

Harnessing Census Microdata

Dr Barry Leventhal,
BarryAnalytics Limited

MRS CGG Seminar
5th November 2014



Agenda

- Introduction to Census Microdata
- Microdata products from the UK Census
- Case study applications
 1. Profiling and targeting
 2. Demographic segmentation
 3. Estimating market sizes
- Use cases in market research
- Accessing Census Microdata
- Conclusions

Census Microdata are...

- Samples of anonymised records selected at random from the Census output database
 - Containing information captured from census forms
 - But no personal identifiers (names, addresses, DOBs)
 - And no potentially disclosive data
 - Individual and household files
- Individual file – random sample of people
- Household file – random sample of households
 - Includes data on all persons in each sampled household

Microdata vs. Traditional Census Output

Traditional
Census Output:
Sets of tables

LC6124EW - Approximated social grade by sex by age
ONS Crown Copyright Reserved [from Nomis on 12 August 2014]

Social Grade	All categories: Age 16 to 64	Age 16 to 24	Age 25 to 34	Age 35 to 44	Age 45 to 54	Age 55 to 64
All categories	62,884	10,484	12,328	14,289	14,465	11,318
AB Higher	18,403	2,412	3,789	4,951	4,370	2,881
C1 Supervisory	21,775	3,998	4,110	4,619	5,106	3,942
C2 Skilled	12,036	2,056	2,397	2,682	2,763	2,138
DE Semi-skilled	10,670	2,018	2,032	2,037	2,226	2,357

Microdata:
Variables for
each case,
similar layout
to sample
survey data

Record Number	Sex	Age	Marital Status	Economic Activity	Occupation	Industry	Hours Worked Per week	Approximate Social Grade
000001	F	25-34	Married	Employee	Caring Profn	Health	FT	C1
000002	M	55-64	Married	Self employed	Professional	Financial	FT	AB
000003	F	0-15	Single	Missing	Missing	Missing	Missing	Missing
000004	F	25-34	Separated	Employee	Secretarial	Transport	PT	C1
000005	F	45-54	Married	Employee	Secretarial	Education	PT	C1
000006	M	45-54	Married	Unemployed	Manager	Real Estate	FT	AB
000007	F	0-15	Single	Missing	Missing	Missing	Missing	Missing
000008	M	65-74	Married	Retired	Skilled	Agriculture	Missing	C1
000009	M	65-74	Divorced	Retired	Skilled	Transport	Missing	C1
000010	M	75+	Widowed	Retired	Sales	Construction	Missing	AB

Pro's and Cons of Microdata

Pro's:

