INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

NUMBER OF LEVELS EFFECT IN ADAPTIVE AND CHOICE-BASED CONJOINT ANALYSIS

Giuseppina Virelli

A Thesis

In

The John Molson School of Business

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Science in Administration at
Concordia University
Montreal, Quebec, Canada

December 2001

© Giuseppina Virelli, 2001

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68430-X

Canada

ABSTRACT

Number of Levels Effect in
Adaptive and Choice-Based Conjoint Analysis

Giuseppina Virelli

Conjoint analysis is a widely used marketing research tool for analyzing consumer preferences for a set of product/service attributes and estimating part-worths for the levels of each attribute. Although numerous studies have confirmed the reliability and validity of part-worth estimates, several studies have reported what was called the Number of Levels (NOL) effect, especially in full-profile conjoint designs. The NOL effect occurs in conjoint analysis when the estimated relative importance weight of an attribute increases as a function of the number of levels by which it is defined. This study investigates the NOL effect for Adaptive and Choice-Based Conjoint Analysis (ACA and CBC), two increasingly popular data collection methods. Three experiments were conducted with 684 physicians in Canada to test the hypothesis that NOL effect will be observed for both ACA and CBC but the effect will be stronger for CBC than ACA. The findings suggest that while CBC suffers from the NOL effect, ACA does not. The results also suggest that a less important attribute may be more susceptible to the NOL effect. Given the findings, it is recommended that when CBC is used it is better to have equal number of levels across all attributes in the conjoint design whenever possible. Also, relative importance weights from two different studies where the number of levels differ should not be compared. ACA does not seem to have these limitations.

ACKNOWLEDGEMENTS

A sincere thank-you to my supervisor, Dr. Kemal Büyükkurt, for the guidance, support and motivation he has shown me throughout the thesis process.

I would also like to thank P\S\L Consulting for allowing me to conduct the experiments with physicians, making the experience even more rewarding. Thanks especially to Nora, who helped tremendously in the data collection process.

To my family and friends, for their invaluable advice and suggestions, for their love and support – and patience – Thank You!

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES vii

SECTION 1: INTRODUCTION 9

SECTION 2: LITERATURE REVIEW 14

2.1. Conjoint Analysis in the Literature 15
2.2. Data Collection Methods 16
2.3. Number of Levels Effect in the Literature 18
2.4. Hypotheses of the Study 28

SECTION 3: METHODOLOGY 30

3.1. Overview 31
Study 1 31
Study 2 33
Study 3 34

SECTION 4: DATA ANALYSIS AND RESULTS 36

4.1. Data Analysis 37

4.2. Hypotheses Testing 39
Study 1 39
Study 2 42
Study 3 45

SECTION 5: DISCUSSION 48

5.1. Summary of Findings 49
5.2. Conclusions and Future Research 51

TABLES AND FIGURES 53

REFERENCES 78

APPENDIX A 84
How ACA estimates Respondent Utilities

APPENDIX B 94
Hierarchical Bayesian Multinomial Logit Regression Model

LIST OF TABLES AND FIGURES

Tables:

Table 1 54
NOL Effect in the Literature

Table 2 55
Summary of Experimental Design (Study 1)

Table 3 56
Summary of Experimental Design (Study 2)

Table 4 57
Summary of Experimental Design (Study 3)

Table 5 58
Average Relative Importance Weights - Study 1 (CBC)

Table 6 59
Regression Coefficients for Equations Predicting Estimated
Importance Weights – Study 1 (CBC)

Table 7 60
Average Relative Importance Weights - Study 2 (ACA)

Table 8 61
Regression Coefficients for Equations Predicting Estimated
Importance Weights – Study 2 (ACA)

Table 9 62
Average Relative Importance Weights - Study 3 (ACA & CBC)

Table 10 63
Regression Coefficients for Equations Predicting Estimated
Importance Weights – Study 3 (ACA & CBC)

Table 11 64
F and p-values for the Analysis of Variance (ANOVA)
Study 3 (ACA & CBC)

Table 12 65
Average Relative Importance Weights – Patient Type 1 (ACA & CBC)

Table 13 66
F and p-values for the Analysis of Variance (ANOVA)
Patient Type 1 (ACA & CBC)

Figures:

Figure 1 67
Sample Conjoint Screen for CBC (Study 1)

Figure 2 68
Sample Conjoint Screen for CBC (Study 1)

Figure 3 69
Sample Introduction Screens for ACA (Study 3)

Figure 4 70
Sample Screen for Ranking Levels of an Attribute in ACA (Study 3)

Figure 5 71
Sample Rating Screen for ACA (Study 3)

Figure 6 72
Sample Conjoint Screen for ACA (Study 3)

Figure 7 73
Sample Calibration Screen for ACA (Study 3)

Figure 8 74
Sample Conjoint Screen for CBC (Study 3)

Figure 9 75
Average Relative Importance Weights for Study 1

Figure 10 76
Average Relative Importance Weights for Study 2

Figure 11 77
Average Relative Importance Weights for Study 3

SECTION 1

INTRODUCTION

Conjoint analysis is a popular decompositional method for analyzing consumer preferences for a predetermined set of product/service attributes. It is a decompositional model because the analyst needs only to know a respondent's overall utility of an object and the characteristics of that object. Based on a respondent's overall evaluation of product/service concepts, part-worths for each attribute and its levels are determined to understand consumer trade-offs (Green and Srinivasan, 1978; 1990; Carroll and Green 1995). Conjoint analysis is based on the assumption that complex decisions are based not on a single factor or criterion, but on several factors considered jointly, hence the term *conjoint* (Johnson 1974).

The estimated part-worths in conjoint analysis are often used in further analysis such as segmentation and new product design (Green and Krieger 1993). Therefore, the reliability and validity of conjoint analysis and part-worth estimation has been the focus of numerous studies (Bateson, Reibstein, and Boulding 1987; Green, Helsen, and Shandler 1988; Leigh, MacKay and Summers 1984; Reibstein, Bateson, and Boulding 1988; Wittink, Reibstein, Boulding, Bateson, and Walsh 1989).

Although these past studies have confirmed the reliability and validity of conjoint analysis and part-worth estimates in general, several studies have shown that one prominent issue affecting conjoint analysis results is the Number of Levels (NOL) effect (Creyer and Ross 1988; Currim, Weinberg, and Wittink 1981; Mishra, Umesh and Stem 1989; Steenkamp and Wittink 1994; Wittink, Krishnamurthi and Nutter 1982; Wittink, Krishnamurthi and Reibstein 1989; Wittink, Huber, Fiedler, and Miller 1991; Wittink, Huber, Zandan, and Johnson 1992; Wittink, McLauchlan, and Seetharaman 1997; Wittink and Seetharaman 1999). The NOL effect occurs in conjoint analysis when attributes with a greater number of levels take on a greater importance than attributes with a lower number of levels. Two main reasons are suggested for the NOL effect: psychological and mathematical (Currim et al 1981: Green and Srinivasan 1990; Steenkamp and Wittink 1994; Wittink and Krishnamurthi 1981; Wittink et al 1989; Wittink et al 1992). According to the psychological explanation, respondents pay more attention to an attribute with a greater number of levels than one with a lower number of levels. However, the mathematical explanation states that if the conjoint judgements are ordinal-like the estimation that is used to calculate part-worths inflates the relative importance of an attribute as the number of levels for that attribute increases.

Several studies in the NOL effect literature focusing in the full-profile data collection method have shown large NOL effects (Wittink et al 1982; Creyer and Ross 1988; Gandais 1994; Wittink et al 1989; Wittink et al 1992). With this method of data collection, profiles are developed based on combinations of levels of all the attributes in the study. Respondents are then asked to rate each profile one at a time or rank all profiles in terms of overall preference.

Although the full-profile method has been one of the most studied data collection methods in the NOL effect, researchers have shown an increasing interest in Adaptive Conjoint Analysis (ACA). Numerous studies have reported the use of ACA as a reliable and valid data collection method for conjoint analysis (Agarwal and Green 1991; Green, Krieger, and Agarwal 1991; Huber, Wittink, Fiedler, and Miller 1991; Huber, Wittink, Fiedler, and Miller 1993). ACA has become one of the most popular data collection methods among market researchers in Europe and the U.S.A. (Wittink, Vriens, and Burhenne 1994; Green, Krieger, and Agarwal 1991).

ACA, introduced to marketing research in the late 1980's (Johnson 1987), collects conjoint data via a "computer-based interactive technique that customizes the stimulus presentations that respondents evaluate." (Green et al 1991, p.215). This is done using a combination of self-explicated rankings and rating scales in addition to a paired profile section of the conjoint session that uses rating scales. The NOL effect has also been studied using the ACA data collection method and studies have shown a smaller NOL effect compared to the full-profile method (Wittink et al 1991; Wittink et al 1992; Wittink et al 1997; Wittink and Seetharaman 1999).

Choice-Based Conjoint (CBC) is another data collection method that has also received some attention by researchers (Elrod, Louviere, and Davey 1992; Huber, Wittink, Johnson, and Miller 1992; Vriens, Oppewal, and Wedel 1998). The CBC method asks the respondent to choose the best profile (among a number of profiles) out of successive choice sets. This method is considered to be more realistic than ratings/ranking because consumers usually just choose a product/service from a set that is available, and do not generally rate alternatives (Huber et al 1992). Despite its realistic choice context, only one study has investigated the NOL effect in CBC (Wittink et al 1997) and found that the NOL effect was greater for the CBC data collection method than for the ACA.

The lack of studies on the NOL effect for the ACA and CBC data collection methods and the increasing popularity of both data collection methods (and the associated software programs) warrant further investigation into the topic to examine the generalizability of the previous findings.

This paper will attempt to investigate whether the NOL effect in conjoint analysis exists for both the ACA and CBC data collection methods. If the effect does exist, its magnitude will also be estimated. An extensive literature review on conjoint analysis and the NOL effect will be presented next, followed by the formulation of hypotheses.

SECTION 2

LITERATURE REVIEW

2.1. Conjoint Analysis in the Literature

Historically, the origin of conjoint analysis can be traced back to an article by Luce and Tukey (1964). It wasn't until 1971, however, that the first consumer research paper on conjoint analysis appeared by Green and Rao (1971), introducing conjoint to the marketing research world.

Two studies by Cattin and Wittink (1982) and Wittink and Cattin (1989) examined commercial use of conjoint analysis in the United States, and Wittink, Vriens, and Burhenne (1994) studied the commercial use of conjoint in Europe. Cattin and Wittink (1982) estimated that approximately 700 commercial applications of conjoint analysis had been used in the U.S.A between 1971 and 1982. The majority of the projects were carried out in the consumer goods category accounting for approximately 60% of all reported uses. New product or concept identification was the most common use of conjoint analysis. Six years later Wittink and Cattin (1989) updated their previous article and found that in the five-year period from 1981-1985, approximately 1062 marketing research projects had been conducted using conjoint in the U.S.A alone. One of the main reasons given for this increase is the introduction of conjoint software for the PC environment around the same time. The authors again found that about 60% of the conjoint analysis projects were in the consumer industry. However, the services sector made the most gains since their 1982 study. The most used data collection method remained the full-profile method.

Wittink, Vriens and Burhenne (1994) reviewed the commercial use of conjoint analysis in Europe. One of the findings was that European firms, in large part, started using conjoint analysis after 1986. Approximately 54% of projects were in the consumer goods industry. However, the primary use of conjoint analysis was different in Europe than the U.S.A. Pricing, rather than new product/concept identification was the objective most cited for using conjoint analysis in Europe whereas pricing was ranked third most popular objective in the U.S.A. The authors also mention a rise in the use of computerized data collection methods in Europe compared to previous periods.

2.2. Data Collection Methods

One of the most cited data collection methods for conjoint analysis is full profile (Wittink et al 1982; Creyer and Ross 1988; Gandais 1994; Wittink et al 1989; Wittink et al 1992). The full profile method presents the respondent with a complete product/service profile based on a combination of the attribute levels of interest. These combinations of levels are produced through a full factorial design. Each profile is then rated or ranked by the respondent. Part-worths are then calculated for each individual using the respondents' ratings or rankings of all the profiles.

ACA refers to Adaptive Conjoint Analysis and it uses adaptive hybrid conjoint modeling, combining self-explicated evaluations with paired conjoint comparisons (Johnson 1987; Green, Kriger, and Agarwal 1991). What sets ACA apart from other data collection methods is that the software interactively adapts the conjoint session to each respondent based on the earlier responses of the respondent in the session. Collected data are analyzed as the interview is in progress so further questions regarding consumer trade-offs can focus on those attributes and levels that the respondent indicates as most important. This reduces the number of questions (compared to full profile) and reduces the time required to complete the survey. Ordinary Least Squares Regression is used to compute utility values (Appendix A).

CBC refers to Choice-Based Conjoint analysis. Although discrete choice conjoint analysis has been discussed by academics since the early 80's, it has not gained widespread commercial use until the past five years (Louviere and Woodworth 1983; Elrod et al 1992; Huber et al 1992; Vriens et al 1998). Its distinguishing feature is that respondents express their preferences by choosing concept descriptions from sets of concepts rather than by rating or ranking each of the concepts individually. This simulates to some degree real-world decisions where products/services are chosen from an existing set. The weights reflecting the part-worth of each attribute and its levels are computed using Multinomial Logit Regression. In recent years, hierarchical bayes (HB) estimation techniques have been used for estimating the coefficients with CBC (Allenby, Arora, and Ginter 1995, Lenk, Desarbo, Green, and Young 1996). Whereas CBC estimates weights on an aggregate level, HB extension of CBC allows the estimation of weights on an individual respondent level. A detailed description of the Hierarchical Bayesian Multinomial Logit Regression model is provided in Appendix B.

