WTP for baby milk formula in China : using attribute non-attendance as a priori information to select attributes in choice experiment

Quan, Shiwen; Zeng, Yinchu; Yu, Xiaohua; Bao, Te

2017

Quan, S., Zeng, Y., Yu, X., & Bao, T. (2018). WTP for baby milk formula in China : using attribute non-attendance as a priori information to select attributes in choice experiment. Agribusiness, 34(2), 300-320. doi:10.1002/agr.21535

https://hdl.handle.net/10356/85742

https://doi.org/10.1002/agr.21535

WTP for Baby Milk Formula in China: Using Attribute Non-Attendance as *a priori* Information to Select Attributes in Choice Experiment

Shiwen Quan, Yinchu Zeng, Xiaohua Yu and Te Bao

Abstract:

It is well-known that these non-attended attributes in choice experiments (CE) could cause some bias. A combination of two successive CEs are designed with focus on consumers' demand for the attributes of baby milk formula in China, where the first CE includes the full set of attributes and the second excludes consumers' self-reported non-attended attributes in the first CE. Significant discrepancies are detected for both WTPs for an individual attribute and the total WTPs (the sum of all WTPs for individual attributes in each CE), and in most cases WTPs decrease as a result of excluding the ignored attributes. Changes in WTP are mainly driven by the variation of choice task complexity. However, there is no significant difference in the influence of excluding non-attended attributes on the WTPs for the same attended attribute in different subsamples. In addition, we also find that consumers in China have significantly higher but dispersed WTP values for imported (11.6-47.0%), organic (6.9-52.5%), and probiotics (7.3-35.5%）and sugar-free (5.9-26.0%) attributes for baby milk formula.

Key words: choice experiments, attribute non-attendance, willingness to pay, attribute selection, choice task complexity, baby milk formula in China.

JEL classification: C90, D12, Q13, Q18

1. Introduction

Choice Experiments (CEs) have been increasingly used to elicit consumer preferences. CE is consistent with Lancaster's utility theory (1966, 1972), where consumers' utility is derived from a bundle of product attributes. Compared with other stated preference approaches, such as contingent valuation methods, CE provides consumers (respondents) with more information (or attributes) for their reference in decision making, and respondents are placed into choice scenarios to make trade-offs among two or more alternatives composed of multiple attributes with different attribute levels. Meanwhile, more information also leads to heavier cognitive load, and respondents must take greater cognitive efforts to acquire and deploy the information (Gao, House and Yu, 2010; Swait and Adamowicz, 2001a; Louviere et al., 2008; Yu, Gao and Shimokawa 2016). CE designers typically face the trade-offs between completeness and complexity (Martinsson et al., 2001; Trine, 2005; Muhammad and Abdulai 2016) However, the limit of attributes number does not tell us how to select attributes in CE design (Louviere et al., 2000; Louviere et al., 2010). A plausible criteria is that priority should be given to more relevant and measurable attributes so that the result is more meaningful to consumers and policy makers (Bennett and Blamey, 2001; Martinsson et al., 2001; Blamey et al., 2002). Exclusion of important or relevant attributes implies misspecification of utility function and will most likely result in biased estimates and inaccurate welfare measures (Trine, 2005; Islam et al., 2007, Muhammad and Abdulai, 2016).

Many food quality attributes in CEs are found to be correlated in empirical studies (e.g. Carlsson et al., 2005; Grolleau and Caswell, 2007; Gao and Schroeder, 2009). For instance, consumers use cue attributes such as country-of-origin and organic certification as a proxy for other unobserved attributes such as safety, tastes and nutrition (Loureiro and Umberger, 2003; Naspetti and Bodini, 2008; Gao, House and Yu, 2010； Yu, Gao and Shimokawa, 2016). This may possibly explain the correlation among attributes because some of the selected attributes may act as proxy for some unobserved ones.

Excluding relevant attributes will certainly induce serious bias, and the more relevant the attributes are, the more serious the bias is. However, as Tonsor (2011) states, there is no *a priori* guarantee that presented attributes (which reflect researcher and/or funding agency interests) are truly the ones of importance to respondents. In practice, economists frequently use qualitative methods involving literature reviews, focus group studies, expert interviews, pilot surveys and other possible methods to pre-identify the importance and relevancy of attributes (Trine, 2005; Louviere et al., 2010; Coast et al., 2012; de Bekker-Grob et al., 2012; Muhammad and Abdulai 2016). Ranking of importance are also used as *a priori* information to select the most relevant attributes (de Bekker-Grob et al., 2012; Hiligsmann et al., 2013).

Another strand of literature concentrate on attribute non-attendance (ANA), i.e. where respondents ignore one or several provided attributes of the alternatives when making decisions. Since the seminal research of Hensher et al. (2005), an increasing number of literature have confirmed that ANA universally exists in CEs (see e.g. Alemu et al., 2013; Muhammad and Abdulai 2016). Currently, two general approaches identifying ANA are respectively specified as stated non-attendance (SNA) and inferred non-attendance (INA)[1], the former directly asking respondents debriefing questions whether they have ignored attributes, and the latter using probabilistic methods to infer ANA for every individual respondent (Hess and Hensher, 2010).

All ANA studies, no matter SNA or INA, actually follow the logic of ex post analysis. That is, researchers use supplementary information or extended model specification after the experiment to conduct the analysis. However, in a thoroughly well-defined CE, researchers should ensure in advance that the provided attributes will be attended by respondents as completely as possible, and this implies ANA should be abated to an extent as low as possible[2]. Thus, there is a dilemma in CE design between the theoretical assumption that all selected attributes are attended by respondents and the practical reality that respondents do ignore some attributes for certain reasons. Despite all of that, the ANA literature have illuminating implications for attributes selection in experiment design, in that attributes ignored by respondents or not selected by designers are intuitively of less importance or relevance to respondents, and excluding them is supposed to facilitate a more precise understanding and elicitation of preferences. Hence, the ANA information can be used as an

---

[1] Recently some more advanced approaches, such as eye-tracking (e.g. Balcombe et al., 2014a&b), have been developed to identify ANA.

