

DATA INTEGRATION

Possibilities and Examples

Corrine Moy

NOP World

31st March 2003



NOP World

Unitec Business Media

Integrating Data

- It is often the case that survey data alone can not completely answer a client's research objectives
- Instead, we need to combine data sources to fully understand the problem

→ Data Integration



NOP World

Unlock Business Media

Data Integration Possibilities

- There are 2 alternative approaches
 - Integrating data from 2 or more different research surveys
 - Integrating survey data and customer database information



Integrating Survey Data

- Again, there are 2 alternative approaches
 - 2 or more customer (consumer or business) surveys
 - Customer satisfaction
+
Mystery shopping
+
Employee satisfaction



Integrating Customer Surveys

- Data integration can be done in 2 different ways
 - aggregation by unit or time period to create a common data file
 - data fusion



Integration By Aggregation

- A datafile is created which includes all information from both or all surveys
- This is done by producing aggregate measures. The aggregation can be done by:
 - time
 - business unit (car dealerships)
 - market
 - any sensible combination of factors



An Aggregated Data Set

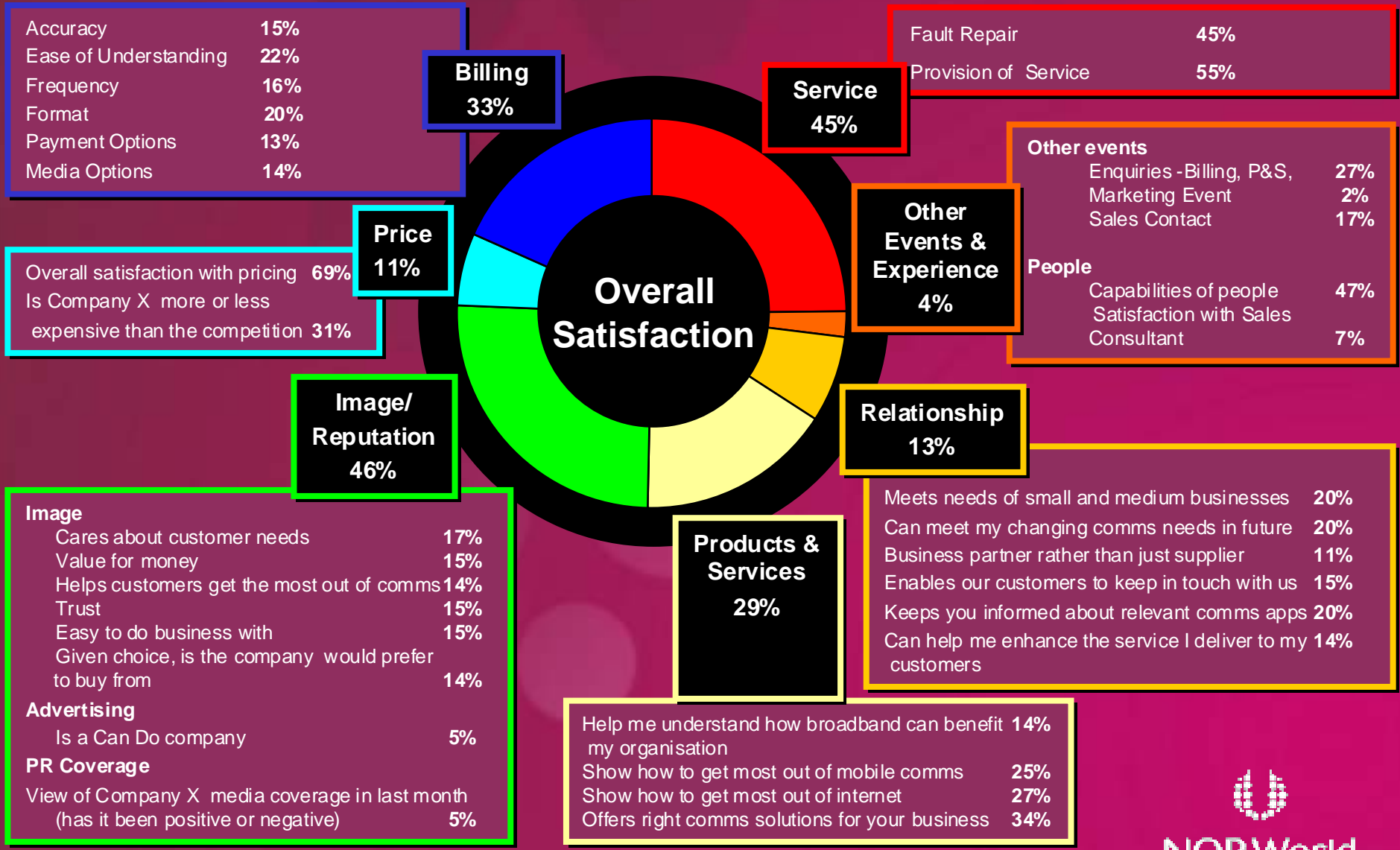
Composite	Driver	Source	Aug 08	Sep 08	Oct 08	Nov 08	Dec 08	Jan 09	Feb 09	Mar 09	Apr 09	May 09	Jun 09	Jul 09	Aug 09	Sep 09	Oct 09	Nov 09	Dec 09	Jan 10	Feb 10	Mar 10	Apr 10	May 10	
Company X Service	Fixed	Event Driven CSM	7.72	8.01	7.96	7.93	7.34	7.88	8.10	8.08	7.94	7.87	7.86	7.80	8.32	8.10	8.81	8.43	7.87	7.47	7.40	8.14	8.76	8.20	
	Provision		7.25	6.51	7.35	7.25	7.10	7.66	8.14	7.61	7.57	7.51	7.87	7.82	7.45	7.80	7.20	7.80	8.84	8.81	8.87	7.84	8.16	7.44	
Other Events and Experiences	Equities, listing, PE & COFFERCO	Relationship Monitor	6.21	6.20	6.00	6.60	6.42	6.45	6.46	6.22	6.18	5.58	5.14	6.51	4.08	7.08	6.36	5.90	5.97	5.96	6.70	5.85	5.76	6.20	
	Sales Contact		5.99	7.14	6.25	6.50	7.15	8.01	7.38	8.60	7.15	7.27	8.25	5.58	7.31	7.08	7.68	6.18	6.20	6.20	7.50	7.85	7.83	7.80	
	Capabilities of Company X people		4.86	6.89	6.88	7.26	7.08	7.05	7.08	7.08	7.08	6.97	7.12	7.58	6.38	7.08	6.98	7.54	6.98	7.17	7.12	6.89	7.46	7.36	
	Satisfaction with Company X Sales Consultant		6.26	7.36	6.80	6.81	7.87	7.82	6.86	6.87	6.86	6.84	7.06	6.86	4.95	6.34	6.07	6.21	6.09	6.94	6.31	6.38	6.48	6.58	
Relationship	Company X meets the needs of small and medium	Image Tracker	5.35	5.36	5.36	5.36	5.37	5.37	5.37	5.35	5.35	5.35	5.30	5.30	5.30	5.67	5.48	5.36	5.90	5.50	5.23	5.55	5.16	5.76	
	Company X can meet my changing needs		6.06	6.77	6.47	6.06	6.53	6.33	5.75	5.98	6.73	6.08	6.07	6.18	6.13	5.68	5.95	5.79	5.90	5.90	5.55	6.24	5.79	5.89	
	Company X is a business partner rather than just		4.55	5.01	4.40	4.60	4.21	4.61	4.47	4.65	4.68	4.65	4.48	4.58	4.61	4.28	4.64	4.55	4.50	4.70	4.76	5.10	4.57	4.86	
	Company X enable my customers to keep in touch		5.23	5.26	5.34	5.39	5.44	5.50	5.55	5.80	5.86	5.71	5.76	5.82	5.87	5.82	5.88	5.88	6.16	6.26	6.00	6.23	5.89	6.20	6.80
	Company X keeps me informed about relevant		6.25	6.85	6.20	6.36	6.94	6.99	5.98	6.68	6.08	6.38	6.08	6.38	6.08	6.38	6.78	6.22	6.08	6.10	6.82	6.73	6.32	5.86	5.76
Products and Services	Company X helps the customer to save	Image Tracker	6.74	6.66	6.82	6.66	6.91	6.45	6.39	6.35	6.26	6.22	6.16	6.11	6.05	6.28	5.68	6.26	5.50	5.90	5.77	5.50	5.50	5.70	
	Company X offers the best value for money		6.47	6.45	6.45	6.40	6.30	6.36	6.34	6.31	6.29	6.27	6.38	6.18	6.18	5.90	6.28	5.94	6.30	6.10	6.20	6.37	6.02	5.80	
	Company X shows me how to get the most out of mobile		4.67	4.99	5.00	4.90	4.90	4.53	4.37	5.52	4.97	4.94	5.11	5.22	5.08	4.74	4.68	4.77	4.90	4.80	4.89	4.86	4.86	4.85	4.75
