Analyzing dietary data, with a focus on estimation of distributions of *usual intake*

June 18, 2018

SCHOOL OF PUBLIC HEALTH AND HEALTH SYSTEMS
APPLIED HEALTH SCIENCES
UNIVERSITY OF WATERLOO

UNIVERSITY OF WATERLOO

Agenda and logistics

- Considerations for analyzing dietary intake data, specific to surveillance – *Sharon Kirkpatrick*
- Estimation of distributions of usual intake – *Kevin Dodd*
- Break
- Estimation of distributions of usual intake, cont'd – *Kevin Dodd*
- Q&A
- Early experience applying the NCI method to CCHS 2015 – *Karelyn Davis* and *Cunye Qiao*
- Reflections from *Kevin Dodd*
- Q&A
- Afternoon: in-person or webcast

Agenda and logistics

- Slides, other materials, and a link to a Google Doc for posting questions (for remote viewers):

https://tinyurl.com/June19CCHS

What do we want to measure?

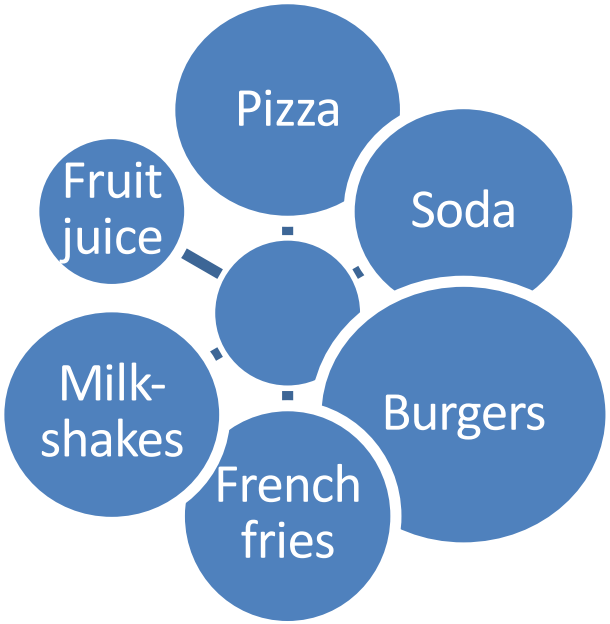
- What was eaten?
- How much was eaten?
- How often is this eaten?

→ Translate into nutrients, food groups, diet quality

\+ Contextual information

What makes capturing these aspects of diet difficult?

- Diet is multidimensional

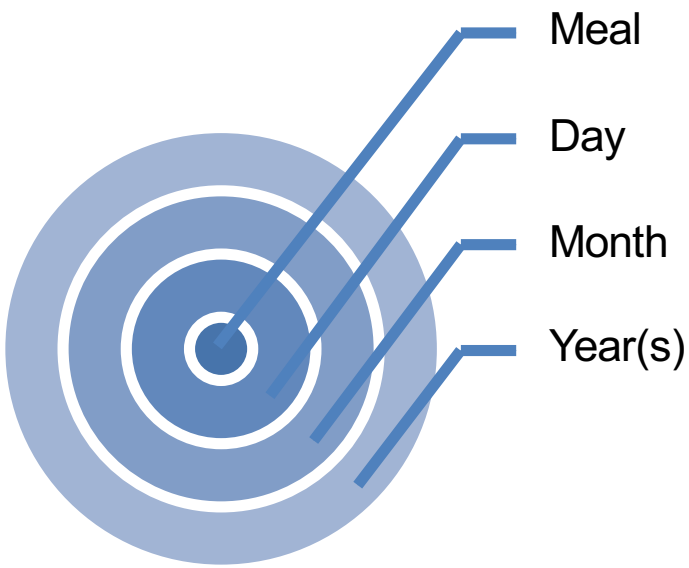


Eating patterns: " . . . the quantities, proportions, variety, or combination of different foods, drinks, and nutrients (when available) in diets, and the frequency with which they are habitually consumed"

Nutrition Evidence Library. A Series of Systematic Reviews on the Relationship Between Dietary Patterns and Health Outcomes. U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, March 2014.

What makes capturing these aspects of diet difficult?