[2] Suppose an individual does value a particular attribute, but because of some resource constraint (e.g. cognition and time), overlook that attribute when making choices. Then a random utility model will imply the marginal utility of that attribute is zero (Cameron and DeShazo, 2010). That means once ANA due to task simplification exists, there is information loss to elicit the real preference.

ex ante indicator of attributes selection when designing CEs (Muhammad and Abdulai 2016). The reliability of the non-attendance information determines to what extent the indicator is useful. ANA is different from other information for attributes selection like a direct importance ranking since it is obtained after respondents participating in a former CE, so that their awareness on choice tasks are truly enhanced.

In principle, the self-reported non-attended attributes by an individual respondent are supposed to have no effect on the specific respondent's trade-offs among alternatives, implying that the respondent should have no preference on the non-attended attributes. Many researches however have recognized that respondents do not really ignore the attributes which they stated they have ignored. The discrepancy has been firstly detected in the SNA literature, where the attributes acclaimed to be ignored are still statistically significant in the unrestricted model[3] (Campbell and Lorimer, 2009; Carlsson et al., 2010). And subsequent INA studies have further discovered the weak concordance between stated and inferred non-attendance (Hess and Hensher, 2010; Kragt, 2013; Scarpa et al., 2013).

Different from the literature, we adopt an innovative approach by using ANA as *a priori* information for attributes selection in the design of CEs. Since it seems impossible to conduct the subsequent CEs after inferring the non-attendance with complicated calculation based on econometrical models, we consider only a typical SNA approach. More specifically, SNA questions are placed in the middle of two successive CEs. The former CE consists of the full set of attributes while the latter excludes the stated non-attended attributes[4]. Different individual consumer may ignore different food quality attributes, the full sample consumers

---

[3] To restrict the parameters of ignored attributes to zero is the typical standard way to deal with ANA (e.g. Hensher, 2005; Carlsson et al., 2010).

[4] Directly excluding the non-attended attributes, albeit not identical, seems similar to the standard approach of restricting the parameters of non-attended attributes to zero in SNA studies. Many authors, as mentioned above, realize that this approach is inappropriate. However, they also recognize that the non-attended attributes are at least less important to respondents. When reconsidering the principle of selecting attributes, it comes back to the question whether attributes of less importance or relevance should be excluded in CE design. As discussed earlier, with respondents' limited information processing ability, the complexity of experiment must be properly controlled. Thus, if attributes of less importance or relevance to the full sample can be excluded, there should be no doubt that for a split sample, their specific non-attended attributes can also be excluded to design a more appropriate subsample-specific CE.

are therefore split into several groups in the second CE, which is group-specific. We consider the situation where consumers ignore only one attribute, thus with an identical design on other dimensionalities (i.e. same number of attribute level, alternatives, etc.), and the second multiple CEs have a same extent of complexity. Standard random parameter logit (RPL) model is adopted afterwards to estimate consumer utility function for each split group and each CE, and further consumer WTPs for every selected attribute are calculated. Finally, we investigate the effects of excluding non-attended attributes on WTPs for other remaining attributes. Such an approach is used to empirically study consumers' willingness to pay for the attributes of baby milk formulas in China.

In recent years, the milk formula market drew a lot attention from Chinese consumers, producers and the governments. The melamine incident in baby milk formula in September 2008 lowered consumer confidence in China's dairy industry dramatically, and hence triggered a sharp drop in the demand for domestic dairy products.

As shown in Figure 1, a significant structural break point around 2008 can be detected for the long-term development of China's baby milk formula industry. In the market of infant milk formula, the squeeze on domestic companies from foreign competitors is even hasher. Statistics from China's Ministry of Commerce show that the market share of imported baby milk formula was lower than 30% before 2008, but it surged to more than 50% in 2013, and in the middle and high-end market the share is as high as 80%. The practical objective of this paper is thus to offer an exact evaluation of Chinese consumers' preference for import label and other food labels on infant milk formula, and our research may shed light on the marketing strategy of the Chinese dairy companies.

2. Model specification

According to Lancaster’s utility theory (1966, 1972), we assume a typical linear summation form of consumer’s indirect utility function, thus utility is defined as

$$(1) \qquad u_{ij}^0 = \alpha_i^0 p_{ij} + \sum_{k=1}^{K} \beta_{ik}^0 x_{ijk} + \sum_{m=1}^{M} \gamma_{im}^0 y_{im}$$

where the subscripts i, j and k denote consumer i, alternative or product j and attribute k respectively. The superscript 0 denotes the first CE in period 0 consisting of the full set of K attributes. p and x denote price and other attributes, while α and β are their corresponding marginal utility. y denotes consumer characteristics and γ is the corresponding parameter. Consumer-specific α and β imply heterogeneous preferences due to variation in consumer tastes.

Consumer i ’s marginal willingness-to-pay for attribute k is the marginal rate of substitution (MRS) of price for attribute k ., and that is

$$(2) \qquad MWTP_{ik}^0 = MRS_{i(p-k)}^0 = \frac{dp_i}{dx_{ik}} = -\frac{\beta_{ik}^0}{\alpha_i^0}$$

Without loss of generality, consider in a new choice experiment, one attribute t is excluded from the full set of K attributes, the utility function then becomes

$$(3) \qquad u_{ij}^t = \alpha_i^t p_{ij} + \sum_{\substack{k=1 \\ k \neq t}}^{K} \beta_{ik}^t x_{ijk} + \sum_{m=1}^{M} \gamma_{im}^t y_{im}$$

in the second CE at period t. Correspondingly, the change in marginal WTP for attribute k in the two experiments is

$$(4) \qquad \Delta MWTP_{ik}^t = MWTP_{ik}^0 - MWTP_{ik}^t = \frac{\beta_{ik}^t}{\alpha_i^t} - \frac{\beta_{ik}^0}{\alpha_i^0}$$

A standard assumption in traditional literature is that consumers hold stable and consistent preference for every individual attributes, and that implies every consumer’s preference can be regarded as determined by a set of exogenous variables which are independent on the choice contexts. To simplify the description, we rewrite β_{ik}/α_i as β_{ik}^*, and with the assumption of stable and consistent preference, β_{ik}^* can be defined as $\beta_{ik}^* = \beta_{ik}(y_i)$. Since consumer characteristics y_i are exogenous, β_{ik}^* should stay constant. As such, there will be no change of marginal WTP across CEs.

