Data requirements for statistical evaluation in nutrition research

Áine Hearty, M.Sc.
Institute of Food & Health
UCD, Dublin
Ireland



Basically, what is required in terms of the data is completely dependent on the research question



Overview

- Dietary data - survey sample, data characteristics
- Exposure assessments
  - Specific data requirements
  - Types of exposure assessments
  - Few examples

Sample characteristics

- Population demographics – representative of population
- Dietary intake and exposure varies with age and physiological status
        - Infants
        - Children
        - Adolescents
        - Pregnancy and Lactation
        - Adults
        - Elderly

- Dietary assessment methods
- Under-reporting
- Survey duration

Dietary assessment methods

1. **Indirect methods:** Economically-derived indicators
   - Household budget
   - Food balance sheets
     - Relatively inexpensive
     - No individual data, no consumer-only data

     ? Limited value estimating chemical exposure

Dietary assessment methods

2. **Direct methods**: Survey based

- Qualitative: FFQ
- Retrospective: 24/48-hour recall, diet history
- Prospective: Food diary
  - Can determine proportion of population who are consumers

Under-reporting

- Common feature of dietary surveys
- May need to account for this by application of equations e.g. Goldberg cut-off based on EI:BMR$_{est}$ (Goldberg *et al*., 1991)
  - Ireland *(McGowan et al., 2001):* 25%
  - Sweden *(Becker, 1999):* 26%
  - UK NDNS *(Gregory et al., 1990):* 49% women, 29% men
- If required can remove subjects who have mis-reported their energy intake
- Best solution – improve study protocol to reduce this being a main feature of the dataset

Survey duration

- Food consumption data influenced by duration
- Study: survey duration 14 days, 948 adolescents, 5 EU countries

	Total population g/day			**Consumers only g/day (%cons)**		
No. days	**1**	**7**	**14**	**1**	**7**	**14**
Apples 	42	34	34	161 (26%)	59 (58%)	49 (68%)
Fish 	13	13	12	112 (12%)	24 (49%)	18 (63%)

Lambe J et al., (2000) EJCN 54, 166-173

Data characteristics

- Food categorization system
- Food composition data (e.g. fortified products, supplements)
- Access to raw data per subject in dataset
  - These type of data most accessible via food diaries
  - FFQs, diet recall methods usually more aggregated
- Extreme values and outliers
- Total population *vs* consumers only

Food Categorization

Raw data should be as disaggregated as possible

- Example 1. Additive X in Wafer-biscuits only
  - Level 1: Snacks
  - Level 2: Biscuits
  - Level 3: Wafer-Biscuits
  - Level 4: Brand of Wafer-Biscuits

- Example 2. Colour Y in Red Lemonade only
  - Level 1: Beverages
  - Level 2: Carbonated Beverages
  - Level 3: Red Lemonade
  - Level 4: Brand of Red Lemonade

Exposure assessment

Purpose: to provide a quantitative evaluation of the likely exposure to risk sources from one or more media

Allows risk assessors to characterise the hazard in the context of real life. Evaluates probability of occurrence of adverse effects

Required:

- information about sources & routes of exposure
- levels & duration of exposure

Sources of data for use in exposure assessments

- Food consumption data
- Recipe data
- Food ingredient databases
- Crop conversion databases
- Market share, brand loyalty
- Concentration databases
- Processing factors data

Methods for exposure assessment

- Dependant on the data available
- Number of initiatives aimed at standardising principles & terminology *(E.C., 1998; Kroes et al., 2002)*
- Stepwise tiered approach. Can stop when possibility of concern ruled out

- Budget method
- Back calculations
- Theoretical maximum daily intake (TMDI)
- Total diet studies/Market basket
- Deterministic – individual level
- Probabilistic models

Scientific Model

Chemical Intake =

$\sum$ **Raw Food Amount** × **Presence Probability** × **Chemical Concentration**

Raw Food Amount:

Actual Intake
*
Recipe Fraction Correction
*
Edible Portion Correction
*
Raw Product Correction

Presence Probability:

Lab data
*
Literature
*
Industry

Chemical Concentration:

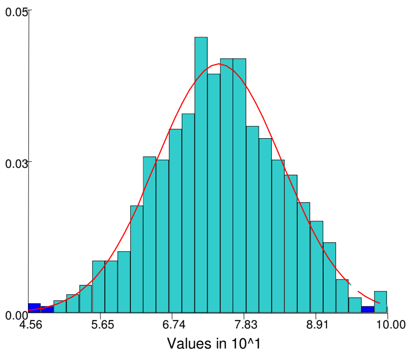
Assay Concentration
*
Processing Correction
*
Limit of Reporting Correction
*
Limit of Detection