- Diet is dynamic



In Utero	Infancy and Childhood	Adolescence	Adulthood	Older Adulthood

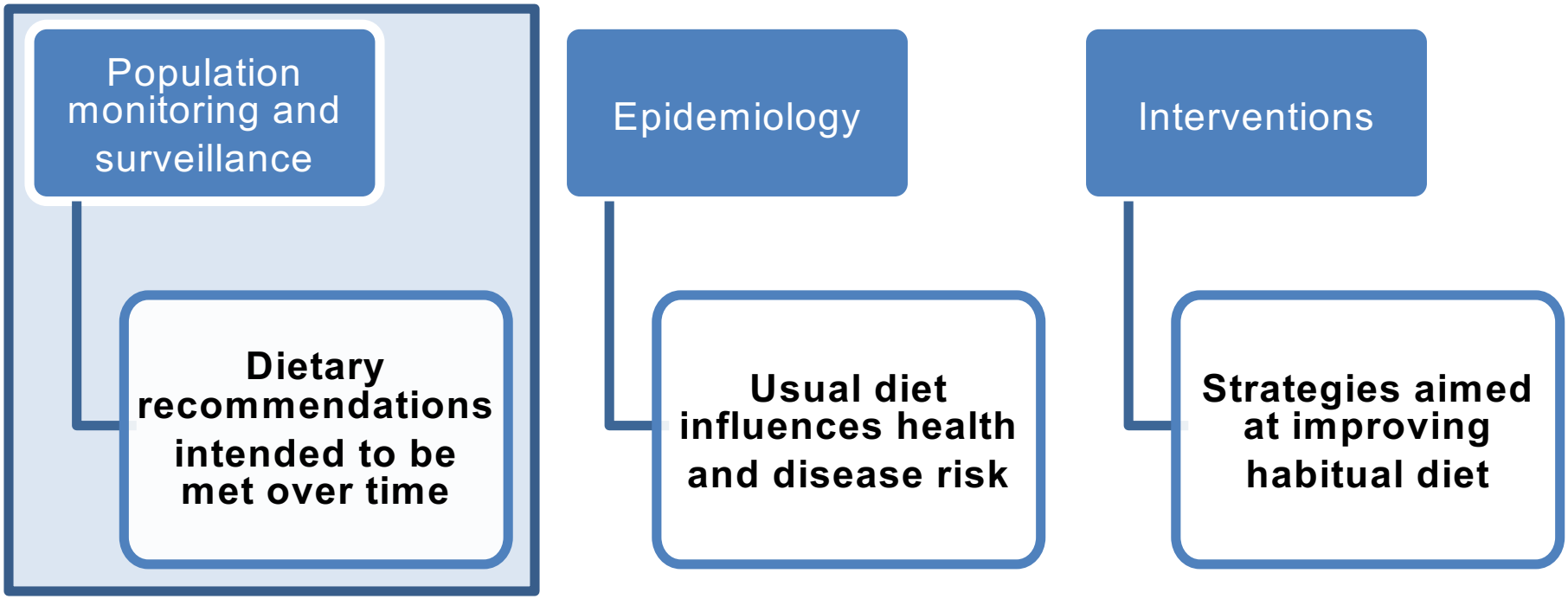
Source: Jill Reedy, National Cancer Institute

