# What's Behind the Numbers? Statistical Sampling and Analytical Methodology

## Beltsville Human Nutrition Research Center

# What's Behind the Numbers? Statistical Sampling Analytical Methodology

Joanne Holden, M.Sc.

Nutrient Data Laboratory

James Harnly, Ph.D.

Food Composition Laboratory

Beltsville Human Nutrition Research Center

# USDA Collaborators

**Pamela Pehrsson, Ph. D.**

**David Haytowitz, M.Sc.**

**Nutrient Data Laboratory**

**Charles Perry, Ph. D.**

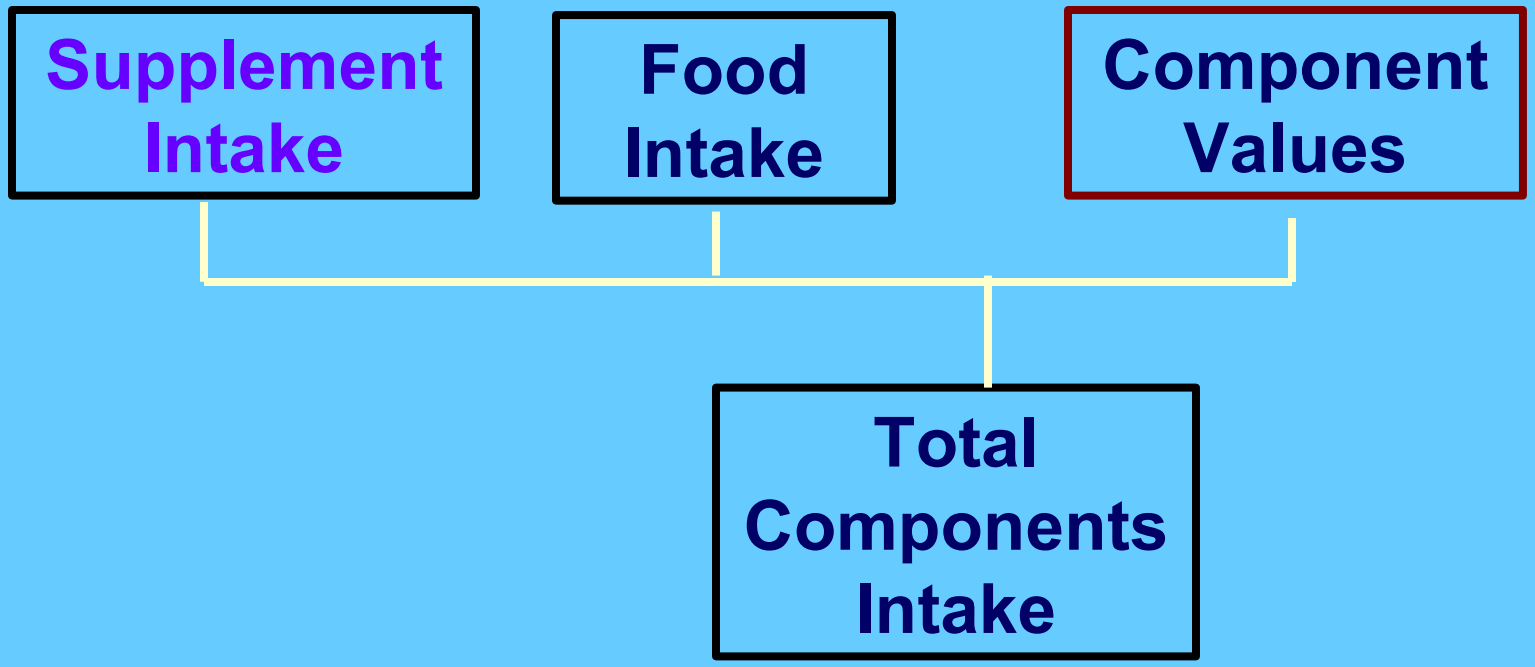
**National Agricultural Statistics Service**

*Acknowledge funding for NFNAP from USDA and NIH, FDA, and CDC

***A knowledge of the chemical composition of foods is the first essential in dietary treatment of disease or in any quantitative study of human nutrition.***

***-McCance & Widdowson, 1940***

# Intake Assessment of Dietary Components



# Components in foods may:

- Occur naturally (physiological function)
- Be generated in response to biological stress
- Be added to foods (fortification)
- Be the result of recipes or formulations (e.g. manufacturing)
- Be contaminants (direct or indirect)
- May accumulate due to feeding/fertilization