2.3. Number of Level Effects in the Literature

The first article to report the NOL effect was by Currim, Weinberg and Wittink (1981). They conducted a conjoint study to determine the impact of six factors on the demand for a subscription series to performing arts events. Three of the factors of the conjoint design were defined on three levels and the other three factors were defined on two levels. They found that three-level attributes all ranked higher in importance than two-level attributes within the same study using ordinary least squares analysis of preference rank orders. The two-level attributes had importance weights ranging from 0.2 to a maximum of 0.6. The three-level attributes had importance ranging from 0.4 to a maximum of 0.8.

This is a significant in conjoint analysis because one of the objectives of conjoint analysis is to assess the preference shares that would be achieved by certain combinations of attribute levels in choice simulations. If the number of levels of certain attributes affect the relative importance of those attributes, this can have significant consequences on the decisions taken based on choice simulations. The inflation of part-worths as the number of levels increases obviously introduces a bias into decisions based on part-worths.

Currim et al (1981) gave two possible explanations for this occurrence: Psychological and Mathematical. The psychological explanation states that a respondent may pay more attention to the attributes with the higher number of levels than (s)he otherwise would, and this results in a higher weight for that attribute. The mathematical explanation argues that if the conjoint judgements made by the subject are ordinal or ordinal-like in nature, ordinary least squares applied to such data increases the relative importance of attributes with higher number of levels. The authors concluded that direct comparisons of conjoint analysis results can be made only across attributes with the same number of levels (Currim et al 1981).

The first study to fully document the NOL effect under experimental conditions was Wittink, Krishnamurthi and Nutter (1982). In a conjoint study of MBA students' preferences for hypothetical summer jobs, they used four attributes: functional activity (auditing, consulting, investment, banking, and marketing research), job location (Chicago, Dallas, New York, San Francisco), monthly salary ($1,200, $1,600, $2,000 and $2,400), and organization climate (flexible and easy to work with or competitive and cut-throat). While one group of students were exposed to four levels of job location and two levels of monthly salary, the other group was exposed to only two levels of job location and four levels of monthly salary. The range for monthly salary was kept constant for both groups. The study also involved two data collection methods: trade-off approach and concept-evaluation method.

The results show that the relative importance weight of the two-level monthly salary was 0.27 for the concept evaluation method and 0.29 for the tradeoff method. For the four-level monthly salary, the importance weight was 0.40 for the concept-evaluation method and 0.43 for the tradeoff method. Furthermore, the authors note that, "if salary is the most critical attribute for a respondent, then the distance between average ranks for the most-preferred and least-preferred levels increases as the number of intermediate levels increases." (Wittink et al 1982, p. 472). In effect, the number of levels by which an attribute is defined has a direct impact on the resulting importance of that attribute.

Creyer and Ross (1988) looked into the effect of range-frequency manipulations on the attribute levels in a conjoint setting. Their work is based on Parducci's range-frequency model, which states that a person's judgement process can be influenced by two fundamental features of the stimulus context: the range principle and the frequency principle. A respondent can use sub-ranges to divide the psychological range of an attribute (range principle). When a respondent uses each sub-range for a specific proportion of his or her judgement, this is the frequency principle (Creyer and Ross 1988).

The authors used seven automobile-related attributes: country of origin, body style, transmission, top speed, price, wheel size and miles per gallon (MPG). The experimental manipulation of the NOL effect involved MPG. While one version had MPG at four levels: 25, 30, 35, 40, the other version of MPG had two levels: 25 and 40, keeping the range constant. The authors tested the relative importance between these two versions as well as differences between part-worths for the 25 and 40 MPG levels. Thirty-two profiles were developed from an orthogonal, fractional factorial design. The profiles were rated by 136 undergraduate students. Each profile contained seven attributes and was rated on a 7-point scale.

The results show significant differences for relative importance between the four-level and two-level attributes, 0.147 and 0.107. They did not find any statistically significant difference between the part-worths of the 25 and 40 MPG levels (0.114 and 0.108 respectively), however the 5% change was in the expected direction. The authors conclude that estimates of the importance of an attribute are affected by the number of levels used to describe that attribute.

Another study to test the NOL effect was Mishra, Umesh and Stem (1989). They conducted a Monte Carlo simulation to test the influence of design factors on bias and precision of estimated importance weights. Their simulation included various estimation algorithms, evaluation strategy used, level of judgement error, number of levels, and number of attributes and profiles. With regards to NOL, the authors used three-level vs. two-level attributes. They found that the two-level attributes have a statistically significant lower mean absolute bias than the three level attributes ($p<0.05$). They also found that importance weights were upward biased for three-level attributes and downward biased for two-level attributes and concluded that relative importance is affected by the number of levels used to define an attribute.

In 1989, Wittink, Krishnamurthi and Reibstein confirmed the NOL effect in another study by examining the findings of a previous study conducted by Reibstein, Bateson and Boulding (1988). Reibstein et al (1988) attempted to compare data collection methods in conjoint analysis as well as the impact on reliability of the NOL effect. They manipulated price using three or five levels while keeping the range of variation in the levels constant. Using this study, Wittink, Krishnamurthi and Reibstein (1989) found that in all three data collection methods (trade-off matrix, full-profile and paired-profile comparisons), when price was defined on five levels it was deemed significantly more important (7%) than when it was defined on three levels.

Wittink, Huber, Fiedler and Miller (1991) examined the possibility of the NOL effect in Adaptive Conjoint Analysis (ACA) and full-profile conjoint analysis. Using refrigerators as the stimuli of their conjoint design, they collected data using both the full-profile and the adaptive conjoint analysis judgements with each subject. The experiment contained either five or nine attributes and manipulated the number of levels in four of the attributes: capacity, energy cost, compressor noise and price. These attributes had either two or four levels. They also manipulated the order of the attributes, the number of attributes, and the order of the data collection method.

Wittink et al's (1991) findings suggested significant NOL effects. The four-level attributes had significantly higher relative importance weights than the attributes with two levels only. As well, the NOL effect was greater in the full-profile preference judgements than in the ACA. In the full-profile method the increase in relative importance from two to four levels was somewhere between 9 and 12 percent, whereas in the ACA it was between 5 and 6 percent. The magnitude of the NOL effect for ACA is about half that of the NOL effect for full-profile. One possible explanation for this difference is that the ACA data collection method uses a combination of self-explicated (compositional) data, for which there is no NOL effect, and preference intensity judgement (decompositional) data, for which there is a NOL effect. In contrast, the full-profile data collection method is entirely decompositional, therefore the NOL effect may be expected to be greater. The authors conclude that "more research is needed to better identify the source of the effect in ACA." (Wittink et al 1991, p. 60).

Wittink, Huber, Zandan and Johnson (1992) also looked into the NOL effect, and related these findings to the commercial use of conjoint. They conducted a study using the ACA method to collect conjoint data on notebook computers using the following six attributes: brand name, size, weight, battery life, performance, and price. Only size, weight, battery life and price were manipulated to generate the NOL effect. Half the respondents saw two levels of notebook size and battery life, but four levels of weight and price. The other half of the respondents saw four levels of notebook size and battery life, but two levels of weight and price. To verify whether there could be a behavioral explanation of the NOL effect, the authors included tutorials and prompting to half the respondents. The tutorial information provided definitions of the attributes to increase the respondents' understanding of those attributes. This is based on the assumption that respondents who do not fully understand the attributes may interpret attributes with more levels as having more importance. A tutorial was expected to reduce the NOL effect (Wittink et al 1992). Prompting, on the other hand, was used randomly throughout the exercise to account for boredom and random answering by respondents. The respondent was prompted with the message, "Are you sure about your answer of x? Please think some more about the strength of your preference. Press any key to answer the question again." (Wittink et al 1992, p. 4). The authors hypothesized that prompting the respondent from time to time about their responses might produce better responses.

Wittink et al (1992) also manipulated the ACA program by "unbalancing" the conjoint section of the program. ACA ordinarily tries to generate pairs of profiles in the conjoint session that "are balanced", that is, have nearly equal utility. The authors argue that because of "balancing" the paired profiles appearing in the conjoint session, ACA decreases the magnitude of the NOL effect. They expected that the respondents who received the "balanced" pairs would display a smaller NOL effect than those who received "unbalanced" pairs.

In ACA, "balanced" pairs are generated by eliminating dominated pairs of attributes from appearing together in the pairs section. For example, assume that Cost and Brand Name were both rated highly by a respondent in the self-explicated portion of

ACA (the first portion). For the paired profiles section, ACA would not select a profile to be better than another on both Cost and Brand Name because that profile would be more likely to "dominate" other potential profiles and this would be reflected in the rating of the individual. For example, if the profile on the right side of an ACA screen (see Figure 6) for pair comparisons dominated the profile on the left, the respondent's rating will be close to the right end of the scale (8 or 9). For nondominant or "balanced" pairs, however, the respondents have a greater propensity to rate the profiles toward the midpoint of the scale which shows relative indifference. This tendency to rate at the midpoint of a scale somehow reduces the NOL effect for the ACA data collection method.

Wittink et al (1992) found that the NOL effect was higher for those respondents who saw pairs that were not "balanced" (the modified ACA), however, only two of the attributes showed statistically significant results: size and price ($p < .01$). The NOL effect was between 1.34 and 4.00 when the pairs were balanced, and between 4.60 and 6.76 when they were not. The authors found no significant differences due to the tutoring and prompting. Therefore, they concluded that the source of the NOL effect was algorithmic in nature and not behavioral.

Gandais (1994) also looked into the NOL effect in metric conjoint analysis. He tested the relationship of NOL with the number of attributes used to construct conjoint profiles, the number of response categories of the rating scales (seven vs. eleven point scales), the nature of the attribute (discrete vs. quasi-continuous), and the overall relative importance of the attribute. The experiment was conducted with 248 undergraduate students using corded residential telephones as the stimuli and full-profile conjoint method. His findings suggest that the NOL effect exists for both quasi-continuous and discrete attributes, but is not affected by the number of response categories of the rating scale. The author also found that NOL effect is stronger when the conjoint study involves a smaller number of attributes. Overall, a 10% difference between relative importance weights of two- versus four-level attributes was found.

In an attempt to empirically test the mathematical explanation for the NOL effect, Steenkamp and Wittink (1994) examined if the NOL effect depended upon the lack of metric quality in the preference judgments in full-profile conjoint analysis. The authors argued that although the NOL effect was initially seen in studies with rank-order preferences, approximately the same size NOL effect is observed for rating data as well (Wittink et al 1989). They noted that preference ratings may be closer to ordinal- than to interval-scaled measurements and therefore lack the assumed metric quality. This in turn may result in inflated part-worths for the same mathematical reason the same effect is observed for ranking data. Steenkamp and Wittink (1994) use magnitude estimation techniques, as well as rating scales for the measurement of preferences. They hypothesize that magnitude estimation techniques will be closer to interval-scaled measurements than simple ratings and, therefore, should provide a reduction in the NOL effect compared to rating scales.

Their study consisted of two products, student apartments and color televisions. Five attributes were used to describe these products, each having two to four levels. The conjoint profiles were constructed using a standard fractional factorial design producing a total of 16 conjoint profiles. Two of the attributes for each product were manipulated, alternating between two and four levels.

The authors found that significantly smaller NOL effects for respondents who satisfied the criteria for metric quality (8.6 percent) than those who did not (13.6 percent). However, no significant NOL effect existed between the two data collection methods: magnitude estimation and rating scale. The authors conclude (p.285) that the "subjects who provided metric responses showed a smaller effect than other subjects". Furthermore, they note that individual differences in response quality are more important in explaining the NOL level effect than the response mode employed (Steenkamp and Wittink (1994), p.285).

Wittink, McLauchlan, and Seetharaman (1997) is the only study to test the NOL effect for both the ACA and CBC data collection methods within the same study. Brand name, speed, hard drive, RAM, and price were the five attributes used to describe personal computers. The authors use three different conjoint designs that they named: version A, version B, and version C. In version A, respondents completed both ACA and CBC with customized numbers of levels depending on each respondent's self-explicated importance ratings. In version B, respondents completed both ACA and CBC where all the attributes contained three levels each. In Version C, respondents also completed both ACA and CBC, but saw three levels for Brand Name, two levels for Speed and RAM, and four levels for Hard Drive and Price.

Wittink et al's (1997) findings suggest the NOL effect does in fact exist for CBC and that CBC had a larger NOL effect than ACA. The average increase in relative importance index was 4 percent and 6 percent for ACA and CBC, respectively, when version B and version C of the conjoint results were compared. Version A was the customized version which varied the levels of attributes depending on each individual respondent's importance of the attributes. The results showed that the NOL effect between Version A and B was only 3%, whereas it was 14% between version A and C for the ACA collection method thereby suggesting that the NOL effect is algorithmic in nature. Unfortunately, Wittink et al (1997) presented no results for CBC data collection in comparisons between version A and B and versions A and C.