However, it is by now widely recognized that consumer preference in choice experiment is related to the design dimensionality of experiment. Firstly, consumer preference for attributes is jointly determined, hence β_{ik}^* can be affected by other attributes provided in the experiment. Secondly, with the constraint of limited cognitive resources or cognitive ability, consumers will take attribute processing strategies with a given size of choice task complexity. So consumer preference can be redefined as $\beta_i^* = \beta_i\,[y_i, C_i(D)]$, where D denotes the design dimensionality of CE, and C_i denotes consumer i's cognitive process of D[5]. Specifically, with regard to an individual attribute k, the preference can be written as

(5) $$\beta_{ik}^* = \beta_{ik}[y_i, \Sigma_k, C_i(D)]$$

where Σ_k reflects the correlation of attribute k with other attributes in CE. Substituting equation (5) into equation (4), we obtain,

(6) $$\Delta MWTP_{ik}^t = \beta_{ik}[y_i, \Sigma_k^t, C_i(D^t)] - \beta_{ik}[y_i, \Sigma_k^0, C_i(D^0)]$$

As shown in equation (6), the alteration of marginal WTP derives from two inducements, the changes in preference correlation and the readjustment of attribute processing strategies due to the complexity change[6].

Unlike previous studies, this paper tries to explore the influence on WTP when excluding a non-attended attribute rather than a relevant one. Since the non-attended attribute is of little

---

[5] With this definition, we still assume preference consistency within one CE, but consumer preference can vary across CEs with different design dimensionalities.

[6] In the process of econometric analysis, the effect of complexity on choice model performance can also be ascribed to estimation-based reasons. As Dellaert et al. (2012) state, the presence of more attributes lowers the amount of information available per attribute, which reduces statistical power and harms estimation precision.

importance for consumers, it is supposed to have little correlations with other attributes (i.e. Σ_k keeps constant across CEs), and the removal of non-attended attributes also have little impacts on budget allocation. Hence, there should be no effect from correlation structure change when excluding an irrelevant attribute. However, the shift of WTP may still arise purely due to consumers' rearrangement of strategic performance. We could derive three hypotheses to test for removing non-attended attributes.

Hypothesis I. *Exclusion of non-attended attributes in CE influences the WTPs for other attributes.*

Hypothesis II. *The influences of excluding non-attended attributes in CE on WTPs are heterogeneous across attributes of different importance for consumers.*

Hypothesis III. *The influences of excluding non-attended attributes in CE on WTPs for the same attribute have no differences among subsamples with an identical change of design dimensionality.*

3. Experiment design, data and empirical model

3.1 Experiment Design

We choose infant milk formula as the hypothetical product in our choice experiments. We believe infant food buyers are higher-involvement consumers who will pay much more attention on the quality information due to their inherent concern for the health of babies, and thus they may potentially put more cognitive resources in CEs in order to identify the complex information more precisely.

According to extensive focus group interviews and market investigations and a pilot survey, we select price attribute and four other attributes, namely, Imported Food, Organic Certificated, Probiotics Contained and Sugar-free Food, to compose alternatives in CEs (See Table 1). All these four attributes are labeled on the surface of infant milk formula packs as signals to manifest product value. Imported and organic label are used by consumers as cue attributes to indicate the overall quality and safety of milk formula. Parents prefer milk containing multiple probiotics because they believe probiotics help enhance immune system and promote digestion. Many parents also believe that sugar-free food can lower the risk of tooth decay and obesity.

Although brand is also confirmed to be a powerful cue being predictive for food quality on the basis of both the literature (e.g. Grunert, 2005; Taglioni et al., 2011) and our interviews, we do not select brand as an attribute in our CE design because it connotes too much information from history, and often downgrades the role of currently observed (actual or perceived) attribute levels (Hensher et al., 2005). In this paper, we treat all milk attributes as generic rather than alternative-specific. Furthermore, as consumers in the interviews reflect, seldom of them can make clear awareness of complicated nutrition indicators of milk formula (e.g. energy, protein, vitamin, calcium, etc.) listed in the ingredient table. And basic nutrition elements are highly similar across brands, making them indifferent to consumers.

The price attribute is specified with 4 levels, where the base price is 238¥/900g, averaged from a pilot survey with 117 respondents in Beijing[7]. We utilize a similar method to that of Gao and Schroeder (2009) to generate two higher prices by increasing the base price by 33% and 66% and a lower price by decreasing the base price by 33%. All the other four attributes (i.e. Imported Food, Organic Certificated, Probiotics Contained and Sugar-free Food) have a two-level 'Yes-or-No' style, i.e. whether imported or domestic, whether organic or not, whether containing probiotics or not, whether sugar free or not. Consumers use the corresponding food labels to identify the levels of these four attributes.

We then design two successive choice experiments in order to empirically examine the

---

[7] There are two main types of packages for infant milk powder existing in China's market, i.e. 400-gram box and 900-gram can. In order to make it easier for consumers to identify, we unify the unit of price as 'yuan per 900-gram can' in the experiment. The average price in the final survey is 252 ¥/900

impact of excluding non-attended attributes on consumer WTP estimates. The full sample consumers take part in the first experiment consisting of all the five attributes. Subsequently, typical debriefing SNA questions are asked for every consumer. The question reads:

"Is there at least one attribute irrelevant to you when making the choices? If so, please click on the specific attribute you concern least."

We assume full attendance to the price attribute, thus only the four other attributes are listed as options. Particularly, consumers are allowed to choose only one attribute which is the most irrelevant[8]. Thus, as shown in Table 1, the full sample can be split into five subsamples, where group 1 includes consumers attending to all five attributes, and group 2 to group 5 respectively includes those ignoring one specific attribute.

In Table 2, the first experiment is marked with the letter "F", while the second experiment marked with "S". Consumers belonging to group 1 will not participate in the second experiment. All other consumers will proceed to participate in the second experiment including four attributes. Note that in Table 2, CEs in F set are from an identical choice experiment with the only difference of participated respondents, while CEs in S set are four experiments including different attributes. All Designs are generated with D-efficiency criterion. Because all experiments in "S" set can be turned into identical generic orthogonal codes, they share the same four-attribute design. Both designs generate a set of eight alternatives or treatment combinations defined by Hensher et al. (2005). And then a random pairing approach proposed by Gao and Schroeder (2009) to magnifying the differences between alternatives is adopted to generate eight choice sets consisting of two alternatives

---

[8] This constraint may not confirm to the fact that some consumers do ignore more than one attribute. However, with regard to the four attributes mentioned above of infant milk powder, according to the interviews on focus groups, only a small proportion of consumer state they ignore more than one attribute. Secondly, from the SNA literature, we believe the stated information on the most irrelevant one attribute has the highest credibility. Furthermore, with this design, we can ensure the results of between-group comparisons reflect the preference differences due to the same extent of complexity.

each. Given that the suitable extent of task complexity in the literature is to have two experiments both including less than ten choice sets, we do not exert blocking technique to split consumers into groups and present members in each group only a subset of the choice sets. Then the choice sets consisting of two ‘coding’ alternatives are transformed to attribute-level labeling hypothetical products in each choice experiment. And the two hypothetical products are labeled ‘Milk formula A’ and ‘Milk formula B’ respectively. In addition, a ‘no-choice’ alternative is added to each choice set in order to represent the situation that consumers prefer buying neither A or B (or consumers prefer the status-quo) in a choice scenario[9].