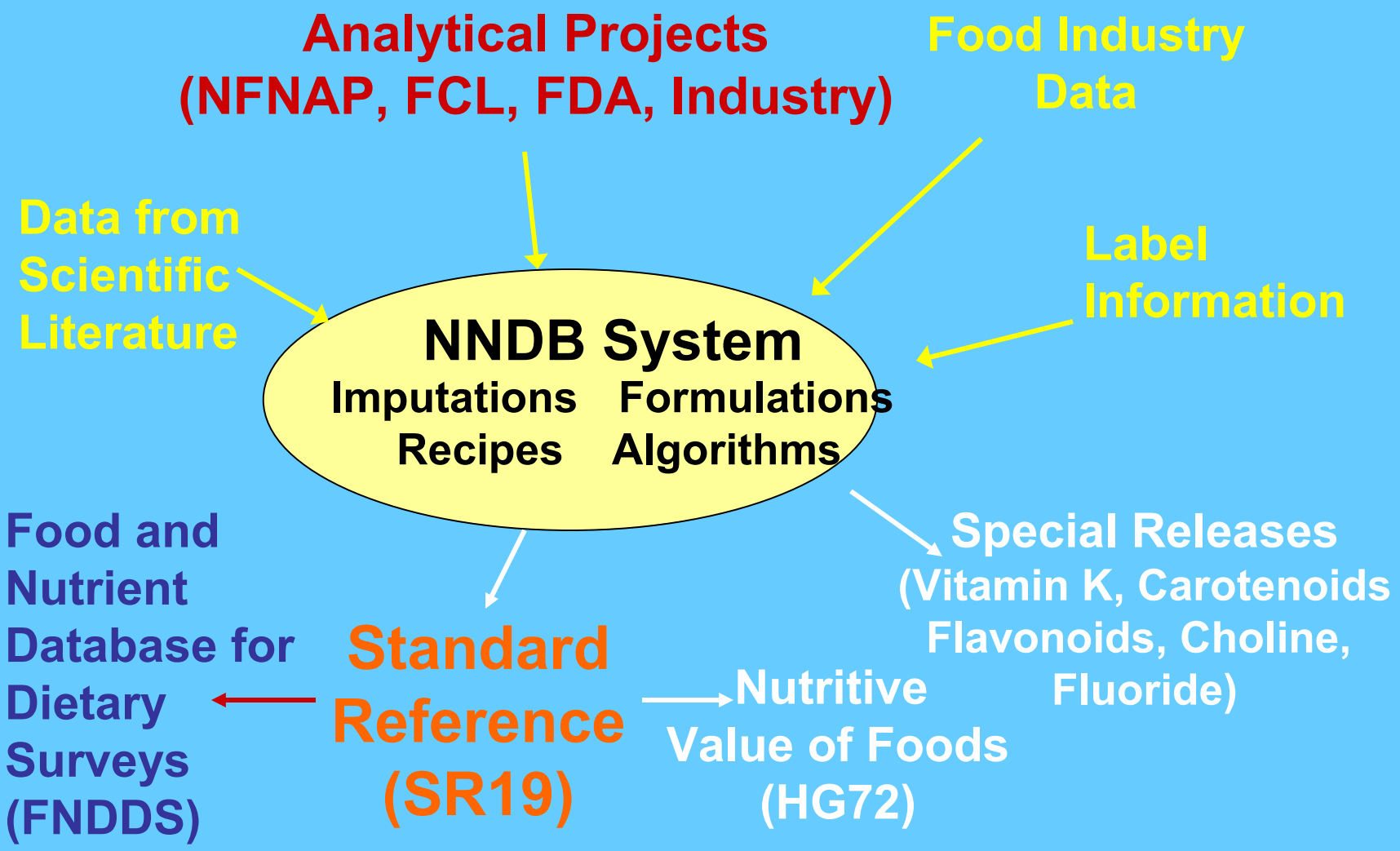
# Today's Objectives

- To introduce Quality Parameters for food composition data
- To describe the National Food and Nutrient Analysis Program (NFNAP)
- To describe the sampling strategy for NFNAP
- To discuss the importance of valid analytical methodology

# USDA's Nutrient Database for Standard Reference (SR19)

- Authoritative source of U.S.food composition data
- 7200 foods
- Values for up to 128 nutritional components
- Agricultural commodities, formulated foods
- Generic estimates for agricultural commodities, processed, and prepared fruits, vegetables, meats, poultry, grains
- Brandname and generic estimates for RTE cereals, selected fast foods, candies, beverages

# Sources of Data and Information Flow



# What Do the Data Represent?

- Estimates of means or central tendency
- Estimates of variability for selected foods
- Nationally representative of food supply?
- Based on high quality analytical data
- Recognized algorithms for calculation and estimation
- Other statistical parameters?