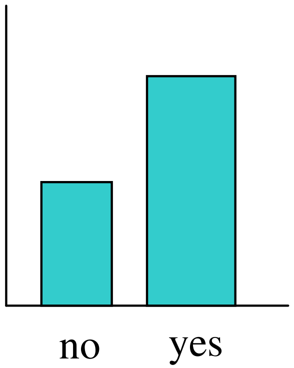
Deterministic V Probabilistic Modelling

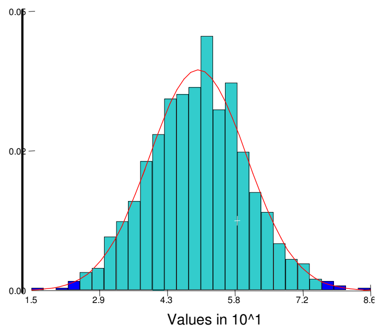
Food intake ☆ Presence probability ☆ Chemical concentration = Exposure

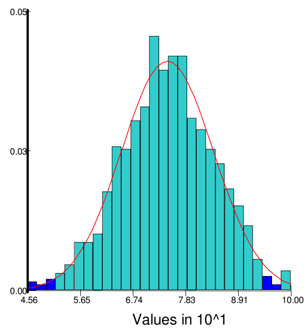
Deterministic approach: Fixed — 100% — Fixed — Fixed

Probabilistic approach









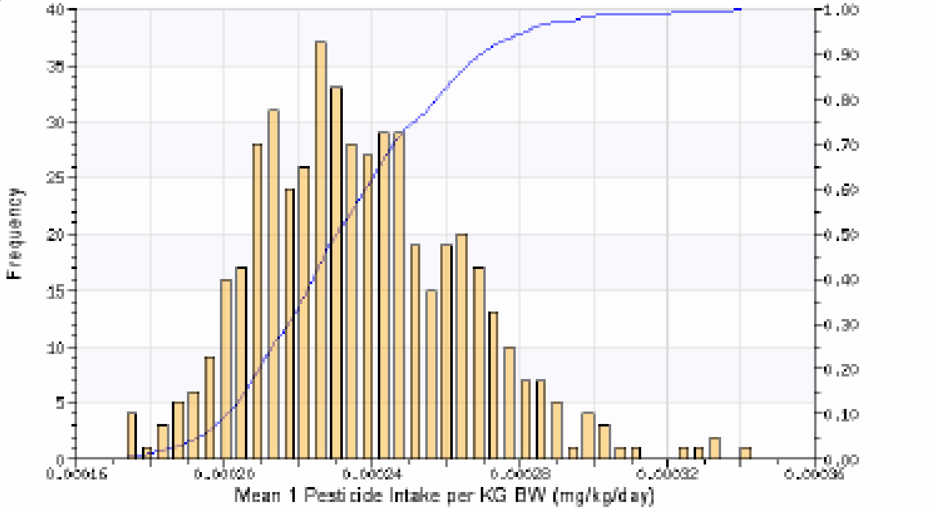
Probabilistic techniques

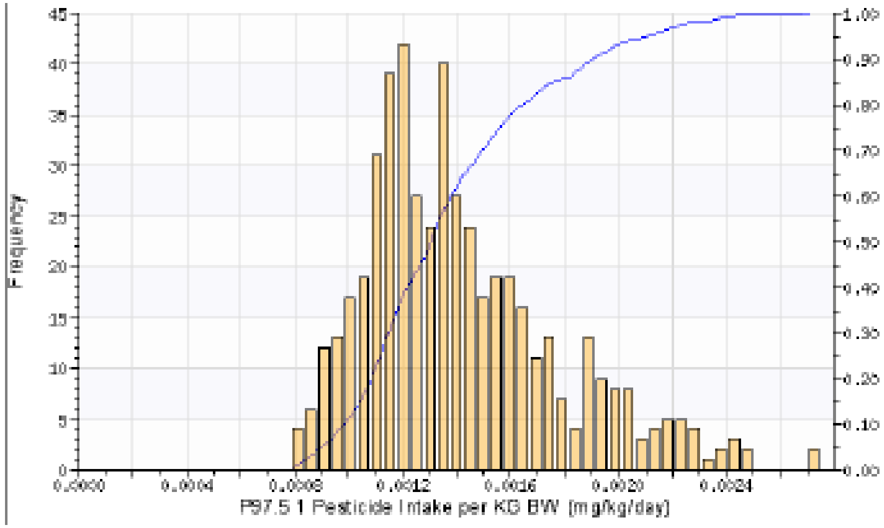
Take into account variability and uncertainty