In order to check the internal consistency within each CE, we add one supplementary choice set in each experiment to test for transitivity. Combined with the two corresponding choice sets in the eight generated choice sets, we can make a simple transitivity test based on three two-alternative choice sets. Then the eight choice sets are randomly sequenced to avoid the sample-level order effects.

3.2 Survey and Data

The formal surveys were conducted in December 2013, in the city of Beijing and Xi’an in China. Because infant milk formula buyers are special groups, it is undoubtedly inappropriate to investigate consumers who do not have any intent to buy this product. Thus a two-step mixed-mode survey design was implemented.

Firstly, we adopted a face-to-face approach to identify the target consumers. Investigators were assigned to several stores specializing in mother and baby products, while the selected stores were randomly distributed in the city. With the assistance from the stores, consumers who purchased infant milk formula in a store were given an exquisitely printed invitation

---

[9] The milk formula in this study is assumed to be stage-2 baby milk. As most moms in China still use breastfeeding, the stage-1 milk does not have a big market. The stage-3 and stage-4 milk formulas are only used for children older than 1 years old, most of them have stopped milk formula.

card with information about our experiment. On top of the card was a brief introduction of the survey including questionnaire content, investigation objective, milk formula attributes information, predictive time required to complete the questionnaire and rewarding mechanism. The website of our online questionnaire was printed at the bottom of the card. In order to make it easier for the subjects to log on to the website, we also offered two other approaches, the first being a linkage from the university homepage, and the second being a two-dimensional QR code of the website. As such, subjects could use a computer or smartphone to log in the online questionnaire by directly inputting the website address or clicking the link from a homepage or scanning the QR code. Finally, at the bottom was a randomly generated card-specific identification code, which was required before answering the questionnaire and automatically became invalid as soon as the questionnaire was submitted. Investigators in the store were also responsible for answering questions and providing explanations to the consumers.

Consumers receiving the invitation cards composed the original sample pool and they could answer the online questionnaire at any time within a given period. The questionnaire consisted of three parts: the first part was questions for respondents' demographics and milk formula consumption, and the other two parts were the first and second choice experiments each consisting of eight choice sets. Between the two experiments was the above-mentioned ANA question, and according to the chosen option, the respondent was assigned automatically by system to a corresponding second experiment. Prior to all choice tasks, we provided respondents with details of attributes and attribute levels, and a choice set example was also presented. This measure provides respondents a better knowledge of choice tasks, facilitates learning processes and was demonstrated by Bateman et al. (2008) to be more incentive compatible.

We sent out 2400 invitation cards in total, 1200 each for Beijing and Xi'an. We successfully collected 619 and 743 submitted questionnaires from the two cities, with a response rate of 51.58% and 61.92% respectively. Summary statistics of demographics by city and group are presented in Table 3. In general, women respondents take up a larger proportion, which is consistent with some previous studies (e.g. Lusk et al., 2003), primarily because women did more household shopping, especially for baby products. The average age of respondents in Beijing was lower than in Xi'an, which was in accordance with the lower percentage of grandparent buyer in Beijing than in Xi'an. Beijing consumers in the sample had, on average, higher levels of education and income than Xi'an consumers. The numbers of respondents ignoring different attributes varied across groups. In both Beijing and Xi'an, consumers in group1 constituted the largest proportion, where there were respectively 250 and 330 respondents (with a proportion of 40.39% and 44.41%) who stated full attendance to all attributes. This proportion approximates the average of existing empirical SNA studies (see e.g. Alemu et al., 2013 for an overview). 156 Beijing consumers in group 2 ignored the attribute of Sugar-free Food, which was more than twice the number of consumers in other groups. By contrast, more Xi'an consumers ignored the attribute of Imported Food. Comparisons of the statistics among groups reveal some differences. We conduct mean or proportion comparison tests for all statistics between different groups, and some results reject the null hypothesis of equivalence at 5% significance level, implying there are some differences across consumer groups for demographic variables. For example, in Beijing, 42.03% consumers in group 4 have a monthly income over 15,000¥, compared with only 14.86% in group5. Because the allocation of consumer groups is based on consumers' own stated non-attendance information rather than a random mechanism, the differences across groups are acceptable and it is a signal that consumer preferences may vary across groups.

3.3 Empirical Model

We use standard random parameter logit models to estimate the utility function for each consumer group. Consumers' utility function in experiment "F" can be written as

$$(7) \qquad u_{ijt}^{fg} = \alpha^{fg} Price_{ijt} + \beta_{i1}^{fg} Imp_{ijt} + \beta_{i2}^{fg} Org_{ijt} + \beta_{i3}^{fg} Pro_{ijt} + \beta_{i4}^{fg} Sug_{ijt} + c^{fg} + \varepsilon_{ijt}$$

$$\beta^{fg} \sim N(b^{fg}, \Sigma^{fg}) \quad \forall g = 1, 2 \cdots 5$$

where $Price_{ijt}$ is the price of alternative j in scenario t that consumer i faces, Imp_{ijt}, Org_{ijt}, Pro_{ijt} and Sug_{ijt} are attribute variables respectively denoting Imported Food, Organic Certificated, Probiotics Contained and Sugar-free Food, and α^{fg} and β^{fg} are corresponding parameters representing consumer preferences. The superscript f and g respectively denote the experiment F and group number. c^{fg} is alternative-specific constant specified for the 'no-choice' option, and we assume c^{fg} as nonrandom to simplify the analysis. ε_{ijt} is a random error distributed i.i.d. extreme value. As mentioned earlier, consumer i's WTP for attribute k is thus $WTP_{ik}^{fg} = -\beta_{ik}^{fg} / \alpha^{fg}$, $\forall k = 1, 2, 3, 4$. In order to get a normally distributed WTP, we typically assume α to be fixed and β to be normally distributed (e.g. Lusk et al., 2003; Gao and Schroeder, 2009). Vector β^{fg} follows normal distribution with mean vector b^{fg} and covariance matrix $\sum^{fg}$, where b^{fg} denotes the mean valuation of consumers in group g in experiment "F", and $\sum^{fg}$ captures the variation of consumer tastes and their preference correlation across attributes. Hence, in the group level, consumers' mean WTP for attribute k is $WTP_{k}^{fg} = -b_{k}^{fg} / \alpha^{fg}$.