	Company X shows me how to get the most out of internet		4.22	5.00	4.81	4.84	4.08	4.44	3.88	4.61	4.63	4.93	4.88	4.93	5.16	4.62	4.88	4.80	4.80	4.80	4.40	4.70	4.84	4.88	4.83
Image/Reputation	Company X offers the right pricing solutions for your	Image Tracker	6.49	6.89	6.65	6.45	6.14	6.91	5.91	6.88	6.38	6.38	6.38	6.38	6.48	6.08	6.23	6.08	6.20	6.13	6.37	6.24	6.37	5.87	
	Company X cares about customer needs		5.77	6.25	5.86	5.61	5.53	5.72	5.03	5.58	5.48	5.58	5.48	5.68	5.58	5.24	5.63	5.36	5.30	5.44	5.20	5.53	4.94	5.26	
	Company X offers good value for money		5.46	5.75	5.69	5.19	5.30	5.49	4.91	5.23	5.38	5.58	5.38	5.28	5.28	5.08	4.93	5.08	5.30	4.93	5.12	5.29	5.80	5.23	
	Company X helps you get the most from your		5.53	6.09	5.68	5.45	5.25	5.61	4.88	5.28	5.78	5.68	5.58	5.68	5.68	5.84	5.78	5.47	5.46	5.39	5.47	5.72	5.34	5.42	
	Company X is a company you can trust		6.48	7.24	6.80	6.47	6.41	6.68	6.28	6.28	6.28	6.38	6.68	6.78	6.68	6.14	6.62	6.43	6.30	6.37	6.76	6.20	6.13	6.26	
	It is easy to do business with Company X		5.46	6.27	6.01	5.90	5.42	5.39	5.07	5.28	5.58	5.68	5.48	5.78	5.68	4.98	5.58	5.28	5.20	5.36	5.11	5.47	5.85	5.10	
	Given the choice, Company X is a liability you would prefer		6.15	6.63	6.56	6.30	6.00	6.34	5.97	5.96	6.28	6.28	6.58	6.48	6.28	6.11	6.43	6.14	6.10	6.06	6.90	6.37	5.80	5.83	
Company X media coverage in the last month has been	6.63	6.84	6.84	6.84	6.85	6.85	6.85	6.85	6.85	6.85	6.85	6.67	6.67	6.67	5.41	6.94	6.34	5.89	5.44	6.10	6.81	6.84	5.82		
Price	Overall satisfaction with pricing	Image Tracker	4.73	4.74	4.74	4.74	4.76	4.76	4.76	4.76	4.76	4.88	4.84	4.62	4.98	4.48	4.81	4.89	4.70	4.82	4.80	4.87	4.81	5.87	
	Is Company X more or less expensive than the		3.16	3.16	3.14	3.13	3.12	3.11	3.10	3.89	3.89	3.48	3.14	2.98	2.94	2.94	2.86	2.84	3.27	2.95	2.95	3.83	3.80	3.87	
Billing	Accuracy	Billing survey	9.70	9.69	9.37	9.55	9.73	9.73	9.73	9.55	9.73	9.55	9.64	9.55	9.55	9.55	9.55	9.64	9.64	9.64	9.64	9.26	9.10	9.20	
	Ease of understanding		7.46	7.51	7.54	7.54	7.50	7.42	7.68	7.72	7.68	7.63	7.68	7.63	7.68	7.63	7.78	7.75	7.74	7.80	7.83	7.75	7.82	7.87	
	Frequency required		8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.33	8.31	8.31	8.38	8.38	8.33	8.26	8.22	8.20	8.34	8.27	8.38	8.26	8.18	8.45
	Format required		7.87	8.01	8.04	8.05	8.12	8.16	8.19	8.25	8.28	8.34	8.34	8.34	8.47	8.61	8.48	8.43	8.38	8.58	8.67	8.70	8.80	8.70	
	Payment options required		6.36	6.36	6.30	6.30	6.40	6.40	6.41	6.41	6.28	6.13	6.45	6.40	6.54	6.58	6.46	6.58	6.61	6.72	6.74	6.16	6.20	6.34	
Media options required	6.91	6.82	6.83	6.84	6.85	6.85	6.85	6.85	6.87	7.98	8.04	8.28	8.11	8.08	8.16	8.06	8.26	8.22	6.45	6.20	6.94	6.80	6.20		
Company X Satisfaction	Overall satisfaction with Company X	Image Tracker	5.79	6.00	5.95	5.42	5.47	5.15	5.33	5.12	5.05	5.78	5.35	5.28	5.28	5.53	5.99	5.83	5.88	5.85	5.56	5.85	5.50	5.80	
Composites	Company X Service		7.46	7.18	7.04	7.37	7.54	7.70	8.30	7.93	7.98	7.67	7.78	7.74	7.84	7.49	7.56	8.13	8.16	7.77	7.79	7.97	8.42	7.78	
	Other Events & Exp		6.39	6.92	6.68	7.61	7.27	7.09	6.93	6.75	6.93	6.61	6.60	6.70	5.56	6.87	6.67	6.40	6.72	6.83	6.80	6.67	6.84	6.95	
	Relationship		5.77	6.05	5.83	5.81	5.78	5.79	5.99	5.87	5.89	5.79	5.72	5.81	5.81	5.67	5.71	5.75	5.81	5.71	5.80	5.84	5.54	5.84	
	Products & Services		5.95	5.94	5.77	5.72	5.42	5.95	5.15	5.97	5.95	5.67	5.70	5.72	5.75	5.40	5.62	5.37	5.53	5.42	5.51	5.62	5.46	5.37	
	Image/Reputation		5.07	6.42	6.06	5.85	5.75	5.92	5.39	5.64	5.79	5.81	5.79	5.87	5.75	5.46	5.86	5.66	5.59	5.81	5.52	5.69	5.39	5.52	
	Price		4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.23	4.00	4.38	4.30	4.09	4.04	3.99	4.17	4.17	4.24	3.99	4.31	4.28	4.43	
	Billing		8.28	8.29	8.29	8.29	8.32	8.32	8.38	8.37	8.37	8.33	8.42	8.43	8.49	8.49	8.40	8.44	8.46	8.56	8.56	8.37	8.41	8.47	