- More refined and more realistic estimate of exposure
- Eliminates creeping conservatism
- Generates a distribution of exposures
- Assess sensitivity of exposure to input variables
- Allows complete use of available data

Probabilistic output

Exposure to pesticide Iprodione mg/kg body wt/day





Distribution around the mean

Distribution around 97.5th percentile

Variability & Uncertainty

Ignoring the presence of variability and uncertainty may generate exposure estimates that do not adequately reflect true chemical intakes

A good survey design allows to reduce uncertainty and variability

Variability

- Variability represents diversity that is irreducible by additional measurements
- **Food intake**: Natural variability in food intakes, both within and between individuals
- **Body weight**: Variability between individuals' body weight
- **Contaminant**: Variability in target chemical concentrations between individual foods

Uncertainty

Uncertainty represents partial ignorance or lack of knowledge that may be reduced by further measurement

Measurement uncertainties

- Food intakes, under/over-reporting
- Effects of cooking, processing
- Knowledge re presence of contaminant in food

Sampling uncertainty

- Small sample
- Representative of region, demographics etc
- Variation in season, year etc

Market share & Brand loyalty

Market share: proportion of the consumption level of a brand with respect to all brands of the same product.

Brand loyalty: consumers’ tendency to repeat the purchase of a brand.

Consumer behaviour with respect to brands has an impact on exposure *(Arcella et al., 2003)*

Modelling

Modelling is a catch-all phrase that usually means any type of activity where you are trying to create a representation of a real life situation so you can analyse it