Similarly, in experiment S consumers in groups 2 to 5 respectively have their utility specified as

$$(8) \qquad u_{ijt}^{s2} = \alpha^{s2} Price_{ijt} + \beta_{i1}^{s2} Imp_{ijt} + \beta_{i2}^{s2} Org_{ijt} + \beta_{i3}^{s2} Pro_{ijt} + c^{s2} + \varepsilon_{ijt}$$

$$u_{ijt}^{s3} = \alpha^{s3} Price_{ijt} + \beta_{i1}^{s3} Imp_{ijt} + \beta_{i2}^{s3} Org_{ijt} + \beta_{i4}^{s3} Sug_{ijt} +$$

$$u_{ijt}^{s4} = \alpha^{s4} Price_{ijt} + \beta_{1i}^{s4} Imp_{ijt} + \beta_{3i}^{s4} Pro_{ijt} + \beta_{4i}^{s4} Sug_{ijt} + \varepsilon_{ijt}$$

$$u_{ijt}^{s5} = \alpha^{s5} Price_{ijt} + \beta_{2i}^{s5} Org_{ijt} + \beta_{3i}^{s5} Pro_{ijt} + \beta_{4i}^{s5} Sug_{ijt} + \varepsilon_{ijt}$$

$$\beta^{sg} \sim N(b^{sg}, \Sigma^{sg}) \qquad \forall g = 2, 3, 4$$

Likewise, the mean WTP for attribute k in group g in experiment "S" can be written as $WTP_k^{sg} = -b_k^{sg} / \alpha^{sg}$. The objective of this paper is to explore the impacts of excluding non-attended attributes on WTP estimates of other remaining attributes, thus we mainly focus on group 2 to group 5 in which consumers participate in two CEs.

In the first place, we test the null hypothesis $WTP_k^{fg} = WTP_k^{sg}$, $\forall k = 1, 2, 3, 4$ and $g = 2, 3, 4, 5$ to examine whether the group-level mean WTP significantly changes when a stated non-attended attribute is excluded. In order to conduct the tests, we utilize two methods to obtain the mean and standard deviation of WTP. Firstly, Krinsky and Robb (1986) bootstrapping method is used to simulate WTP estimates for every attribute in every CE on the basis of estimates from random parameter logit model. Secondly, we derive the individual consumer's conditional distribution based on their own choices and take the mean as individual-specific preferences, which allows us to obtain every surveyed consumer's specific WTP (see e.g. Train, 2003, Chapter 11 for more details about individual-level parameters). Each of these two methods has some advantages and drawbacks. The bootstrapping approach ensures large sample but has to assume independence across CEs. This assumption is plausible when conducting within-group comparison (Gao and Schroeder, 2009) . While the method of deriving individual specific parameters internalizes the potential correlation between CEs due to the matched (or paired) observations, but may not precisely capture the preference distribution of the population due to limited sample size. We believe that these two WTP estimates may offer complementary information.

Define $WTP_{ik}^{dg} = WTP_{ik}^{fg} - WTP_{ik}^{sg}$ as observation i 's change in WTP for attribute k from experiment F to experiment S. We assume WTP_{ik}^{dg} has a (or an approximate) normal distribution[10], with a mean of $WTP_{k}^{dg} = b_{k}^{sg} / \alpha^{sg} - b_{k}^{fg} / \alpha^{fg}$, and then we respectively test the null hypothesis $WTP_{k}^{dg} = WTP_{m}^{dg}$ and $WTP_{k}^{dg} = WTP_{k}^{dh}$, $k \neq m$ and $g \neq h$ to examine whether the changes in WTP are significantly different across attributes and groups.

4. Results

4.1 Estimates of Random Parameter Logit Model

Table 4 and Table 5 report the estimates of random parameter logit model, including means for every attribute and constant for none-option, standard deviations for random attributes capturing tastes variation, and lower covariance matrix capturing tastes correlation. Interestingly, all estimates are generally larger in “S” set experiments than in the “F” set[11]. As explained by Train (2003), this difference in scale means the unobserved portion of utility has less variance in experiment S than F, revealing consumers have less uncertainty in experiment S possibly (due to greater awareness after the learning process in experiment F) and/or less complexity (fewer attributes) of experiment S[12]. Estimation-based reasons with regard to different task complexity may also systematically affect the scale (Dellaert et al., 2012). The scale difference make it inappropriate to directly compare estimates from the two successive CEs, however, we could ignore this problem when comparing WTPs, since the ratio of two coefficients is not affected by scale parameter.

---

[10] The bootstrapping method assumes independence between WTP_{ik}^{fg} and WTP_{ik}^{sg} , thus WTP_{ik}^{dg} follows normal distribution due to the axiom that the difference of two independent normal distribution is also normally distributed. As shown below, with regard to simulated WTP, observations are randomly paired in order to get the observation level WTP change.

[11] Comparison across cities also shows that estimates of Beijing consumers have a smaller scale than that of Xi’an consumers.

[12] Respondent fatigue will also increase choice uncertainty and thus utility scale (Swait and Louviere, 1993). In present research, if consumers fatigue do increase from experiment F to S, our results show the influence of fatigue on scale is surpassed by above mentioned opposite effects.

As expected, all price coefficients are negative and other attribute coefficients are positive, indicating that the population distributions of WTP on all attributes have positive means, i.e. more than 50% consumers would like to pay a premium for every attribute listed. The significance of most standard deviation estimates confirmed that there is heterogeneity in consumer preference for almost all attributes. We calculate the predicted shares of preferences based on typical method. With the estimated mean of 1.276 and estimated standard deviation of 1.346, we can predict that 82.84% of Beijing consumers in group 1 prefer organic milk.

