



# Accommodating Big Data In TAM Analysis Systems

## -The Probability Is Something Has To Give

Venice, November 2013

Steve Wilcox, Managing Director, RSMB

---



# Traditional TAM Panels

- Standard model for design and operation
  
- Continuous measurement of individuals:
  - known demographics
  - viewing across events, channels, platforms and devices
  
- Allows simple counting algorithms to estimate:
  - currency audiences
  - programme duplication
  - advertising reach and frequency



# Site Centric Data

- Already significant development of analytics solutions
- Continuous measurement of devices:
  - unknown demographics
  - viewing across events and channels
  - within, not across, devices and platforms
- Models and other data sources are required to estimate:
  - currency audiences
  - programme duplications
  - advertising reach and frequency
- Sophisticated modelling and analysis systems may not be transparent



# The TAM Integration Challenge

- Convert site centric views to demographic audiences
- Estimate reach and frequency across:
  - site centric sub-universes defined by device and platforms
  - site centric and TAM panel traditional TV universes
- Configure the audience analysis system:
  - specific vs. multi-purpose analysis requirements
  - database structures
  - gold standard calculation conventions



# Are We Close To Breaking Point?

- In principle, TAM panels provide a transparent respondent level database and multi-purpose analysis potential
- In reality, complex gold standard calculation conventions are required to deal with panel dynamics and consistency across a range of reporting requirements
- Can this TAM panel counting system be extended to embrace integrated SC data?
- What is the potential role of probability based analysis methodology?



# Probabilities In Other Media Currencies

- Press - personal AIR probabilities for schedule reach and frequency
- Radio - aggregated probability model for extended reach beyond the diary week
- Posters - to configure travel survey reach and frequency data for integration with “billboard” site centric traffic counts



# TAM Probabilities In TouchPoints

- Personal probabilities approximate BARB gold standards in multi-media reach and frequency
- For each BARB panel member calculate:
  - average commercial GRP
  - for each channel
  - for each time segment
- Calibrate to gold standard GRP's by factoring all personal probabilities up or down
- Binominal expansion to create TV schedule personal reach and frequency
- Combine with other media reach and frequency before aggregating the sample



# Probabilities in TAM Reach and Frequency

- Frequency distribution
  - for each BARB panel member, count the number of exposures to an advertising schedule
  - aggregate the sample
- Under estimates GRPs
  - guest viewing
  - panel turnover
- Fit a probability model (NBD) to frequency distribution
  - scale parameter = GRPs
  - reach build parameter = derived from 1+ frequency
- Create calibrated frequency distribution
  - increase scale parameter to gold standard GRPs
  - keep same reach build parameter



# Channel 4 – Integration of “Big VoD Data”

- Use an external survey to calculate relative rates of Channel 4 VoD viewing for different segments of the population
- Make a proportional allocation of site centric VoD advertising schedule GRPs to each segment
- Use the BARB TAM panel to calculate reach and frequency for the traditional TV schedule
  - standard methodology
  - for each segment
- Include Vod Schedule by increasing the probability model (NBD) scale parameter to:
  - traditional TV + VoD GRPs
  - for each segment



# So What's The Prognosis So Far?

- Don't be concerned about using probabilities in TAM analysis system?
    - they are widely used in other media
    - they are already in the background in TAM analysis systems
    - existing models (e.g. NBD) can be used to bridge the gap between TAM panel reach and frequency counting and site centric audiences
  - What's the downside?
    - probability models tend to support analysis specific rather than general purpose solutions
    - a move away from transparent, respondent level data may put the possibility or validity of non-standard analysis at risk
    - there may be a move to proprietary, black box modelling and analysis systems
- 



# How Else Can Probabilities Help?

- Data Integration and modelling solutions are designed to predict the probabilities of a person viewing a series of programme or commercial events
- Conversion of probabilities to familiar binary data limits the potential of calibration to site centric data
- Can we keep the probabilities and still do reach and frequency?
  - yes, but the bureaux won't like it

# Counting Frequency For One Person

## Commercial Audience

Spot Binary Probability

A 1 0.8

B 0 0.2

C 1 0.7

## Frequency Distribution

Freq Binary Probability

0 0 0.04

1 0 0.32

2 1 0.53\*

3 0 0.11

$$\begin{aligned} *Freq (2) &= 0.8 \times 0.2 \times (1 - 0.7) \\ &+ 0.8 \times (1 - 0.2) \times 0.7 \\ &+ (1 - 0.8) \times 0.2 \times 0.7 \end{aligned}$$



# The Ideal Solution

- A respondent level database:
  - looks like traditional TAM panel viewing file
  - increases granularity of measurement of all events, channels, platforms and devices
  - achieves all SC calibration goals
  - not too big
- Calculation procedures
  - preserve all objectives of existing gold standard calculations for traditional TV
  - intuitive and convenient to program



# The Ideal Solution

- Is it achievable?
  - at all?
  - without being too contrived?
  - without being spurious
  
- What about the factory operation to create the output?



# The Probability Is That Something Has To Give

- Replication of existing, transparent database structures plus full calibration to site centric data is an ambitious objective
- Probability models embedded in the calculation conventions provide an attractive solution but run the risk of supporting only a subset of audience analysis requirements
- Holding viewing data as probabilities could improve panel calibration to site centric data - but what about computer run-time?
- Existing TAM calculation conventions are already complex and future expectations are even higher
- Something really has to give



