




# Addressing Gaps in Market Level Databases

**Shu Wen Ng**  
**University of North Carolina at Chapel Hill**  
March 2012



THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL



UNC Food Research Program  
Robert Wood Johnson Foundation  
University of North Carolina at Chapel Hill



## UNC Food Research Program Team

- Barry Popkin, PhD
- Shu Wen Ng, PhD
- Meghan Slining, PhD
- David Guilkey, PhD
- Phil Bardsley, PhD
- Izabela Annis, MS
- Donna Miles, PhD
- Dan Blanchette
- Lauren Butler, RD
- Emily Ford, RD
- Peggy Rupert, RD
- Doctoral Students

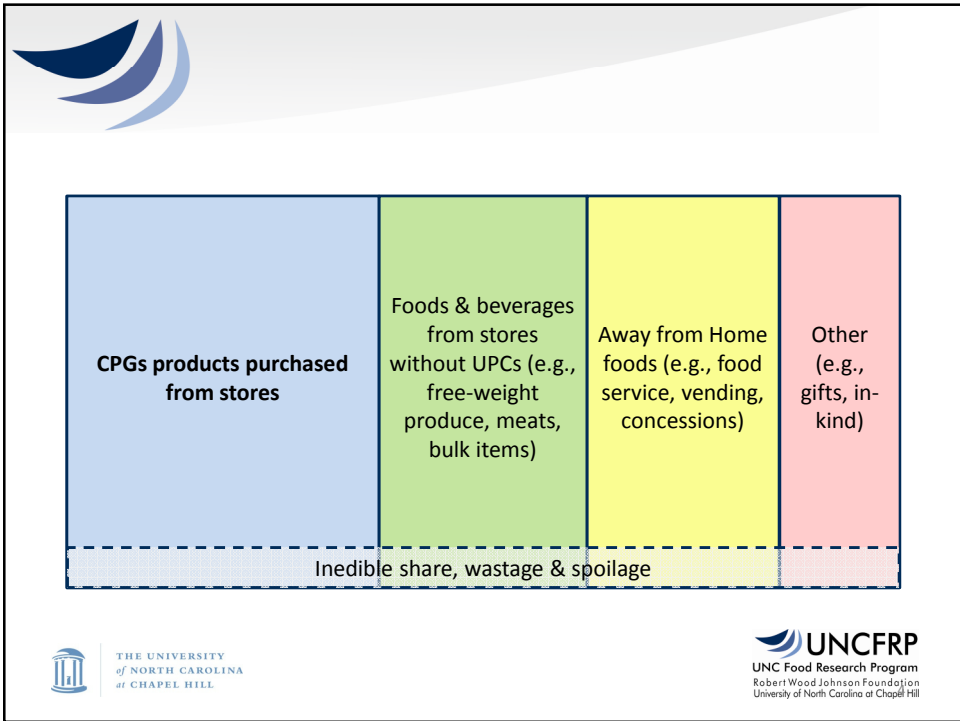
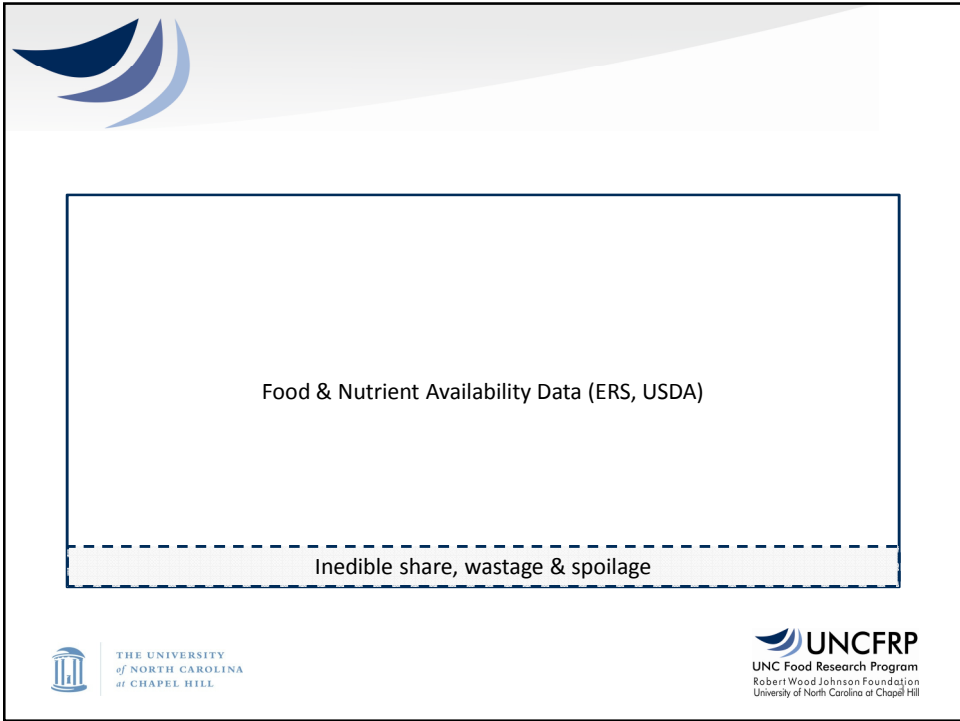
**Funding Sources:** Robert Wood Johnson Foundation, National Institutes of Health



THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL



UNC Food Research Program  
Robert Wood Johnson Foundation  
University of North Carolina at Chapel Hill



Scantrack + Homescan,  
The Nielsen Co.

Food Sales/ Purchase

**Homescan**

- Commercially available barcode/UPC level data on CPGs purchased by households (key respondent scans food purchases with UPCs)
- Data history back to mid-1990s; ongoing
- Nationally representative survey of 35,000-60,000 households in 'static' panel per year.
- Targeted direct recruitment by mail and online
- Cross-sectional (but 30-40% are retained from year to year)
- Raw data is at the household-shopping episode-UPC level
- Includes >400,000 UPCs/ year
- Costs: Raw disaggregated UPC data is about \$100,000/year
- Key measures:
  - Demographics (gender, age, race/ethnicity, income, education, employment, marital status, household composition)
  - Geography (market & census region)
  - Purchase details such as quantity, price, date, store
  - UPC details such as brand, product description, category type
  - Can be linked to TD-links (containing store information)

Scantrack + Homescan,  
The Nielsen Co.

Food Sales/ Purchase

**Homescan**

- Key Advantages:
  - Large sample; potential longitudinal design
  - Analyses of household purchases, including prices and quantities, patterns & frequency of purchases
  - Captures purchases from all stores
- Key Limitations:
  - No detailed random-weight items since 2008
  - No information about away-from-home food and beverage purchases that were never brought into the house by all members of the household
  - Do not know what is actually eaten (large bias if spoilage is high or systematically different)
  - Do not know proportions of consumption among household members
  - No nutrition information already linked

Scantrack + Homescan,  
The Nielsen Co.

Food Sales/ Purchase

**Scantrack**

- Point-of-sale, UPC level data on volume & dollar sales of food stores, food/drug combinations, drug stores, mass merchandisers (F/D/M), and convenience stores (C-stores)
- Data history for F/D/M goes back 3-yrs; C-stores only 2 yrs; ongoing
- Stratified systematic probability sample, designed to measure consumer sales across major markets, region, and projectable to the US
- Raw data is at the market (or total US)- weekly (or quarterly or annual)-UPC level
- Includes >400,000 UPCs/ yr
- Costs: Raw market level, weekly data is about \$200,000/yr
- Key measures: Brand, Product description, Category type, Dollar sales, Unit or volume sales, Vendor, Percentage of stores in sample selling each product, Price (regular & promotion)

Scantrack + Homescan,  
The Nielsen Co.

