Growing Up in Ireland
National Longitudinal Study of Children

Combining nutritional data from two surveys to augment dietary intake estimates

M.Crowe, M.O'Sullivan,O.Cassetti, A.O'Sullivan
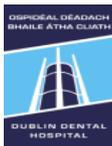
Introduction

1. Rationale for combining survey data

2. The data mapping process

3. Results for Foods 'Covered' / 'Not Covered'

4. Results for Sugar Analysis

5. Conclusions and Future Work

Rationale for combining survey data

- Increase information - limited resources

- Augment database with additional information from another source

- Improve precision

- Synergies from data combination

- Multidisciplinary benefits - mixed methods research

'Data Science'

Growing Up in Ireland
National Longitudinal Study of Children

- **Scientific query?**
- **What we want to estimate or predict?**
- **Goal if we had all the data?**

- **Data sampling**
- **Data Linkage?**
- **Data Quality**
- **Augment?**

Question?

Obtain Data

Modelling

Exploratory Data Analysis

- **Select technique**
- **Build model**
- **Evaluate model**
- **Train Model**
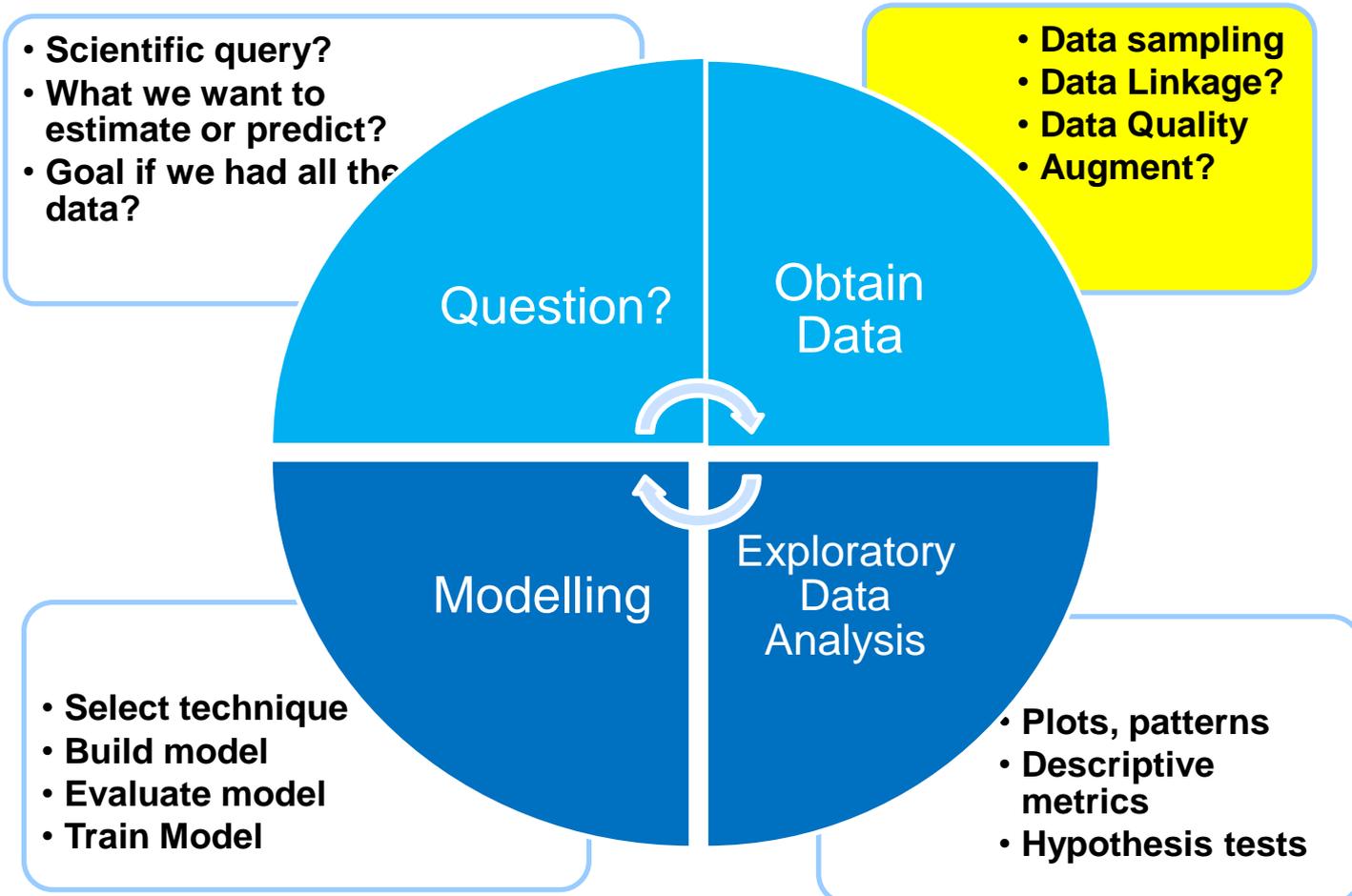
- **Plots, patterns**
- **Descriptive metrics**
- **Hypothesis tests**
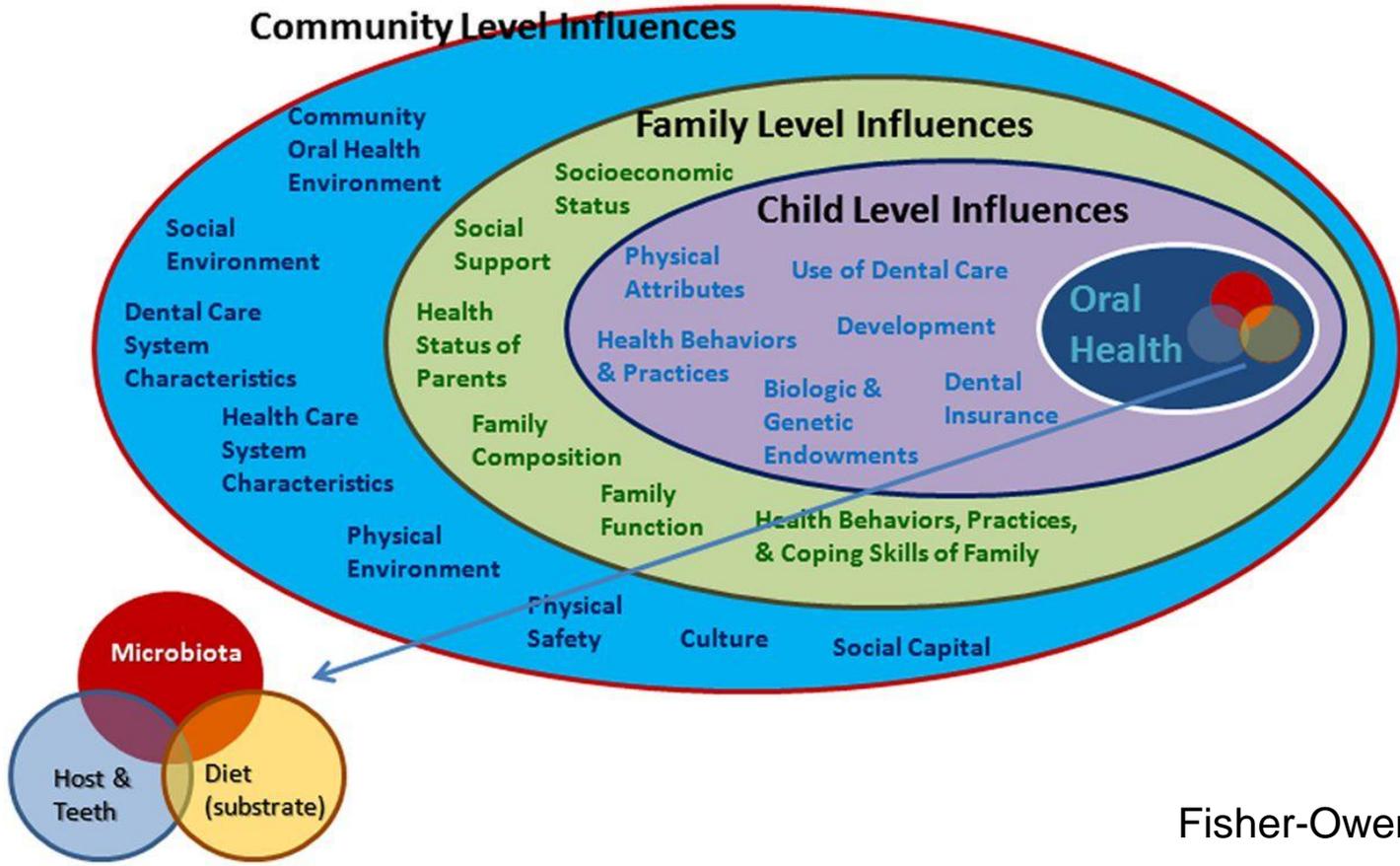
Fisher-Owens, 2007