Modelling Aggregated Data

- Aggregated dataset can then be used as if it were single-source
- Complex analysis techniques can then be used
- eg. regression techniques can be used to produce standard satisfaction models
- Need a reasonable number of data points!



Key Drivers of Customer Satisfaction



Impact on Overall Satisfaction Mean Score

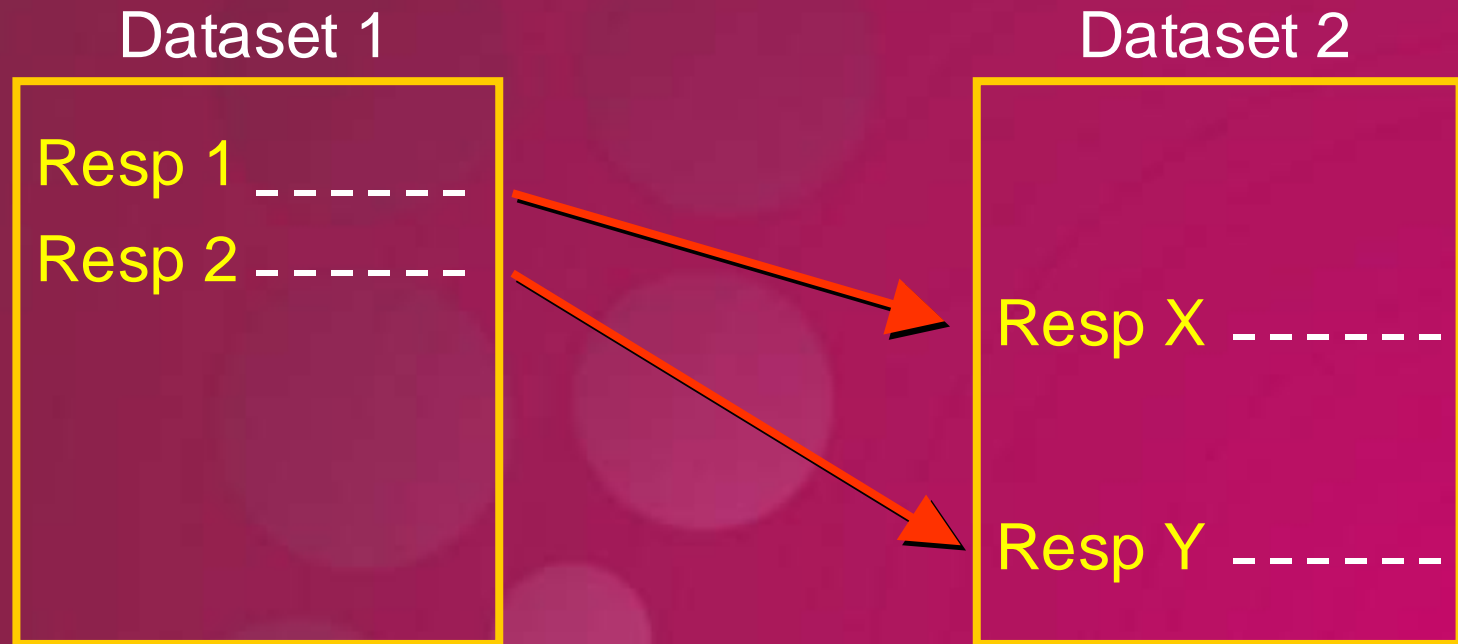
Integrating By Data Fusion

- 2 or more datasets are 'fused' together at the respondent level
- Matching is done on a number of key criteria
 - age
 - sex
 - income etc