- Need to outline what inputs affect your exposure to a chemical

- Need to have data for these inputs

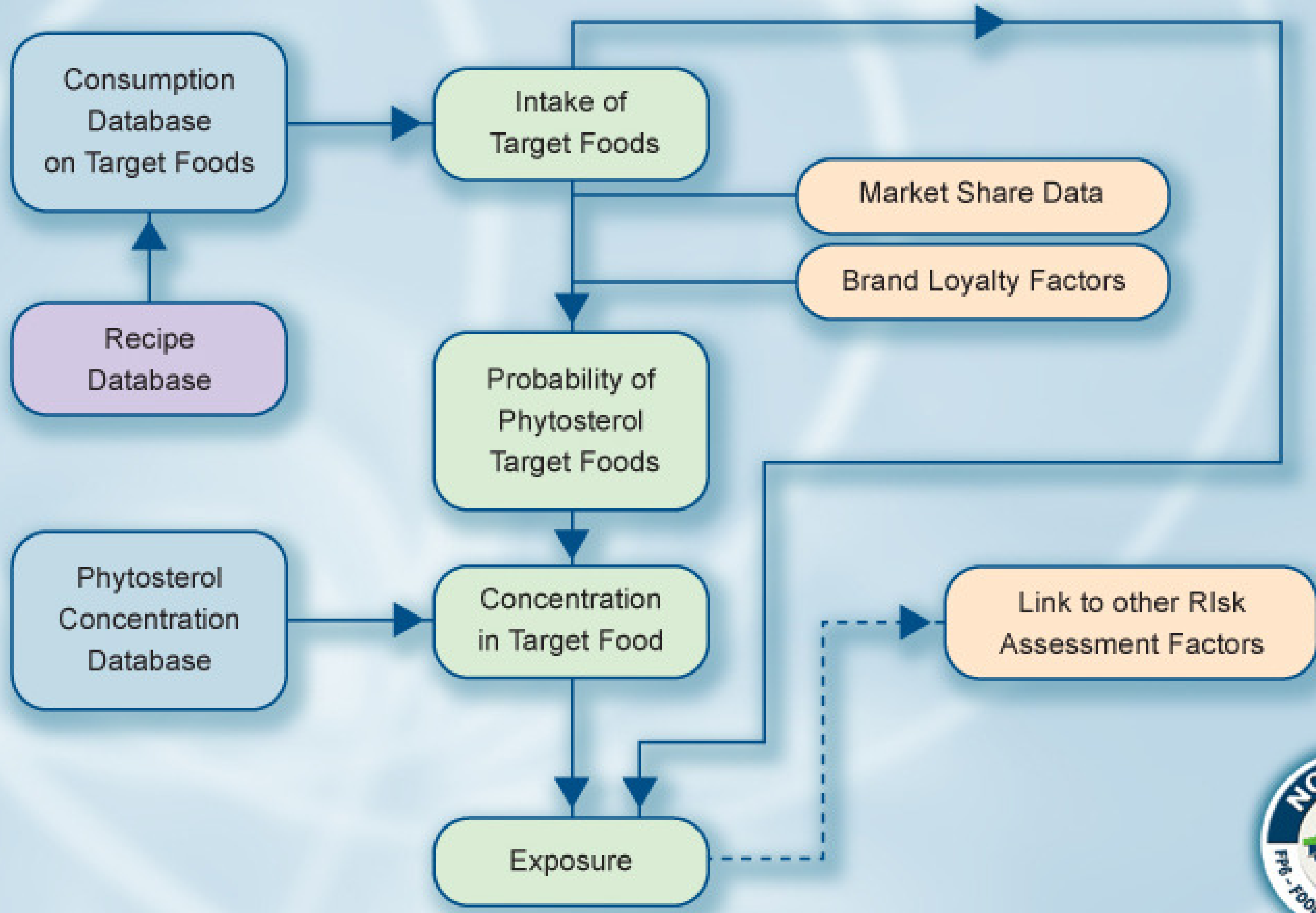
Conceptual models

‘a set of rules governing the manner in which independent datasets are linked in the computation of food chemical exposure’

To develop conceptual models:

- ✓Nutritionists
- ✓Food intake experts
- ✓Food technologists
- ✓Statisticians
- ✓Analytical chemists