Why augment?

- Diet-Health relationships[1]

- Decision Trees - food categories – GUI[2]

- Common Risk Factors: dental caries and obesity

- Improve accuracy of food intake data and reduce attenuation

[1] Crowe, M., et al. "Early Childhood Dental Problems Classification Tree Analyses of 2 Waves of an Infant Cohort Study." *JDR Clinical & Translational Research* (2016).
[2] Crowe, M., et al. "Dental problems and weight status in early childhood: classification tree analysis of a national cohort" (submitted)

Why augment?

- Foods were found to be low level predictors in Classification tree analysis for GUI infants at 3 years – why was this?

- Is the frequency or amount of food more important?

- Sugar – is there a link between dental caries and obesity?

- Aim of study- e.g. GUI v IUNA-NPNS

- Comparability of data, population, time frame

- All self report dietary instruments contain measurement error

- Describe usual daily mean intake distributions- frequency AND Weight

- Short term (24-HR) V long-term (FFQ)

Data sources

	NPNS	GUI
Sample size (n)	500 (126=3yo)	9,793
Study type	Cross- sectional	Longitudinal
Nationally representative	Yes	Yes
Date of survey	Oct 2010-Sept 2011	Dec 2010-July 2011
Food measurement tool	4 day weighed food diary	Modified FFQ

Methods-1

1. Primary data - GUI and NPNS (IUNA)

2. FFQ in GUI 15 food groups, NPNS had 77

3. Features were selected for food mapping using shallow Natural Language Processing (NLP)

4. Foods not covered by the GUI FFQ- part of risk

Methods-2

- GUI frequency of consumption defined for 0, 1, >1

- BMI, social class, food frequency categories chi-square proportion test and equivalence tests ($p<0.05$)

- Data files were imported from SPSS (IBM) and csv file formats to R (version 3.2.2) for linkage and analysis

C24. How old was <child> [in months] when he/she completely stopped being breastfed? ____Months
[Int: Only Accept answer in Months]

C25. [Card C25] In the last 24 hours has <child> had the following foods and drinks once, more than once, or not at all?

	Not At all	Once	More than Once
A. Fresh fruit	$\square_0$	$\square_1$	$\square_2$
B. Cooked vegetables	$\square_0$	$\square_1$	$\square_2$
C. Raw vegetables or salad	$\square_0$	$\square_1$	$\square_2$
D. Hamburger, hot dog, sausage or sausage roll, meat pie,	$\square_0$	$\square_1$	$\square_2$
E. Hot chips or French fries	$\square_0$	$\square_1$	$\square_2$
F. Crisps or savoury snacks	$\square_0$	$\square_1$	$\square_2$
G. Biscuits, doughnuts, cake, pie or chocolate	$\square_0$	$\square_1$	$\square_2$
H. Sweets	$\square_0$	$\square_1$	$\square_2$
I. Full fat cheese/yoghurt/ fromage frais	$\square_0$	$\square_1$	$\square_2$
J. Low fat Cheese/ low fat yoghurt	$\square_0$	$\square_1$	$\square_2$
K. Water (tap water / still water/ sparkling water)	$\square_0$	$\square_1$	$\square_2$
L. Fizzy drinks / minerals / cordial / squash (diet)	$\square_0$	$\square_1$	$\square_2$
M. Fizzy drinks / minerals / cordial / squash (not diet)	$\square_0$	$\square_1$	$\square_2$
N. Full cream milk or full cream milk products	$\square_0$	$\square_1$	$\square_2$
O. Skimmed/Semi-skimmed milk or Skimmed/Semi skimmed milk products	$\square_0$	$\square_1$	$\square_2$