Interest is typically in *usual intake*

- Usual dietary intake
  - Average or long run or habitual intake

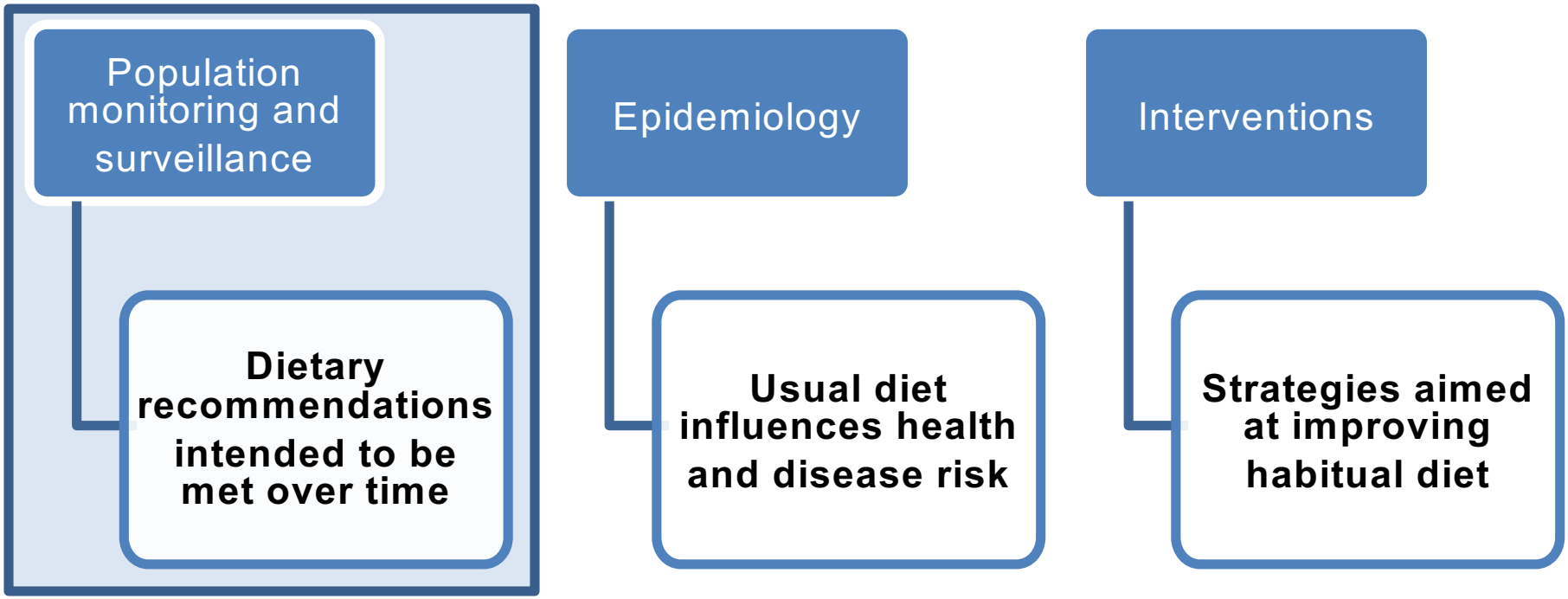
Interest is typically in *usual intake*

- Usual dietary intake
  - Average or long run or habitual intake



Interest is typically in *usual intake*

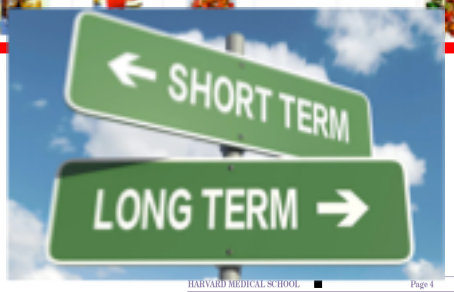
- Usual dietary intake
  - Average or long run or habitual intake

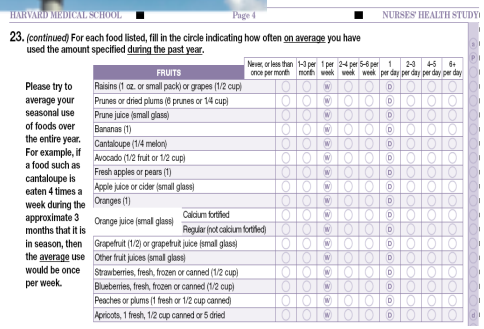


→ Observable?

Main types of self-report assessment instruments

- **Short-term instruments**
  - Snapshot in time
    - 24-hour recalls
    - Food records/diaries





HARVARD MEDICAL SCHOOL — Page 4 — NURSES' HEALTH STUDY

23. *(continued)* For each food listed, fill in the circle indicating how often on average you have used the amount specified during the past year.

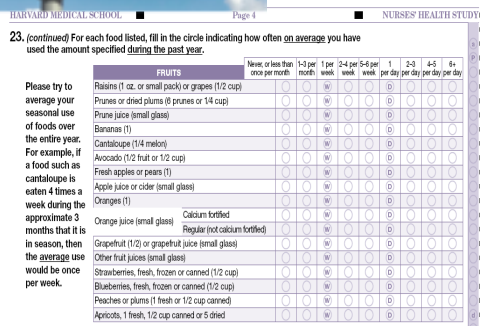
Please try to average your seasonal use of foods over the entire year. For example, if a food such as cantaloupe is eaten 4 times a week during the approximate 3 months that it is in season, then the average use would be once per week.