Although likelihood ratio tests suggest for some CEs (i.e. S3 in Beijing, F5 in Xi’an at significance level 0.05, F4 and F5 in Beijing, F2 in Xi’an at significance level 0.01) RPL models with independent preferences across attributes appear more appropriate, there is little difference of the mean and standard deviation estimates. In order to capture the correlation change, we report all models allowing correlated preference across attributes. The negative correlation between Imported Food and Organic Certificated is found extensively for consumers in group 1, group 2 and group 3, implying these two labels act as substitutes to each other. They are both used by Chinese consumers as overall indicator for milk formula safety and quality. But for group 4 and group 5 in which consumers respectively ignore Imported Food and Organic Certificated, there is no evidence for their negative correlation. All other significant correlations estimated across attributes are positive except for the one between Probiotics Contained and Organic Certificated in Beijing S2 model.

Comparing the covariance estimates reveals that more significant correlations are discerned in experiment S than F (i.e. group 2, group 4 and group 5 in Beijing, group 2 and group 5 in Xian). This suggests excluding non-attended attributes makes consumers more concentrated on the relation across attributes, which is consistent with Gao and Schroeder (2009) in the reverse direction.

Transitivity test has then been conducted based on responses from three choice sets (including the one supplementary choice set). Comparisons among the responses show that, in total, there are 89 and 37 consumers who made inconsistent choices in experiment F (with a proportion of 6.53% in 1362 participants) and experiment S (with a proportion of 4.73% in 782 participants), respectively including 30 and 13 Beijing consumers and 59 and 24 Xi’an consumers. In contrast to other CE applications, we believe such a small proportion of inconsistent responses do not appear to be a serious intransitivity problem. And we have tried to estimate the models in Table 4 and Table 5 after removing the inconsistent responses, and we obtain very similar results.

4.2 Estimates of Willingness to Pay

Table 6 and Table 7 respectively report the statistics of simulated and individual-level WTP estimates respectively. WTP estimates are very similar from these two approaches. But the standard deviations of individual-level WTP estimates appear to be less than that of simulated WTP estimates. This is probably because individual-level parameters conditional on their choices are strictly ‘same-choice-specific’ parameters, and the homogeneity within subsamples that made the same choices decreases the overall taste variation of full sample. In general, we find that consumers in China have significantly high but dispersed WTP values for imported (11.6-47.0%), organic (6.9-52.5%), probiotics（7.3-35.5%）and sugar-free (5.9-26.0%) attributes for baby milk formula.

All the means of WTP estimates, including those for stated non-attended attributes in F set, are significantly above zero, which is consistent with the findings of empirical SNA studies that respondents do not really ignore the attributes that they have stated to have ignored (e.g. Campbell and Lorimer, 2009; Carlsson et al., 2010). However, the percentage values also show that stated non-attended attributes in every group is of the least importance, except for group5 in Beijing, in which the ignored attribute, Imported Food, still takes a larger share in total WTP (defined as the sum of all WTPs for individual attributes in each CE) than Sugar-free Food. This supports the viewpoint of e.g. Hess and Hensher (2010) that stated non-attended attributes are actually of less importance rather than zero importance.

Though in the present paper the WTP estimates may be overstated in hypothetical scenarios (Lusk et al., 2003; Lusk and Schroeder, 2004; Gao and Schroeder, 2009), as Lusk et al. (2003) acclaim, we can be more confident on the comparisons of WTPs than the absolute values of WTPs with the assumption that hypothetical bias is of similar magnitude across groups and experiments. Comparisons across percentage values indicate consistent findings that country of origin always plays the most important role for consumers to choose food (e.g. Mennecke et al., 2007; Gao and Schroeder, 2009). As discussed in section 2, the difference is that Chinese consumers prefer imported milk formula to domestic ones due to food safety concerns. In addition, organic label generally appears to be the second important attribute except for group 4. And in group 5 in Beijing and group 2 in Xi'an, WTP for organic label even takes the largest share. By contrast, attributes of Probiotics Contained and Sugar-free Food generally play less important roles.

Table 6 and Table 7 also report the within-group comparisons testing null hypothesis of equal means of WTPs. The P values are respectively for unpaired and paired *t*-tests with alternative hypothesis of unequal means (i.e. two-tailed p-value) in simulated data set and individual-level data set[13]. Comparison of these two methods show that paired t-tests in individual-level data set have a lower rate to reject the null hypothesis (i.e. 22 in 32 at 0.10 significance level) than unpaired t-tests in simulated data set (i.e. 27 in 32 at 0.10 significance level)[14]. An important explanation to this difference is the preference correlation between experiment F and S. Actually, correlation tests show that WTPs for every individual attribute

---

[13] Reported unpaired *t*-tests in Table 6 are with equal variances of two samples, assuming stable tastes variation for every specific sub-sample. Results of tests with unequal variances are very similar to those with the assumption equal variances.

[14] Methodologically, due to the relatively smaller variance, it should be easier in individual-level data set to reject the null hypothesis of mean equality if WTP series are independent of each other.

in each group and city are highly significantly correlated between these two successive experiments. Another possible reason is the relatively smaller sample size in individual-level data set which results in higher standard errors of the mean differences.

The *t*-test on both data sets reveals significantly changes in WTPs in a majority of cases. All total WTPs and the majority of individual attribute WTPs are lower in experiment S than in F, and the decrease generally appears greater in Xi’an than in Beijing. However, in most cases the relative importance ranking of attributes do not change except for group 2 and group 5 in Xi’an. These findings can be summarized to Result 1:

Result 1: *Our findings support Hypothesis 1. The total WTP and most WTPs for each individual attributes deceases in the Experiment S compared to the Experiment F.*

Now consider the reasons for such decreases of WTP from experiment F to S. As discussed in section 2, when the provided attributes in CE are replaced, theoretically both correlation structure change and complexity change give rise to shifts of WTP, but the former is supposed to take no effect when excluding non-attended attributes. Although the stated non-attended attributes are virtually of less importance rather than no importance, all covariance between non-attended attribute and other attributes in Table 4 and Table 5 are not significant. We believe the decrease of WTP is mainly driven by the change of design dimensionality, which is consistent with the theoretical analysis in section 2.

Another possible explanation to the results of decreasing WTP from Table 6 with full attributes to Table 7 for the design without ANA is that consumers may ignore the values of ANA when the attribute does not appear. When consumers see the attribute in the full design, they are willing to pay some for it, even though it is not important for them.