Food and drink FFQ GUI

Food and drink items

Legend: Not at all / Once / More than once

Food and drink item	Not at all	Once	More than once
Fizzy drinks/minerals/cordial/squash (diet)	73.2%	11.4%	15.3%
Tap water, still, sparkling	17.1%	13.3%	69.6%
Skimmed/Semiskimmed milk or milk products	88.5%	4.4%	7%
Full cream milk or full cream products	15.5%	19.3%	65.1%
Low fat cheese/low fat yoghurt	91.7%	5.6%	2.6%
Full fat cheese/low fat yoghurt	15.1%	40.9%	43.9%
Raw vegetables or salad	74.8%	18.6%	6.6%
Cooked vegetables	15.2%	48.7%	36%
Fresh fruit	10.8%	29.9%	59.2%
Crisps or savoury snacks	52.8%	39.1%	7.9%
Hot chips or French fries	71.7%	26.3%	1.9%
Hamburger, hot dog, sausage, meat pie	63.7%	32.2%	4%
Fizzy drinks/minerals/cordial/squash (not diet)	70.3%	13.2%	16.4%
Sweets	51.3%	37.8%	10.9%
Biscuits, doughnuts, cake, pie or chocolate	25.8%	52.3%	21.8%

% consumptions

Data processing steps

Results

	SUBJECID	SURVDAY	non_gui_ct	non_gui_fwt	day_ct	day_fwt	ratio_ct	ratio_fwt
1	108	1	9	366	23	1025	0.3913043	0.35707317
2	108	2	12	203	22	1074	0.5454545	0.18901304
3	108	3	8	212	25	979	0.3200000	0.21654750
4	108	4	5	115	15	770	0.3333333	0.14935065
5	109	1	6	231	23	1062	0.2608696	0.21751412
6	109	2	7	138	23	925	0.3043478	0.14918919
7	109	3	9	484	24	1732	0.3750000	0.27944573
8	109	4	7	525	21	1520	0.3333333	0.34539474
9	158	1	11	359	22	1054	0.5000000	0.34060721
10	158	2	10	488	27	1086	0.3703704	0.44935543
11	158	3	10	482	26	1163	0.3846154	0.41444540
12	158	4	8	341	22	1134	0.3636364	0.30070547
13	159	1	9	489	22	1448	0.4090909	0.33770718
14	159	2	17	625	33	1627	0.5151515	0.38414259
15	159	3	11	482	22	1238	0.5000000	0.38933764
16	159	4	12	332	29	1504	0.4137931	0.22074468

Growing Up in Ireland
National Longitudinal Study of Children

Unmapped food weight

Unmapped food items
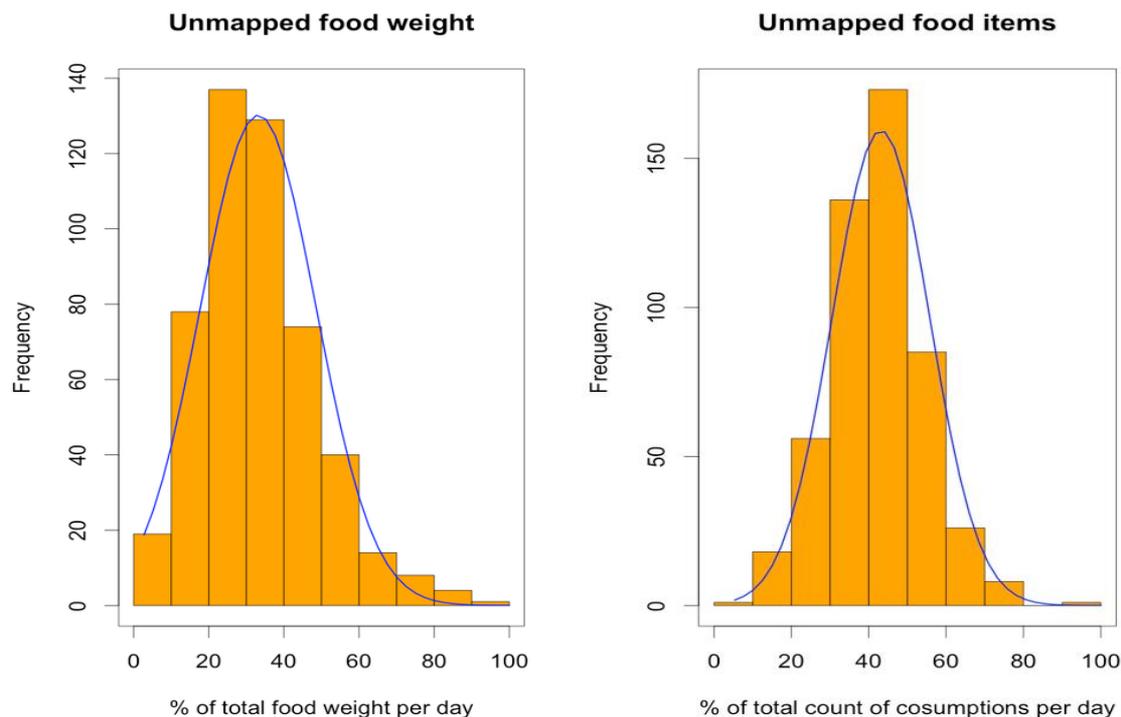
Histograms represent the distribution of the ratio of consumption counts* or weight of a food item consumed in IUNA that were not mapped by GUI.

* number of food consumptions not represented in GUI divided by the total number of foods consumed in a given day.