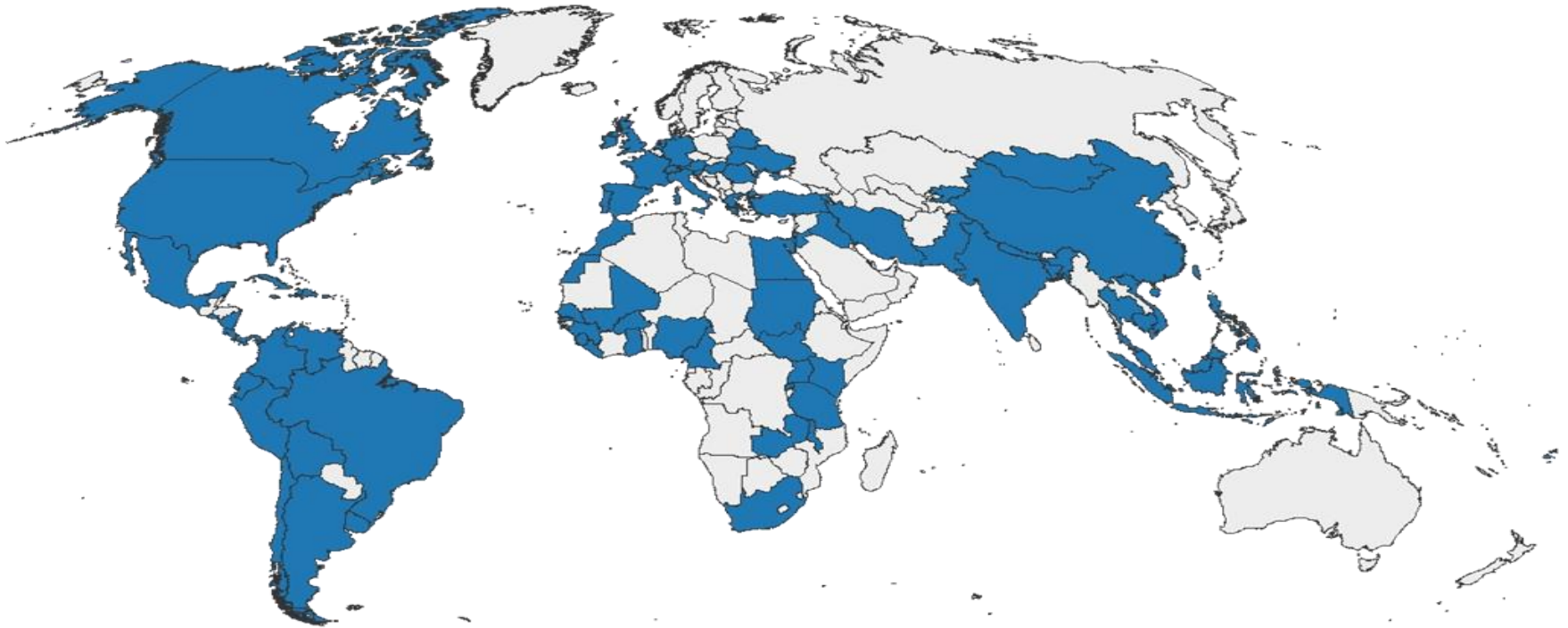
- More flexible analysis possible, e.g. multivariate
- Can produce tables that aren't available in published census output
- Large sample sizes – good for small subpopulations
- Can use alongside other Census output
- Can use alongside market research surveys

Cons:

- Amount of geographical detail is limited due to potential disclosure issue
- Analyses on microdata are subject to sampling errors
- Tabulation results will not agree perfectly with same table produced on 100% Census
- Needs to be analysed using a statistics package

Census Microdata are available worldwide

79 countries as at July 2014, with 20 further candidates currently



Source: Integrated Public Use Microdata Series, International
www.ipums.org/international

History in the UK

- 1991 Census - Samples of anonymised records first become available:
 - 2% individual file (SAR areas)
 - 1% household file (Region)
- 2001 Census - greater concern about confidentiality
 - Restriction of some detail
 - Diversification of file types
 - Extra perturbation to licensed files
 - Added access controls on some data
 - More data and detail available in 2001 than in 1991, but in a more complex 'family' of products
 - Household file effectively unavailable to business users

2001 Census Microdata products

File	Sample type	Geography	Availability
Individual licenced	3% sample of individuals	UK GOR (+ Wales, Scot, NI, Inner/Outer London)	EUL CCSR
Small area microdata	5% sample of individuals	UK: LA (or consituecy in NI)	EUL CCSR
Household licensed	1% hierarchical file	None: England & Wales only	Special licence UKDA
Individual CAMS	Same sample as Individual licenced SAR	LA (GB) or Constituency (NI) IMD info for SOA	In house at ONS
Household CAMS	1% hierarchical file	All of UK	In house at ONS

The 2011 family of Census microdata products

Access Level	Method of Access	Individual File	Household File
Public	Openly available - Online download	Teaching File 1%	X
Safeguarded	End User Licence for academics and LAs; VML for commercial users	5%	???
Secure	Secure environment (VML) for approved researchers only	10%	10%

- A safeguarded household file would be of most interest to business users, but would have greater disclosure risks

The 1% public teaching file contains...