Data Fusion

- One dataset acts as the 'donor' to the second dataset

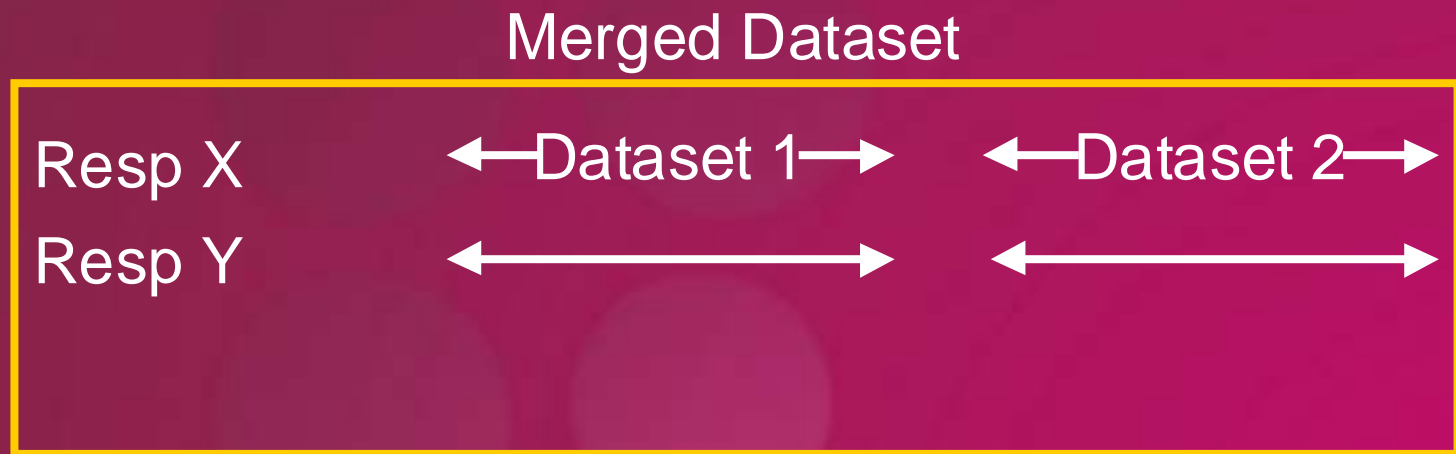


- Once a match is found, all data for respondent 1 is given to respondent X, to form one merged dataset



Data Fusion

- Merged dataset



- Matches may be as close as **'Love at first sight'** or as distant as **'Shotgun wedding'**
- The resultant dataset can then be used as if it is from a single source



Fusion Example

- Fusion of two surveys
 - Buyer survey
 - Image survey
- Objective was to add Buyer data to Image data
 - Image Survey → recipient
 - Buyer Survey → donor



Fusion Example

- **Critical variables** (must match exactly)
 - product ownership
 - gender
 - age (5 year bands)
- **Matching variables** - as close as possible
 - occupation
 - income
 - size of household
 - region etc



Example of a match

- A recipient on Image Study is a male, aged 48, with four people in his household, including a child, living in the South East. He is a company director, earns £90,000 or more and is the owner of a flat screen plasma television
- He is matched with a donor from Buyer Survey who is a male, aged 46, married, with four people in his household, including a child, living in the South East. He is the owner of a company with 20 or more employees, has an higher education, earns £60,000 - £69,999 and is the owner of a flat screen plasma television.



Frequency Distribution of Scores

	<u>%</u>	<u>Cum.%</u>
Perfect match (score 9)	5%	5%
Very good match (score 8 to 8.9)	42%	47%
Good match (score 7 to 7.9)	32%	79%
Moderate match (score 6 to 6.0)	14%	93%
Match (score less than 6)	7%	100%



Validation Of Fusion

- Fusion is validated by examining variables which existed on both datasets, but were not used for fusion process
- We found high levels of alignment between the 'real' and 'modelled' variables

Do you own a PC ?