The above literature review suggests that conjoint analysis has become a popular tool of research in marketing and numerous studies provided empirical evidence regarding the reliability of the estimated part-worths associated with the levels of attributes used in conjoint designs. Numerous studies, however, showed the existence of what is later called the "number of levels" (NOL) effect, an inflation of the part-worths as a function of the number of intermediate levels along an attribute. As the reviewed studies suggest, NOL effect is generalizable since it exists for different samples, stimuli, attributes and data collection methods. In general, the magnitude of the NOL effect seems to be the highest for the full-profile method of data collection, lower for CBC and lowest for ACA (these results are summarized in Table 1).

As a result of this review, it can be concluded that a majority of the previous studies on the NOL effect focused on the full-profile and ACA methods of data collection and ignored CBC analysis with the exception of a study by Wittink et al (1997). Because of the increasing popularity of ACA and CBC as data collection methods, and related software, a further examination of the NOL effect and a comparison of its magnitude for ACA and CBC are needed.

2.4. Hypotheses of the Study

Since several studies empirically verified the NOL effect for the ACA data collection method (Wittink et al 1991; Wittink et al 1992; Wittink et al 1997; Wittink and Seetharaman 1999), the first hypothesis states that the NOL effect will be observed in ACA.

Hypothesis 1: Increasing the number of intermediate levels of an attribute while keeping the end levels constant will increase the relative importance of that attribute in ACA.

Although only one study (Wittink et al 1997) has shown an NOL effect for the CBC data collection method, we expect this effect to appear in this study as well. Therefore the second hypothesis states that the NOL effect will also exist in CBC.

Hypothesis 2: Increasing the number of intermediate levels of an attribute while keeping the end levels constant will increase the relative importance of that attribute in CBC.

Adaptive Conjoint Analysis, as explained earlier, is a hybrid conjoint method combining self-explicated evaluations with paired conjoint comparisons. CBC, on the other hand, is choice-based and involves the choice of a conjoint profile from a set of manipulated conjoint profiles. Based on Wittink et al (1997), a smaller NOL effect is expected for ACA than for CBC.

Hypothesis 3: Increasing the number of intermediate levels of an attribute while keeping the end levels constant will have a smaller increase in the relative importance of that attribute in ACA than in CBC.

SECTION 3

METHODOLOGY

The hypotheses of the study were tested using experimental design. Three studies were conducted; Studies 1 and 2 were based on the same five attributes, whereas Study 3 was based on a different set of five attributes. Studies 1 and 2 were related to Dyslipidemia drugs that were described in terms of five attributes. Data were collected from 592 general practitioners and cardiologists. Study 3 pertained to Schizophrenia and was conducted with 92 Psychiatrists. The random sample of physicians was collected from 5 regions, British Columbia, the Prairies, Ontario, Quebec and the Maritimes in proportions representative of the Canadian physician population. The data were collected in collaboration with PSL Consulting, a pharmaceutical marketing research firm situated in Montreal. Computerized conjoint sessions on average took 15 minutes for ACA and 10 minutes for CBC.

3.1. Overview

Study 1

Study 1 used CBC data collection method and involved two patient types, both relating to Dyslipidemia. A patient type is a Dyslipidemic described to the respondents in terms medical test results on several parameters. These parameters are also the attributes used to describe the Dyslipidemia medications. Based on previous Dyslipidemia studies, the five attributes/parameters used were (in random order), Total cholesterol, Triglycerides, HDL cholesterol, LDL cholesterol, and Fibrinogen. In the conjoint task, each physician is presented with a patient type description (based on the levels of the medical test parameters) and is asked to choose from among three medication profiles, the one that would best treat this patient type. This choice task is repeated ten times for each respondent. After the first patient type conjoint analysis is complete, the physician is given a description of the second patient type and completes the conjoint exercise for this patient type. An example of this conjoint analysis is presented in Figures 1 and 2 for patient types 1 and 2. Although the attribute names are mentioned above, their levels in the patient description are not disclosed because of confidentiality.

As it is summarized in Table 2, the experimental manipulations of the NOL effect for patient types 1 and 2 were different although the same five attributes were involved for both types of patients. For patient type 1, given a base conjoint design (called version C) of five attributes and corresponding levels of 5,3,3,5 and 2 for attributes 1 through 5, the number of levels for one or two of the important attributes was increased to five levels to create conjoint design version B and A, respectively. Patient Type 2 also involved a similar kind of NOL manipulation in terms of the number of levels but the manipulated attributes were different. While the NOL manipulation involved attributes 2 and 3 for patient type 1, they included attributes 3 and 4 for patient type 2. In both cases, the manipulated attributes were the two most important attributes as indicated by previous conjoint studies conducted by the market research firm that collected the data for the current study. Because of the differences in the manipulation of the NOL effect, it is not possible to regard patient types 1 and 2 as two levels of an experimental factor. Therefore, the data associated with each patient type needs to be analyzed separately.

Study 2

Study 2 was identical to Study 1 except that it was conducted using only the first patient type and using the ACA data collection method. The five attributes mentioned above, as well as the levels of those five attributes, were identical in both studies. The patient type description for this study was identical to the first patient type description in Study 1. Again, the NOL effect was manipulated by increasing the number of levels from three to five for either attribute 3 or for attributes 2 and 3 (Table 3). Both attributes are the two most important attributes as in Study 1.

The ACA data collection method is different from CBC because it incorporates a self-explicated importance section followed by the conjoint section and then the calibration section. After the introduction screens (Figure 3), ACA asks each respondent to rank the levels within each attribute, one at a time (Figure 4). The second step consists of rating the perceived importance of the difference between the highest and lowest level of each attribute (Figure 5). The rating scale used is a four-point importance scale. The respondent must choose whether the difference between the highest and lowest level within each attribute is (4) Extremely important, (3) Very important, (2) Somewhat important, or (1) Not important at all.

The conjoint section of ACA consists of showing the respondent two potential medication profiles each defined by the levels of the five attributes (Figure 6). The respondent is asked to state his/her preference for either of the two concepts based on a nine-point scale. The respondents indicate whether they (1) Strongly prefer the profile on the left, (5) Don't care, or (9) Strongly prefer the profile on the right or any number in between depending on the direction and the strength of the their preference.

Finally, calibration concepts are added in the end to determine if respondents' judgements in previous sections are congruent with each other (Figure 7). This exercise consists of showing the respondent, one at a time, five potential medication profiles and asking their likelihood to prescribe each medication profile using a percentage from 0 to 100.

Study 3

Study 3 involved prescription drugs for Schizophrenia and the respondents were Psychiatrists. The experiment was a between-subjects design and involved two factors: (1) Method of data collection (ACA or CBC), and (2) Manipulation of the NOL effect (three vs. two levels on attribute 1). As in Studies 1 and 2, NOL effect was manipulated on attribute 1, which was found to be the most important attribute in previous conjoint studies. A summary of the experimental design is provided in Table 4.

Psychiatrists completed the computerized exercise with either ACA or CBC data collection methods. Based on previous Schizophrenia studies, the following five attributes (in random order) were determined to be important in this disease area and constituted the attributes of the conjoint sessions: anticholinergic side effects, onset of action, positive symptoms, extrapyramidal side effects, and negative symptoms. Because of confidentiality, levels of the attributes are not disclosed.

For the CBC data collection method, the physicians were simply asked to choose one of three profiles of medications indicating the drug they most likely would prescribe to a Schizophrenia patient in general. Such choice scenarios were repeated ten times for each respondent. An example of this conjoint exercise is presented in Figure 8. The

ACA data collection method, as described in Study 2 above, consisted of three sections: self-explicated importance, conjoint profiles, and calibration (see Figures 3-7).

SECTION 4

DATA ANALYSIS AND RESULTS

4.1. Data Analysis

The data collected from 684 physicians who completed conjoint tasks using the ACA or CBC data collection methods (computerized) were analyzed using the ACA and CBC programs by Sawtooth Software. Both programs produce part-worths (utility values) for each level of each attribute. CBC uses the Hierarchical Bayesian approach to multinomial logit regression and ACA uses ordinary least squares to estimate part-worths for each of the levels (please see Appendices A and B for algorithmic details).

Given the part-worth estimates, the importance of an attribute can be expressed as the difference in part-worths between the most and least preferred levels for that attribute. To be able to compare these importance weights across the respondents, the relative importance of attributes is calculated as follows (Wittink et al, 1991):

$$RIMP_i = (MAX_i - MIN_i) / \sum_{1}^{t} (MAX_k - MIN_k)\ ,\ \ k=1,2,\ldots,t$$

Where:

RIMPi is the relative importance of the ith attribute, (Σ $RIMP_i = 100$),
i denotes the attribute number ($i = 1,2,\ldots,t$),
MAX_i is the largest part-worth of the ith attribute,
MIN_i is the smallest part-worth of the ith attribute.

In order to test the hypothesis related to the number of levels effect, linear regression and ANOVA were used. Fourteen different linear regressions of the estimated relative importance weights on the number of levels were conducted. Following Wittink et al (1991) each regression equation can be expressed as follows:

$$ERIMP_{ij} = f\,(K_j, LEV_{ij})$$

Where:

$ERIMP_{ij}$ is the estimated relative importance of the jth attribute for the ith subject.

K is the constant term of the regression equation.

LEV_{ij} is the code reflecting the NOL manipulation for the ith subject (-1 indicates the number of levels in the base design. +1 indicates the increased number of levels).

Univariate Analysis of Variance (ANOVA) was used to test for the effects of number of levels and data collection method on the relative importance weights. The equation can be written as follows:

$$ERIMP_{ij} = f\,(K_j, LEV_{ij}, COL_{ij}, LEV_{ij} * COL_{ij})$$

Where:

$ERIMP_{ij}$ is the estimated relative importance of the jth attribute for the ith subject.

K is the constant.

LEV_{ij} is the code reflecting the NOL manipulation for the ith subject (no manipulation = -1, manipulation = +1).

COL_{ij} is the data collection method of the jth attribute for the ith subject (ACA = -1, CBC = +1),

$LEV_{ij} * COL_{ij}$ is the interaction term between number of levels and data collection method.

In the following discussion, the results are presented for each study. The estimated relative importance weights will be discussed first, followed by number of levels effect tests, which include the regression equations and ANOVA.

4.2. Hypothesis Testing

Study 1

The average relative importance weights for the manipulated attributes in Study 1 are presented in Table 5. For patient type 1 attribute 2 was the most important attribute for all three versions of the conjoint design ($\mu_{att2a} = 46.2$, $\mu_{att2b} = 43.2$, $\mu_{att2c} = 46.7$), followed by attribute 3 ($\mu_{att3a} = 26.3$, $\mu_{att3b} = 25.3$, $\mu_{att3c} = 21.4$). This confirms previous findings for this patient type that attribute 2 was the most important attribute followed by attribute 3. For patient type 2, attribute 4 was the most important for conjoint designs A and C but not B ($\mu_{att4a} = 41.7$, $\mu_{att4b} = 36.2$, $\mu_{att4c} = 42.5$), followed by attribute 3 ($\mu_{att3a} = 39.6$, $\mu_{att3b} = 41.5$, $\mu_{att3c} = 36.8$). In conjoint design B, average relative importance weight for attribute 3 was higher than for attribute 4. Previous findings for patient type 2 have shown that attribute 4 was most important followed by attribute 3. In this study, versions A and C confirm this, however, version B reverses the order of importance for these two top importance attributes.

Figure 9 represents graphically the relative importance weights for the manipulated attributes across the three versions of the conjoint design for patient types 1 and 2. The relative importance weight is consistently higher when an attribute is defined on five levels than when it is defined on three levels, with a couple of exceptions. In patient type 1, when comparing attribute 2 in versions C (three levels) and A (five levels), the average relative importance weight was almost identical (μ_{att2c} = 46.7 and μ_{att2a} = 46.2). For patient type 2, when comparing attribute 4 in versions C (three levels) and A (five levels), the average relative importance weight was slightly higher for version C (μ_{att4c} = 42.5 and μ_{att4a} = 41.7).

To test for the effects due to the number of levels, four regression analyses were conducted for each patient type. For Study 1, the parameters for the four regression equations, along with the F-test results, are presented in Table 6 for each patient type. For patient type 1, the results show a significant NOL effect for attribute 3 when the importance weights in conjoint design C and B are compared ($F = 7.862$, $p < 0.006$) and when the importance weights in conjoint design C and A are compared ($F = 14.051$, $p<0.000$). This confirms that the estimated relative importance weight is significantly greater when attribute 3 is defined on five rather than three levels.

The regression coefficients reflecting the NOL effect for the two equations just mentioned are 0.020 and 0.025 for attribute 3 (Table 6). If the effect of the independent variable coding of –1 and +1 is taken into account, the estimated relative importance increases by about 4% when the number of levels of attribute 3 is increased. If the number of levels of both attribute 3 and attribute 2 are increased, the relative importance weight for attribute 3 increases by about 5%. Although both these coefficients are significant and support Hypothesis 2, the increases are slightly lower than the ones reported in Wittink et al (1997) who reported a 6% increase in relative importance weights due to the number of levels effect for the CBC data collection method and approximately 4% in the ACA method.

The NOL effect was not statistically significant for comparisons of conjoint designs involving manipulation of attribute 2. The F and associated p-values for conjoint designs B vs. A and C vs. A were ($F = 2.116$, $p < 0.147$) and ($F = 0.064$, $p < 0.800$) respectively. It is also important to note that the difference in the relative importance weights between attribute 2 and attribute 3 was about 20% in all three conjoint designs (A, B and C) for patient type 1. Attribute 2 was clearly the most important attribute and was not affected by the number of levels affect, whereas attribute 3, a distant second in importance, was indeed affected by the number of levels effect.