# Quality Parameters for Food Composition Data

- Food and Sample Description
- What was the sampling strategy?
- How were the samples handled?
- Were valid analytical methods used?
- What was the quality control program?
- How many samples were analyzed?

# National Food and Nutrient Analysis Program

- Sets priorities for food analysis for the National Nutrient Databank
- Defines a unique plan for each food
- Procures samples units from 12-48 locations
- Processes samples per std. protocol
- Analyzes foods for nutrients, other
- Reviews and disseminates data in SR

# Food Sampling

## “I came, I saw, I sampled…”

# Objectives of Food Sampling

- Develop representative values for the U.S.
- Estimate central value(s) of food composition
  - Mean
  - Median/ Geometric mean
- Estimate variation(s) of food composition
  - Standard deviation
  - Relative standard deviation/ Coefficient of variation
  - Confidence Intervals

Holden, Joanne:

# Definition of Sampling

- The selection and collection of items of foods defined in number, size and nature to represent the food under consideration
- The physical operation of removal of items from lots or fields
- The consolidation and reduction of the collected items to form the portion for analysis

from Horwitz et al., 1978

# When to Sample?

- What foods to choose?
- What nutrients or components should be measured?
- Where to sample?

# What Foods to Choose?

- What new data are needed?
- Current entries in SR and USDA's FNDDS?
- Ranking of Key Foods by Nutrient and Consumption
- Changes in the food supply?
- Quality and status of existing data
- Specificity of nutrients and foods

# What are Key Foods?

**The list of foods which provide about 80% of the intake of a specific component to the diet.**

- Frequency vs. Concentration
- Foods alone and as ingredients
- Some foods are “key” for several nutrients

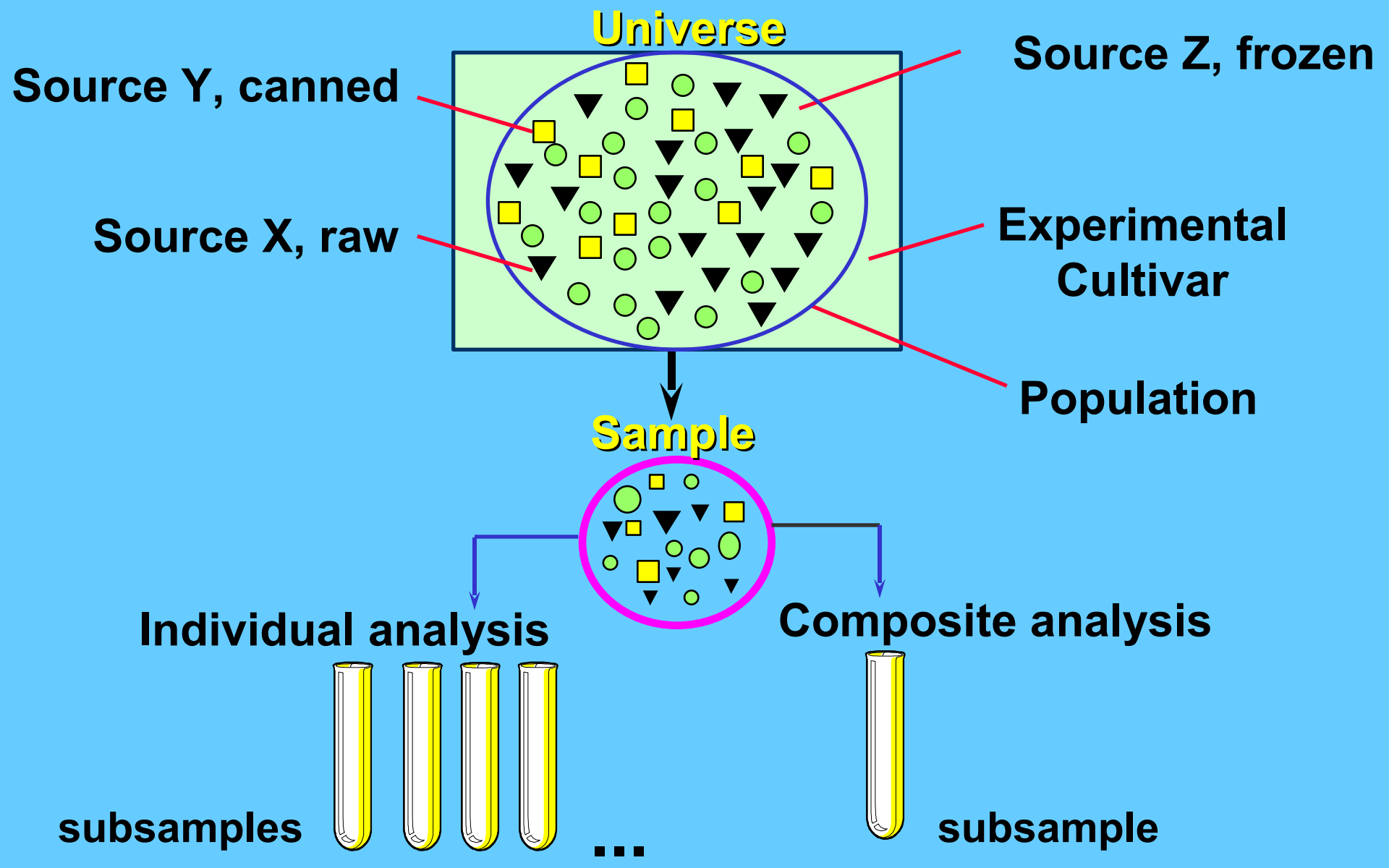
# What components should be measured?

