

Design A Evaluation

Sarah Nusser

Department of Statistics

Center for Survey Statistics and Methodology

Iowa State University

Questions

- What are the design's strengths and weaknesses?
- Which users is this design optimal for?
- What will the output of data collection look like?
- What statistical procedures will need to be applied?

Users for Design

- Target users: CPI
- Federal and state programs that use information at the level of the CPI or in aggregation
- Economic research, but not in the ideal form

Strengths / priorities unique to A

- Accurate reporting of concurrent expenses using records to the extent possible
 - Relatively short, but intensive collection to minimize respondent fatigue
 - Focus on relatively easy modes of recording items in natural units and in real time to minimize recall bias

Strengths / priorities unique to A

- Keep sample design and data collection simple
 - Focus on reducing measurement error and increasing response rates
 - Keep costs low
 - Simplify logistical operations of a new survey
 - Simplify post-data collection statistical processing

Questions

- Is 2 weeks too long for accurate reporting for all household members?
 - When does respondent fatigue start affecting quality?
 - What is the statistical gain of the 2nd week in a single reporting period?
 - Some models to consider: Dietary intake surveys, maybe some media surveys that require recording prospectively (radio listening survey)

Questions

- What does “recall” mean for the larger or recurring items?
 - Many of these items often have records associated with them
 - Develop methods to encourage record use here
- Is there sufficient sample size for special CPI calculations?
 - Survey has tons of data for modeling
 - Small area estimation might mitigate the need for direct estimates in smaller domains

Structural constraints

- Does not provide a direct and complete picture of household spending and income over the time periods of interest to economic researchers
 - With CPI as a goal, effort is devoted to expense detail, compromising time span of data collection
 - Statistical modeling can be used to bridge this gap
 - Survey has lots of data to support longitudinal modeling of aggregate categories of interest to researchers

Data for 2-wk data collection event

- Household or person-level
 - Demographics, life events (person)
 - Income, assets, labor force participation (person)
 - Recurring expenses (household, maybe person)
 - Larger purchases (household, maybe person)
- Expenditure-level by person
 - Individual purchases for items by person
- All expense and income data are in natural units for accurate reporting amounts

Data structure over time

Qtr	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	
1	2 wks	2 wks							
1		2 wks	2 wks						
1			2 wks	2 wks					
2				2 wks	2 wks				
2					2 wks	2 wks			
2						2 wks	2 wks		
3							2 wks	2 wks	
...									...

Minimum statistical processing

- Translate the entered data into units that are comparable across individuals and households
 - Could be done by survey instrument
 - Could offer this as a check to respondents
- Diagnostics, editing for microdata
- Sample and nonresponse weighting, variance estimation variables
- Calculate CPI budget shares
 - Need a model to estimate quarter, annual params

Some additional opportunities

- More detailed nonresponse adjustment
- Small area estimation for specific CPI domains
- Measurement error modeling
 - Strong interest in seasonality → repeated observations
 - Evaluate quality of information (possibly for comparison within an experimental setting)
 - Potential bridge for uses beyond the CPI and state/federal programs

Example: dietary intake surveys

- Focus is on parameters from the “usual intake” distribution across individuals
 - Distn of indiv means: annual mean daily consumption of a dietary component for an indiv
- Food frequency: estimate of typical daily intake as a direct measure of usual intake
 - Shown to have considerable measurement error
- 24-hr dietary intake record or recall: more accurate information on a short period of time
 - Better quality, wrong concept
 - Can be related to usual intake via a model