The results for patient type 2 (Table 6) show a significant effect due to the number of levels for attribute 3, when importance weights for conjoint designs C and B are compared ($F=4.294$, $p < 0.040$) and for attribute 4 when importance weights for conjoint designs B and A are compared ($F=8.926$, $p < 0.003$). The regression coefficients associated with the NOL effect were 0.024 and 0.027 respectively for the regression equations. Taking into account the independent variable coding of -1 and $+1$, as the number of levels increases for each of attributes 3 and 4 separately, the estimated relative importance weights will increase by 4.8% and 5.4% respectively. These percentage increases in estimated relative importance weight are closer to what was previously reported by Wittink et al (1997) for the CBC data collection method (6%). It is interesting that when the number of levels for both attribute 3 and attribute 4 are increased jointly, there is no statistically significant NOL effect as evidenced in the comparison of importance weights for conjoint designs C vs. A for attribute 3 ($F= 2.108$, $p < 0.147$) and attribute 4 ($F=0.187$, $p < 0.666$).

A potential behavioral explanation for this interesting finding related to the experimental design for patient type 2 is that both attribute 3 and attribute 4 are highly important to the subjects, with average importance weights of approximately 38.2% and 42% respectively. Because of their relatively close importance, the subjects pay close attention to both attributes when they express their judgements. Therefore, increasing the number of intermediate levels for both these attributes does not attract any more attention than already paid by the physicians. For patient type 1, however, the difference in average relative weights between the most important attribute and the second (attributes 2 and 3 respectively) are approximately 20%. NOL manipulation may attract more attention in the case of the relatively less important attribute 3 in comparison to attribute 2. This in turn may be reflected in terms of higher importance in respondents' judgements. Gandais (1994) reported similar findings for increased NOL effect for relatively unimportant attributes in full profile metric conjoint analysis.

Study 2

As mentioned above, Study 2 replicated the patient 1 part of the experimental design for Study 1 but substituted CBC for ACA data collection and estimation method. Study 2 had three versions of the conjoint design, A, B, and C (Table 3). Version C of the conjoint design had no manipulations and can be considered the base design. In version B, Attribute 2 was manipulated to have five instead of three levels. In version A, Attribute 2 and Attribute 3 were manipulated to have five instead of three levels. Attribute 2 was considered the most important attribute for this patient type from previous conjoint studies, whereas Attribute 3 was the second most important attribute.

The average relative importance weights for the manipulated attributes in Study 2 are presented in Table 7. For this study, attribute 2 was most important in all three versions ($\mu_{att2a} = 31.2$, $\mu_{att2b} = 29.8$, $\mu_{att2c} = 33.5$), followed by attribute 3 ($\mu_{att3a} = 23.2$, $\mu_{att3b} = 22.0$, $\mu_{att3c} = 20.8$). This is consistent with previous findings for this patient type using the CBC data collection method, however, the estimated relative importance weights for both attribute 2 and 3 are lower using the ACA data collection method than the CBC data collection method. The difference in relative importance between these two attributes in Study 1 was approximately 20% whereas in Study 2, using ACA data collection, the difference in relative importance was about 10%. This indicates that approximately a 10% difference in relative importance weights was obtained by conducting the same study with two different conjoint data collection methods (ACA and CBC). This may be due to the fact that ACA uses self-explicated importance measurements along with conjoint task to produce the part-worths. CBC only uses conjoint tasks.

Figure 10 shows graphically the relative importance weights for the manipulated attributes in Study 2. The relative importance weights are higher when an attribute is defined on five levels versus three levels, except in one case. The only exception was for attribute 2 when comparing versions C and A, the relative importance weight is slightly higher for the three level version ($\mu_{att3c} = 33.5$) versus the five level version ($\mu_{att3a} = 31.2$). Another notable observation in this study is that the average differences in relative importance weights between the three- and five-level versions were lower using the ACA data collection method (1.67%) than using the CBC data collection method in Study 1 (3.9%).

In Study 2, four regression equations were estimated to test for the number of levels effect. The regression coefficients and the F-Test results are presented in Table 8. The results show no significant number of levels effect for Attribute 2 when the results for conjoint design B and A were compared ($F=0.247$, $p< 0.620$), and when conjoint designs C and A were compared ($F=0.741$, $p< 0.392$). Similarly, no significant NOL effects were observed for conjoint design comparisons involving C and B ($F=0.576$, $p< 0.450$) or designs C and A ($F=1.866$, $p< 0.176$) for Attribute 3. Hypothesis 1 stated that an NOL effect would occur in the ACA data collection method, however, the results of Study 2 cannot confirm any NOL effect.

Patient type 1 was used in both Studies 1 and 2 with the same attributes and levels but with different data collection methods, CBC and ACA respectively. As reported above, in Study 1 the NOL effect was strongest when the manipulation occurred on the second most important attribute, keeping in mind that there was a 20% difference between the first and second most important attributes. In Study 2, there were no significant NOL effects found on either of the two manipulated attributes, noting that there was only about a 10% difference between the first and second most important attributes (Table 12).

To test the significance of these differences in relative importance weights between the ACA and CBC data collection methods, an ANOVA was performed on number of levels and data collection method for patient type 1 as summarized in Table 13. The results indicate that for the most important attribute, Attribute 2, the data collection method is statistically significant when conjoint designs B and A were compared ($F=46.591$, $p<0.000$), and when conjoint designs C and A were compared

(F=50.723, p<0.000). Attribute 2 had a significantly lower relative importance weight using the ACA data collection method than the CBC method. These results indicate that the 10% difference in relative importance weights found between ACA and CBC methods reported above was statistically significant.

The second most important attribute, Attribute 3, did not show any significant effect due to data collection method (Table 13). However, for this attribute, the number of levels was statistically significant when comparing conjoint designs C and A (F=7.460, p<0.007) and when comparing conjoint designs C and B (F=4.463, p<0.035). This confirms that Attribute 3, the second most important, was affected by the number of levels for patient type 1.

Study 3

Whereas the previous two studies related to medication prescription for Dyslipidemia, Study 3 dealt with Schizophrenia. Another important feature of the experimental design in Study 3 is that the data collection method is a factor of the design with two levels: ACA vs. CBC [Table 4]. The subjects of the experiment were exposed to either version A or version B of the conjoint design that manipulated the NOL effect. Attribute 1, which was found to be the most important attribute in previous market research projects with the same attributes of the current study, involved the manipulation of the NOL effect with either two or three levels.

The estimated average relative importance weights for the manipulated attribute, Attribute 1, can be seen in Table 9. The relative importance weight for Attribute 1 increased from 17.9% to 24.4% when the number of levels was increased from two to three for CBC whereas it remained somewhat constant for ACA with respective weights of 26.7% and 25.9% (Figure 11). It is interesting to note that the average relative importance weight was higher for ACA than for CBC at two levels of Attribute 1 with means 26.7% and 17.9% respectively.

As presented in Table 10, two regression equations were estimated to test the number of levels effect for Study 3. The equations are related to the data collection methods, CBC and ACA. Looking at the results, the NOL effect is statistically significant only in the manipulation of Attribute 1 for the CBC data collection method ($F=5.144$, $p<0.028$). The regression coefficient for this equation was 0.032, which indicates that for Attribute 1, as the number of levels increases, the relative importance weight increases by 6.4% (taking into account -1 and $+1$ coding for independent variable). This percentage is similar to the 6% increase in relative importance weight found by Wittink et al (1997).

For the ACA data collection method there was no statistically significant NOL effect ($F=0.066$, $p<0.798$). Hypothesis 3 stated that the CBC data collection method would produce a higher number of levels effect than the ACA method. The results of Study 3 confirm the hypothesis.

An ANOVA was also performed on the number of levels and the data collection method as summarized in Table 11. The results reveal that the data collection method (ACA vs. CBC) has a significant effect on predicting relative importance weights ($F=6.547$, $p<0.012$). The number of levels effect was not statistically significant ($F=2.004$, $p<0.160$). The interaction between number of levels and data collection method was not significant either, but the means were in the expected direction ($F=3.164$, $p<0.079$). The small cell sizes, each under 30, may account for the lack of significance.

The results of Study 3 indicate that the number of levels effect occurred only in the CBC data collection method but not in ACA. Another interesting finding of the study was that Attribute 1 turned out to be the second most important attribute for CBC whereas it was the most important attribute for ACA as suggested by previous studies. This suggests the possibility that the ranking of attributes in terms of relative importance weights may vary across ACA and CBC which needs to be investigated further in other studies.

SECTION 5

DISCUSSION

5.1. Summary of Findings

In this paper, the Number of Levels (NOL) effect in Adaptive Conjoint Analysis (ACA) and Choice-Based Conjoint (CBC) data collection methods was investigated by conducting three experiments with a total of 684 medical doctors as subjects. Because of its emphasis on data collection methods and the related estimation algorithms, the reported experiments examine the mainly algorithmic explanations of the NOL effect.

The first hypothesis which stated that the estimated relative importance of an attribute increases as the number of intermediate levels used to define that attribute increases, in ACA data collection method, was not supported. Although the average relative importance weights of the attributes showed some increase when going from three to five levels, there was no statistically significant finding confirming the hypothesis.

The second hypothesis stated that the estimated relative importance of an attribute would increase as the number of intermediate levels used to define that attribute increased, in the CBC data collection method. This hypothesis was supported in Study 3 and partially supported in Study 1. With a couple of exceptions, in both of these studies the average relative importance weights of the manipulated attributes were higher when defined on more levels. The reported percentage increases in relative importance weights were similar to those previously reported by Wittink et al (1997).

Finally, the third hypothesis, which stated that the number of levels effect would be stronger in the CBC data collection method than in the ACA data collection method was partially supported. In fact, since no level effect was found to be statistically significant in the ACA data collection method, the CBC data collection method certainly had a stronger NOL effect. When the differences in the average relative importance weights for the attributes defined on a fewer number of levels and those defined on a greater number of levels are examined, it is clear that the larger differences are found for the CBC data collection method than for ACA. These differences were in the expected direction and were as high as 6.5% for CBC (Table 9) and only as high as 2.4% for ACA (Table 7).

There were also some findings that suggested that the number of levels effect may be a function of the relative importance of the manipulated attribute in a conjoint design as argued by Gandais (1994). It was found that when the differences in relative importance weights between the two most important attributes were small (10%), the NOL effect occurred on both attributes. However, when the differences in relative importance weights between the two most important attributes was large (20%), the NOL effect occurred only on the less important attribute. This suggests that a less important attribute may be more susceptible to the NOL effect than a more important attribute.

Another important finding was the significant differences found in relative importance weights between the two data collection methods, ACA and CBC. Differences of 10-15% were found between the most important attribute in patient type 1 when comparing ACA to CBC data collection methods (Table 12). The two studies, each using only one of the data collection methods, were not conducted on the same sample, which may account for some of the differences in importance.

5.2. Conclusions and Future Research

The NOL effect was found to exist in the CBC data collection method but not in the ACA data collection method. The estimated regression coefficients indicate that the increase in relative importance weights in CBC are about 5% in Study 1 and about 6.4% in Study 3. These estimates are rather close to the 6% reported by Wittink et al (1997). The NOL effect was not statistically significant in any of the ACA experiments. This may be partially due to the fact that ACA uses self-explicated measurements along with conjoint tasks to compute the part-worths and such combinations of two types of judgements may reduce any potential bias involved in conjoint judgements only. Obviously there are many differences between the ACA and CBC data collection methods and the related estimation algorithms and any one of these differences may be contributing to the observed differences in their relative sensitivity to the NOL effect.

These findings suggest that researchers should refrain from comparing the relative importance weights estimated by CBC when different number of levels are used even when the same attributes are involved in related conjoint designs. If the objective of a conjoint study using CBC data collection method is to use the estimated part-worths and the relative importance weights in comparison with the results of the previous CBC studies. similar number of levels should be involved in both studies. Since ACA seems to be less prone to the NOL effect. comparisons without much sensitivity to the number of levels involved. is more acceptable for ACA than for CBC.

The manipulation of the number of levels all occurred on the more important attributes in all three studies of this paper. For future research. it would be interesting to manipulate an unimportant attribute as well to see if the NOL effect would occur. The findings of this study suggest that number of levels effect may be higher for relatively unimportant attributes. It would be interesting to test whether a less important attribute would have a larger relative importance weight than a more important attribute when defined on a greater number of levels.

In this paper the number of levels manipulation involved three- versus five-level attributes and two- versus three- level attributes in different studies. While the first type of manipulation had an increase of two intermediate levels, the second type had an increase of only one intermediate level. However, in future studies, it may be interesting to examine within the same study whether the number of levels effect is positively related to the number of intermediate levels in the manipulation.

The observed differences in relative importance weights between the two data collection methods, ACA and CBC, also warrants further investigation. Since this study did not use the same sample of physicians for the studies conducted in each of the two methods, this may account for some of the differences. However, it would be interesting to verify whether the two data collection methods report such differences in relative importance weights when the same respondents use both of these data collection methods within the same study.

