

# Exploring the Reliability and Validity of Conjoint Analysis Studies

Christopher N. Chapman  
Microsoft Corporation

James L. Alford  
Volt Information Sciences

Edwin Love  
Western Washington University

## Introduction

There is little reported data on the reliability and validity of conjoint analysis methods akin to traditional psychometric measures. Using a set of 9 conjoint analysis studies conducted at Microsoft Hardware, we attempted to develop:

- A set of statistical procedures to examine conjoint analysis reliability and validity, leading to ...
- A dataset of results that we could use to predict the effect of changing parameters such as sample size in our future studies

## Methods

### General

All methods use Hierarchical Bayes part worth (HBPW) data computed from Choice-Based Conjoint (CBC), Adaptive Conjoint Analysis (ACA), and Adaptive Choice-Based Conjoint (ACBC), from Sawtooth Software SSI/Web system.

### Split-Sample Reliability of HBPW Mean Beta

Split sample into equal halves, and compute HBPWs for each. Compare mean HBPW estimates between halves (*r* of mean beta for sample 1 vs. sample 2).

### Split-Test Reliability of Product Preference

Compute HBPWs separately using random half of CBC trials. Compute and compare preference share for 10000 random product pairs (*r* of 10000 preference shares for Half1 vs. Half2).

### Split-Sample Reliability of Preference Share

Split sample into equal halves, and compute HBPWs for each. Compute and compare preference share for 10000 random product pairs (*r* of 10000 preference shares for Sample 1 vs. 2).

### Validity

- Assess agreement of CBC and ACBC prediction of preference vs.
- Comparative conjoint method (e.g., CBC vs. ACBC)
  - Holdout trial in CBC block (CBC or ACBC vs. Holdout)
  - Actual product selection in similar or different Trial format

## Data

Surveys of PC hardware devices with US adult samples, administered online or in person with Sawtooth Software SSI/Web.

## Results

### Reliability measures

#### Split-Sample: Correlation of group mean beta by sample size

Study #	Method	Attributes/levels	Sample type	N	r
<i>r of mean beta utilities between Focus Group samples</i>					
1	ACA	12/32	Between group	36/38	0.993
3	CBC	6/16+none	Between group	16/16	0.974
4	CBC	6/16+none	Between group	14/15	0.943
5	CBC	6/26+none	Between group	40/39	0.930
<i>r of mean beta utilities between Online Survey samples</i>					
6	CBC	6/21+none	Between group	20/20	0.804
7	CBC	7/24	Between group	20/20	0.655
6	CBC	6/21+none	Between group	40/40	0.743
7	CBC	7/24	Between group	40/40	0.787
6	CBC	6/21+none	Between group	100/100	0.918
7	CBC	7/24	Between group	100/100	0.899
6	CBC	6/21+none	Between group	200/200	0.970
7	CBC	7/24	Between group	200/200	0.974
6	CBC	6/21+none	Between group	504/504	0.996

#### Split-test (within-subject): Approximate test-retest reliability for preference share *s*

*r of preference share estimates across 10000 random product pairs*

6	CBC	6/21, 16 trials (split 8/8)	Between estimate	1008	0.975
7	CBC	7/24, 12 trials (6/6)	Between estimate	400	0.913
8	CBC	13/52, 10 trials (5/5)	Between estimate	1008	0.944
2	CBC	5/16, 12 trials (6/6)	Between estimate	665	0.986

#### Split-sample reliability of preference share estimates

##### Preference Share split-sample agreement

6	CBC	6/21+none	Between group	20/20	0.711
7	CBC	7/24	Between group	20/20	0.577
6	CBC	6/21+none	Between group	40/40	0.601
7	CBC	7/24	Between group	40/40	0.639
6	CBC	6/21+none	Between group	100/100	0.798
7	CBC	7/24	Between group	100/100	0.850
6	CBC	6/21+none	Between group	200/200	0.934
7	CBC	7/24	Between group	200/200	0.939
6	CBC	6/21+none	Between group	504/504	0.988