Conceptual model - phytosterols

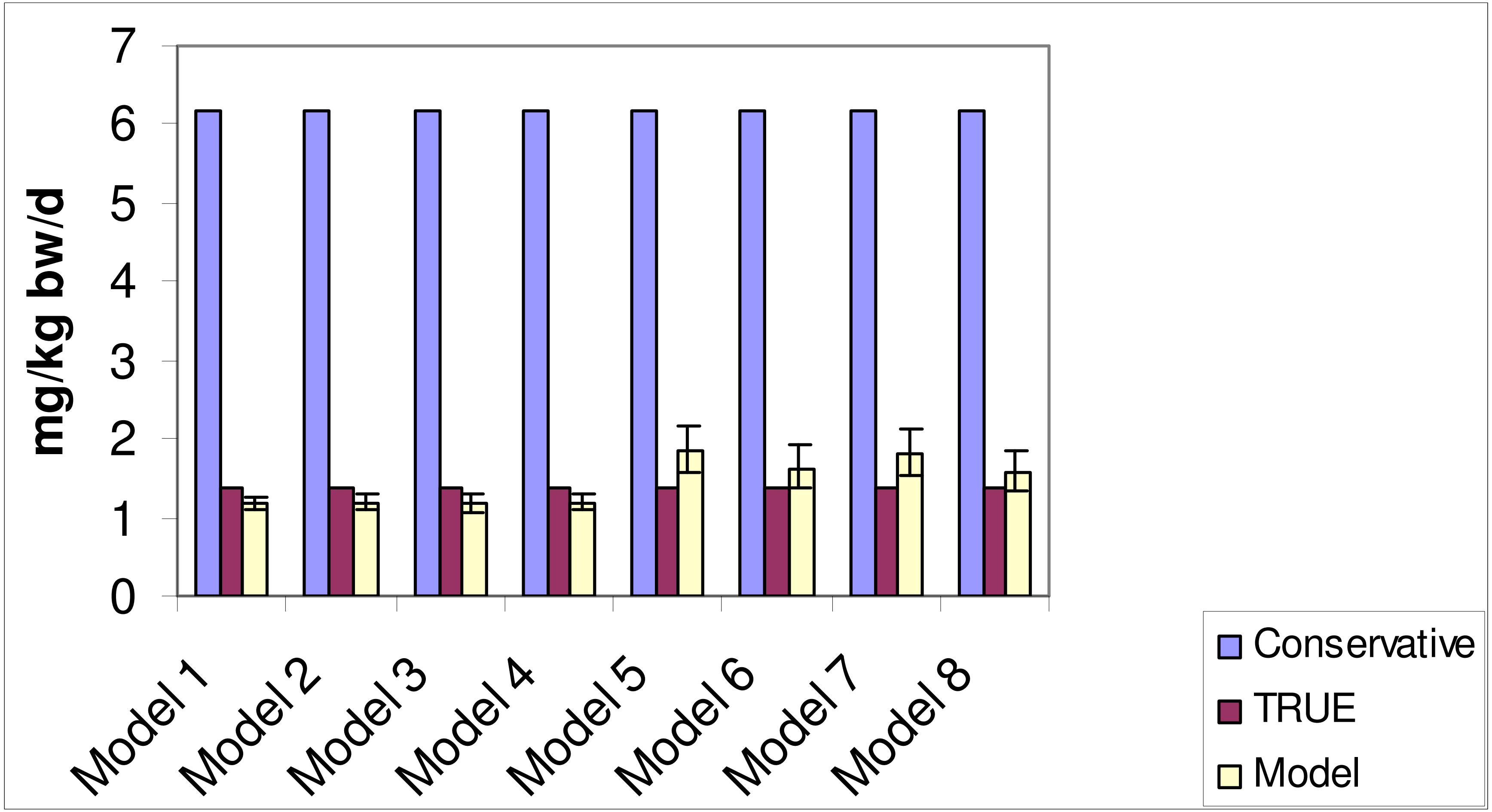


Monte Carlo project: Model validation

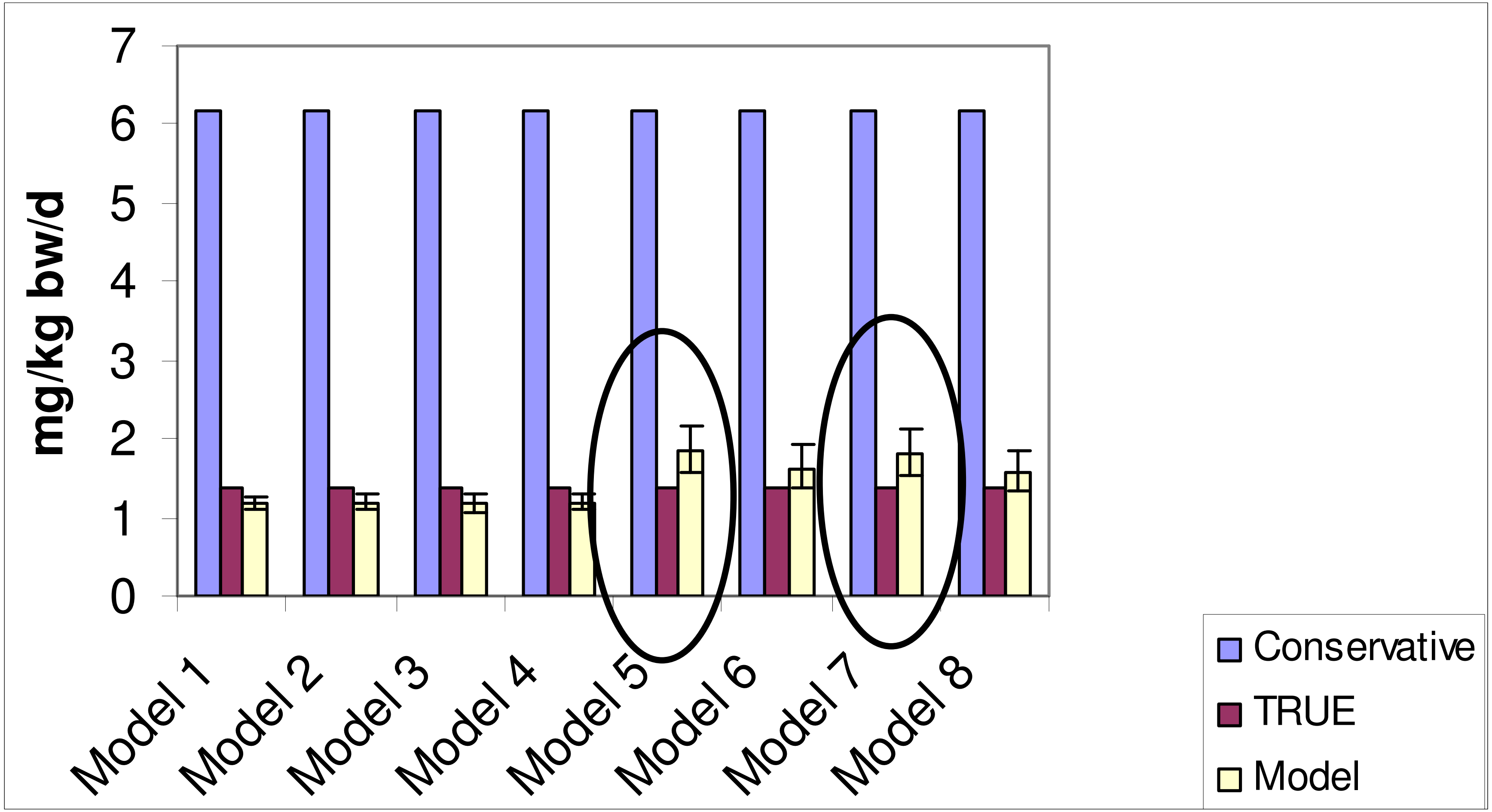
- Create a database which has no uncertainty
- Calculate “true” exposure from this
- Compare modelled exposure
- Accept if “modelled” > “true” but < “deterministic”
- Evaluate on a “fit for purpose” basis