TABLES AND FIGURES

Table 1 – NOL Effect in the Literature

Study	Subjects	Product/ Service	Manipulated Attributes	Number Of Levels	Data Collection Method	Magnitude of NOL effect
Wittink et al. (1982)	MBA students	Summer jobs	Job location, monthly salary	Two vs. four	Full-profile, tradeoff paired comparisons	Monthly salary (full-profile 13%, tradeoff 14%)
Creyer & Ross (1988)	Undergrad students	Cars	MPG	Two vs. four	Full-profile	5%
Wittink et al. (1989)	Consumers	Colour TV, banking, typewriter, yogourt, long-distance service	Price	Three vs. five	Full-profile, tradeoff, paired profiles	7%
Wittink et al. (1991)	Consumers	Refrigerators	Capacity, energy cost, compressor noise, price	Two vs. four	Full-profile, ACA	Full-profile (10%) ACA (5%)
Wittink et al. (1992)	Consumers	Notebook computers	Size, weight, battery life, price	Two vs. four	ACA (balanced and unbalanced)	Balanced (3%), Unbalanced (5.5%)
Gandais (1994)	Undergrad students	Residential telephones	Price, colour, country of origin, memory	Two vs. four	Full-profile	10%
Wittink and Steenkamp (1994)	Students	Student apartments, colour televisions	APT: monthly rent, size TV: length of warranty, price	Two vs. four	Full-profile (rating scale and magnitude estimation)	10% (both)
Wittink et al. (1997)	Consumers	Personal Computers	Brand name, speed, hard drive, RAM, price	Two vs. three, and three vs. four	ACA, CBC	ACA (4%) CBC (6%)

Table 2 – Summary of Experimental Design (Study 1)

Conjoint Design	Patient Type	Sample Size[a]	# Attributes[b]	Attribute	# of levels
A	Patient Type 1	254	5	Attribute 1	5
				Attribute 2	5
				Attribute 3	5
				Attribute 4	5
				Attribute 5	2
B	Patient Type 1	102	5	Attribute 1	5
				Attribute 2	3
				Attribute 3	5
				Attribute 4	5
				Attribute 5	2
C	Patient Type 1	107	5	Attribute 1	5
				Attribute 2	3
				Attribute 3	3
				Attribute 4	5
				Attribute 5	2
A	Patient Type 2	257	5	Attribute 1	5
				Attribute 2	5
				Attribute 3	5
				Attribute 4	5
				Attribute 5	2
B	Patient Type 2	100	5	Attribute 1	5
				Attribute 2	5
				Attribute 3	5
				Attribute 4	3
				Attribute 5	2
C	Patient Type 2	100	5	Attribute 1	5
				Attribute 2	5
				Attribute 3	3
				Attribute 4	3
				Attribute 5	2

[a] Each physician participated in conjoint sessions for Patient Types 1 and 2, but was exposed to version A, B, or C of the conjoint designs. Sample sizes for corresponding versions of the conjoint design for patient types 1 and 2 are not equal because some respondents failed to complete both conjoint tasks.

[b] The attributes, in random order, were: LDL Cholesterol, HDL Cholesterol, Fibrinogens, Total Cholesterol and Triglycerides.

Table 3 – Summary of Experimental Design (Study 2)

Conjoint Design	Patient Type	Sample Size[a]	# Attributes[b]	Attribute	# of levels
A	Patient Type 1	40	5	Attribute 1	5
				Attribute 2	5
				Attribute 3	5
				Attribute 4	5
				Attribute 5	2
B	Patient Type 1	43	5	Attribute 1	5
				Attribute 2	3
				Attribute 3	5
				Attribute 4	5
				Attribute 5	2
C	Patient Type 1	43	5	Attribute 1	5
				Attribute 2	3
				Attribute 3	3
				Attribute 4	5
				Attribute 5	2

[a] Each physician saw only one version, A, B or C.

[b] The attributes, in random order, were: LDL Cholesterol, HDL Cholesterol, Fibrinogens, Total Cholesterol and Triglycerides.

Table 4 – Summary of Experimental Design (Study 3)

Conjoint Design	Data Collection Method	Sample Size[a]	# Attributes[b]	Attribute	# of levels
A	CBC	28	5	Attribute 1	3
				Attribute 2	3
				Attribute 3	3
				Attribute 4	3
				Attribute 5	3
B	CBC	21	5	Attribute 1	2
				Attribute 2	3
				Attribute 3	3
				Attribute 4	3
				Attribute 5	3
A	ACA	24	5	Attribute 1	3
				Attribute 2	3
				Attribute 3	3
				Attribute 4	3
				Attribute 5	3
B	ACA	19	5	Attribute 1	2
				Attribute 2	3
				Attribute 3	3
				Attribute 4	3
				Attribute 5	3

[a] Each physician saw only one version. A or B and the data collection method was CBC or ACA.

[b] The attributes, in random order, were: anticholinergic side effects, onset of action, positive symptoms, extrapyramidal side effects, and negative symptoms

Table 5 – Average Relative Importance Weights - Study 1 (CBC)

	Average Relative Importance		Difference in Average Relative Importance
Patient Type 1			
Attribute 2	43.2 (19.6)	46.2 (16.6)	+3
Number of Levels:	*3*	*5*	
Version:	*B*	*A*	
Attribute 2	46.7 (17.0)	46.2 (16.6)	-0.5
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	21.4 (9.6)	26.3 (12.1)	+4.9
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	21.4 (9.6)	25.3 (10.5)	+3.9
Number of Levels:	*3*	*5*	
Version:	*C*	*B*	
Patient Type 2			
Attribute 4	36.2 (14.2)	41.7 (16.0)	+5.5
Number of Levels:	*3*	*5*	
Version:	*B*	*A*	
Attribute 4	42.5 (17.4)	41.7 (16.0)	-0.8
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	36.8 (18.1)	39.6 (16.0)	+2.8
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	36.8 (18.1)	41.5 (13.7)	+4.7
Number of Levels:	*3*	*5*	
Version:	*C*	*B*	

Standard Deviations in parenthesis

Table 6 – Regression Coefficients for Equations

Predicting Estimated Relative Importance Weights – Study 1 (CBC)

Patient Type 1

Conjoint Design[a]	Manipulated Attribute[b]	df	Constant	Number of Levels	R^2	F	Sig.	Change in Relative Importance[c]
C & B	Attribute 3	208	0.234	0.020	0.037	7.862	0.006	4.0%
B & A	Attribute 2	355	0.447	0.015	0.006	2.116	0.147	3.0%
C & A	Attribute 2	360	0.464	-0.002	0.000	0.064	0.800	-0.4%
C & A	Attribute 3	360	0.239	0.025	0.038	14.051	0.000	5.0%

[a] Version C - no manipulations; Version B – Attribute 2 manipulated; Version A - Attribute 2 and 3 manipulated

[b] Manipulated attribute 3 vs. 5 levels

[c] Since low and high number of levels were coded as –1 and +1 respectively, the change in relative importance weight is twice the beta coefficient for the number of levels.

Patient Type 2

Conjoint Design[a]	Manipulated Attribute[b]	df	Constant	Number of Levels	R^2	F	Sig.	Change in Relative Importance[c]
C & B	Attribute 3	199	0.391	0.024	0.021	4.294	0.040	4.8%
B & A	Attribute 4	356	0.389	0.027	0.025	8.926	0.003	5.4%
C & A	Attribute 3	356	0.382	0.014	0.006	2.108	0.147	2.8%
C & A	Attribute 4	356	0.421	-0.004	0.001	0.187	0.666	-0.8%

[a] Version C - no manipulations; Version B – Attribute 4 manipulated; Version A - Attribute 3 and 4 manipulated

[b] Manipulated attribute 3 vs. 5 levels

[c] Since low and high number of levels were coded as –1 and +1 respectively, the change in relative importance weight is twice the beta coefficient for the number of levels.

Table 7 – Average Relative Importance Weights - Study 2 (ACA)

	Average Relative Importance		Difference in Average Relative Importance
Patient Type 1			
Attribute 2	29.8 (10.8)	31.2 (13.2)	+1.4
Number of Levels:	*3*	*5*	
Version:	*B*	*A*	
Attribute 2	33.5 (11.0)	31.2 (13.2)	-2.3
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	20.8 (5.9)	23.2 (9.7)	+2.4
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	20.8 (5.9)	22.0 (8.9)	+1.2
Number of Levels:	*3*	*5*	
Version:	*C*	*B*	

Standard Deviations in parenthesis

Table 8 – Regression Coefficients for Equations

Predicting Estimated Relative Importance Weights – Study 2 (ACA)

Patient Type 1

Conjoint Design[a]	Manipulated Attribute[b]	df	Constant	Number of Levels	R^2	F	Sig.	Change in Relative Importance[c]
C & B	Attribute 3	85	0.214	0.006	0.007	0.576	0.450	1.2%
B & A	Attribute 2	82	0.305	0.007	0.003	0.247	0.620	1.4%
C & A	Attribute 2	82	0.323	-0.011	0.009	0.741	0.392	-2.2%
C & A	Attribute 3	82	0.220	0.012	0.023	1.866	0.176	2.4%

[a] Version C - no manipulations; Version B – Attribute 2 manipulated; Version A - Attribute 2 and 3 manipulated

[b] Manipulated attribute 3 vs. 5 levels

[c] Since low and high number of levels were coded as –1 and +1 respectively, the change in relative importance weight is twice the beta coefficient for the number of levels.

Table 9 – Average Relative Importance Weights - Study 3 (ACA & CBC)

	Average Relative Importance		Difference in Average Relative Importance
CBC			
Attribute 1	17.9 (8.9)	24.4 (10.4)	+6.5
Number of Levels:	*2*	*3*	
Version:	*B*	*A*	
ACA			
Attribute 1	26.7 (10.3)	25.9 (8.3)	-0.8
Number of Levels:	*2*	*3*	
Version:	*B*	*A*	

Standard Deviations in parenthesis

Table 10 – Regression Coefficients for Equations

Predicting Estimated Relative Importance Weights – Study 3 (ACA & CBC)

Data Collection Method	Manipulated Attribute[a]	df	Constant	Number of Levels	R^2	F	Sig.	Change in Relative Importance[b]
CBC	Attribute 1	48	0.211	0.032	0.099	5.144	0.028	6.4%
ACA	Attribute 1	42	0.236	-0.004	0.002	0.066	0.798	-0.8%

[a] Manipulated attribute 2 vs. 3 levels

[b] Since low and high number of levels were coded as –1 and +1 respectively, the change in relative importance weight is twice the beta coefficient for the number of levels.

Table 11 – F and p-values for the Analysis of Variance (ANOVA)

Study 3 (ACA vs. CBC)

Attribute	Number of Levels [a] (1)	Data collection method [b] (2)	(1) * (2)
Attribute 1	2.004 (p< 0.160)	6.547 (p< 0.012)	3.164 (p< 0.079)

[a] Three vs. two levels
[b] ACA or CBC

Table 12 – Average Relative Importance Weights – Patient Type 1 (CBC & ACA)

	Average Relative Importance		Differences
CBC (Study 1)			
Attribute 2	43.2 (19.6)	46.2 (16.6)	+3
Number of Levels:	*3*	*5*	
Version:	*B*	*A*	
Attribute 2	46.7 (17.0)	46.2 (16.6)	-0.5
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	21.4 (9.6)	26.3 (12.1)	+4.9
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	21.4 (9.6)	25.3 (10.5)	+3.9
Number of Levels:	*3*	*5*	
Version:	*C*	*B*	
ACA (Study 2)			
Attribute 2	29.8 (10.8)	31.2 (13.2)	+1.4
Number of Levels:	*3*	*5*	
Version:	*B*	*A*	
Attribute 2	33.5 (11.0)	31.2 (13.2)	-2.3
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	20.8 (5.9)	23.2 (9.7)	+2.4
Number of Levels:	*3*	*5*	
Version:	*C*	*A*	
Attribute 3	20.8 (5.9)	22.0 (8.9)	+1.2
Number of Levels:	*3*	*5*	
Version:	*C*	*B*	

Standard Deviations in parenthesis

Table 13 – F and p-values for the Analysis of Variance (ANOVA)

Patient Type 1 (ACA vs. CBC)

Attribute	**Conjoint Design**	**Number of Levels**[a] **(1)**	**Data collection method**[b] **(2)**	**(1) * (2)**
Attribute 2	B & A	1.050 (p< 0.306)	46.591 (p< 0.000)	0.181 (p< 0.671)
Attribute 2	C & A	0.498 (p< 0.481)	50.723 (p< 0.000)	0.214 (p< 0.644)
Attribute 3	C & A	7.460 (p< 0.007)	1.883 (p< 0.171)	0.857 (p< 0.355)
Attribute 3	C & B	4.463 (p< 0.035)	2.647 (p< 0.105)	1.243 (p< 0.266)

[a] Three vs. five levels
[b] ACA or CBC

Figure 1 – Sample Conjoint Screen for CBC (Study 1)

Which drug would you most likely choose?

Patient Type 1: (description here)

Attribute 1 - Level 2 Attribute 2 - Level 4 Attribute 3 - Level 3 Attribute 4 - Level 1 Attribute 5 - Level 2	Attribute 1 - Level 3 Attribute 2 - Level 5 Attribute 3 - Level 2 Attribute 4 - Level 4 Attribute 5 - Level 1	Attribute 1 - Level 5 Attribute 2 - Level 2 Attribute 3 - Level 5 Attribute 4 - Level 3 Attribute 5 - Level 1
1	2	3

Press a key between 1 and 3

Note: Attributes used – Total Cholesterol, Triglycerides, Fibrinogens, LDL Cholesterol, HDL Cholesterol

(Because of confidentiality, patient type description and the levels of the attributes are not disclosed)

Figure 2 – Sample Conjoint Screen for CBC (Study 1)

Which drug would you most likely choose?