- Advantages and disadvantages of using FFQs

- Quantify bias in results of diet - health outcome

- Sufficient to analyse specific food category fully covered but need to establish foods uncovered

- Focus of this group is on sugars, in particular from a dental/weight status perspective

NPNS Total sugar by "GUI codes"

GUI food and daily total sugar
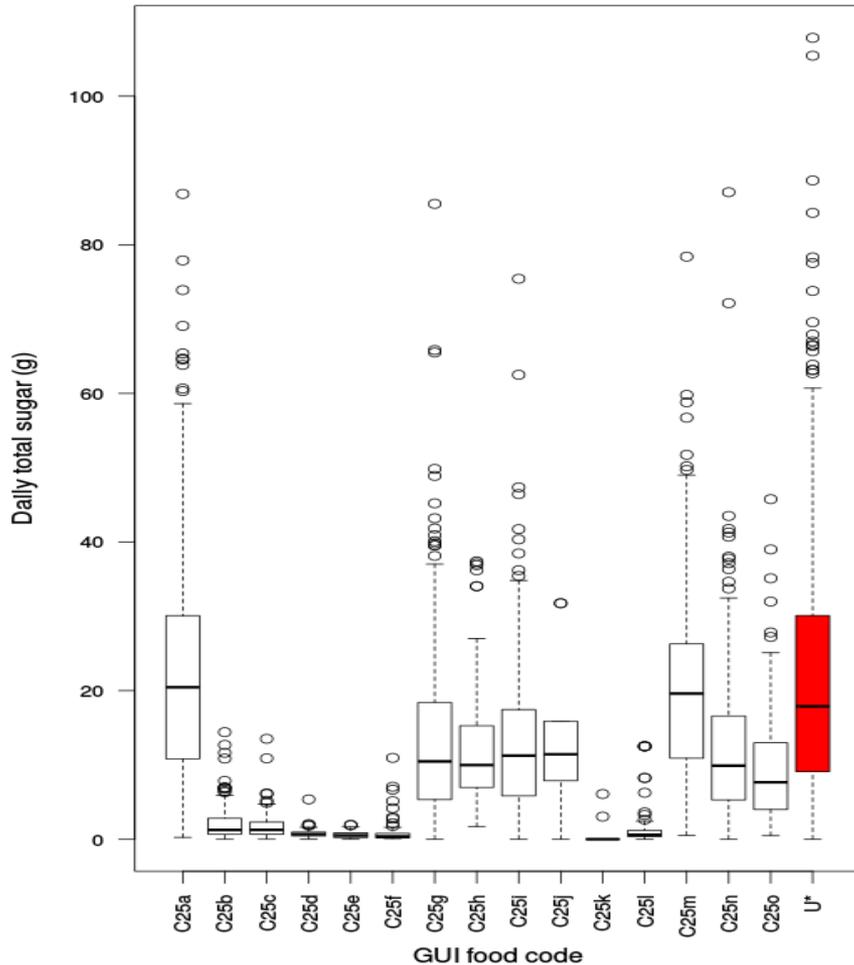
GUI food and daily total sugar

Total sugar-groups
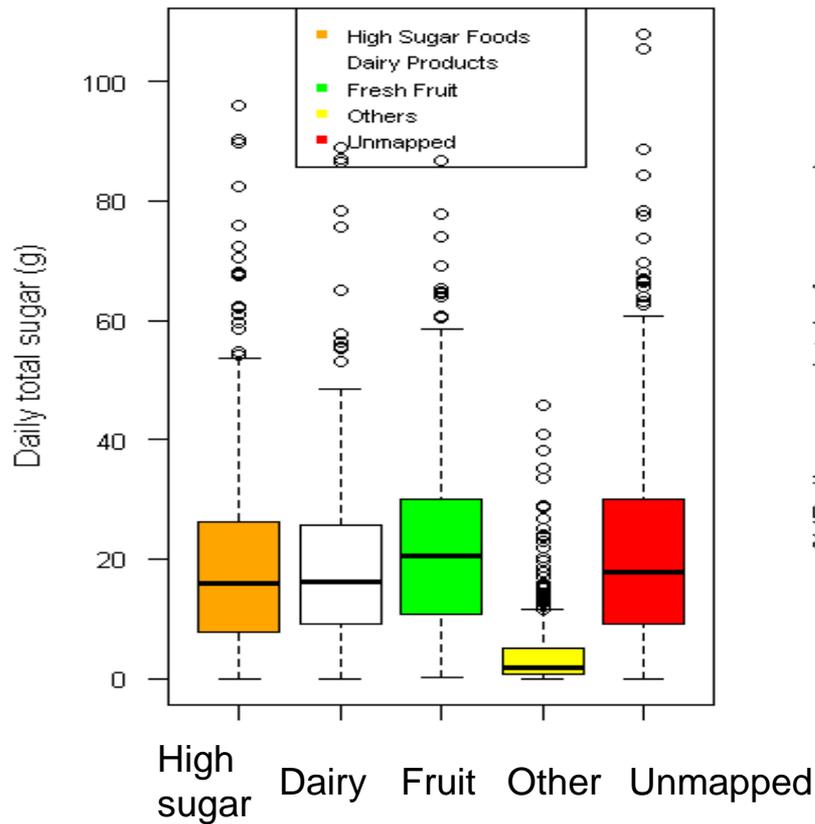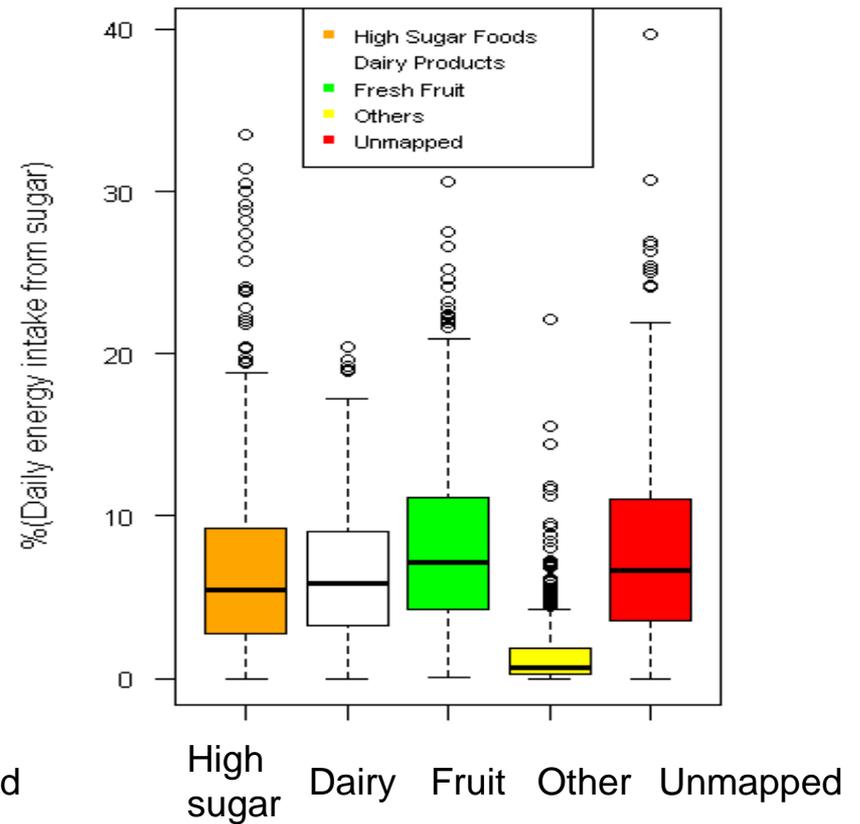
Conclusions