Variable	Number of Categories
Person ID	Unique Identifier
Region	10
Residence Type	2
Family Composition	6 (+ no code)
Population base	3
Sex	2
Age	8
Marital Status	5
Student	2
Country Of Birth	2 (+ no code)
Health	5 (+ no code)
Ethnic Group	5 (+ no code)
Religion	9 (+ no code)
Economic Activity	9 (+ no code)
Occupation	9 (+ no code)
Industry	12 (+ no code)
Hours Worked Per week	4 (+ no code)
Approximate Social Grade	4 (+ no code)

For business users, **safeguarded** microdata files are likely to contain more detail and be of greater interest – however, access will be more difficult

Application 1: Profiling and Targeting

- Census microdata may be applied to profile a target audience - to understand its characteristics and regional dispersion
- The data can be modelled using the analyst's tool of choice e.g. regression, decision tree or neural network
- Possible 2011 Census example: profile and model for internet form completion

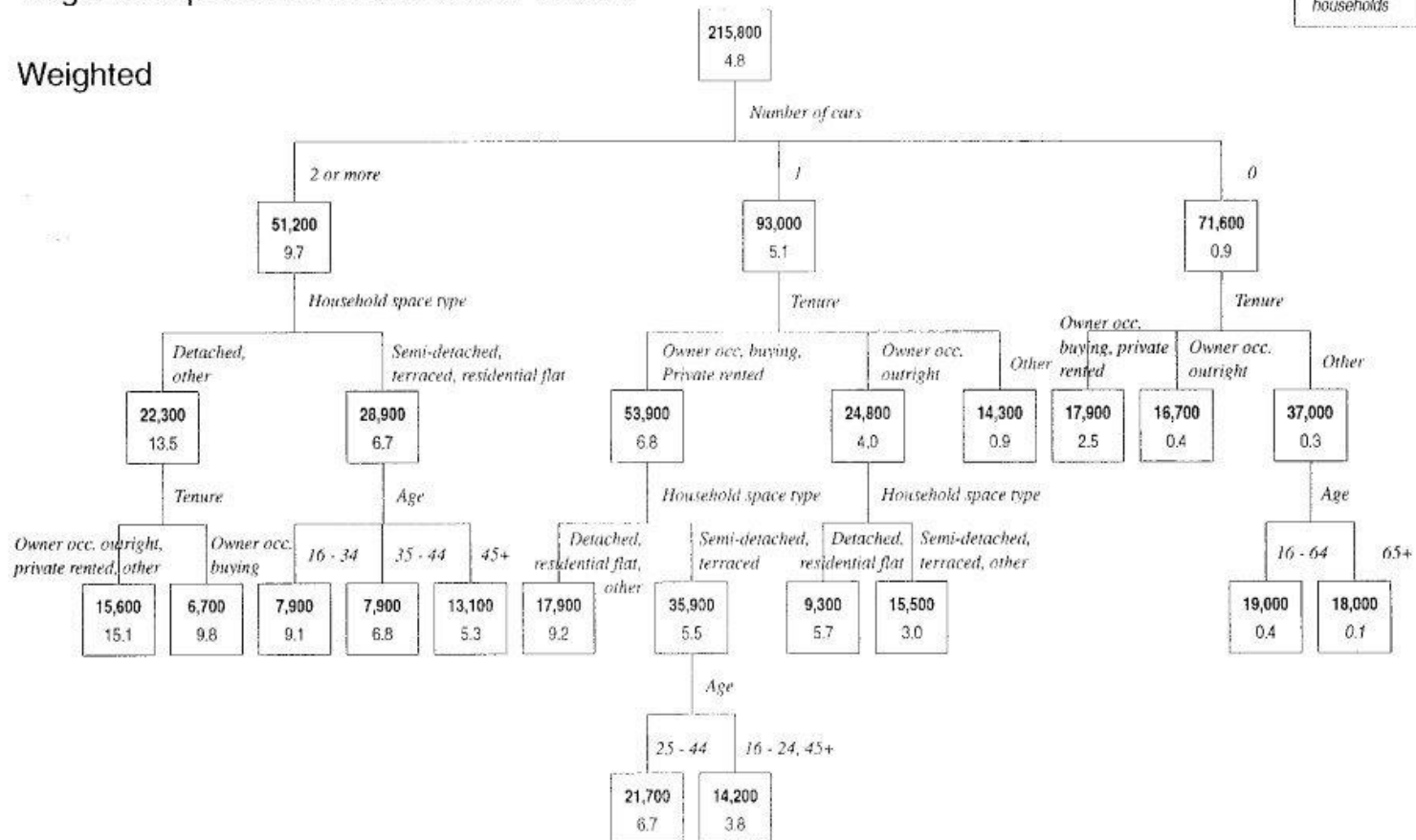
Example – Decision Tree Model for households in Social Class I