Patient Type 2: (description here)

Attribute 1 - Level 4 Attribute 2 - Level 1 Attribute 3 - Level 4 Attribute 4 - Level 2 Attribute 5 - Level 2	Attribute 1 - Level 1 Attribute 2 - Level 3 Attribute 3 - Level 1 Attribute 4 - Level 5 Attribute 5 - Level 1	Attribute 1 - Level 5 Attribute 2 - Level 2 Attribute 3 - Level 2 Attribute 4 - Level 4 Attribute 5 - Level 2
1	2	3

Press a key between 1 and 3

Note: Attributes used – Total Cholesterol, Triglycerides, Fibrinogens, LDL Cholesterol, HDL Cholesterol

(Because of confidentiality, patient type description and the levels of the attributes are not disclosed)

Figure 3 – Sample Introduction Screens for ACA (Study 3)

During this computerized interview. we will ask you to rate characteristics of drugs used in the treatment of:

SCHIZOPHRENIA

and to express your preferences between fictitious antipsychotic agents. The selections you make will help us judge the relative value you place on the attributes of the drugs.

Please press any key now.

First. I will now show you sets of drug attributes or features which vary slightly from one another.

For each of these features or attributes. I will ask you to choose the one that you would like most. then second most. etc.. until you have ranked them all.

Please press any key to continue.

Figure 4 – Sample Screen for Ranking Levels of an Attribute in ACA (Study 3)

Type the number of your choice.

1. Attribute 1 – Level 1
2. Attribute 1 – Level 2
3. Attribute 1 – Level 3

Type number | ESC to back up | CTRL END to quit

Type the number of your choice.

1. Attribute 1 – Level 1

3. Attribute 1 – Level 3

Type number | ESC to back up | CTRL END to quit

Type the number of your choice.

3. Attribute 1 – Level 3

Type number | ESC to back up | CTRL END to quit

Note: Attributes used – anticholinergic side effects, onset of action, positive symptoms, extrapyramidal side effects, and negative symptoms (Because of confidentiality levels of the attributes are not disclosed)

Figure 5 – Sample Rating Screen for ACA (Study 3)

If two drugs (ABC & XYZ) were both acceptable in all other ways.

HOW IMPORTANT WOULD THIS DIFFERENCE BE?

To answer, type a number from the scale shown below.

ABC: Attribute 4 – Level 1

versus:

XYZ: Attribute 4 – Level 3

4 = Extremely Important (I could almost never prescribe XYZ)

3 = Very Important (This would give ABC a key advantage over XYZ)

2 = Somewhat Important (But not a major advantage for ABC)

1 = Not Important at All

Note: Attributes used – anticholinergic side effects, onset of action, positive symptoms, extrapyramidal side effects, and negative symptoms (Because of confidentiality levels of the attributes are not disclosed)

Figure 6 – Sample Conjoint Screen for ACA (Study 3)

If two drugs only differed as shown below. WHICH WOULD YOU PREFER?

Type a number from the scale below to indicate your preference.

Attribute 1 – Level 2 Attribute 4 – Level 3 Attribute 2 – Level 1 Attribute 5 – Level 3 Attribute 3 – Level 2	or	Attribute 1 – Level 1 Attribute 4 – Level 2 Attribute 2 – Level 3 Attribute 5 – Level 1 Attribute 3 – Level 3

Strongly Prefer Left				Don't Care				Strongly Prefer Right
1	2	3	4	5	6	7	8	9

Type number | ESC to back up | CTRL END to quit

Note: Attributes used – anticholinergic side effects. onset of action. positive symptoms. extrapyramidal side effects. and negative symptoms (Because of confidentiality levels of the attributes are not disclosed)

Figure 7 – Sample Calibration Screen for ACA (Study 3)

Very
Likely

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

Not at
All
Likely

HOW LIKELY WOULD YOU BE TO USE THIS DRUG IF IT WERE AVAILABLE?
Answer by typing a percentage from the thermometer scale.

Attribute 1 – Level 3

Attribute 2 – Level 2

Attribute 3 – Level 1

Attribute 4 – Level 3

Attribute 5 – Level 1

Type a number from 0 to 100, then press ENTER.
Previous Answers: ___ ___ ___ ___ ___

Type Number (0-100) | ESC to back up | CTRL END to quit

Note: Attributes used – anticholinergic side effects, onset of action, positive symptoms, extrapyramidal side effects, and negative symptoms (Because of confidentiality levels of the attributes are not disclosed)

Figure 8 – Sample Conjoint Screen for CBC (Study 3)

In the following scenarios. please select the drug

which you would most likely prescribe for the treatment of:

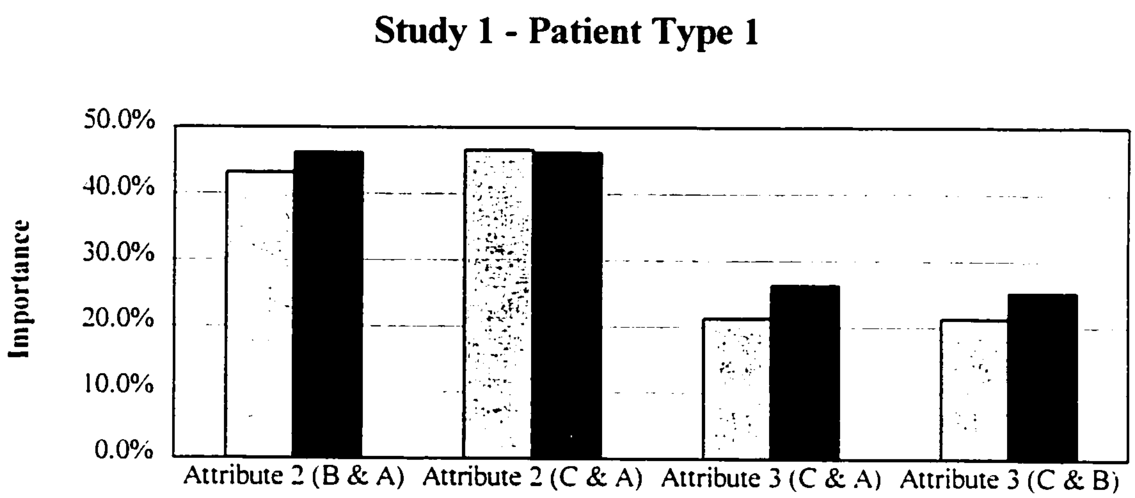
SCHIZOPHRENIA

Attribute 1 – Level 2 Attribute 2 - Level 3 Attribute 3 - Level 1 Attribute 4 - Level 1 Attribute 5 - Level 1	Attribute 1 - Level 1 Attribute 2 - Level 1 Attribute 3 - Level 3 Attribute 4 - Level 3 Attribute 5 - Level 3	Attribute 1 - Level 1 Attribute 2 - Level 2 Attribute 3 - Level 2 Attribute 4 - Level 2 Attribute 5 - Level 2
1	2	3

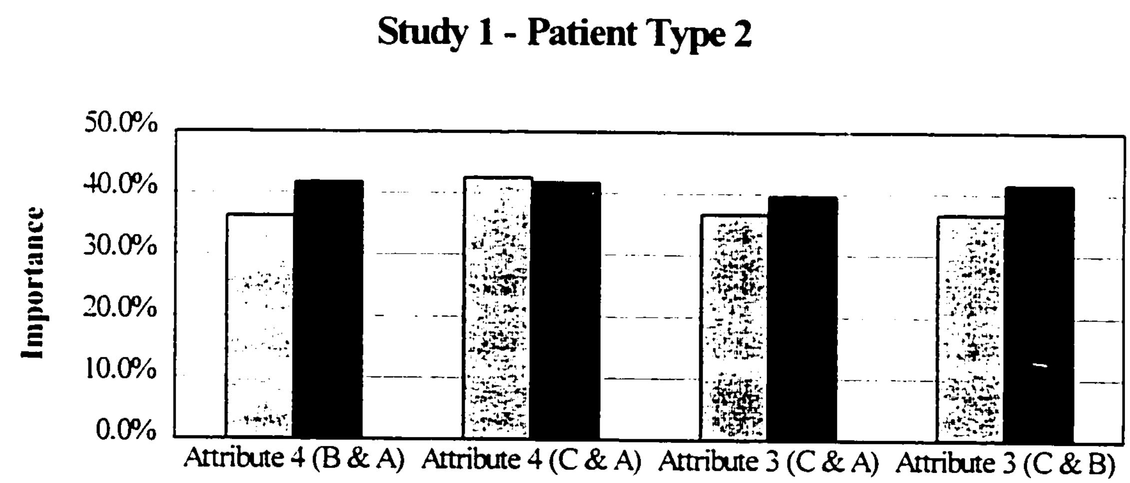
Press a key between 1 and 3

Note: Attributes used – anticholinergic side effects. onset of action. positive symptoms. extrapyramidal side effects. and negative symptoms (Because of confidentiality levels of the attributes are not disclosed)

Figure 9 – Average Relative Importance Weights for Study 1



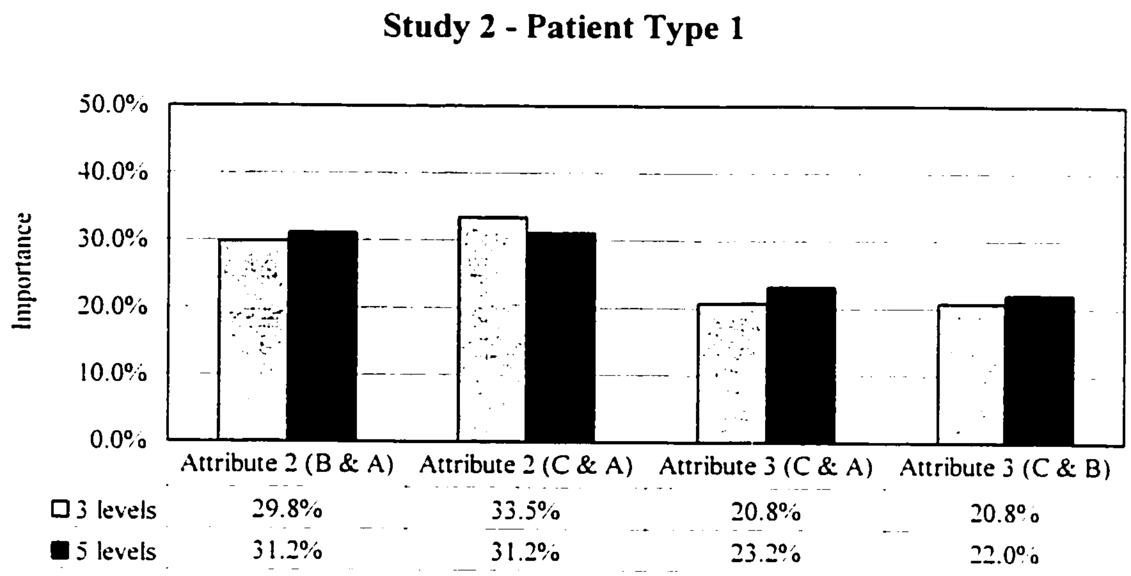
	Attribute 2 (B & A)	Attribute 2 (C & A)	Attribute 3 (C & A)	Attribute 3 (C & B)
□ 3 levels	43.2%	46.7%	21.4%	21.4%
■ 5 levels	46.2%	46.2%	26.3%	25.3%



	Attribute 4 (B & A)	Attribute 4 (C & A)	Attribute 3 (C & A)	Attribute 3 (C & B)
□ 3 levels	36.2%	42.5%	36.8%	36.8%
■ 5 levels	41.7%	41.7%	39.6%	41.5%

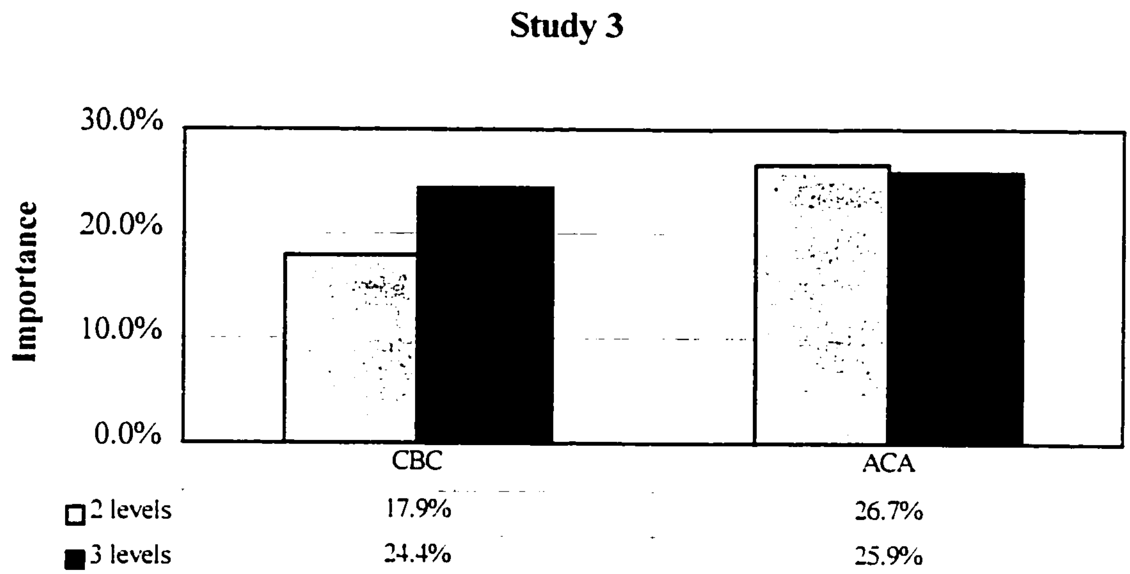
Note: A, B and C refer to the different conjoint designs reflecting the manipulation of the number of levels effect (please see Table 2)

Figure 10 – Average Relative Importance Weights for Study 2



Note: A, B and C refer to the different conjoint designs reflecting the manipulation of the number of levels effect (please see Table 3)

Figure 11 – Average Relative Importance Weights for Study 3



REFERENCES

Agarwal, Manoj K., and Paul E. Green (1991), "Adaptive Conjoint Analysis versus Self-Explicated Models: Some Empirical Results," *International Journal of Research in Marketing*, 8, 141-146.