- Needs for Public Health-Assessment, research, policy, education, regulation
- Increased level of specificity?
- Emerging components
- Nutrients with new health implications
- Changes in units of measure?
- Changes in food formulation?

# Elements of a Sampling Strategy?

- Sample units representative of the
- geographic/consumption/distribution pattern for food
- Statistical sampling plan is needed
- Sampling strategy related to demographic plan and population distribution and food description

# Carrots


Universe
Source Y, canned
Source Z, frozen
Source X, raw
Experimental Cultivar
Population
Sample
Individual analysis
Composite analysis
subsamples
...
subsample

# Representative Sampling

**If the population contains**

**Type A: 40% Type B: 20%**

**Type C: 30% Type D: 10%**

**Then the sample should contain**

**Type A: 40% Type B: 20%**

**Type C: 30% Type D: 10%**

# Where to Sample?

- Supermarkets
- Restaurants-fast food, family-type,
- Ethnic foods-markets, tribes, food fiestas
- Street vendors and food courts
- Farmer’s markets or central markets
- Manufacturers or distributors
- Households

# Current NFNAP Retail Sampling Strategy

- Probability proportional to size (p.p.s.)
- 4 regions (~ equal population)
- 3 strata per region (diverse population sizes)
- 1 gCMSA* per strata
- 2 counties per gCMSA (rural and urban)
- Retail stores with sales >$2m
- Brands or varieties: market share or consumption data

*generalized Consolidated Metropolitan Statistical Area (Census)

# Design of Sampling Protocols

"To composite or not to composite…that is the question?"

# Design of Sampling Protocols

## What is the objective?

- To estimate the mean nutrient content
- To estimate the variability of nutrient in the food