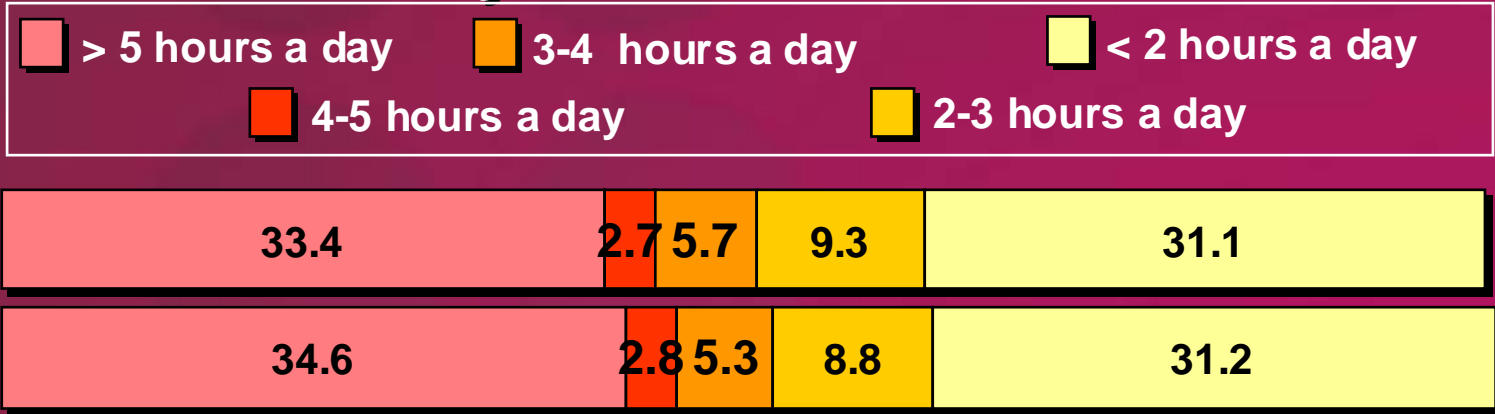


 Yes  No  Planning a purchase

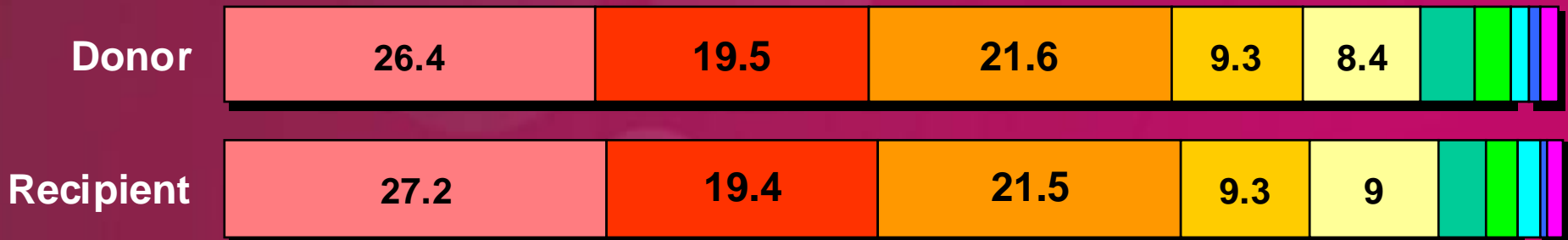


Validation Of Fusion

How much television do you watch in a week?



Overall satisfaction with your new television



What Can We Do With New Database?

- There are many questions on the Buyer Survey that are not on the Image survey
 - method of purchase
 - satisfaction with sales experience
- We can now have access to these responses using Image Survey respondents
 - additional information/data
- We can run cross tabs using Image questions against Buyer questions
- Modelling at respondent level is also possible



Major Data Fusion Projects

- There have been a number of fusions between major industry surveys, which bring great benefit to users
- TGI has been fused with National Readership Survey and BARB
- Other major industry surveys, such as NOP's FRS, are in the process of being fused with such audience measurement surveys



Major Data Fusion Projects

- These fusions allow the cross-analysis of
 - Readership data with brand preference
 - TV viewing with financial holdings
 - Shopping behaviour with magazine readership
- To answer questions such as
 - Which advertising slots should a pensions company buy?
 - In which magazines should a supermarket advertise a new proposition



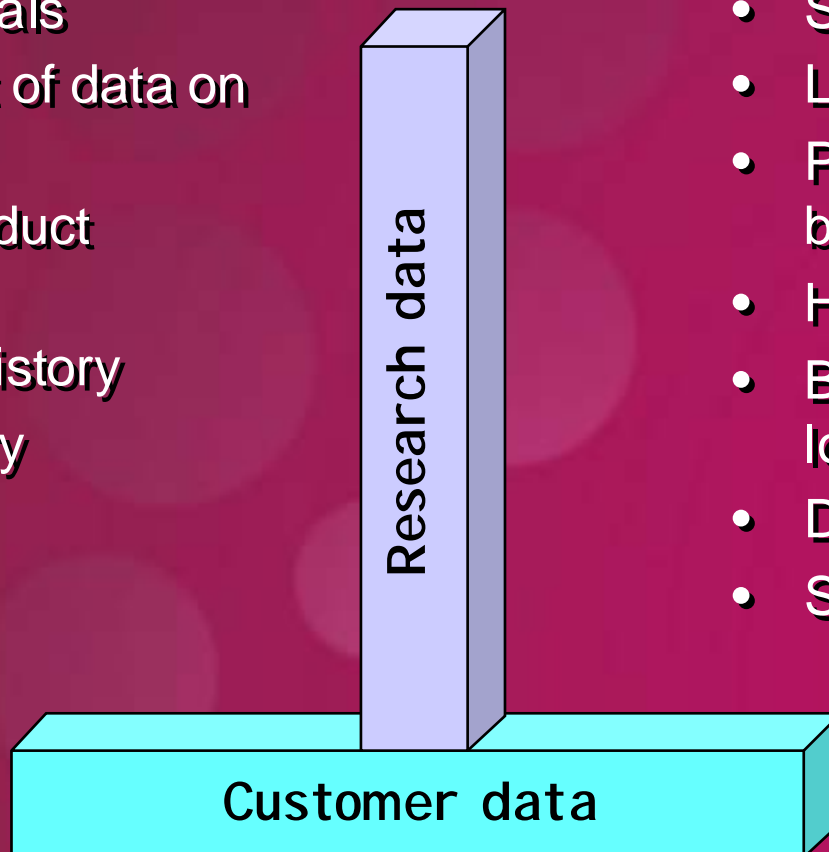
Combining Survey Data & Customer Databases

Customer data

- Many individuals
- Small amount of data on each
- Company product holdings
- Transaction history
- Contact history

Research data

- Smaller number
- Large amount of data
- Product holdings and brands
- Holding value
- Brand perceptions and loyalties
- Demographics
- Segmentation



NOP World

Unlock Business Media

The Benefits

- Gain greater understanding of customers
 - allows us to explore an individual's behaviour within a market
 - we can measure competitor activity, loyalty
 - we can uncover more information about customers
- Combine *depth* of market research with the *breadth* of population databases
- Apply to recruitment and relationship management
- Improved targeting on individual data



The Benefits

- Improve hit-rate for outbound calling/ direct mail
- Identify at-risk customers to focus account activity
- Identify customers who are at different stages of switching or moving spend to competitor
- Identify prospects for new products
- Develop mechanisms for tailored one-to-one marketing



NOP World

United Business Media

How it's done

- Market research is conducted on a sample of a company's database
- The sample is drawn to meet the research objectives e.g. active customers
- The sample may consist of 2,000 customers from a database of 2 million
- Information from the database is attached to the sample



How it's done

- Research is conducted on the sample to establish market behaviour which is not contained within the database
 - competitor holdings
 - financial situation
 - switching activity
- Develop model of loyalty/ likelihood to switch using database information on sample
- Link back to database, thus producing a measure of vulnerability on database
- Target action to protect vulnerables



Case Study - Mortgage Retention



- Sample of mortgage customers drawn from database
- Research carried out to establish mortgage outcome and reasons for change
- Lost customers were categorised into ‘switchers’, ‘movers’ and ‘paid-offs’
- Model built to predict switching, moving or paid-off for whole sample



Case Study - Mortgage Retention



- Model overlaid onto database to predict outcome for each customer
- **‘Mortgage contribution value’** attached to each customer
- Retention activity designed to minimise loss of profitable customers