- Combining data surveys by mapping is useful

- Complex protocol - 'covered': food item dependant

- Mapping food categories allows us to increase the precision of food estimates

- Survey design and instrument selection should reflect priorities and anticipated outcomes

Conclusions

- Mapping of sugars will allow targeting of specific cariogenic foods

- Diet-disease relationships can be explored using continuous data

- Data linkage (Unique identifier)

- Inform policy food and oral health strategy

- Generate synthetic data (Monte Carlo simulation) with improved accuracy

- Re-run regression/CTA analyses with GUI data focusing on obesity and dental problems

- Predictive modelling long term goal

- 5 year old FFQ (dental problems-16%)

- Ability to use statistical modelling to investigate role of free sugars in dental problems and obesity

Acknowledgments

Thanks to:

- GUI infants and parents
- ESRI/GUI team
- IUNA/NPNS

Questions?

"I can prove it or disprove it! What do you want me to do?"

References

- Watt RG, Sheiham A. 2012. Integrating the common risk factor approach into a social determinants framework. Community dentistry and oral epidemiology. 40(4):289-296.

- Sheiham A, James WP. 2015. Diet and dental caries: The pivotal role of free sugars reemphasized. J Dent Res. 94(10):1341-1347.

- Schenker N, Raghunathan TE. 2007. Combining information from multiple surveys to enhance estimation of measures of health. Statistics in medicine. 26(8):1802-1811.

- Newens KJ, Walton J. 2016. A review of sugar consumption from nationally representative dietary surveys across the world. Journal of human nutrition and dietetics : the official journal of the British Dietetic Association. 29(2):225-240.

- Louie JCY, Moshtaghian H, Boylan S, Flood VM, Rangan A, Barclay A, Brand-Miller J, Gill T. 2015. A systematic methodology to estimate added sugar content of foods. European journal of clinical nutrition. 69(2):154-161.

- Hooley M, Skouteris H, Boganin C, Satur J, Kilpatrick N. 2012. Body mass index and dental caries in children and adolescents: A systematic review of literature published 2004 to 2011. Syst Rev. 1(1):57.

- Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, Tooze JA, Krebs-Smith SM. 2006. Statistical methods for estimating usual intake of nutrients and foods: A review of the theory. Journal of the American Dietetic Association. 106(10):1640-1650.