Example: dietary intake surveys

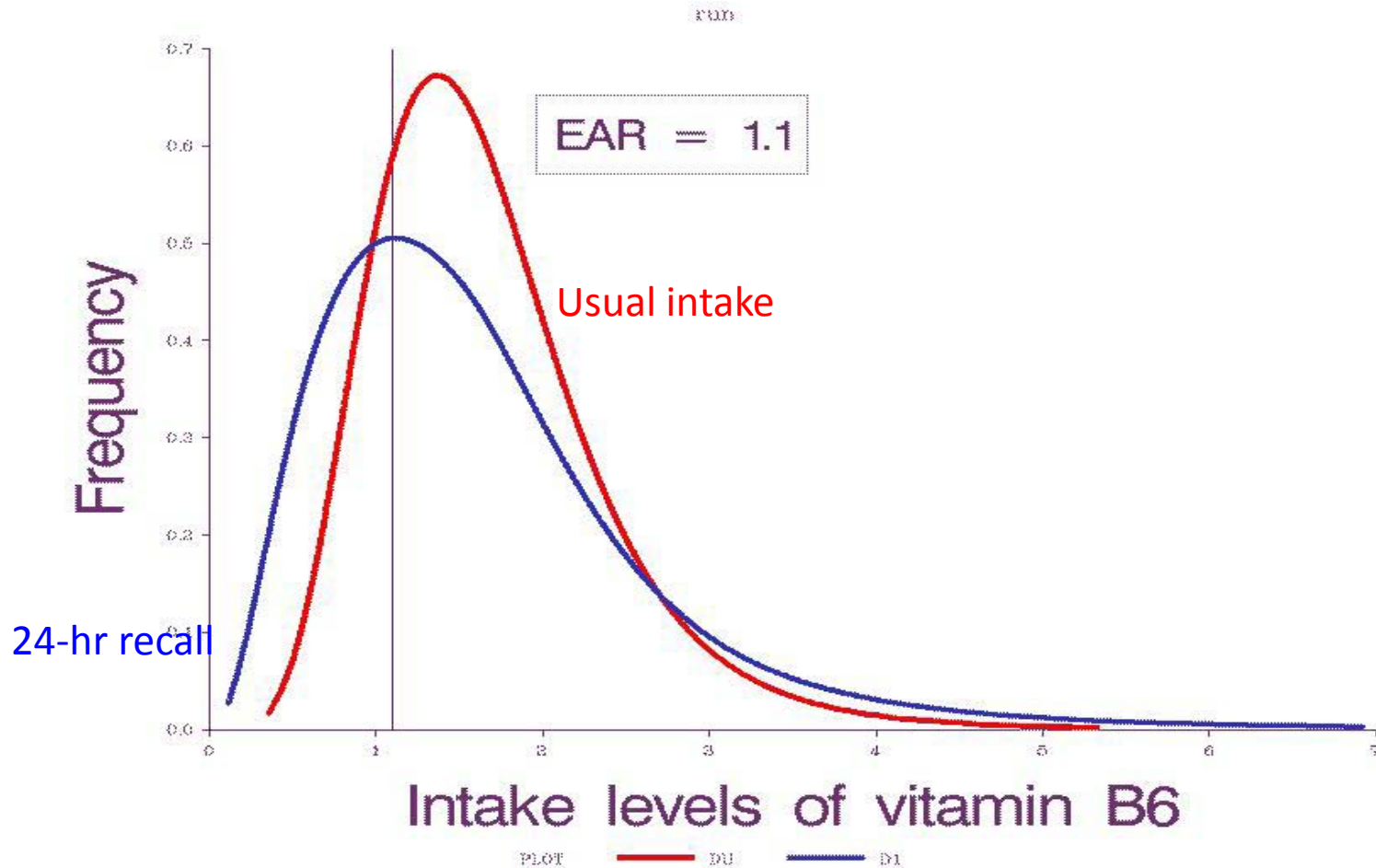
24hr recall = usual intake + error

$$D_{ij} = U_i + e_{ij}$$

D_{ij}	24 hr recall for indiv i on day j (NHANES: 2 days)
U_i	usual intake for indiv i (indiv mean over days)
e_{ij}	error in 24hr recall as an est of usual intake

- Interest is in parameters for distn of usual intake
→ estimate this distribution

Vitamin B₆ (mg/d) , women 19-50 yrs



Example: dietary intake surveys

- Lots of different nutrients, foods, etc.
 - Goal of methodology development was to serve the full range of behaviors
- Food intakes, for example, have many 0 values
 - Is this a structural 0 (not a consumer of the food) or a 0 on that day (consumer)
 - Mixture model to allow for a parameter that
- Ratios
 - Dietary component in relation to calorie intake
- Policy analysis
 - Food fortification alternatives