NOP World

Unfoc Business Media

Case Study - Tolerance To Direct Mail

- Project for major bank
- Objectives were to
 - Measure tolerance to direct mail
 - Produce segmentation on attitudes to direct mail
 - Overlay segmentation onto database

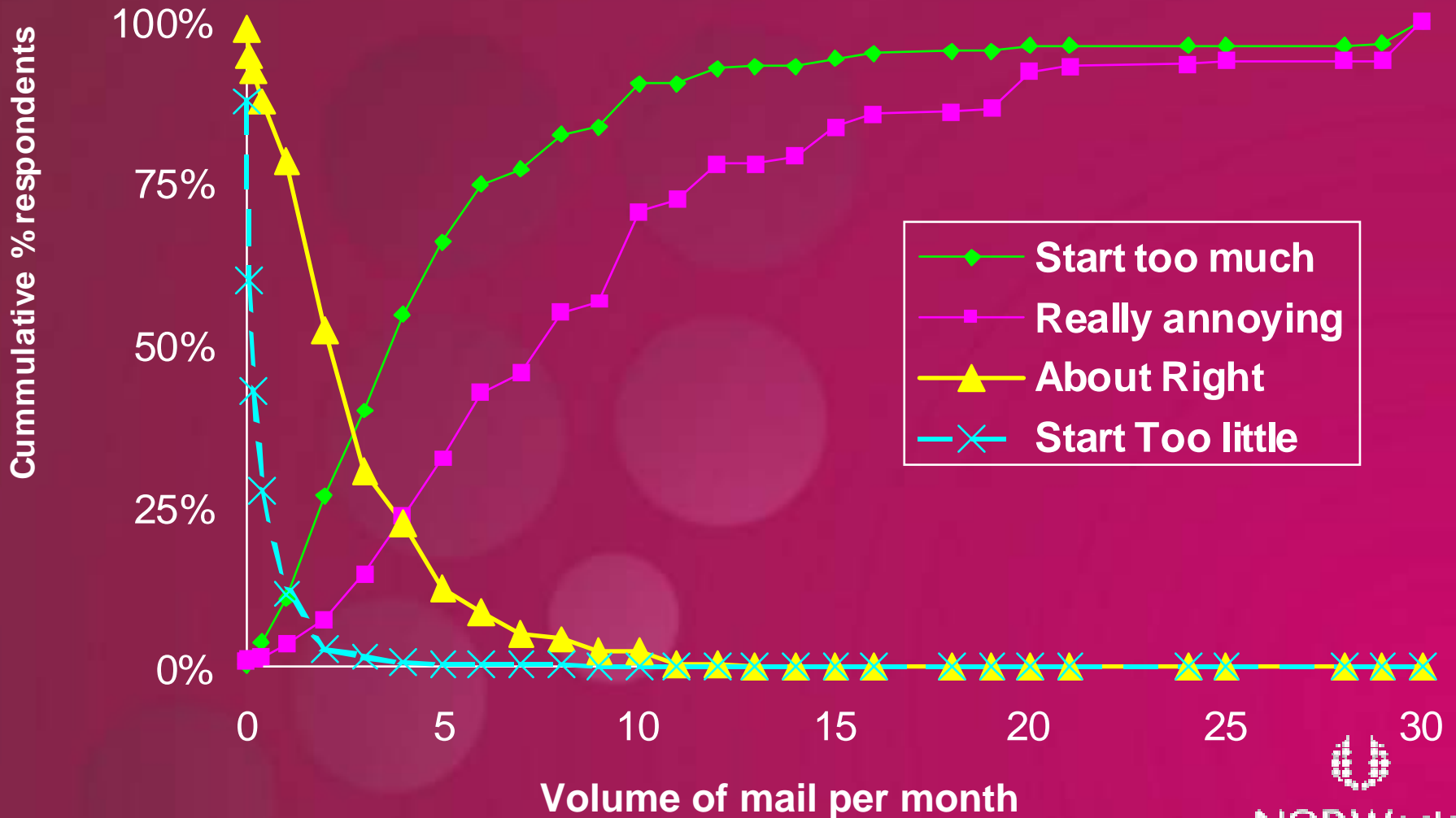


Case Study - Tolerance To Direct Mail

- Sample of 1,000 customers drawn from database
- Asked what about of mail per month would
 - start to be too much
 - be really annoying
 - be about right
 - start to be too little
- Graph shows the elasticity of tolerance with optimum level around 3 per month



Case Study - Tolerance To Direct Mail



Case Study - Tolerance To Direct Mail

- Segmentation performed on 28 attitudinal questions
 - treated as a number, not an individual
 - hate receiving mail from Co's I deal with
 - none of the direct mail I get is relevant to me
 - expect direct mail - Co's need to use them to get business
 - waste of time and money
 - don't mind, they might be interesting



Case Study - Tolerance To Direct Mail

Five segments were produced:

Moderates

- direct mail about right
- ‘average’ customers

Mail-o-philes

- downmarket, 25-34
- positive to direct mail, high tolerance

Hands-off

- more 45+, upmarket, low risk
- cynical to bank, too much mail
- throw away without reading, very negative

Anti-bank cynics

- mostly would hold one product
- very cynical about bank
- happy with level of mail, more would be very annoying
- not willing to give personal details to improve targeting

Neglected

- young, downmarket, renting
- perceive less mail than receive
- receive right amount or too little



NOP World

Unlock Business Media

Case Study - Tolerance To Direct Mail

- Segments profiled by demographics, behaviour
- Differential perception of amount of mail sent and amount desired
- Differential perception of relevance of mail
- Mail-o-philes have far greater tolerance than any other segment
- More mail would seriously upset Hands-off
- Differential usefulness of mail across segments



Case Study - Tolerance To Direct Mail

- Segmentation was modelled from database information sample
- Model was be applied to database
- Customers were categorised into segments
- Mailing activity was then differentiated by segment
- Validation amongst subset of database was undertaken to measure effectiveness of model



Case Study – NOP Fresco

- NOP Financial Research Survey
 - 60,000 interviews per year
 - Extensive measurement of people's financial holdings and behaviour
- Used by clients for
 - Research and marketing
 - Strategic and tactical
 - Savings and lending behaviour
 - Applicable to client's customers, the market (including competitors) and prospects