FRUITS	Never, or less than once per month	1–3 per month	1 per week	2–4 per week	5–6 per week	1 per day	2–3 per day	4–5 per day	6+ per day
Raisins (1 oz. or small pack) or grapes (1/2 cup)			W			D			
Prunes or dried plums (6 prunes or 1/4 cup)			W			D			
Prune juice (small glass)			W			D			
Bananas (1)			W			D			
Cantaloupe (1/4 melon)			W			D			
Avocado (1/2 fruit or 1/2 cup)			W			D			
Fresh apples or pears (1)			W			D			
Apple juice or cider (small glass)			W			D			
Oranges (1)			W			D			
Orange juice (small glass) — Calcium fortified			W			D			
Orange juice (small glass) — Regular (not calcium fortified)			W			D			
Grapefruit (1/2) or grapefruit juice (small glass)			W			D			
Other fruit juices (small glass)			W			D			
Strawberries, fresh, frozen or canned (1/2 cup)			W			D			
Blueberries, fresh, frozen or canned (1/2 cup)			W			D			
Peaches or plums (1 fresh or 1/2 cup canned)			W			D			
Apricots, 1 fresh, 1/2 cup canned or 5 dried			W			D			

Main types of self-report assessment instruments

- **Short-term instruments**
  - Snapshot in time
    - 24-hour recalls
    - Food records/diaries

- **Long-term instruments**
  - Aim to capture usual intake
    - Food frequency questionnaires
    - Brief dietary instruments (screeners)


HARVARD MEDICAL SCHOOL ■ Page 4 ■ NURSES' HEALTH STUDY

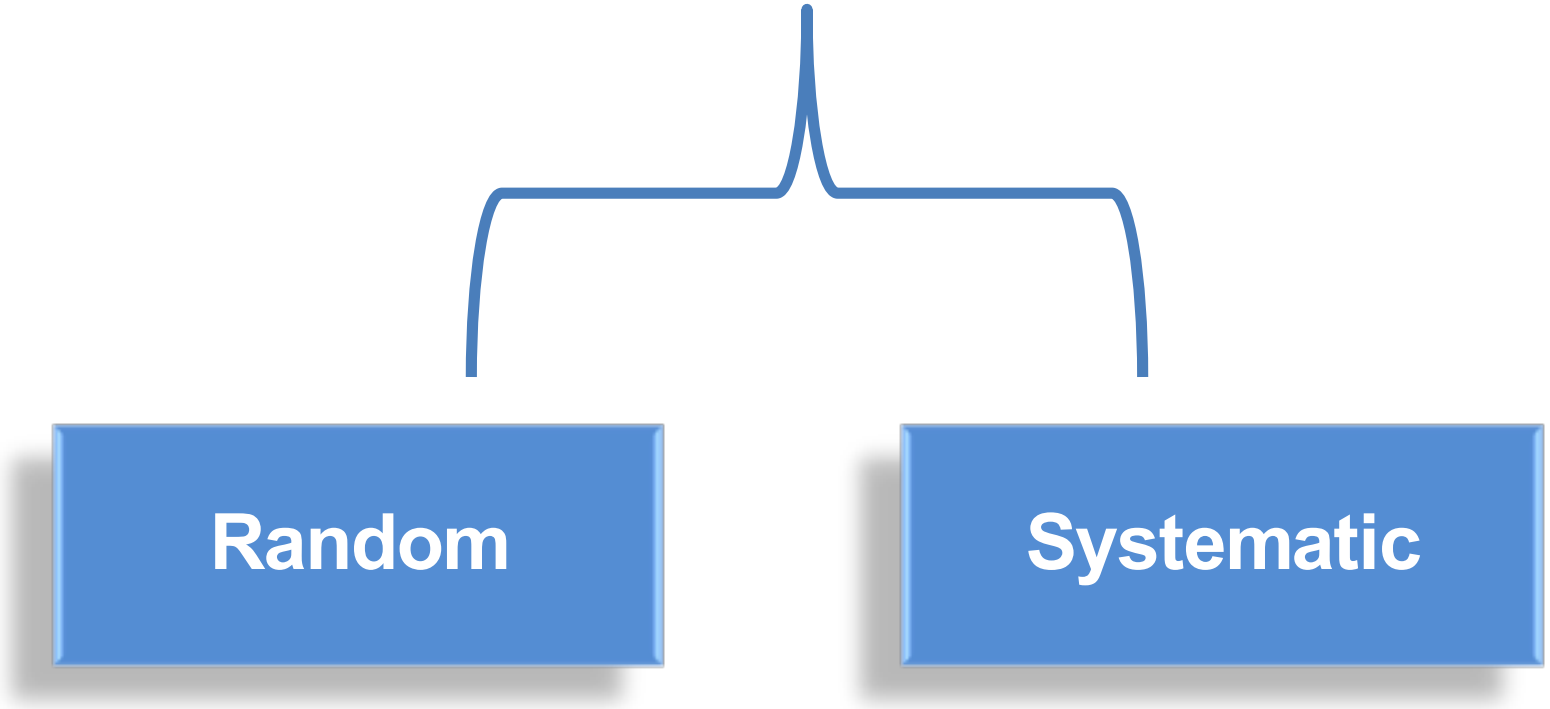
23. *(continued)* For each food listed, fill in the circle indicating how often on average you have used the amount specified during the past year.