TREE SEGMENTATION ANALYSIS

Target Group: Households in Social Class I

Key: Sample Size
% of Class I households

Weighted



Base: Household SAR

Analysis grossed up to represent all SAR households

Source: Case study examples on use of SARs in Marketing Analysis
Census Microdata Unit, University of Manchester, Occasional Paper 5, 1994

Application 2. Demographic segmentation

- A prototype household segmentation was developed on SAR data
- 115 distinct household types were identified, e.g...

Ill poor, non-related pensioners, sharing	0.25%
Affluent sharing youngsters	0.67%
Unmarried, single parent council house dwellers	1.10%
Affluent couple youngest kid 5-10	1.63

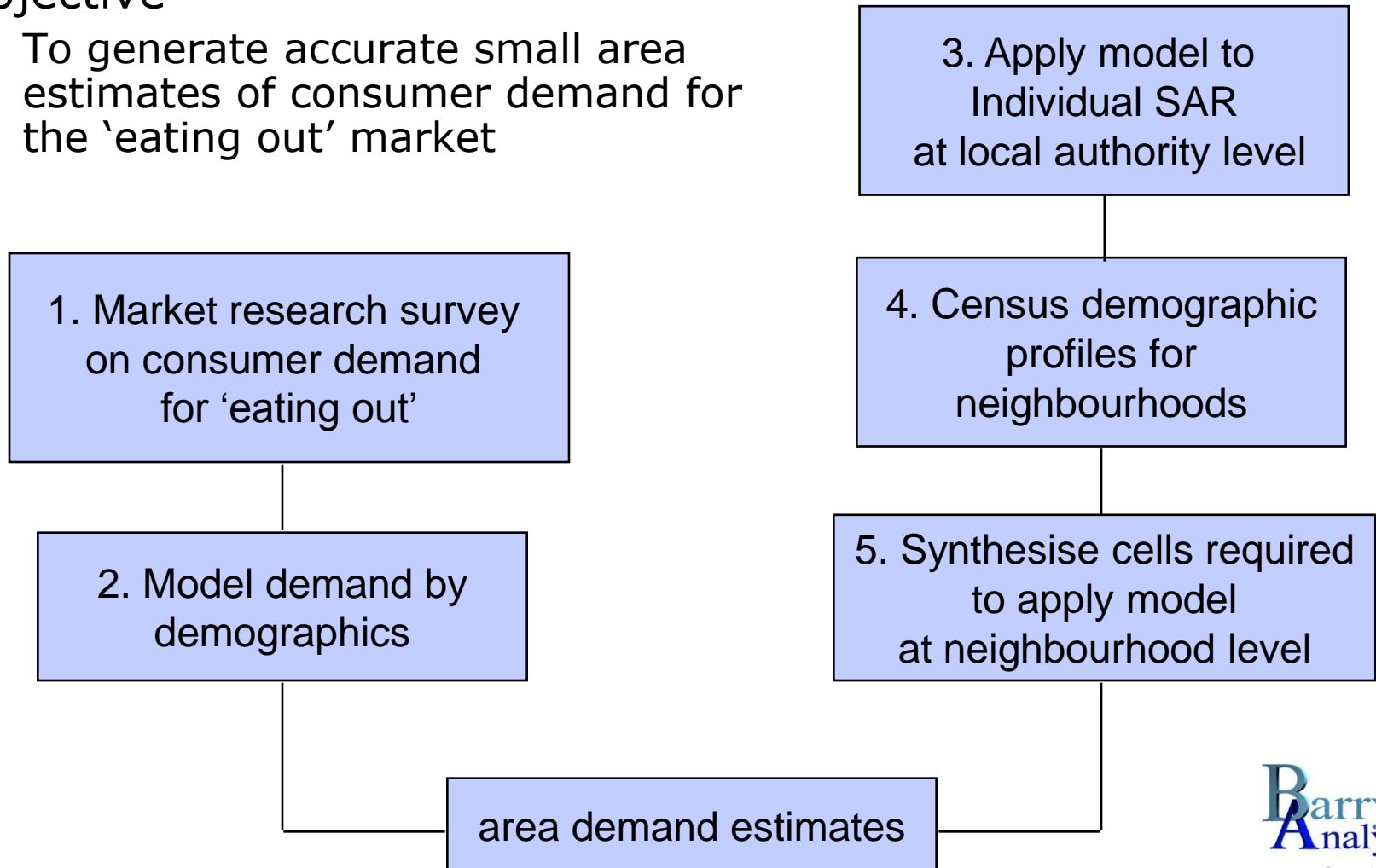
- These 115 types could be aggregated into major segments via hierarchical analysis
- Market involvement rates could be obtained by applying the same questions and algorithms to a market research survey
- Applying the algorithms to the Census database would enable segment profiles to be produced by Output Areas
- Combining the OA profiles and market involvement rates, accurate small area market estimates would be obtained