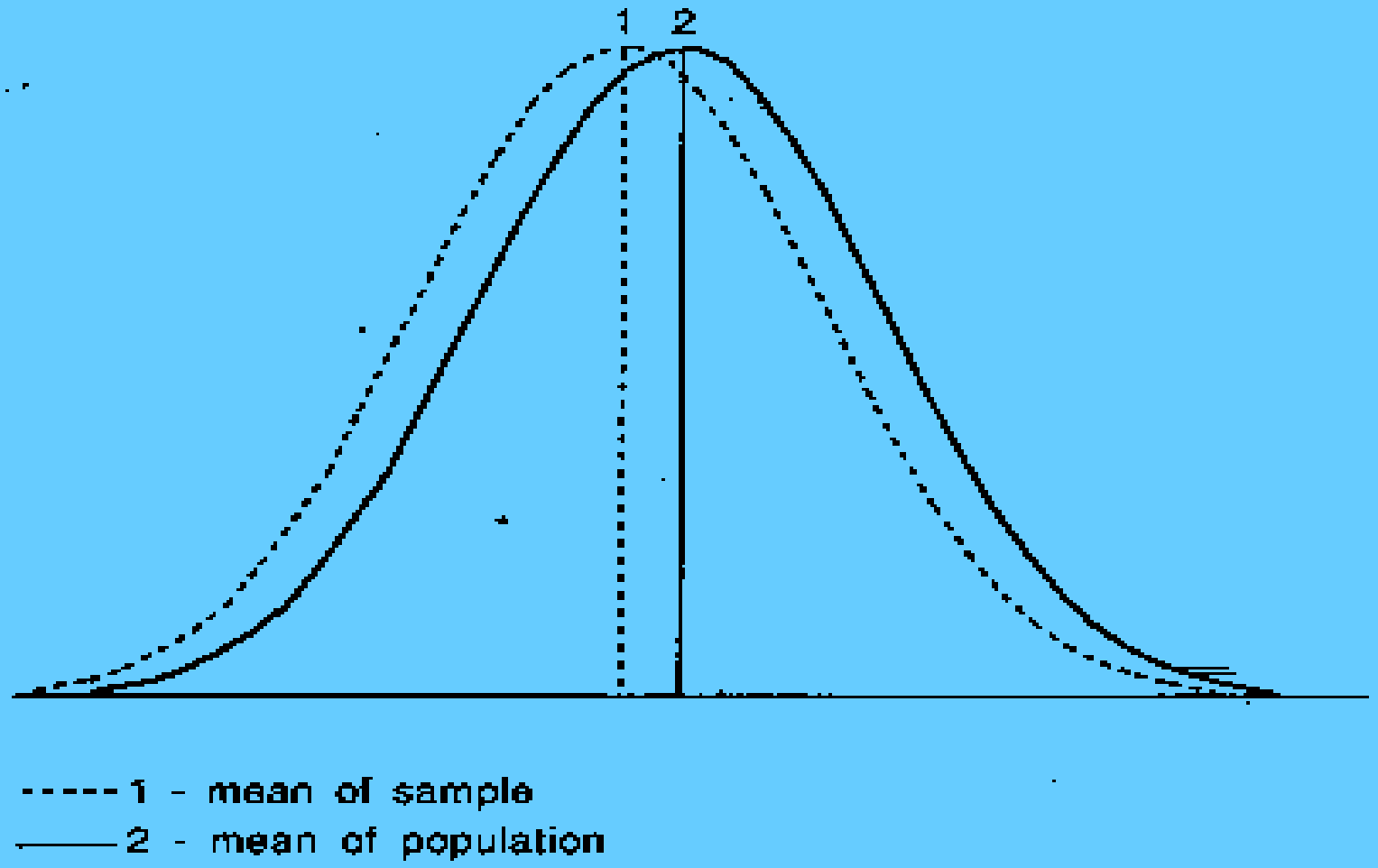
# Calculation of Sample Size

- How secure do you want to feel?
- What is the variability of a nutrient in a food?
- What is the level of the component in food?
- What kind of assessment tool will be used?
- What resources are available?

# What Do the Data Represent?

- Estimates of means or central tendency
- Nationally representative of food supply
- Based on high quality analytical data?
- Recognized algorithms for calculation and estimation
- Preliminary estimates of variability and sample size

# Impact of Bias on Nutrient Means



# Revision of Folate Values for Enriched Grain Products

| Enriched Grain Products | Previous Values (mg/100g) | Fortification Levels [1] (mg/100g) |
|---|---|---|
| Flour | 26 | 154 |
| Cornmeal, Grits | 48 | 187 |
| Farina | 24 | 173 |
| Rice | 8 | 231 |
| Macaroni, Noodles | 18, 29 | 231 |
| Bread, Rolls, Buns | 34, 30, 27 | 95 |

[1] For item where the fortification level is a range, the midpoint is used

# Dietary Components in SR

- Proximate components
- Fatty Acids and Amino acids
- Total sugars and starch
- Dietary Fiber
- Vitamins
- Minerals
- Individual Carotenoids
- Selenium, Fluoride, Choline (to be in SR20)