Please try to average your seasonal use of foods over the entire year. For example, if a food such as cantaloupe is eaten 4 times a week during the approximate 3 months that it is in season, then the average use would be once per week.

FRUITS	Never, or less than once per month	1–3 per month	1 per week	2–4 per week	5–6 per week	1 per day	2–3 per day	4–5 per day	6+ per day
Raisins (1 oz. or small pack) or grapes (1/2 cup)									
Prunes or dried plums (6 prunes or 1/4 cup)									
Prune juice (small glass)									
Bananas (1)									
Cantaloupe (1/4 melon)									
Avocado (1/2 fruit or 1/2 cup)									
Fresh apples or pears (1)									
Apple juice or cider (small glass)									
Oranges (1)									
Orange juice (small glass) — Calcium fortified									
Orange juice (small glass) — Regular (not calcium fortified)									
Grapefruit (1/2) or grapefruit juice (small glass)									
Other fruit juices (small glass)									
Strawberries, fresh, frozen or canned (1/2 cup)									
Blueberries, fresh, frozen or canned (1/2 cup)									
Peaches or plums (1 fresh or 1/2 cup canned)									
Apricots, 1 fresh, 1/2 cup canned or 5 dried									

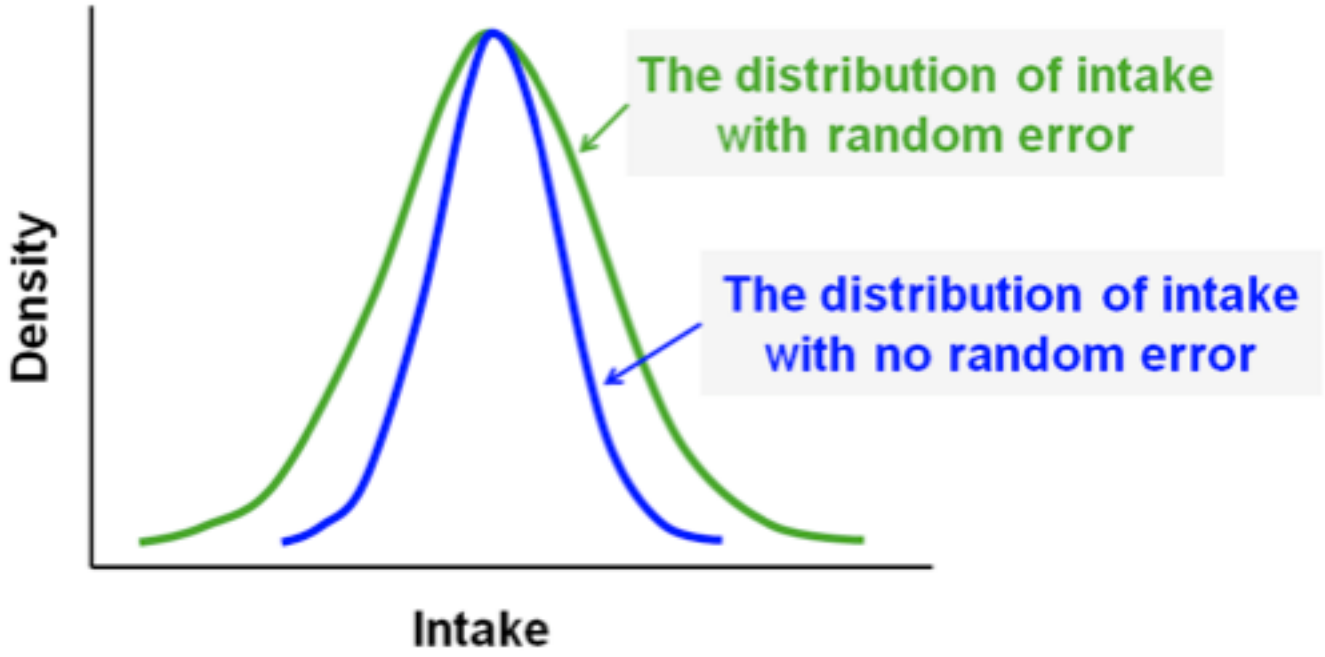
What is measurement error?