In each group, WTP for non-attended attributes in experiment F only constitute a proportion of less than 10%, except that in group 5 a little higher. It seems more appropriate to ascribe WTP shifts to changes of respondents' strategic performance resulting from task complexity change. The decrease of total WTP may be a combined effect of removing payment for excluded attributes and more explicit awareness on choice tasks, while the latter is probably the main driving force accounting for changes in WTP for individual attributes. Although the pattern is not very consistent, it seems WTP for individual attributes more often decreased when excluding an attribute, which is consistent with the case of comparisons between CEs respectively including four and five attributes in Gao and Schroeder (2009) and Muhammad and Abdulai (2016).

4.3 Changes in WTP across Attributes and Groups

Table 8 reports comparisons of the magnitude of WTP variation across attributes within each group, which is useful to evaluate hypothesis II. Both paired and unpaired *t*-tests are performed for simulated and individual-level WTP estimates, and results of these two tests exhibit high consistency. And, similar to earlier discussion, tests in simulated data set are more likely (i.e. 22 in 24 at 0.05 significance level) to reject the null hypothesis of equal mean changes in WTP than those in individual-level data set (i.e. 15 to 16 in 24 at 0.10 significance level). Cases for which we fail to reject the null hypothesis are mainly from group 3, which corresponds to test results in Table 6 and Table 7 that WTPs for more individual attribute in group 3 do not significantly changed. But beyond that, in most cases, test results in Table 8 substantiate significant differences of WTP shifts across attributes. We can summarize these findings to Result 2.

Result 2: *Our experimental results support Hypothesis II that effects of excluding non-attended attributes on WTPs varies across attributes.*

However, as previously discussed, the evidence on whether WTP changes greater for more important attributes is ambiguous. Supportive argument is that more important attributes are more sensitive to experiment design due to more extensive correlation with other attributes (Gao and Schroeder, 2009). Opposite view stems from ANA literature that task simplifying strategies are mostly focused on less important attributes. In order to further examine the relationship between WTP changes and attribute importance, Table 8 also reports the differences of absolute mean WTP values for the two attributes listed in the second column, which are ordered by relative importance measured by percentage values of experiment F in Table 6 or Table 7. Among the total twenty-four difference values of absolute mean WTPs, thirteen display a positive sign and eleven negative, offering no direct conclusions on the relationship between changes in WTP and attribute importance. However, a further examination gives more illuminations. Consider the by-importance-ordered two attributes in the second column, Sugar-free acts as the less important attribute in every pair. And among the twelve pairs including attribute of Sugar-free, nine (75%) difference values are negative. Another two negative values both arise in pairs with Probiotics as the second attribute, taking a proportion of 40% in five pairs including Probiotics as the less important attribute. For all other pairs, including the total four pair of Imported and Organic, three in four pair of Imported and Probiotics, three in four pair of Organic and Probiotics and another three pair including Sugar-free, the differences values are positive. Such a structure reveals a U-shaped relation between WTP shifts and attribute importance. When excluding non-attended attributes, WTPs for remaining attributes both of most and least importance are more sensitive to the change of experiment dimensionality, albeit with possibly different reasons, than those for attributes of middle importance.

Table 9 reports comparisons of WTP shifts across groups for each comparable attribute to gauge Hypothesis III. But the two data sets represent obviously contradictory results, where tests for simulated WTP estimates mostly reject the null hypothesis of equal WTP shifts across groups (i.e. 34 in 36 at 0.05 significance level), but tests for individual-level WTP estimates in most cases cannot reject the null hypothesis (i.e. only 12 in 36 rejected at 0.05 significance level and another 3 at 0.10 level). Reasons for this difference presumably lie in the underlying assumption about whether WTP estimates are correlated across CEs and different sample size of data sets. Although further analysis is required to examine the relationship of WTP shifts and task complexity variation at population level, we can hardly reject at the surveyed sample level that WTP changes at a same amount across groups due to a same magnitude of task complexity change. We can formulate this finding to Result 3:

Result 3*: Our findings also support Hypothesis 3. There is no significant difference in the influence of excluding non-attended attributes on the WTPs for the same attended attribute in different subsamples.*

Many potential factors including difference of consumer cognitive ability across groups, complexity measurement conditional on lexicographic information of attributes, endogeneity of information processing strategies may contribute to the rejection of equal WTP shifts across groups. Actually, since consumers in different groups may have different preference structures for attributes, it seems inappropriate to assume different group consumers take homogeneous information processing strategies on a same attribute. For example, attribute Probiotics in group 2, Beijing takes a proportion close to 20% of total WTP in both experiment F and S, in contrast, the corresponding proportion in group4, Beijing increase from 27% to 39%. As such, the total WTP shift ruling out attribute-specific preference difference may be a better indicator for overall effect of complexity change. Among the twelve comparisons of total WTP shifts, ten tests reject the null hypothesis of equal shifts across groups at 5% significance level, providing more explicit evidence that the overall effects of excluding one non-attended attribute on WTPs do not differ across the sample groups.

5. Conclusions and Implications

This paper used Attribute Non-Attendance (ANA) as *a priori* information for attributes selection in CEs engaged to elicit Chinese consumers' preference of infant milk formula. Particularly, two successive CEs were conducted, where the former included the full set of attributes and the latter excluded the one stated non-attended attribute by the respondents in the former CE. ANA information is different from those obtained from other methods because consumer awareness is truly enhanced after participating in the former CE. And such a two-stage experiment design is also partly consistent with the suggestion of Tonsor (2011) on split-sample analysis since the ignored attributes are different across respondents. Multiple CEs consisting of different subsets of attributes were designed and for each split sample the corresponding CE was supposed to be capable of depicting the focus of consumer preference more succinctly and precisely.

Our conclusion is consistent with SNA literature that consumers actually put less weight on their stated non-attended attributes rather than not weight (totally ignore them). But the SNA information is still of high reliability since almost all non-attended attributes take a much smaller share of total WTP than other attributes in each group. This conclusion calls further reconsideration on attributes selection. It is acknowledged in existing literature that omitting relevant attributes give rise to seriously biased estimates, but little attention has been paid to the impacts of excluding less relevant or irrelevant attributes.

We then take a systematic investigation on how excluding non-attended attributes affected WTPs for other remaining attributes. Similar to Muhammad and Abdulai (2016), significant discrepancies are detected for both WTPs for an individual attribute and the total WTPs, and in most cases WTPs decrease as a result of excluding the ignored attributes. Another finding is that almost all estimates display a larger size in the latter CE, implying diminished choice uncertainty after excluding the ignored attributes.