Total sugars

Mean sugar intake

Mean sugar intake (g)
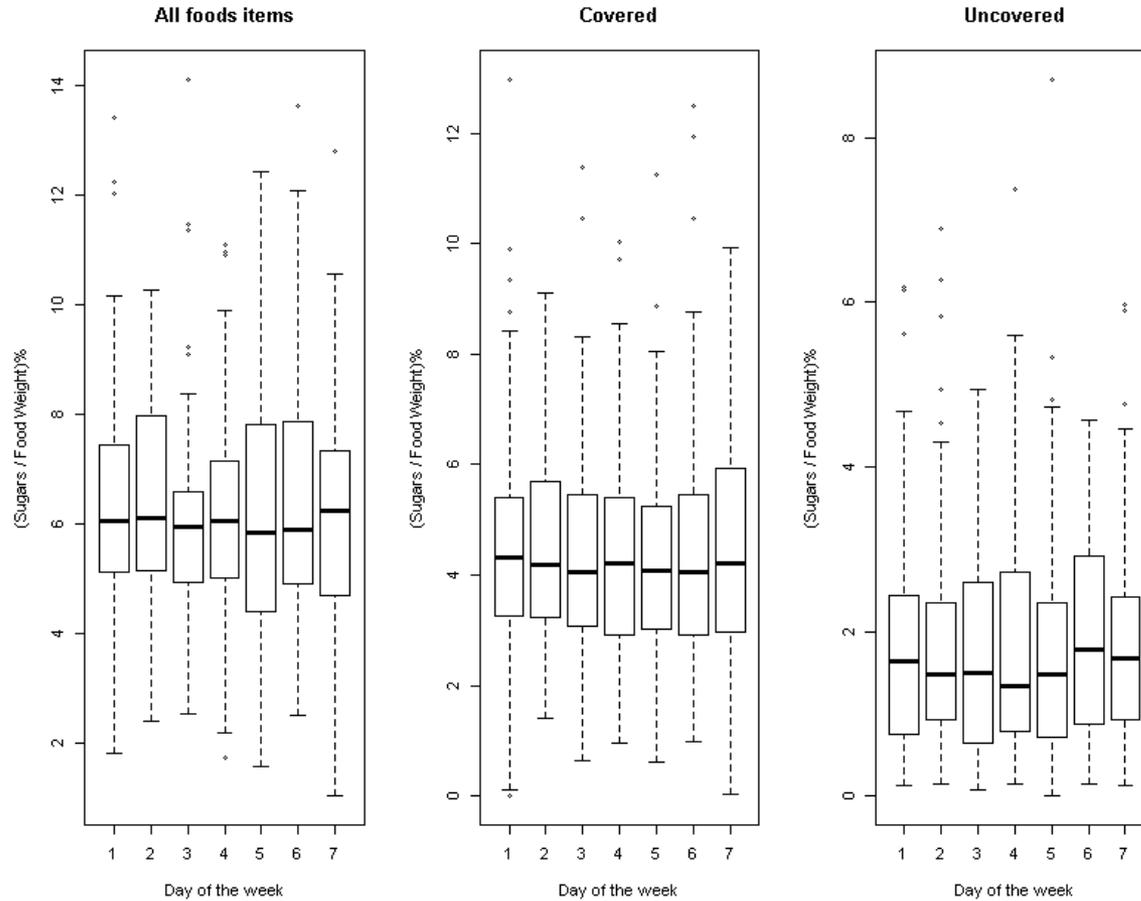
Mean % energy from sugars
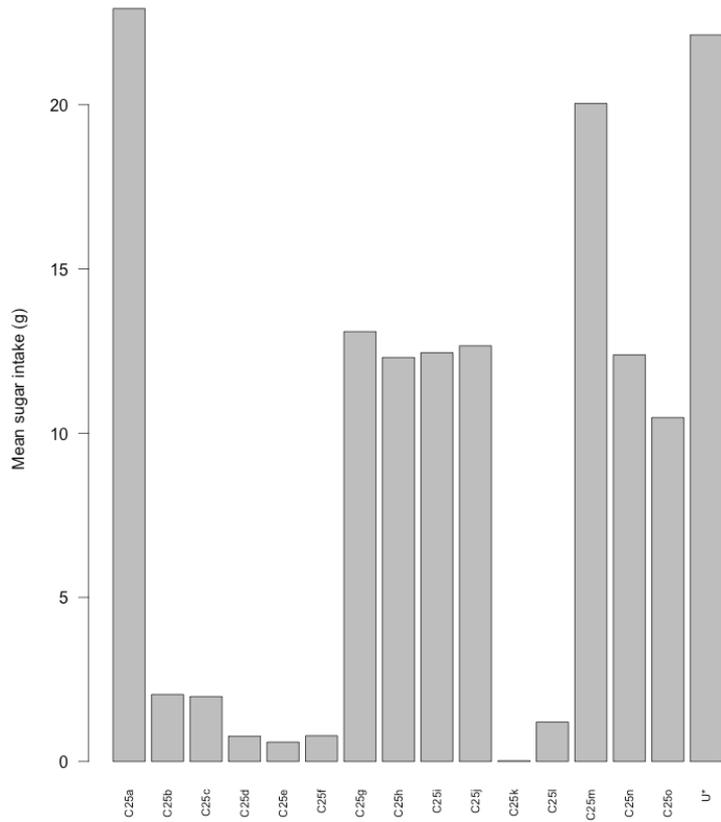
Mean % energy from sugars
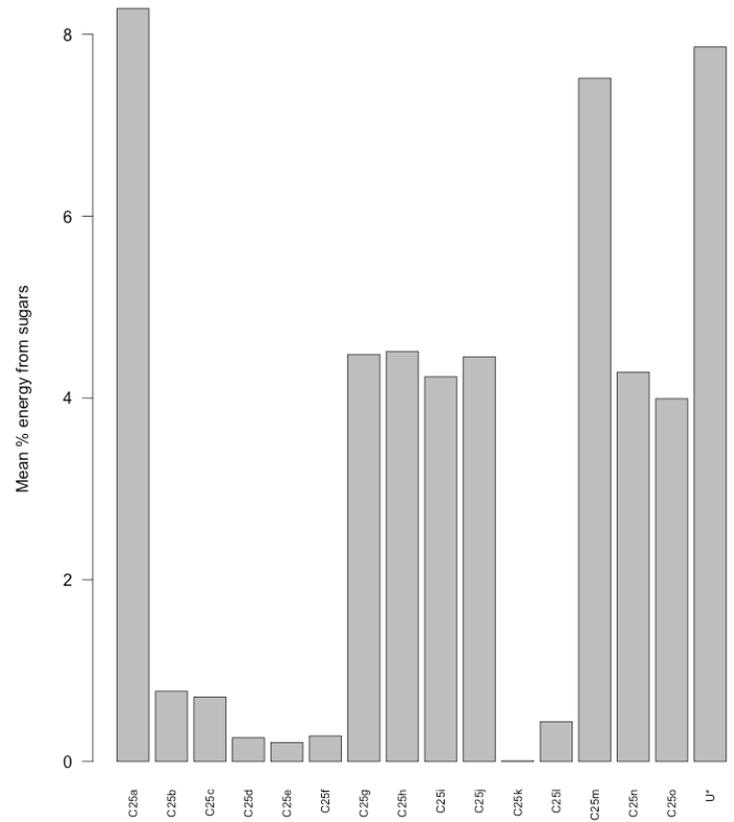
Sugar frequency consumptions