- Difference between the true value and the value obtained from a measure



Random error

- An unpredictable source of error that contributes variability
  - Instrument may be accurate (i.e., unbiased), but may not be precise



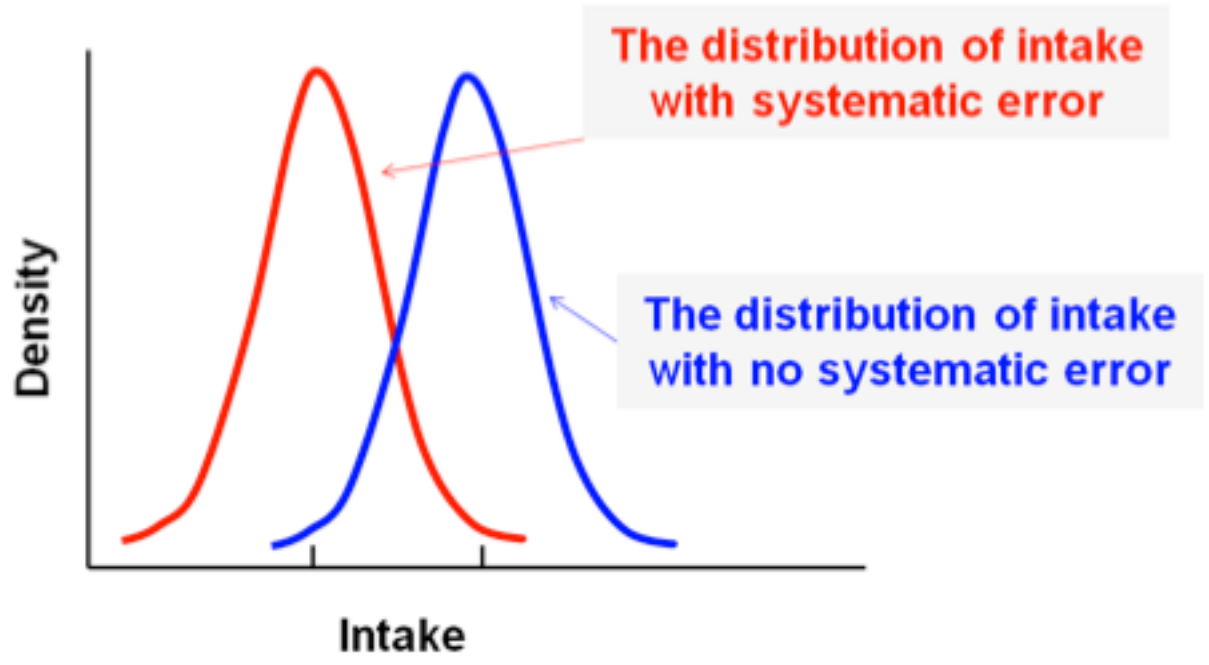
Random error

- An unpredictable source of error that contributes variability
  - Instrument may be accurate (i.e., unbiased), but may not be precise

If an instrument has only random error, the average of many repeat measures approximates the true value

Systematic error

- A source of error in which measurements consistently depart from the true value in the same direction
  - Instrument is inaccurate (i.e., biased*)*