# Analytical Methods

## Development, Validation, and Application

# Food Composition Laboratory

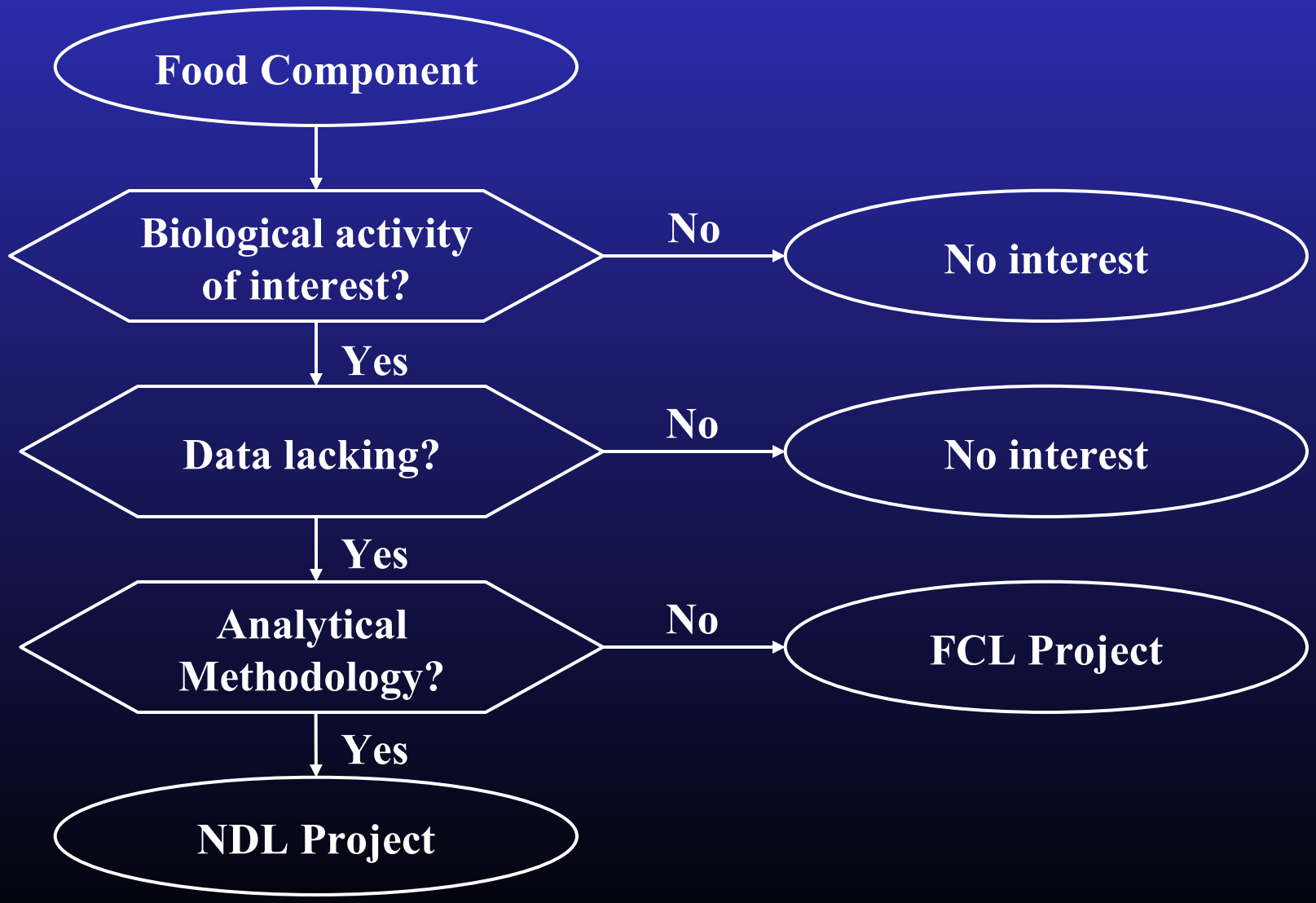
## Mission

To develop measurement systems for the determination of food components that influence human health

## Goals

- Design and develop improved analytical methodology for dietary components
- Transfer developed analytical technologies to scientific communities
- To improve the quality of food composition data