Food Sales/ Purchase

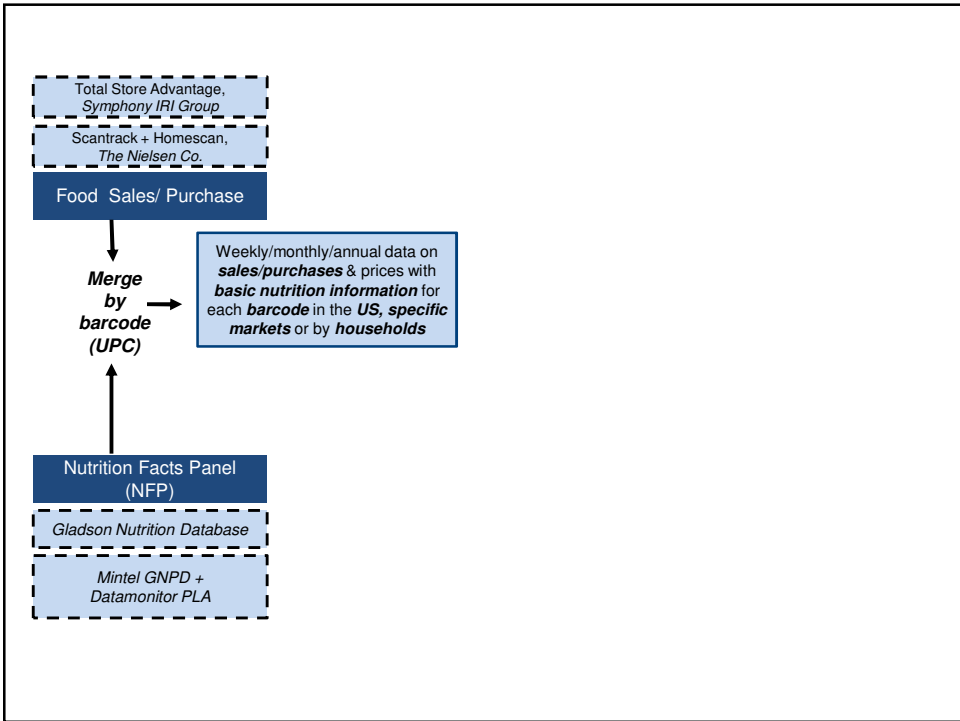
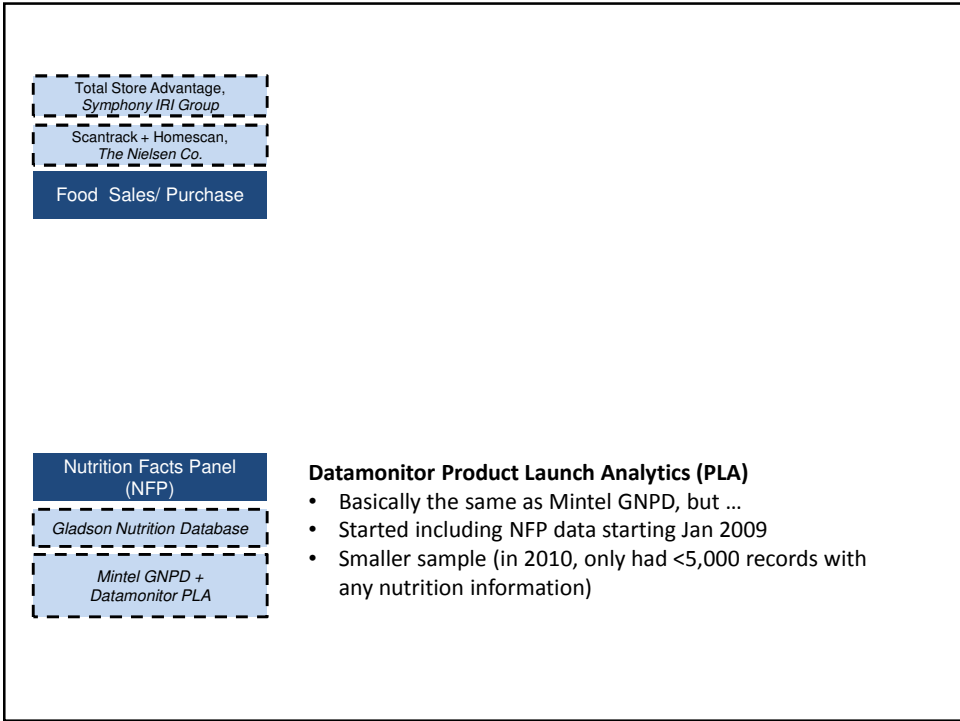
**Scantrack**

- Key Advantages
  - Analyses of sales, prices, and quantities by brand and UPC, sold from stores within major markets and nationally
  - Retailer-specific information also is available for purchase (using TD-Links)
  - More accurate measure of what products are sold/purchased
- Key Limitations
  - Only representative of the stores included in their sample
  - No information about sales from smaller F/D/M & C-stores, vending machines, restaurants, food trucks, mom and pop stores, ethnic markets, specialty markets, etc.
  - Does not include non-UPC coded products
  - No nutrition information

<p>Total Store Advantage, Symphony IRI Group</p>	<p><b>Total Store Advantage</b></p> <ul style="list-style-type: none"> <li>• Similar to Scantrack, but larger samples of stores</li> <li>• Data history goes back 5-yrs</li> <li>• Market coverage across channels are better than Scantrack</li> <li>• Costs: Raw market level, weekly data is about \$400,000/yr</li> <li>• Key Advantages: same as Scantrack</li> <li>• Key Limitations                             <ul style="list-style-type: none"> <li>– Mostly the same as Scantrack</li> <li>– Claims to contain some NFP data for about 20% of UPCs</li> </ul> </li> </ul>
<p>Scantrack + Homescan, The Nielsen Co.</p>	
<p>Food Sales/ Purchase</p>	