Systematic error

- A source of error in which measurements consistently depart from the true value in the same direction
  - Instrument is inaccurate (i.e., biased*)*

If an instrument has systematic error, the average of many repeat measures *does not* approximate the true value

Implication

- Random error can be dealt with by averaging across repeat measures
- This is not the case for systematic error

The ideal instrument has only random error

Biomarker-based validation studies

24-hour recall (24HR)

- **Larger within-person random error**
- Smaller systematic error

Food frequency questionnaire (FFQ)

- Smaller within-person random error
- **Larger systematic error**

24HR as main instrument

- Intake distributions closer to truth because within-individual variation (i.e., random error) can be accounted for
- (Less bias and more power to detect diet-health relationships)

Use the instrument with the smallest systematic error (i.e., 24HR) – **assume it is *unbiased***

Sources of bias

- Recall bias
- Social desirability bias
- Limitations in food and nutrient databases

Working assumption

- Unbiasedness of 24HR is a **working assumption**
- Required to proceed with development of methods
- May be more or less justified depending on dietary component of interest

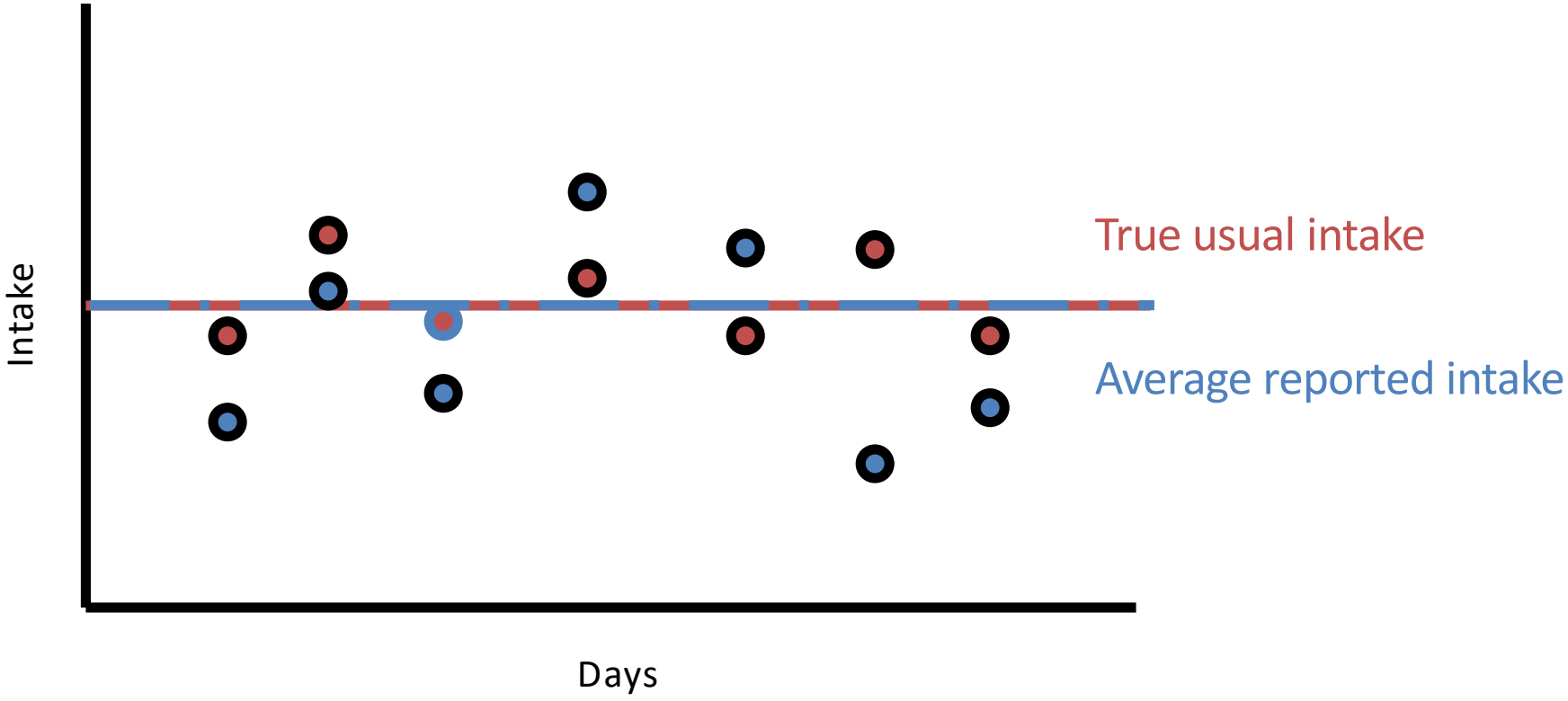
The extent of bias

- **Freedman et al., Validation Studies Pooling Project**
  - Energy underreporting
    - FFQs: 24-33%
    - 24HRs:
      - 6-16% for young and middle aged adults
      - 25% for elderly women
  - Underreporting of other nutrients
    - Protein: 10% FFQ, 5% 24HR
