text{average change}}{\text{SD of change}} \quad \text{Responsiveness} = \frac{\text{difference in means}}{\text{SD}_{\text{pooled}}}
   \]

   assesses how big the treatment effect is relative to the variability of the magnitude of the treatment effect in a population
Factors to consider in choosing an indicator of nutritional status

Characteristics of a responsive indicator:

1. Must be part of the causal chain, that is, treatment must cause the indicator to change

2. Useful for screening for treatment/interventions

3. Useful for estimating prevalence for monitoring/evaluating change

4. Useful for analysis of congruence or for examining etiology
It is best to have multiple indicators of what you are interested in: congruence

- Measurement of multiple indicators that assess various dimensions of underlying nutriture of interest
- One can hypothesize a priori the amount and/or direction of change in each indicator depending on the underlying biology/mechanism/conceptual framework (one can also hypothesize the absence of change!)
- Supportive evidence from multiple indicators (congruency of information) strengthens one’s case regarding etiology/pathway/mechanism…. 
Can values of this indicator accurately identify those who are malnourished? Why or why not?
How about this indicator? Why or why not?
How about this indicator? Why or why not?
Here the values of the indicator (test result) are higher in those with disease. That distribution is the Sensitivity (Se) distribution.

The distribution of values of the indicator for those without disease. That is the Specificity (Sp) distribution.

Varying the cut-point or criterion changes the Se and Sp of the indicator for diagnosing disease presence or absence.
<table>
<thead>
<tr>
<th>Diagnosis of:</th>
<th>Malnourished</th>
<th>Not malnourished</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malnourished</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Not malnourished</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>
The famous truth table

<table>
<thead>
<tr>
<th>Diagnosis of:</th>
<th>Malnourished</th>
<th>Not malnourished</th>
<th>TOTALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malnourished</td>
<td>TP</td>
<td>FP</td>
<td>TP + FP</td>
</tr>
<tr>
<td>Not malnourished</td>
<td>FN</td>
<td>TN</td>
<td>FN + TN</td>
</tr>
<tr>
<td>TOTALS</td>
<td>TP + FN</td>
<td>FP + TN</td>
<td></td>
</tr>
</tbody>
</table>
Some definitions

Sensitivity is the proportion of the malnourished who were correctly diagnosed as malnourished
i.e., Sens = TP / (TP + FN)

Specificity is the proportion of non-malnourished who were correctly diagnosed as non-malnourished
i.e., Spec = TN / (TN + FP)

Positive predictive value is the proportion of those who are diagnosed as malnourished who are truly malnourished
i.e., V+ = TP / (TP + FP)

Negative predictive value is the proportion of those who are diagnosed as non-malnourished who are truly non-malnourished
i.e., V- = TN / (TN + FN)
Evaluate the performance of and indicator or compare across indicators

- Sensitivity and Specificity

- Values of Se and Sp range from 0-1 or from 0-100 (%) depending on the scale chosen

- What is the best indicator?
  - Maximize sum of sensitivity and specificity (max value is 2 or 200 (%)) (Youden index)
  - Maximize sensitivity or specificity
  - Maximize V+ or V-
  - “Choose highest sensitivity at high specificity”
Receiver operating characteristics curves (ROC) illustrate the accuracy of the indicator across cut-off points, and across indicators.
Typical ROC Curve published today: Diagnostic test comparison of HbA1c and fasting plasma glucose (FPG) in diagnosing NIDDM (using 2-hr OGTT)

Youden Index: greatest vertical distance from line of indifference – best cut point

Diagnostic tests have high Se and Sp (low false positive rate)
Typical ROC Curve published today: Se (true positive rate) versus 1-Sp (false positive rate)

Many indicators commonly discussed have some diagnostic value (indicate risk) but do not diagnose the condition.

Here you can see trade-offs in Se and Sp are fairly monotonic.