### Validity measures

#### Method agreement

*Method agreement of choice preference (random product pairs)*

7	CBC & ACBC	7/24	N = 400	0.638	0.534	0.196	Slight
7	CBC & ACBC	7/24	400	0.695	0.533	0.347	Fair

#### Holdout trial prediction

*CA method's prediction of holdout trial choice*

7	CBC	7/24	400	0.705	0.494	0.417	Moderate
7	ACBC	7/24	400	0.655	0.549	0.236	Fair
9	CBC	3/13	479	0.808	0.551	0.572	Moderate

#### Prediction of actual choice, within-subject

*Holdout trial prediction of actual product choice (same format)*

9	CBC	3/13	82	0.793	0.612	0.465	Moderate
---	-----	------	----	-------	-------	-------	----------

*CA prediction of actual product choice (same format)*

9	CBC	3/13	82	0.793	0.623	0.450	Moderate
---	-----	------	----	-------	-------	-------	----------

*Holdout trial prediction of actual product choice (extended description)*

7	CBC	7/24	90	0.489	0.481	0.014	Slight
---	-----	------	----	-------	-------	-------	--------

*CA prediction of actual product choice (extended description format)*

7	CBC share	7/24	100	0.560	0.490	0.136	Slight
7	ACBC share	7/24	100	0.530	0.490	0.080	Slight

CBC split-sample reliability for mean beta exceeded 0.90 in **focus group** settings with  $N \geq 14$

Online CBC surveys needed  $N \geq 100$  to achieve split-sample reliability for mean beta of 0.90

Within-subject split-test reliability was very high for CBC with as few as 5 trials per block.

Split-sample reliability for **group-level preference share** needed  $N \approx 200$ .

Conjoint methods predicted choices presented in *identical formats* well but were hardly better than chance at within-subject accuracy when product presentation format *differed*.

We suggest to use group-level models when possible, and to use other validity measures in addition to holdout trials.

## Discussion

**Reliability.** We found that conjoint analysis (CA) methods demonstrate strong psychometric reliability for group-level part worth and preference share estimates.

The results suggest that, for products and categories similar to PC accessories:

- For online CBC,  $N \approx 100$ -200 is needed for group-level studies
- For in-person CA,  $N \leq 15$  may be adequate with good sampling
- CBC surveys may be shortened, possibly to  $K \approx 5$ -8 trials

**Validity.** Validity of CA is more difficult to establish. CA methods were good at predicting *within-survey holdout tasks* presented in *identical format*. When format differed, prediction accuracy *within-subject* was little better than chance.

However, this validity limitation was observed for *within-subject* results, which may be of lower interest to many researchers than *group-level* results. The authors have separately demonstrated high validity for CBC and ACBC in a group-level market share study (Chapman et al 2009). Still, the limitations here may be important for subject-level analyses such as segmentation or small population research.

- Future.** In future research, we hope the CA community will:
- Consistently report reliability measures for CA
  - Extend these analyses to additional product categories
  - Report measures other than holdout trials for CA validity
  - Further investigate respondent-level reliability & accuracy

## References

Chapman, C.N., Alford, J.L., Johnson, C., Lahav, M., and Weidemann, R. (2009). Comparing results of CBC and ACBC with real product selection. *Proceedings of the 2009 Sawtooth Software Conference*, Del Ray Beach, FL, March 2009.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, vol.20, no.1, pp. 37–46. [kappa statistic]

R Development Core Team (2008). R: A language and environment for statistical computing [Computer software]. Version 2.8.1. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>. [analysis platform]

Sawtooth Software (2008). CBC 6.0 Technical Paper. [survey platform]

## Corresponding author

Chris Chapman  
1 Microsoft Way (cchap), Redmond, WA 98052  
chris.chapman@microsoft.com

*R analysis scripts available from author by email request*