Allenby, Greg M., Neeraj Arora, and James L. Ginter (1995), "Incorporating Prior Knowledge into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32 (May), 152-162.

Bateson, John E. G., David Reibstein, and William Boulding (1987), "Conjoint Analysis Reliability and Validity: A Framework for Future Research," in *Review of Marketing*, Ed. Michael J. Houston, Chicago: American Marketing Association, 451-481.

Carroll, Douglas J. and Paul E. Green (1995), "Psychometric Methods in Marketing Research: Part I, Conjoint Analysis," *Journal of Marketing Research*, 32 (November), 385-391.

Cattin, Philippe and Dick R. Wittink (1982), "Commercial Use of Conjoint Analysis: A Survey," *Journal of Marketing*, 46 (Summer), 44-53.

Creyer, Elizabeth and William T. Ross (1988), "The Effects of Range-Frequency Manipulations on Conjoint Importance Weight Stability," in *Advances in Consumer Research*, 15, Ed. Michael J. Houston, 505-509.

Currim, Imran S., Charles B. Weinberg, and Dick R. Wittink (1981), "Design of Subscription Programs for a Performing Arts Series," *Journal of Consumer Research*, 8 (June), 67-75.

Elrod, Terry, Jordan J. Louviere, and Krishnakumar S. Davey (1992), "An Empirical Comparison of Ratings-Based and Choice-Based Models," *Journal of Marketing Research*, 29 (August), 368-377.

Gandais, Alban F. An Experimental Study of the Number of Levels Effect in Metric Conjoint Analysis. Montreal: Concordia University, 1994.

Green, Paul E. and Vithala R. Rao (1971), "Conjoint Measurement for Quantifying Judgmental Data," *Journal of Marketing Research*, 8 (August), 355-363.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5 (September), 103-123.

Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice," *Journal of Marketing*, 54 (October), 3-19.

Green, Paul E., Kristiaan Helsen, and Bruce Shandler (1988), "Conjoint Internal Validity Under Alternative Profile Presentations," *Journal of Consumer Research*, 15 (December), 392-397.

Green, Paul E., A.M. Krieger, and M.K. Agarwal (1991), "Adaptive conjoint analysis: Some Cautions and Caveats," *Journal of Marketing Research*, 28 (May), 215-222.

Green, Paul E. and A.M. Krieger (1993), "Conjoint Analysis with Product-Positioning Applications," in *Handbooks in OR & MS*, Vol. 5, J.Eliashberg and G.L. Lilien (eds.), New York: Elsevier Science Publications.

Huber, Joel C., Dick R. Wittink, John A. Fiedler, and Richard L. Miller (1991), "An Empirical Comparison of ACA and Full Profile Judgements," in *Proceedings of the Sawtooth Software Conference*, Sawtooth Software, 189-202.

Huber, Joel C., Dick R. Wittink, Richard M. Johnson, and Richard Miller (1992), "Learning Effects in Preference Tasks: Choice-Based versus Standard Conjoint," in *Proceedings of the Sawtooth Software Conference*, Sawtooth Software.

Huber. Joel C., Dick R. Wittink, John A. Fiedler, and Richard Miller (1993). "The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice," *Journal of Marketing Research*, 30 (February), 105-114.

Johnson, Richard M. (1974), "Trade-Off Analysis of Consumer Values," *Journal of Marketing Research*, 11 (May), 121-127.

Johnson, Richard M. (1987), "Adaptive Conjoint Analysis," in *Proceedings of the Sawtooth Software Conference*, Sawtooth Software, July, 253-265.

Leigh, T. W., David B. MacKay, and John O. Summers (1984), "Reliability and Validity of Conjoint Analysis and Self-Explicated Weights: A Comparison," *Journal of Marketing Research*, 21 (November), 456-462.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young (1996), "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs," *Marketing Science*, 15(2), 173-191.

Louviere, Jordan J. and G. Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments," *Journal of Marketing Research, 20 (November), 350-367.*

Luce, Duncan R. and John W. Tukey (1964), "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement," *Journal of Mathematical Psychology*, 1, 1-27.

Mishra, Sanjay, U. N. Umesh, and Donald E. Stem, Jr. (1989), "Attribute Importance Weights in Conjoint Analysis: Bias and Precision," in *Advances in Consumer Research*, 16, Ed. Thomas K. Srull, 605-611.

Reibstein, David, John E. G. Bateson, and William Boulding (1988), "Conjoint Analysis Reliability: Empirical Findings," *Marketing Science*, 7, 3, 271-286.

Sawtooth Software (1994). Adaptive Conjoint Analysis (ACA) - Version 4.0. Evanston Illinois: Sawtooth Software.

Sawtooth Software (1995). Choice Based Conjoint (CBC) - Version 1.2. Evanston Illinois: Sawtooth Software.

Steenkamp, Jan-Benedict E. M., and Dick R. Wittink (1994), "The Metric Quality of Full-Profile Judgements and the Number-of-Attribute-Levels Effect in Conjoint Analysis," *International Journal of Research in Marketing*, 11, 275-286.

Vriens, Marco, Harmen Oppewal, and Michael Wedel (1998), "Ratings-Based versus Choice-Based Latent Class Conjoint Models – An Empirical Comparison," *Journal of Market Research Society*, 40 (3), 237-248.

Wittink, Dick R. and Lakshman Krishnamurthi (1981), "Rank Order Preferences and Part-Worth Model: Implications for Derived Attribute Importances and Choice Predictions," in *Proceedings of the Third ORSA/TIMS Special Interest Conference on Market Measurement and Analysis*, Ed. John W. Keon, 8-21.

Wittink, Dick R. and Philippe Cattin (1989), "Commercial Use of Conjoint Analysis: An Update," *Journal of Marketing*, 56 (July), 91-96.

Wittink, Dick R. and P.B. Seetharaman (1999), "A Comparison of Alternative Solutions to the Number-of-Levels Effect," in *Proceedings of 1999 Sawtooth Software Conference*, Sawtooth Software, 269-281.

Wittink, Dick R., Lakshman Krishnamurthi, and Julia B. Nutter (1982), "Comparing Derived Importance Weights Across Attributes," *Journal of Consumer Research*, 8 (March), 471-474.

Wittink, Dick R., Lakshman Krishnamurthi, and David J. Reibstein (1989), "The Effect of Differences in the Number of Attribute Levels on Conjoint Results," *Marketing Letters*, 1, 2, 113-123.

Wittink, Dick R., William g. McLauchlan, and P.B. Seetharaman (1997), "Solving the Number-of-Attribute-Levels Problem in Conjoint Analysis," in *Proceedings of 1997 Sawtooth Software Conference, Sawtooth Software*, 227-240.

Wittink, Dick R., Marco Vriens, and Wim Burhenne (1994), "Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections," *International Journal of Research in Marketing*, 11, 41-52.

Wittink, Dick R., Joel Huber, John A. Fiedler, and Richard L. Miller (1991), "Attribute Level Effects in Conjoint Revisited: ACA versus Full Profile," in *Proceedings of Second Advanced Research Techniques Forum*, Ed. Rene Mora, 51-61.

Wittink, Dick R., Joel Huber, Peter Zandan, and Richard M. Johnson (1992), "The Number of Levels Effect in Conjoint: Where Does it Come From, and Can it be Eliminated?" in *Proceedings of 1992 Sawtooth Software Conference*, Sawtooth Software, 1-11.

Wittink, Dick R., David J. Reibstein, William Boulding, John E. G. Bateson, and John W. Walsh (1989), "Conjoint Reliability Measures," *Marketing Science*, 8, 4, 371-375.

APPENDIX A

How ACA estimates Respondent Utilities*

(ACA Manual, version 4.0, 1994, Appendix G & H)

* Reprinted with the permission of Sawtooth Software

Initial Estimates

Before the conjoint (Pairs) section, ACA constructs prior utility estimates (self-explicated portion) for each attribute level as follows:

- For the rank orders of preference, ACA converts them to relative desirabilities by reversing them. For example, ranks of 1, 2, and 3 would be converted to values of 3, 2 and 1, respectively.
- The average for each attribute is subtracted to center its values at zero. For example, desirability values 3, 2 and 1 would be converted to 1, 0 and –1 respectively.
- The values for each attribute are scaled to have a range of unity. For example, desirability values of 1, 0 and –1 would be converted to 0.5, 0, and –0.5.
- The importance ratings for each attribute are scaled to range from 1 to 4, and then used as mulipliers for the unit-range desirability values. Thus, if an attribute has desirabilities of 0.5, 0, and –0.5 and an importance of 3, the result is 1.5, 0, -1.5.

The resulting values are initial estimates of utilities, with these characteristics"

- For each attribute the range of utility values is proportional to stated importance, and attribute importances differ by at most a factor of 4

- Within each attribute. the values have a mean of zero. and differences between values are proportional to differences in desirability rank orders of preference

Conjoint Section and Updating

Estimates of the respondent's utilities are updated after each Pairs response. First. consider the general case of how least squares regression coefficients can be updated to include the effect of an additional observation.

Let X be a matrix of predictor variables with a row for each of n observations and a column for each variable.

Let y be a vector of responses for the first n observations.

Let z' be a row vector of predictor values for a new observation. appended as a row to X.

Let r be a response for the new observation.

Considering only the first n observation. we have the regression equation:

$$Xb_n \sim y,$$

where

$$B_n = (X'X)^{-1}\,(X'y) \qquad (1)$$

is the vector of coefficients that would be obtained by least squares estimation based on the first n observations.

Now consider adding one observation. The expanded layout is:

$$\begin{bmatrix} X \\ z' \end{bmatrix} b_{n+1} \sim \begin{bmatrix} y \\ r \end{bmatrix}, \qquad (2)$$

where

$$b_{n+1} = (X'X+zz')^{-1} (X'y+zr)$$

is the least squares estimate based on n+1 observations. Suppose b_n, X, y, z, and r exist and want to obtain b_{n+1}.

First consider an identity. Let

$$v = (X'X)^{-1} z. \qquad (3)$$

Then it can be shown that

$$(X'X+zz')^{-1} = (X'X)^{-1} \frac{vv'}{1+v'z}. \qquad (4)$$

Substituting into equation (2)

$$b_{n+1} = b_n + v \frac{r-z'b_n}{1+v'z}. \qquad (5)$$

Equation (5) gives a formula for updating the estimate of utilities following each response, a relatively easy computation since the numerator and denominator on the right are both scalars. One must also update the inverse as in equation (4). That is also fairly easy since the vector v is already available. If one is dealing with k attribute levels, then an updating cycle requires about $2k^2$ multiply and add operations. This is a significant savings when compared to the cost of re-estimating "from scratch" after each response, and the final results are identical.

Now consider how this scheme is applied to the specific situation in ACA:

Before the first updating X equals to the identity matrix and both b_n and y equal to the initial utility estimates.

The vector z consists of plus and minus 1's and 0's. An element equals 1 if the corresponding attribute level appeared in the concept on the right of the screen, -1 if in the concept on the left of the screen, and 0 if that level did not appear in either concept.

The response r is coded so that +4 means "strongly prefer right," -4 means "strongly prefer left," and 0 means "indifference".

Updating Initial Estimates with Responses to Conjoint Section (Pairs Questions)

In previous versions of ACA, information from the first part of the interview, summarized by the initial estimates of utilities, was weighted equally with information from the conjoint section of the interview. In Version 4, those two types of information are given separate weights, estimated by regression involving responses to the calibration concepts.

To understand what was done in previous versions as well as the additional steps taken in Version 4, first consider the regression layout (in which the notation is altered slightly from above):

$$\begin{bmatrix} I \\ X \end{bmatrix} U - \begin{bmatrix} P \\ Y \end{bmatrix} = \begin{bmatrix} E1 \\ E2 \end{bmatrix} \qquad (6)$$

where:

I is the identity matrix
X is the design matrix for the pairs
U is the vector of final utility estimates
P is the vector of initial utility estimates
Y is the vector of responses to the conjoint, coded as deviations from mid-scale values.
E1 and **E2** are "error" vectors, the sum of whose squares is to be minimized.