**Fig. 2.** Receiver operating characteristic curves for identifying metabolic syndrome risk in women. BMI, body mass index; WC, waist circumference; WHtR, waist to hip ratio; WHtR, waist to height ratio.
Compare the performance of competing indicators of nutritional status

- Statistical tests comparing the Se and Sp distributions
  - Area under the curve (A_{2z}) (AUC)
  - Standardized distance (d’ or “d prime”)
  - Index of detectability (d_a or “da”)

Statistical measures of performance

• $d^\prime$ ("d prime") – Normalized distance

$$d^\prime = \frac{|X_d - X_n|}{s}$$

- $X_s$ are means; $d$ is diseased; $n$ is non-diseased
- Values typically vary between 0 – 4
- Assumes variance are equal
- Operationally defined equality: ratio of variances between 0.8 – 1.2

• $d_a$ ("da") – Index of detectability

$$d_a = \frac{|X_d - X_n|}{\sqrt{0.5(S_d^2 + S_n^2)}}$$

- $X_s$ are means
- Values typically vary between 0 - 4
To statistically compare the performance of two indicators

$Z\text{-statistic (}d'\text{)} = \frac{d'(X) - d'(Y)}{\sqrt{\text{Var } d'(X) + \text{Var } d'(Y)}}$

Note: Can also account for covariance when comparing indicators measured on the same subject
Example of methods

• The next slides come from the paper by Ruel et al on the reading list.....

• The goal of the paper was to try to identify the best anthropometric indicator during infancy for identifying children who would experience growth faltering leading to stunting by age 3 y
FIGURE 2 Receiver operating characteristics curves comparing the performance of four screening indicators to predict stunting at 3 y of age among a sample of Guatemalan children. LTA = length-for-age; WTA = weight-for-age.
Table shows calculation of da, variance (da) and compares indicators

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Stunted (LTA ≤-2 sd) [n = 250]</th>
<th>Nonstunted (LTA ≥-2 sd) [n = 150]</th>
<th>da (A)²</th>
<th>Vda (A)³</th>
<th>t test⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTA Z-scores (6 mo)</td>
<td>-2.08 0.85</td>
<td>-1.07 0.68</td>
<td>1.30</td>
<td>0.012</td>
<td>13.07</td>
</tr>
<tr>
<td>LTA Z-scores (3 mo)</td>
<td>-1.61 0.86</td>
<td>-0.76 0.69</td>
<td>1.09</td>
<td>0.012</td>
<td>10.85</td>
</tr>
<tr>
<td>WTA Z-scores (6 mo)</td>
<td>-1.24 0.94</td>
<td>-0.46 0.81</td>
<td>0.89</td>
<td>0.011</td>
<td>8.77</td>
</tr>
<tr>
<td>WTA Z-scores (3 mo)</td>
<td>-0.68 0.86</td>
<td>-0.08 0.74</td>
<td>0.74</td>
<td>0.011</td>
<td>7.38</td>
</tr>
<tr>
<td>Head circumference (3 mo)</td>
<td>38.55 1.42</td>
<td>39.25 1.17</td>
<td>0.54</td>
<td>0.011</td>
<td>5.34</td>
</tr>
<tr>
<td>Head circumference (6 mo)</td>
<td>41.20 1.61</td>
<td>41.95 1.20</td>
<td>0.53</td>
<td>0.010</td>
<td>7.65</td>
</tr>
<tr>
<td>Arm circumference (6 mo)</td>
<td>12.71 1.10</td>
<td>13.20 0.97</td>
<td>0.48</td>
<td>0.011</td>
<td>4.65</td>
</tr>
<tr>
<td>Weight velocity (3-6 mo)</td>
<td>1.27 0.48</td>
<td>1.51 0.51</td>
<td>0.45</td>
<td>0.011</td>
<td>4.66</td>
</tr>
<tr>
<td>Length velocity (3-6 mo)</td>
<td>5.50 1.63</td>
<td>5.71 1.34</td>
<td>0.35</td>
<td>0.010</td>
<td>3.39</td>
</tr>
<tr>
<td>Arm circumference (3 mo)</td>
<td>12.03 1.12</td>
<td>12.31 0.97</td>
<td>0.26</td>
<td>0.010</td>
<td>2.63</td>
</tr>
<tr>
<td>WTLT Z-scores (3 mo)</td>
<td>0.81 0.93</td>
<td>0.81 0.78</td>
<td>0.13</td>
<td>0.010</td>
<td>0.00</td>
</tr>
<tr>
<td>WTLT Z-scores (6 mo)</td>
<td>0.53 0.95</td>
<td>0.46 0.95</td>
<td>0.09</td>
<td>0.019</td>
<td>0.71</td>
</tr>
</tbody>
</table>