<p>Total Store Advantage, Symphony IRI Group</p>	<p><b>Gladson</b></p> <ul style="list-style-type: none"> <li>• Database that includes NFP labels, ingredient list and claims for products with UPCs (national brands and private label items)</li> <li>• Since 1999; claims to be updated weekly</li> <li>• Sample grows as new products enter the market (claims to have ~ 2,000 new or reformulated UPCs each week)</li> <li>• Key Measures:                             <ul style="list-style-type: none"> <li>– NFP information</li> <li>– Full ingredient listings, warning, claims on the packaging, date of last update</li> <li>– Manufacturer &amp; brand by UPC</li> </ul> </li> <li>• Costs depends on vintage of data, snapshot or annual subscription</li> </ul>
<p>Scantrack + Homescan, The Nielsen Co.</p>	
<p>Food Sales/ Purchase</p>	
<p>Nutrition Facts Panel (NFP)</p>	
<p>Gladson Nutrition Database</p>	







## How valid is NFP data?

- Nutrition information limited to FDA requirements
- Outside box vs. inside box
  - Chemical analyses is the best way to validate this
  - NDL work
- Comparing commercial source of NFP data to what is on the outside of boxes
  - Need geographically diverse field work to compare most current NFP data with what exists in the stores
  - Can food industry assist by creating centralized database with their NFP information?



THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL



UNC Food Research Program  
Robert Wood Johnson Foundation  
University of North Carolina at Chapel Hill



## Estimating Nutrients

- One way to derive missing nutrient information
  - USDA
  - University of Minnesota NCC
- Linear programming at the UPC level
  - Order of ingredients
  - Each UPC has basic nutrition information
  - Link ingredients to a database of foods including commercial ingredients with more complete nutrient information
  - Minimize error based on nutrient specific error tolerances
- Applications: Added sugars; Potassium
- Even better: more information required on NFP

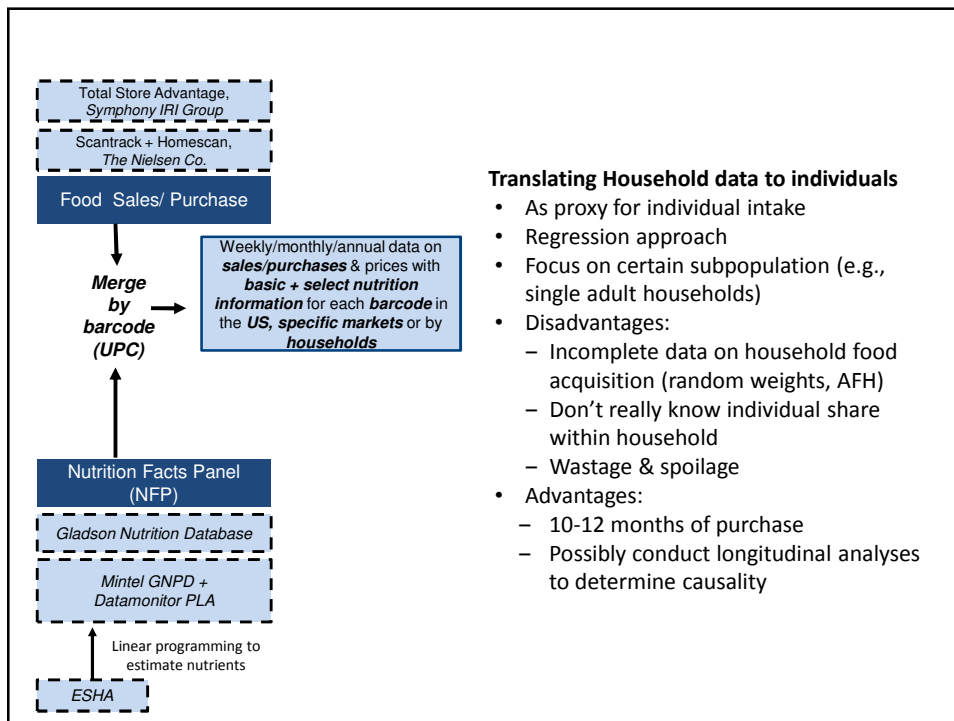
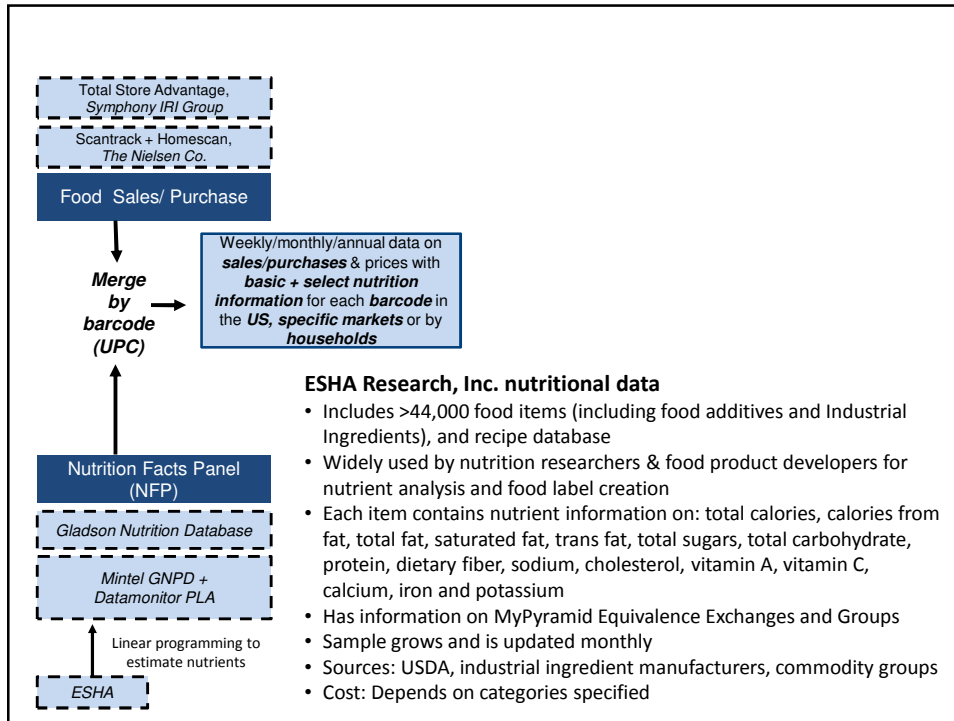