Monte Carlo

P97.5 Total population intakes of sorbates



P97.5 Total population intakes of sorbates



Occurrence data

- Often major source of uncertainty is knowledge of the actual occurrence of the chemical of interest
- Food ingredient databases - provide occurrence data e.g INFID (*Gilsenan et al., 2002*)
- FP7 project - FACET, 2008-2012

Food ingredient database
Irish example: INFID

- Normally, food intake studies do not record brand level data.

- Thus Coca Cola & Pepsi Cola are “Carbonated soft drinks containing sugars”.

- In Ireland, we collect brand level data and brand packaging

- From brand packaging we get brand ingredients and brand specific nutrition information

Field work

Coordinating research centre UCD

Brand

Food label data

INFID database

Application of INFID
1. Children's exposure to food additives

- Considerations:
  - If an additive is legally permitted in a food category
  - Traditionally, we always assume 100% usage in that category
- INFID - true probability of occurrence
- Case Study: children's exposure to food additives and to food contact materials

Food additives and hyperactive behaviour in 3-year-old and 8/9-year-old children in the community: a randomised, double-blinded, placebo-controlled trial

Donna McCann, Angelina Barrett, Alison Cooper, Debbie Crumpler, Lindy Dalen, Kate Grimshaw, Elizabeth Kitchin, Kris Lok, Lucy Porteous, Emily Prince, Edmund Sonuga-Barke, John O Warner, Jim Stevenson

Lancet 2007; 370: 1560–67

Published Online September 6, 2007 DOI:10.1016/S0140-6736(07)61306-3

See Comment page 1524

See Department of Error page 1542

School of Psychology (D McCann PhD, A Barrett BSc, A Cooper MSc, D Crumpler BSc, L Dalen PhD, E Kitchin BSc, K Lok MSc, L Porteous BSc, E Prince MSc, Prof E Sonuga-Barke PhD, Prof J Stevenson PhD) and School of Medicine (K Grimshaw MSc), Department of Child Health, University of

Summary

Background We undertook a randomised, double-blinded, placebo-controlled, crossover trial to test whether intake of artificial food colour and additives (AFCA) affected childhood behaviour.

Methods 153 3-year-old and 144 8/9-year-old children were included in the study. The challenge drink contained sodium benzoate and one of two AFCA mixes (A or B) or a placebo mix. The main outcome measure was a global hyperactivity aggregate (GHA), based on aggregated z-scores of observed behaviours and ratings by teachers and parents, plus, for 8/9-year-old children, a computerised test of attention. This clinical trial is registered with Current Controlled Trials (registration number ISRCTN74481308). Analysis was per protocol.

Findings 16 3-year-old children and 14 8/9-year-old children did not complete the study, for reasons unrelated to childhood behaviour. Mix A had a significantly adverse effect compared with placebo in GHA for all 3-year-old children (effect size 0·20 [95% CI 0·01–0·39], p=0·044) but not mix B versus placebo. This result persisted when analysis was restricted to 3-year-old children who consumed more than 85% of juice and had no missing data (0·32 [0·05–0·60], p=0·02). 8/9-year-old children showed a significantly adverse effect when given mix A (0·12 [0·02–0·23], p=0·023) or mix B (0·17 [0·07–0·28], p=0·001) when analysis was restricted to those children consuming at least 85% of drinks with no missing data.