1 Abbreviations used: LTA, length-for-age; WTA, weight-for-age; WTLT, weight-for-length.

2 $da(A) = \overline{A} - \overline{A_0}/\sqrt{[1/2 \{s_1^2(A) + s_0^2(A)\}]},$ where $1 =$ nonstunted and $0 =$ stunted [Brownie et al. 1986].

3 $Vda(A) = da(A)^2 \{1/\delta_A^2 \{1/n_0 + b_A^2/n_1\} + [1/2(1 + b_A^2)^2 \{1/n_0 + b_A^4/n_1\}]\},$ where $\delta_A = \overline{A} - A_0/s_0(A)$ and $b_A = s_1(A)/s_0(A)$ [Brownie et al. 1986].

4 $t$ test $= \overline{A} - \overline{A_0}/\sqrt{(s_1^2/n_1) + (s_0^2/n_0)}$ [Plabicht et al. 1982].
Table presents statistical comparisons of pairs of indicators

**TABLE 2**

*Comparison of indicators for screening Guatemalan children at risk of being stunted at 36 mo of age*

<table>
<thead>
<tr>
<th>Indicators compared</th>
<th>$Z_{da} (\Lambda - B)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTA$^3$ at 6 mo</td>
<td>LTA at 3 mo</td>
</tr>
<tr>
<td>WTA at 6 mo</td>
<td>WTA at 3 mo</td>
</tr>
<tr>
<td>LTA at 3 mo</td>
<td>WTA at 3 mo</td>
</tr>
<tr>
<td>LTA at 6 mo</td>
<td>WTA at 6 mo</td>
</tr>
<tr>
<td>LTA at 3 mo</td>
<td>WTA at 6 mo</td>
</tr>
<tr>
<td>WTA at 3 mo</td>
<td>Weight velocity [3–6 mo]</td>
</tr>
<tr>
<td>WTA at 6 mo</td>
<td>Weight velocity [3–6 mo]</td>
</tr>
<tr>
<td>LTA at 3 mo</td>
<td>Length velocity [3–6 mo]</td>
</tr>
<tr>
<td>LTA at 6 mo</td>
<td>Length velocity [3–6 mo]</td>
</tr>
</tbody>
</table>

1. $Z_{da} (\Lambda - B) = \frac{dA(A) - dA(B)}{\sqrt{\text{var} \, dA(A) + \text{var} \, dA(B) - 2 \text{cov} \, [dA(A), \, dA(B)]}}$ (Brownie et al. 1986).
2. Using the Bonferroni correction, $P$ values <0.05/9 (0.005) were considered significant (indicated by an asterisk).
3. Abbreviations used: LTA, length-for-age; WTA, weight-for-age.
Compare the performance of competing indicators of nutritional status

- Responsiveness of indicators can be compared/tested by utilizing an experimental design (supplementation trial, e.g.,) and examining change in indicators with change in underlying nutriture brought about with supplementation.
- Here the comparison of treatment to control (placebo) at follow-up provides the best measure of responsiveness.
- Compare responsiveness
  - Se/Sp analysis (define truth as response yes or no by some criteria)
  - ROC analysis
  - Compare distributions with d’ or “da”
Compare Responsiveness with Distance Functions

- Responsiveness = average change / SD of change
  Responsiveness = difference in means/SD_{pooled}

  assesses how big the treatment effect is relative to the variability of the magnitude of the treatment effect in a population

- d_a ("da") – Index of detectability

  \[ d_a = \frac{|X_d - X_n|}{\sqrt{0.5 \times (S_d^2 + S_n^2)}} \]

  assesses degree of separation or distance between two distributions

  Xs are means of treatment and control (placebo) at follow up (post treatment); Xs at baseline can be also evaluated to assess comparability