NOP World

Unfoc Business Media

Case Study – NOP Fresco

- Created a segmentation to give insight into the current issues in the market
- Segmentation developed using
 - Channel
 - Value
 - Attitude
- Wide variety of inputs improved segmentation validity & sensitivity across all markets



NOP World

United Business Media

Case Study – NOP Fresco

Objectives for New Segmentation

- Classify FRS respondents in an exact fashion
- Project onto prospect mailing lists
- Model onto customer databases

Partnered with CACI



NOP World

Unlock Business Media

Case Study – NOP Fresco

- We wanted to take maximum advantage of
 - CACI's population databases
 - The depth and breadth of NOP 's FRS data
 - Our shared understanding of consumers' financial behaviour
- To create something that is
 - Intuitive and pragmatic
 - Statistically robust and creative
 - Highly granular and discriminatory



NOP World

Unfoc Business Media

How Was it Built?

- Three distinct modelling stages
 - Classification of the FRS
 - “Filling in” the survey
 - Models for roll-out to prospects & customers
- Multiple models at each stage
- Each stage validated



How Was it Built?

- 60,000 records from FRS 2001
- Over 500 Variables
 - All products
 - Existing and new accounts
 - Value of holdings
 - 11 attitudinal statements
 - Channel usage by product
 - Household and personal income
 - Wealth of other demographics



NOP World

Unfoc Business Media

Dimension Reduction

- Eliminate highly correlated variables
- Focus on highly discriminatory variables
- Select Key behavioural variables including
 - PC banking,
 - Unit trusts,
 - Savings value
 - Have loans > £2,000
- Describe groups of people whose financial behaviour is similar

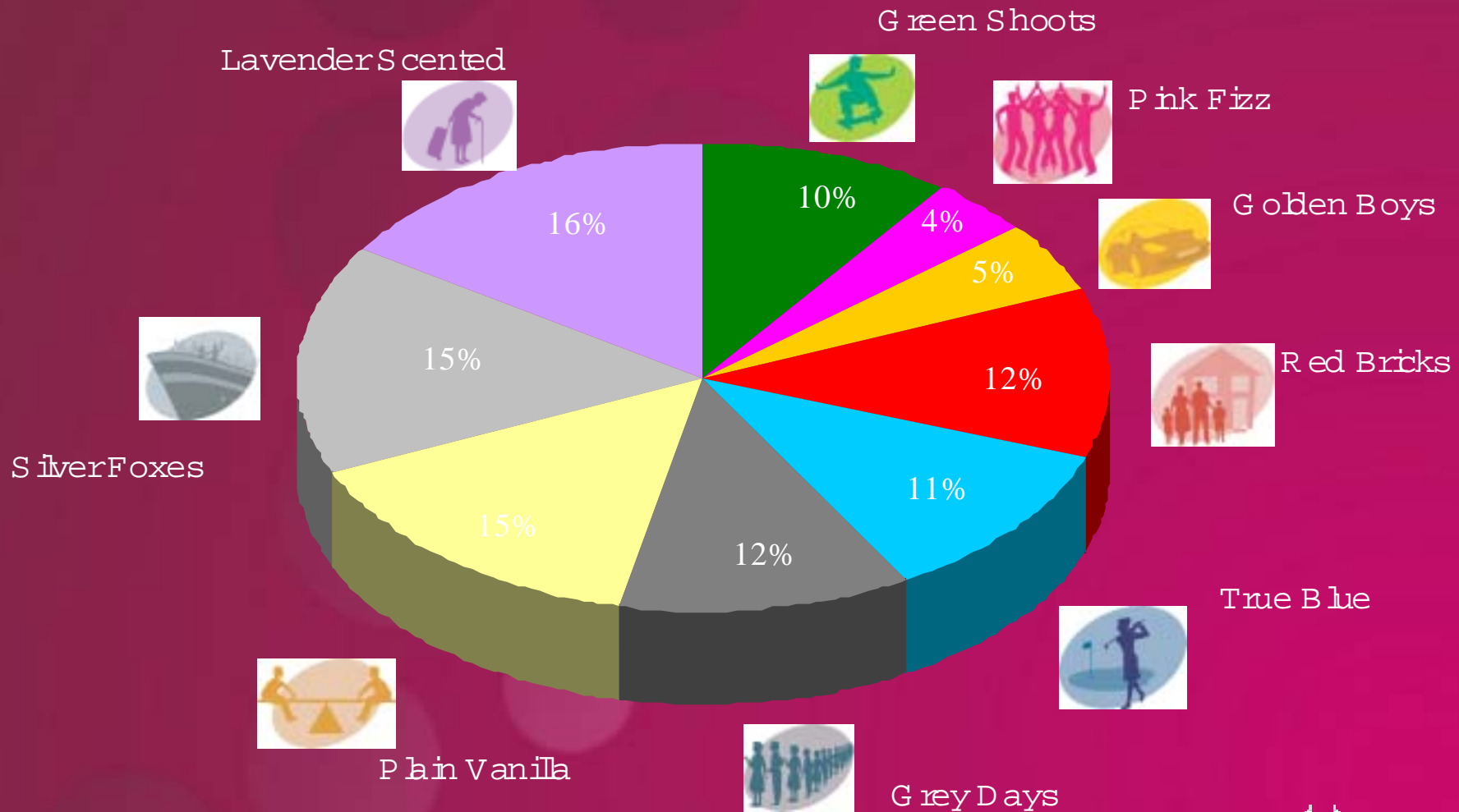


Prospect models

- Predict the key behavioural variables using external demographics
 - Household income
 - Individual income
 - Age
 - Marital status
 - Gender
- Demographic bridge allows us to accurately and effectively assign customers and prospects to Fresco types from the survey



Fresco Segments



Segments and sub-segments differ in terms of...