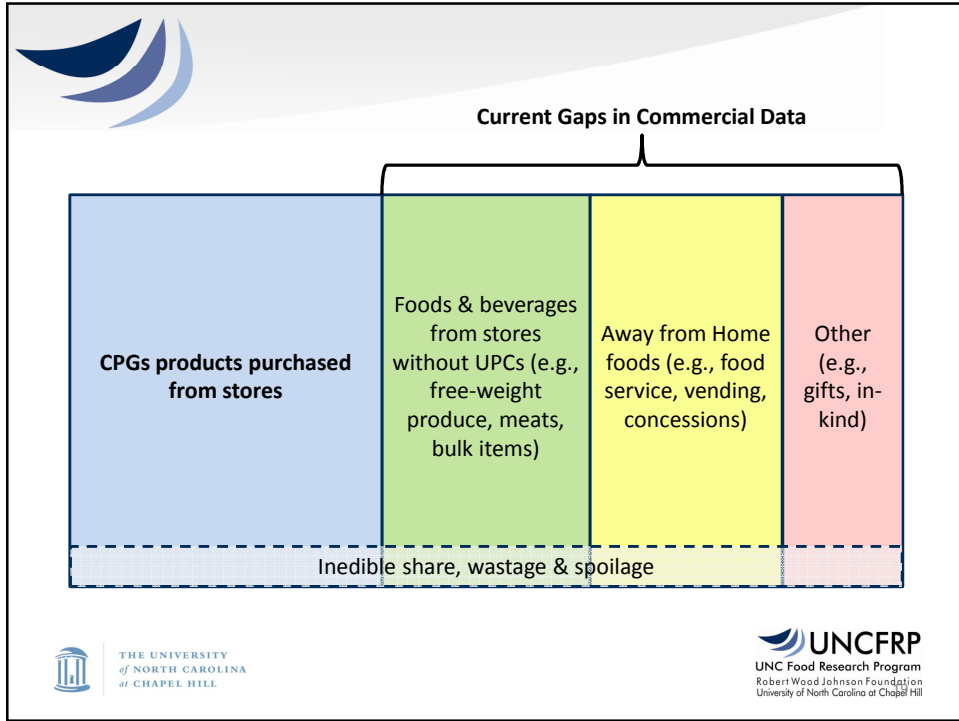
THE UNIVERSITY  
of NORTH CAROLINA  
at CHAPEL HILL



UNC Food Research Program  
Robert Wood Johnson Foundation  
University of North Carolina at Chapel Hill







**Thank you**

Shu Wen Ng: [shuwen@unc.edu](mailto:shuwen@unc.edu)  
Meghan Slining: [slining@unc.edu](mailto:slining@unc.edu)

**THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL**

**UNCFRP**  
UNC Food Research Program  
Robert Wood Johnson Foundation  
University of North Carolina at Chapel Hill