    - Potassium: 5-6% FFQ, 0-4% 24HR
    - Sodium: 28%-39% FFQ, 4%-13% 24HR

Freedman et al., AJE 2014; AJE 2015

Estimating mean (or average) intake

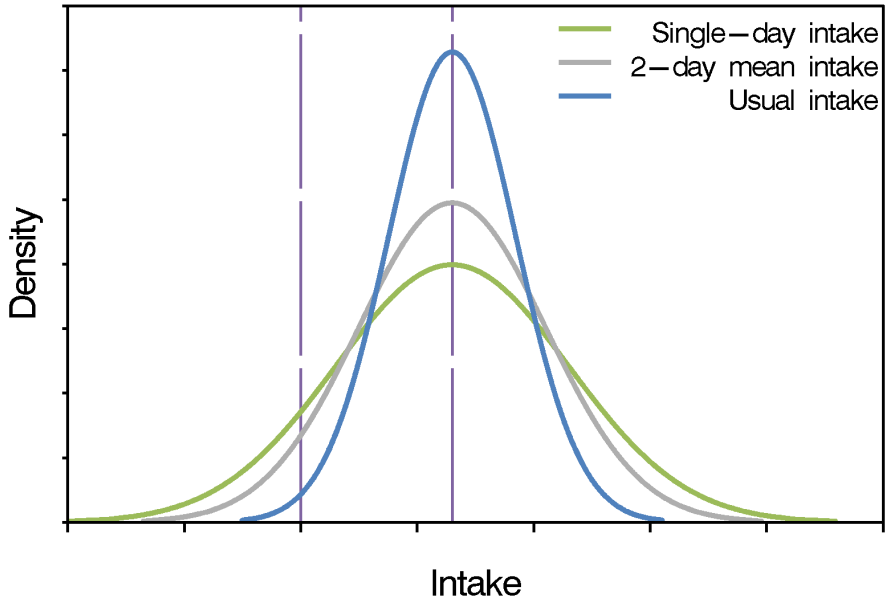
- Unbiasedness assumption: 24HR unbiased for individual-level usual intake



Characteristics of the distribution other than the mean

- With only a few repeats, averaging only removes some of the within-person variation

- Distributions of averages are too wide relative to usual intake distributions



Estimating distributions

- U.S. National Research Council (NRC)/Institute of Medicine (IOM)
- Iowa State University (ISU) Method (SIDE)
- U.S. National Cancer Institute (NCI) Method
- EFCOVAL Consortium Multiple Source Method (MSM)
- Statistical Program for Age-adjusted Dietary Assessment (SPADE)

Scope of today's presentations

- Estimation of distributions of usual intake via the NCI method
  - Non-episodically consumed
  - Episodically consumed
  - Covariates

Scope of today's presentations

- Estimation of distributions of usual intake via the NCI method
  - Non-episodically consumed
  - Episodically consumed
  - Covariates
- Not covered in detail:
  - Bivariate and multivariate methods
  - Incorporating intake from vitamin and mineral supplements
  - Applications for epidemiology

Agenda and logistics

- **Estimation of distributions of usual intake – *Kevin Dodd***
- Break
- Estimation of distributions of usual intake, cont'd – *Kevin Dodd*
- Q&A
- Early experience applying the NCI method to CCHS 2015 – *Karelyn Davis* and *Cunye Qiao*
- Reflections from *Kevin Dodd*
- Q&A
- Afternoon: in-person or webcast