Source: Individual Multivariate Household Classification
by Martin Callingham, 28/11/2000

Application 3: Estimating market sizes

- Objective

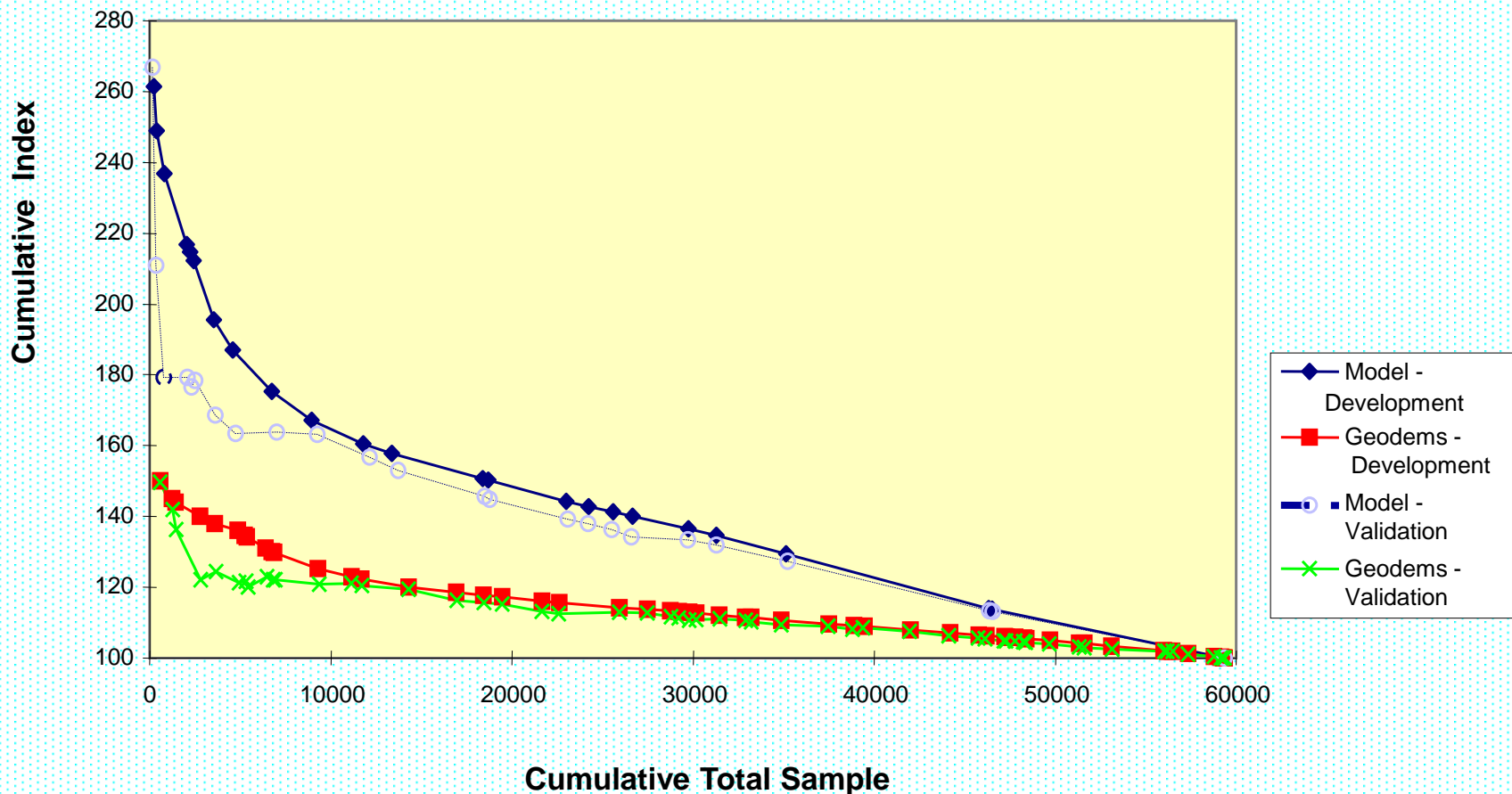
- To generate accurate small area estimates of consumer demand for the 'eating out' market



The individual-level model was highly discriminatory, when compared with a geodemographic approach

Market: Fish & Chips

Individual Model vs Geodemographic Discriminator



Example use cases in Market Research

- SARs have numerous applications to survey design – to plan level of screening to reach certain household types, or decide how to balance samples, e.g...
 - Understanding household composition (e.g. presence of children of particular age groups, households containing only 75+ age group) and using this to optimise sample design
 - Analysing penetrations of households containing combinations of people with limiting long-term illness (LLTI), and using results to structure a survey on disabled people
 - Measuring penetration of a religious group by region, and profiling households belonging to that religion, for designing an attitude survey
 - Identifying areas containing significant proportions of vulnerable people (pensioners, disabled, LLTI, etc) for a survey that measured awareness and preparedness for digital switchover

Accessing Census Microdata

- Access to the 2011 Teaching File:
 - <http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/census-microdata/microdata-teaching-file/index.html>
- Census Microdata at the UK Data Service – including access to 1991 and 2001 files:
 - <http://census.ukdataservice.ac.uk/get-data/microdata.aspx>

Conclusions

- Microdata is highly flexible – enables you to examine unpublished relationships between census variables
- Microdata would be invaluable for survey designers
- And has powerful analysis possibilities, when combined with market research data
- The safeguarded 2011 Census Microdata is likely to be of particular interest to business users
 - But data access will need to be via end user licence, otherwise Microdata is a non-starter!
 - VML access is not practical for most business users

Thank you!

Barry Leventhal

07803 231870

barry@barryanalytics.com