- Demographics
- Income and wealth
- Financial activity
- Attitudes to finances
- Brand preference and loyalty
- Channel usage












Distribution by age and personal income



NOP World

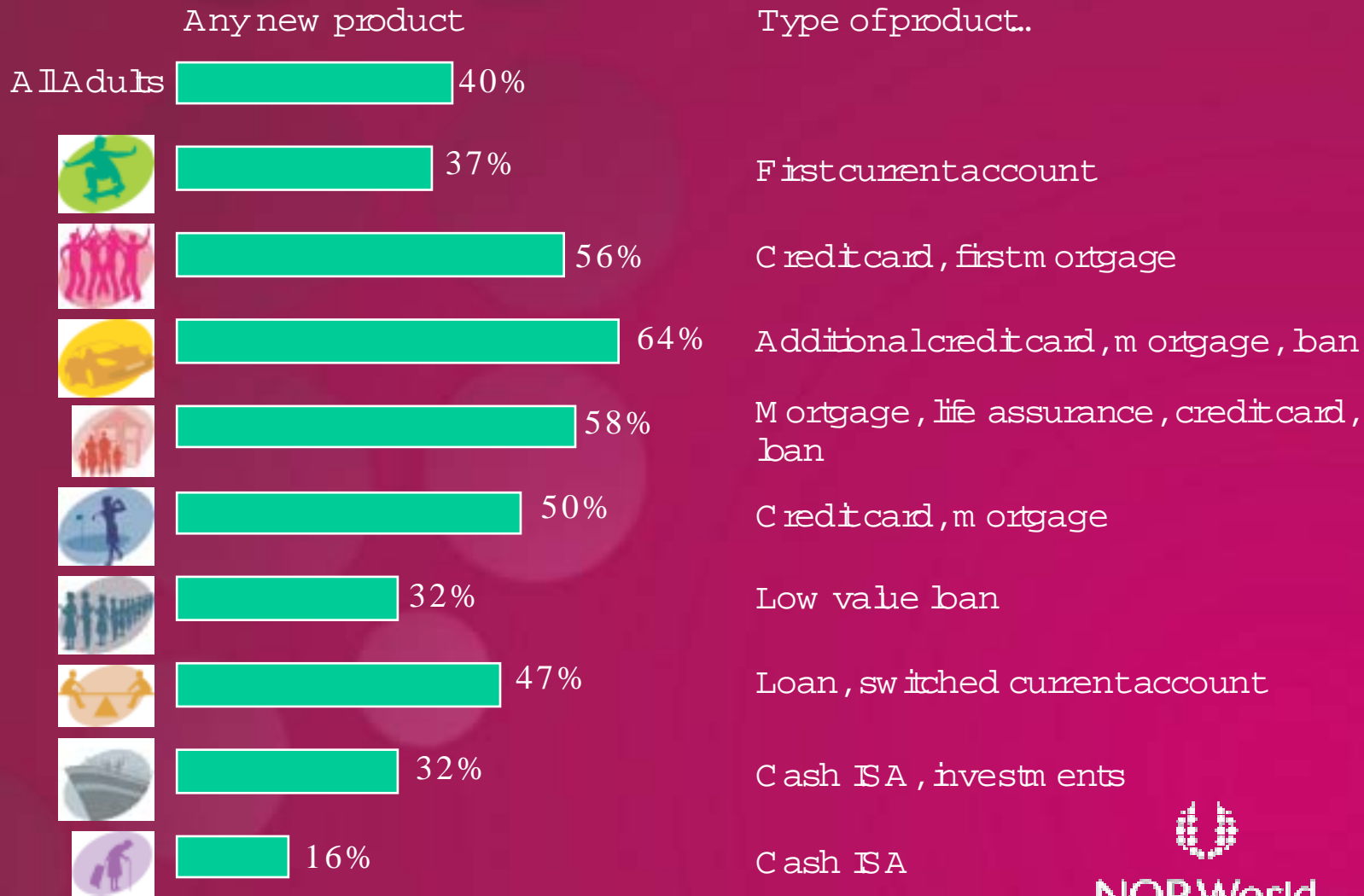
United Business Media

Number of sub-segments

	Number	Key differences
	3	Gender, working status
	2	Gender
	3	Marital status, lifestage
	3	Gender, income
	4	Gender, income, working status, age
	3	Age, lifestage, gender
	5	Wide range of lifestages
	6	Working status, age, gender
	3	Gender, marital status



Financial Activity



Channel Preference

Household Insurance

All in market



1st choice

Phone (42%)



Branch (41%)



Phone (41%)



Phone (51%)



Phone (42%)



Phone (52%)



Branch (48%)



2nd choice

Branch (34%)



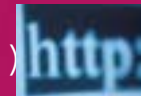
Phone (36%)



Branch (37%)



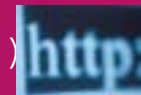
Internet (29%)



Branch (37%)



Internet (20%)



Phone (33%)



Base: All in market for Household ins in the next 12 months



NOP World

Unfoc Business Media

How is Fresco used?

- New dimension to understanding Financial Services and customers
- Fresh insights into...
 - New product development
 - Communication strategy
- Understand and identify gaps in market
- Develop products suited to strengths of client's customer base



How is Fresco used?

I want to get into UK Financial Services -
what's the best bet?

- Best product?
- Best target group?
- What do we know about them
- How do we approach them?

Who's prime for offset mortgages?

- Can we identify who holds this product?
- Who should we then target?



NOP World

Unfoc Business Media

Match to Consumer Register

- Fresco can be coded by simple match to CACI's new Consumer Register
 - Currently 2001 Electoral Roll is not available for marketing purposes
 - Future Electoral Rolls offer consumer opt-out
- CACI has now undertaken massive validation and infill project to create new Consumer Register



Match to Consumer Register

- Selected three partners
 - Dataworks (14 million records)
 - Bounty (10 million records)
 - Claritas (23 million records)
- Large files giving comprehensive coverage of post October 2000 change
- Simple match to CR allows you to add Fresco code simply and quickly



NOP World

Unfoc Business Media

Modelling Approach

- Alternatively if you are an organisation with lots of demographic data
- Take the actual modelling rules and use real data to assign customers to segments
- Where you are missing data on a proportion of your customers, CACI can supply infill models
- Takes advantage of the accuracy of real data



Using Fresco in a Business Environment

- Product development
- Cold customer acquisition via Consumer Register list
- Focusing in-branch sales staff on customer opportunities
- Setting performance targets for branches
- Undertaking local marketing activity