# Selection of Projects



# Continuously Evolving Demand for Methods

Interest in new components as technology advances:

- Research
  - Epidemiological studies
  - Clinical studies
- National Academy reports
  - Dietary Reference Intakes
- Legislation:
  - Nutritional Labeling and Education Act (NLEA)
- New foods, additives, and supplements

# Methods for Food Components

**Every food component has been analyzed, sometime, somewhere, somehow, by someone!**

**Most of these methods and data are not suitable for databases due to limitations regarding:**

**Validation**

**Quality assurance**

**Speed**

**Cost**

# Key Requirements for Method Validation

Specificity – ability to measure what is intended

Accuracy – closeness to true value measured by:

- reference materials
- method of standard additions
- spike recoveries

Precision – degree of agreement of results

Linearity - results proportional to concentration (preferred for optimum accuracy)

Limit of Detection – lowest measurable concentration

Ruggedness – precision under widely varying conditions (different labs, analysts, days, etc.)

# Our Definition of Robust Methods
## (Suitable for support of databases)

**Qualitative:** Specific for each form of target nutrient.

Inclusive of all forms of target nutrient.

Well defined chemistry.

**Quantitative:** Accurate and precise with suitable detection limits.

Reference materials (optimal).

**Economical:** Inexpensive, high throughput, user friendly.

# State of Method Development

(appropriate for food components and composition databases)

| Nutrient | Robust | Limited | Inadequate |
|---|---|---|---|
| Proteins | total nitrogen<br>amino acids (most) | tryptophan<br>cysteine<br>total protein | |
| Lipids | total fat<br>saturated fat<br>*cis*-fatty acids<br>Ω3-*cis*<br>Ω6-*cis* | trans-fatty acids<br>conjugated linoleic | tocotrienols<br>oxidized fatty acids<br>oxidized cholesterol<br>sphingolipids |
| Vitamins | tocopherols (E)<br>vitamin C<br>vitamin A | vitamin D<br>vitamin K<br>pantothenic acid<br>choline<br>niacin<br>thiamin<br>riboflavin<br>vitamin $B_6$ | cobalamins ($B_{12}$)<br>folates<br>biotin |

# State of Method Development

(appropriate for food components and composition databases)

| Nutrient | Robust | Limited | Inadequate |
|---|---|---|---|
| Carbohydrates | starch<br>monosaccharides<br>disaccharides<br>trisaccharides | oligosaccharides<br>dietary fiber<br>pectins/uronic acids<br>beta-glucan | resistant starch<br>amylose/amylopectin<br>polysaccharides<br>lignans |
| Minerals | total Na<br>total K<br>total Ca<br>Total Mg<br>total Fe<br>total Cu<br>total Mn | total Si<br>total Li<br>total As<br>total S<br>total B<br>total Co<br>total Se<br>total F<br>total I | heme/non-heme Fe<br>Se-methionine<br>cobalamins |

# State of Method Development

(appropriate for food components and composition databases)

| Nutrient | Robust | Limited | Inadequate |
|---|---|---|---|
| Phytonutrients | carotenoids | flavonoids | thiosulfinates |
| | | isoflavonoids | indoles |
| | | lignans | isothiocyanates |
| | | phytosterols | inositol phosphates |
| | | proanthocyanidins | phenolics |
| | | | protease inhibitors |
| | | | saponins |
| | | | thiols |

# Current Projects at FCL

| | |
|---|---|
| Phytonutrients: | Phenolics |
| | Flavonoids |
| | Phenolic acids |
| | S-containing amino acids |
| Vitamins | Multivitamin Determinations |
| | Vitamin D |
| | Niacin |
| | Thiamine |
| Mineral species | Selenomethionine |
| | Heme/non-heme iron |
| Fingerprinting | Pattern recognition - extracts analyzed with no prior separation |

# Research Outcomes

- Well documented and representative food composition values
- Strategies for statistical sampling
- Use of valid analytical methods and analytical quality control
- Protocols for sample handling
- Provisional Databases

# Research Outcomes, cont.

- Analytical Methods
- Publications in peer-reviewed journals
- Transfer technology through:
  - Collaborations
  - Official analytical methods
  - New Reference materials
  - New instrumentation
  - Patents

# Nutrient Data Laboratory

Web Site:

**http://www.ars.usda.gov/nutrientdata**