Histogram of uncovered_ct_sugars
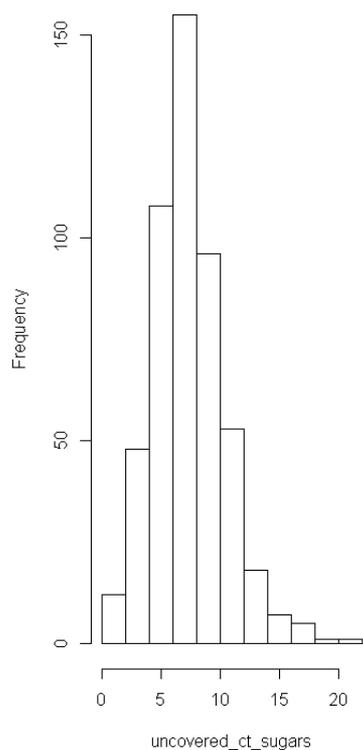
Histogram of covered_ct_sugars
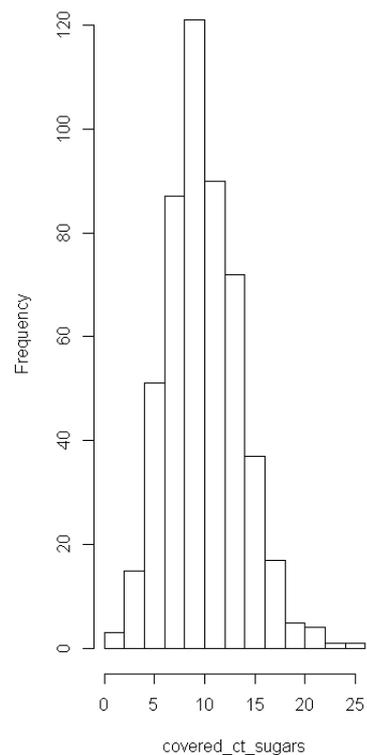
Top contributors by number of subjects
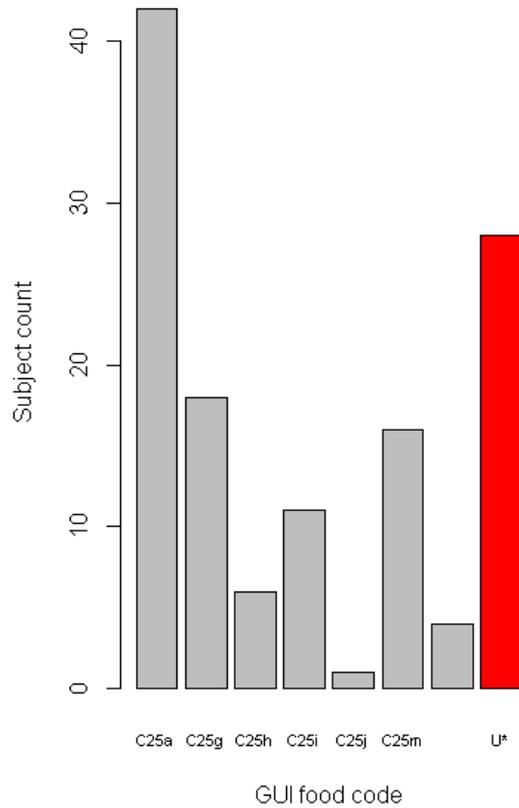
Top contributor by consumption
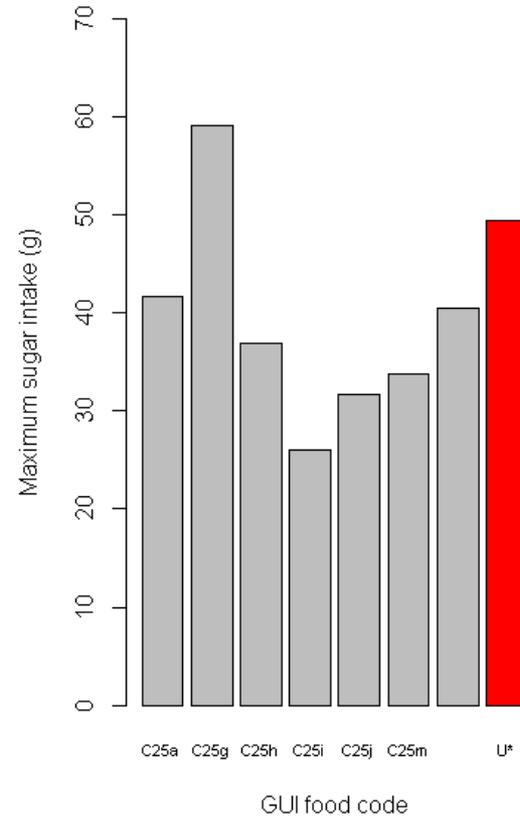
>P95 contributors by number of subjects
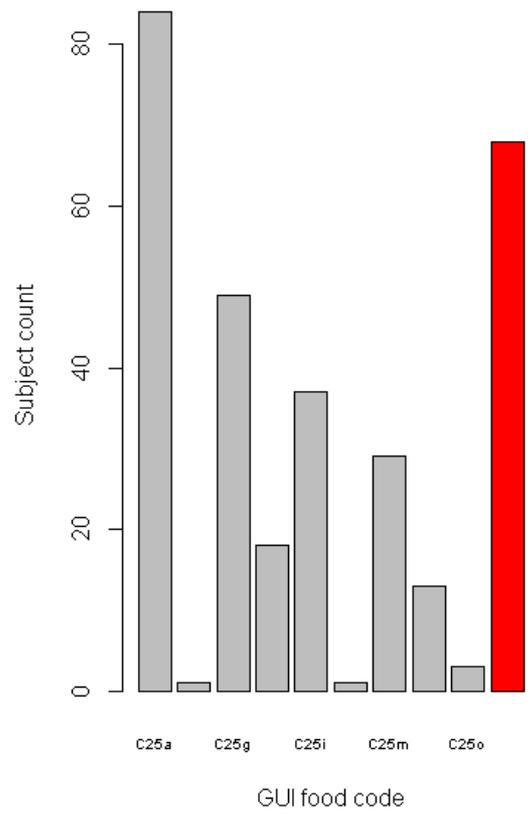
>P95 contributor by consumption
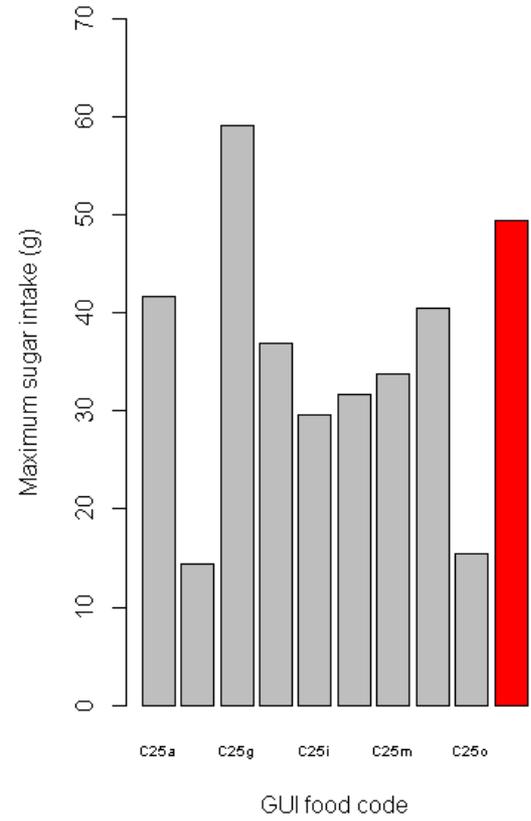
Mean Sugar intake
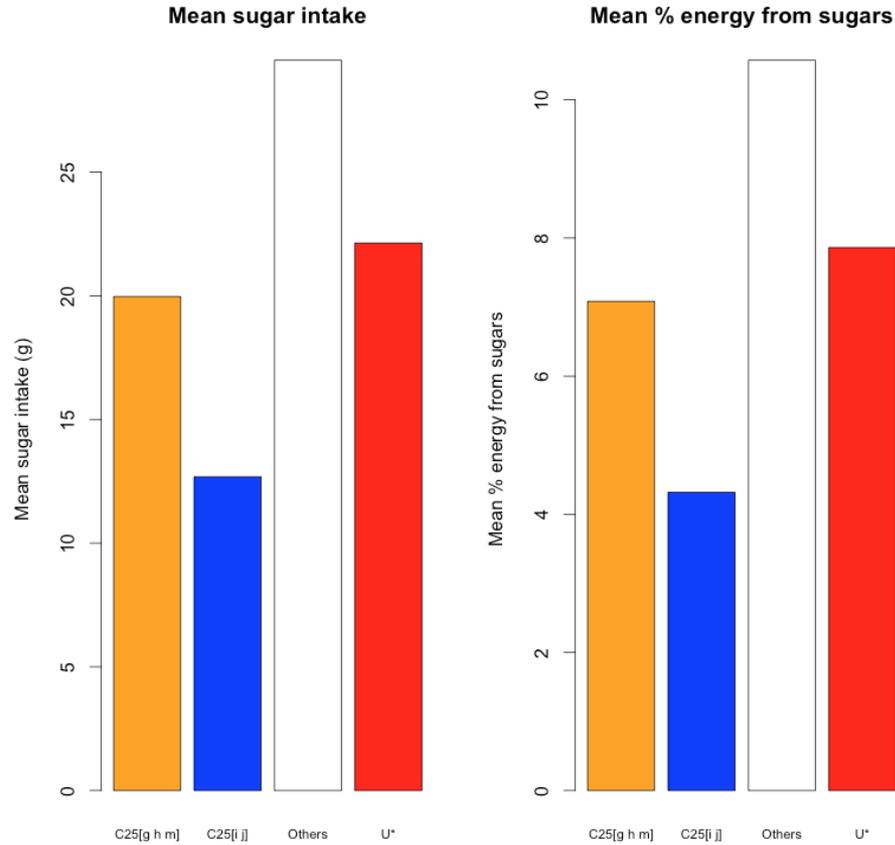
Mapping GUI