The vector of final utilities, U, is chosen to minimize the sum of squares of elements of E1 and E2. This loss function serves two useful purposes. A small sum of squares for E1 insures that the final utilities are similar to the prior estimates. A small sum of squares for E2 insures that the final utilities fit the responses to the paired-comparison (conjoint) questions.

In previous versions of ACA, the final utility estimates were given by the least squares estimate:

$$U = (I+X'X)^{-1} (P+X'Y) \qquad (7)$$

ACA uses least squares as a handy way of getting utilities, but makes no claim that elements of E1 and E2 have equal variance, and provides no confidence intervals or any other statistics that depend on normality or homoscedasticity. This procedure was intuitively appealing, and had the desirable feature that the prior estimates had decreasing influence as the amount of information from the paired-comparisons increased. Of course, one would get different estimates if either half of the data were scaled differently. However, the scaling of prior utility estimates was chosen empirically to make the two parts of the regression layout approximately compatible, which was feasible as long as the rating scales in the various questionnaire sections always had the same numbers of categories.

However, ACA Version 4 lets the author choose from a variety of scales for the various sections of the questionnaire. With different types of scales one is motivated to find a way to combine data from the initial part of the questionnaire and the Pairs section that does not depend on the scaling of either.

Equation (7) can be rewritten as the sum of three components:

$$U = P + Q + R$$

where

$$Q = (I+X'X)^{-1} X'Y$$

and

$$R = -(I+X'X)^{-1} X'X P.$$

(to see this, premultiply equation (7) by (I+X'X), and do likewise for (P+Q+R).)

Utility estimates for earlier versions (and those used while the conjoint section of the questionnaire is in progress) are obtained by adding P. Q. and R together. These three components have distinct "personalities."

- The P component is the vector of priors. which summarizes information obtained during the first part of the interview. It is not sensitive to the scaling of the conjoint section.

- The Q component is based only on data from the conjoint (pairs). and is not sensitive to the scaling of data in the initial part of the questionnaire. Q is a "ridge estimator" of responses to the conjoint responses. with a ridge coefficient of unity. Q may be regarded as a least squares estimate of the respondent's utilities. using pairs. but not data from the first part of the interview. and augmented with "prior" estimates of zero.

- Just as Q depends on responses to the pairs. R depends on predicted responses to the pairs. where the predictions are based solely on the priors. R is the negative of a ridge regression estimator of those values. with ridge coefficient of unity.

One way to think about these three components of utility estimates is that it starts with information from the priors (P). adds information from responses to the pairs (Q). and then subtracts a "correction" term containing information about how responses to the pairs would have been predicted just on the basis of the priors (R). If one could always ask pairwise questions containing concepts with equal utility. then the R component would be zero.

Rather than giving these three components equal weights, as in the past, it may be useful to weight them differently. Several sets of ACA data were re-analyzed to see whether unique weights for the three components could produce improved prediction of holdout concepts. The overall conclusion was that the P and Q components should receive approximately equal weight when ACA questions are presented using scales like those in earlier versions, but that the R component should be given substantially less weight, and in many cases should even receive negative weight.

These analyses suggest retaining P and Q but discarding R entirely. That is the approach taken in ACA version 4 (although the option remains of estimating utilities as has been done in the past, which entails weighting P, Q, and R equally). The weights assigned to P and Q for each respondent are estimated using responses to the Calibration Concepts.

Calibration Concepts

In the Calibration Concepts section, each respondent is first shown what should be the least attractive possible concept followed by the most attractive possible concept, as determined from the respondent's own answers. Those two concepts establish a frame of reference. The remaining concepts are of middling attractiveness, but differ most strongly in the two components of respondent utility from the initial and paired-comparison sections of the questionnaire. ACA determines an intercept and two regression coefficients to apply to utilities to best predict logits of likelihood responses.

Those parameters are then used in a final scaling of utilities, which are therefore no longer arbitrarily scaled. The procedure is as follows:

Let:

- p = the predicted likelihood of buying a concept
- x_1 = the concept's utility scored using the initial or "prior" estimates
- x_2 = the concept's utility scored using the component of utility derived from the conjoint (pairs) section
- b_1 = the coefficient used to weight the prior utilities
- b_2 = the coefficient used to weight the utilities derived from the conjoint (pairs) section
- a = an intercept parameter

The actual likelihood response is a single digit on a scale with n points. Dividing it by n+1 regards the result approximately as the probability p. Then, using logit transformation, the model buying likelihood is built as a function of the respondent's utilities as:

$$\ln[p/(1-p)] \sim a+b_1x_1+b_2x_2$$

If the two regression coefficients differ too widely one assumes the estimation is faulty and uses more conservative values. Also, the user has the option of doing utility estimation and scaling as in previous versions: in that case all utility components are added together with equal weight, and a single regression coefficient is used to scale them collectively.

In any case, utilities are scaled so their sums are logits. The intercept is divided by the number of attributes, and the quotient is added to the utility for every attribute level. The utilities can be added up and antilogs of their sums are predictions of odds ratios for claimed likelihood of purchase of any concept, just as though that concept had been included in the Calibration Concepts section of the questionnaire.

APPENDIX B

Hierarchical Bayesian Multinomial Logit Regression Model*

(Sawtooth Software Technical Papers, May 2000)

*Reprinted with the permission of Sawtooth Software

The Hierarchical Model

The Hierarchical Bayes model used by the CBC/HB Module is called "hierarchical" because it has two levels.

At the higher level, it is assumed that individuals' part worths are described by a multivariate normal distribution. Such a distribution is characterized by a vector of means and a matrix of covariances.

At the lower level it is assumed that, given an individual's part worths, his/her probabilities of choosing particular alternatives are governed by a multinomial logit model.

To make this model more explicit, some notation is defined. One assumes individual part worths have the multivariate normal distribution,

$$\beta_i \sim \mathbf{Normal}(\alpha, \mathbf{D})$$

where:

β_i = a vector of part worths for the ith individual

α = a vector of means of the distribution of individuals' part worths

$\mathbf{D}$ = a matrix of variances and covariances of the distribution of part worths across individuals

At the individual level, choices are described by a multinomial logit model. The probability of the ith individual choosing the kth alternative in a particular task is

$$p_k = \exp(x_k' \beta_i) / \Sigma_j \exp(x_j \beta_i)$$

where:

p_k = the probability of an individual choosing the kth concept in a particular choice task

x_j = a vector of values describing the jth alternative in that choice task

In words, this equation says that to estimate the probability of the ith person's choosing the kth alternative (by the familiar process used in many conjoint simulators) we:

1) add up the part worths (elements of β_i) for the attribute levels describing the kth alternative (more generally, multiply the part worths by a vector of descriptors of that alternative) to get the ith individual's utility for the kth alternative

2) exponentiate that alternative's utility

3) perform the same operations for other alternatives in that choice task, and

4) percentage the result for the kth alternative by the sum of similar values for all alternatives.

The parameters to be estimated are the vectors β_i of part worths for each individual, the vector α of means of the distribution of worths, and the matrix **D** of the variances and covariances of that distribution.

Iterative Estimation of the Parameters

The parameters β, α, and **D** are estimated by an iterative process. That process is quite robust, and its results do not appear to depend on starting values. However, to make the process converge as quickly as possible, it is started with estimates of the parameters that are reasonably close to final values.

The initial estimates of the betas are approximate least-squares estimates, where the dependent variable consists of choices coded as 1 and 0.

The initial estimate of alpha is the average of the initial betas.

The initial estimate of **D** consists of variances and covariances of the initial betas.

Given those initial values, each iteration consists of these three steps:

Using present estimates of the betas and **D**, generate a new estimate of α. Assuming α is distributed normally with mean equal to the average of the betas and covariance matrix equal to **D** divided by the number of respondents, a new estimate of α is drawn from that distribution.

Using present estimates of the betas and α, a new estimate of **D** from the inverse Wishart distribution is drawn.

Using present estimates of α and **D**, generate new estimates of the betas. This is the most interesting part of the iteration, and it is described in detail below. A procedure known as a "Metropolis Hastings Algorithm" is used to draw the betas. Successive draws of the betas generally provide better and better fit of the model to the data, until such time as increases are no longer possible. When that occurs one considers the iterative process to have converged.

In each of these steps one set of parameters is re-estimated (α, **D** or the betas) conditionally, given current values for the other two sets. This technique is known as

"Gibbs sampling," and converges to the correct distributions for each of the three sets of parameters.

Another name for this procedure is a "Monte Carlo Markov chain," deriving from the fact that the estimates in each iteration are determined from those of the previous iteration by a constant set of probabilistic transition rules. This Markov property assures that the iterative process converges.

This process is continued for a typically many thousands of iterations. After confidence of convergence, the process is continued for many further iterations, and the actual draws of beta for each individual as well as estimates of α and **D** are saved to the hard disk. The final values of the part worths for each individual, and also of α and **D**, are obtained by averaging the values which have been saved.

The Metropolis Hastings Algorithm

What follows is a description of the procedure used to draw each new set of betas, done for each respondent in turn. The symbol β_o (for "beta old") is used to indicate the previous iteration's estimation of an individual's part worths. A trial value is generated for the new estimate, which shall be indicated as β_n (for "beta new"), and then tested whether it represents an improvement. If so, it is accepted as our next estimate. If not, it is accepted or rejected with probability depending on how much worse it is than the previous estimate.

To get β_n a random vector **d** is drawn of "differences" from a distribution with mean of zero and covariance matrix proportional to **D**, and let $\beta_n = \beta_o + \mathbf{d}$.

The probability of the data is calculated (or "likelihood") given each set of part worths, β_o and β_n, using the formula for the logit model given above. That is done by calculating the probability of each choice that individual made, using the logit formula for $\mathbf{p_k}$ above, and then multiplying all those probabilities together. Call the resulting values $\mathbf{p_o}$ and $\mathbf{p_n}$, respectively.

The relative density of the distribution of the betas corresponding to β_o and β_n, given current estimates of parameters α and **D** (which serve as "priors" in the Bayesian updating) is also calculated. These values are called $\mathbf{d_o}$ and $\mathbf{d_n}$, respectively. The relative density of the distribution at the location of a point β is given by the formula

$$\textbf{Relative Density} = \exp[-1/2(\beta - \alpha)' D^{-1}[(\beta - \alpha)]$$

Finally the ratio is calculated:

$$\mathbf{r = p_n\, d_n \,/\, p_o\, d_o}$$

Recall from the discussion of Bayesian updating that the posterior probabilities are proportional to the product of the likelihoods times the priors. The probabilities $\mathbf{p_n}$ and $\mathbf{p_o}$ are the likelihoods of the data given parameter estimates β_n and β_o, respectively. The densities $\mathbf{d_n}$ and $\mathbf{d_o}$ are proportional to the probabilities of drawing those values of β_n and β_o, respectively, from the distribution of part worths, and play the role of priors. Therefore, **r** is the ratio of posterior probabilities of those two estimates of beta, given current estimates of α and **D**, as well as information from the data.

If **r** is greater than or equal to unity, β_n has posterior probability greater than or equal to that of β_o, and β_n is accepted as the next estimate of beta for that individual. If **r** is less than unity, then β_n has posterior probability less than that of β_o. In that case a random process is used to decide whether to accept β_n or retain β_o for at least one more iteration. β_n with probability equal to **r** is accepted.

As can be seen, two influences are at work in deciding whether to accept the new estimate of beta. If it fits the data much better than the old estimate, then $\mathbf{p}_n$ will be much larger than $\mathbf{p}_o$, which will tend produce a larger ratio. However, the relative densities of the two candidates also enter into the computation, and if one of them has a higher density with respect to the current estimates of α and **D**, then that candidate has an advantage.

If the densities were not considered, then betas would be chosen solely to maximize likelihoods. This would be similar to conducting logit estimation for each individual separately, and eventually the betas for each individual would converge to values that best fit his/her data, without respect to any higher-level distribution. However, since densities are considered, and estimates of the higher-level distribution change with each iteration, there is considerable variation from iteration to iteration. Even after the process has converged, successive estimations of the betas are still quite different from one another. Those differences contain information about the amount of random variation in each individual's part worths that best characterizes them.

It was mentioned that the vector **d** of differences is drawn from a distribution with mean of zero and covariance matrix proportional to **D**, but did not specify the proportionality factor. In the literature, the distribution from which **d** is chosen is called the "jumping distribution," because it determines the size of the random jump from β_o to β_n. This scale factor must be chosen well because the speed of convergence depends on it. Jumps that are too large are unlikely to be accepted, and those that are too small will cause slow convergence.

An adaptive algorithm is adopted to adjust the average jump size, attempting to keep the acceptance rate near 0.30. The proportionality factor is arbitrarily set at 0.1 initially. For each iteration the proportion is counted of respondents for whom β_n is accepted. If that proportion is less than 0.3, The average jump size is reduced by ten percent. If that proportion is greater than 0.3, the average jump size is increased by ten percent. As a result, the average acceptance rate is kept close to the target of 0.30.

The iterative process has two stages. During the first stage, while the process is moving toward convergence, no attempt is made to save any of the results. During the second stage it is assumed the process has converged, and results for hundreds or thousands of iterations are saved to the hard disk (or averaged on-the-fly if the option to save draws is not in effect). For each iteration there is a separate estimate of each of the parameters. One is particularly interested in the betas, which are estimates of individuals' part worths. Point estimates are produced for each individual by averaging the results from many iterations. The variances and covariances of the distribution of respondents can also be estimated by averaging results from the same iterations.