McCann et al., Methods

- 153 3-year old and 144 8/9 year old UK children given a juice cocktail of 6 colours & 1 preservative or a placebo juice
- Study design: 6 week ‘additive-free’ diet, with fortnightly challenges with either an additive mix (twice) or placebo (once)
- Outcome: Those exposed to the 6 colour & 1 additive cocktail were more “hyper”

Additives Investigated

Colours:

- E129 Allura Red
- E110 Sunset yellow
- E122 Carmosine
- E124 Ponceau 4R
- E102 Tartrazine
- E104 Quinoline Yellow

Preservative:

- E211 Sodium Benzoate

Additives investigated: frequency in the Irish diet?

	Irish children: 72,024 eating events
Colours:	
○ E129 Allura Red	n=343
○ E110 Sunset yellow	n=557
○ E122 Carmosine	n=443
○ E124 Ponceau 4R	n=544
○ E102 Tartrazine	n=138
○ E104 Quinoline Yellow	n=539
Preservative:	
○ E211 Sodium Benzoate	n=2188

Consumption of targeted additives

Of total foods consumed (72,024), how many times were target additives eaten?

0% of any food contained all 7 additives

Of total meals consumed (19,795), how many times were target additives eaten?

Only 16% of all meals/snacks contained >1 additive

Application of INFID 2: Food contact materials

- INFID retains data on food packaging
- National Food Packaging Database
- Packaging usage linked to food consumption
- Globally unique database
- More precise estimates of food packaging migratory compound exposure

FCM in foods consumed by Irish children

Packaging Form	% Occurrence
Metal	7.1
Paper & board	7.4
Plastic	75.9
Combinations	9.5

Duffy E et al., (2006). Food Additives & Contaminants, 23(6): 623-633

FCM – Types of plastics

Plastic packaging	% Occurrence
Polethylene (PE)	41.6
Polypropylene (PP)	35.7
Polethylene terephthalate (PET)	15.8
Polystyrene (PS)	5.6
Polyvinyl chloride (PVC)	0.8
Others	0.5

Acute *vs.* Chronic exposure

Chronic exposure:

Safety limit = Acceptable Daily Intake (ADI)

Intake assessed as mean intake within a long period of time

Prolonged intake of a chemical agent at relatively lower levels may lead to accumulation – chronic toxicity

Acute exposure:

Safety limit = Acute Reference Dose (ARfD)

Exposure assessed within a restricted time interval (i.e. meal or day)

Extremely high levels of contamination required to produce acute adverse response from a single ingestion

Food chemicals - exposure assessments

- Additives
- Pesticides
- Sweeteners
- Contaminants
- Flavouring substance
- Novel ingredients
- Packaging material
- Nutrients

Example - Pesticides

- Should be included in exposure assessment:
  (EPA 2000, Crossley 2000, Peterson 2000)
  - Concentrations per food group
  - Require consideration of non-detects
  - Variability within composite samples
  - Processing factors
  - Crop conversion factors

Pesticides: Limit of reporting (LOR)

- Cannot assume that absolutely no pesticide present as test may not be sensitive enough
- All tests have a specific limit of reporting
- Important because the majority of samples have levels < LOR
- For non-detects, the LOR can be applied
- Assign actual LOR, 0.5xLOR or use distribution of values

Pesticides: Variability

- Variability factors describe the variability of pesticide between units within batches
- Within exposure assessment software, often the capability to deal with accounting for potential variability
- Use estimated values, default (conservative), or use weight ratios to define max variability of residue levels in consumed portions

Pesticides: Crop conversion database

Food name	Conversion factor		Crop
Lemon juice 100g	1.2		Lemon 120g
Cider 100g	2.0		Apple 200g

Pesticides: Effects of processing

Compound	Product	Processing	Proc. Factor
Primicarb	Apple	Peeling	0.63
Iprodione	Grape	Drying	2.63
Iprodione	Grape	Washing with water	0.64

Summary

- Data itself central to use in exposure assessments e.g. quality, methodology, food grouping
- Method of exposure assessment dependant on data and required output
- Conceptual models & model validation key role in process
- Occurrence data often missing and central to accurate assessments
- Certain substances need specific data to run exposure assessments e.g. pesticides

Thank you for listening

QUESTIONS?