	A	B	C	D	E	F
1	FOODNAME	COOKINGMET	IUNA_NPNS_	CMETH	ct	GUI_CODE
33	Potatoes	Combination	25	12	1	
34	Potatoes	Unknown	25	13	1	
35	Processed & homemade potato products	Not Cooked	26	1	1	
36	Processed & homemade potato products	Grilled	26	2	12	
37	Processed & homemade potato products	Boiled	26	3	1	
38	Processed & homemade potato products	Fried/stir-fried	26	6	3	
39	Processed & homemade potato products	Baked	26	9	17	
90	Processed & homemade potato products	Combination	26	12	2	
91	Chipped, fried & roasted potatoes	Not Cooked	27	1	1	
92	Chipped, fried & roasted potatoes	Fried/stir-fried	27	6	13	C25e
93	Chipped, fried & roasted potatoes	Deep-Fat Fried	27	7	31	C25e
94	Chipped, fried & roasted potatoes	Roasted	27	8	21	C25e
95	Chipped, fried & roasted potatoes	Baked	27	9	38	C25e
96	Chipped, fried & roasted potatoes	Combination	27	12	1	C25e
97	Chipped, fried & roasted potatoes	Unknown	27	13	2	C25e
98	Vegetable & pulse dishes	Not Cooked	28	1	8	C25c
99	Vegetable & pulse dishes	Boiled	28	3	12	C25b
00	Vegetable & pulse dishes	Stewed	28	5	1	C25b
01	Vegetable & pulse dishes	Fried/stir-fried	28	6	2	C25b
02	Vegetable & pulse dishes	Roasted	28	8	1	C25b
03	Vegetable & pulse dishes	Baked	28	9	1	C25b
04	Vegetable & pulse dishes	Combination	28	12	3	C25b
05	Peas, beans & lentils	Not Cooked	29	1	7	C25c
06	Peas, beans & lentils	Boiled	29	3	67	C25b
07	Peas, beans & lentils	Steamed/blanche	29	4	5	C25b
08	Peas, beans & lentils	Stewed	29	5	2	C25b
09	Peas, beans & lentils	Fried/stir-fried	29	6	1	C25b
10	Peas, beans & lentils	Microwaved	29	10	5	C25b

- Ethnicity most NB predictor of Dental problem

- Highest prev. Dental Problems: Children, Irish, obese/underweight with longstanding illness and PCG BMI>24.9

- Food: Low fat cheese/yoghurt. Raw veg/salad, Fresh fruit, French fries - levels 3 and 4 predictors

- Sociodemographic: HH Annual Income