Changes in WTPs can be interpreted as a result mainly due to reduced uncertainty for making trade-offs. Influences of attributes interaction and consumer inference demonstrated in existing literature (e.g. Gao and Schroeder, 2009; Tonsor, 2011) with regard to excluding relevant attributes, however, take little effects in this analysis since no significant correlation between non-attended attributes and other attributes have been detected. These findings, consistent with Gao and Schroeder (2009) and Tonsor (2011), further confirm that WTP is conditional on what is presented in CEs, not only the attributes of more relevance, but also those of less or even no relevance.

But unlike Gao and Schroeder (2009), we find greater sensitivity for attributes of both relatively more and less importance to experiment design dimensionality than for those of middle importance in our study. The higher sensitivity of more important attributes probably results from their stronger signal effect as a proxy for overall food quality, while the higher sensitivity of less important attributes possibly arises from a larger variation of strategic performance due to consumers' stronger uncertainty on them. Moreover, so far as the surveyed sample is concerned, little difference is found across groups for changes in total WTP as a result of the same magnitude of complexity change for each split sample. This conclusion provides further evidence that consumers' general awareness and their corresponding information processing strategies are systematically affected by task complexity and independent of their concrete preference structure for attributes.

With regard to future studies adopting CEs, our conclusions call more prudence by practitioners in attributes selection, welfare estimates evaluation and results comparison across CEs. A precondition of ensuring the validity of CEs is to thoroughly and deliberately select attributes. Given respondent's limited cognitive capacity, a rule of thumb is that preference should be given to more relevant attributes which are supposed to be more important to consumers (Bennett and Blamey, 2001; Blamey et al., 2002; Gao, House and Yu

2010). Empirical studies also confirm that excluding relevant attributes will most likely result in biased welfare estimates (e.g. Gao and Schroeder, 2009). However, our study further demonstrate that even excluding ignored attributes will also significantly influence the estimates of remaining attributes. Because estimates derived from every specific CE are potentially affected by subjects' strategic performance determined by the inherent task complexity of the specific experiment design. Our conclusions pose a challenge to the design of CEs.

Albeit without a quantitative criteria, whether to design a more comprehensive or concise CE, as Tonsor (2011) suggested, may lie in the primary objective of studies in practice. For those aiming to capture the full picture of consumer preference and the overall interaction across attributes, more attributes should be selected in case of any information loses. But more frequently for those focusing on one or several particular attributes, a more concise design with main interactions being controlled seems more appropriate to generate estimates of higher validity. Besides the traditional rule of giving priority to relevant attributes, our study provides another attributes selecting rule to compromise completeness and complexity in CEs design. We suggest practitioners to use SNA as *a priori* information to select attributes since the subsequent CE has been confirmed with higher internal validity. However, two continuous CEs imply greater cognitive burden on respondents. As such, assuming that SNA information is consistent across choice sets, practitioners can ask respondents' ignored attributes after only a few choice scenarios instead of a full CE with much more scenarios. Our conclusion is also consistent with Burton & Rigby (2012) that practitioners would better design a set of CEs with different complexity according to respondents' attended attributes and their cognitive capacity.

References

Alemu, M. H. and M. R. Mørkbak, et al. (2013). "Attending to the Reasons for Attribute Non-Attendance in Choice Experiments." Environmental and Resource Economics 54 (3): 333-359.

Balcombe, K. and I. Fraser, et al. (2014a). "Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments." Journal of Applied Econometrics: Published online in Wiley Online Library.

Balcombe, K. and M. Bitzios, et al. (2014b). "Using Attribute Importance Rankings within Discrete Choice Experiments: An Application to Valuing Bread Attributes." Journal of Agricultural Economics: Published Online.

Bateman, I. J. and R. T. Carson, et al. (2008). Choice Set Awareness and Ordering Effects in Discrete Choice Experiments, CSERGE working paper EDM.

Bennett, J. and R. Blamey (2001). The Choice Modelling Approach to Environmental Valuation, Edward Elgar Publishing.

Blamey, R. K. and J. W. Bennett, et al. (2002). "Attribute Causality in Environmental Choice Modelling." Environmental and Resource Economics 23 (2): 167-186.

Cameron, T. A. and J. R. DeShazo (2010). "Differential Attention to Attributes in Utility-Theoretic Choice Models." Journal of Choice Modelling 3 (3): 73-115.

Campbell, D. and V. S. Lorimer (2009). Accommodating Attribute Processing Strategies in Stated Choice Analysis: Do Respondents Do What They Say They Do? European Association of Environmental and Resource Economists Annual Conference. Amsterdam.

Carlsson, F. and M. Kataria, et al. (2010). "Dealing with Ignored Attributes in Choice Experiments on Valuation of Sweden’s Environmental Quality Objectives." Environmental and resource economics 47 (1): 65-89.

Carlsson, F. and P. Frykblom, et al. (2005). "Consumer Preferences for Food Product Quality Attributes from Swedish Agriculture." AMBIO: A Journal of the Human Environment 34 (4): 366-370.

Caswell, J. A. and E. M. Mojduszka (1996). "Using Informational Labeling to Influence the Market for Quality in Food Products." American Journal of Agricultural Economics 78 (5): 1248-1253.

Coast, J. and H. Al-Janabi, et al. (2012). "Using Qualitative Methods for Attribute Development for Discrete Choice Experiments: Issues and Recommendations." Health economics 21 (6): 730-741.

de Bekker-Grob, E. W. and M. Ryan, et al. (2012). "Discrete Choice Experiments in Health Economics: a Review of the Literature." Health Economics 21 (2): 145-172.

Dellaert, B. G. C. and B. Donkers, et al. (2012). "Complexity Effects in Choice Experiment–Based Models." Journal of Marketing Research 49 (3): 424-434.

Gao, Z. and L. O. House, et al. (2010). "Using Choice Experiment to Estimate Consumer Valuation: the Role of Experiment Design and Attribute Information Loads." Agricultural Economics 41 (6): 555-565.

Gao, Z. and T. C. Schroeder (2009). "Effects of Label Information on Consumer Willingness-to-Pay for Food Attributes." American Journal of Agricultural Economics 91 (3): 795-809.

Gao, Z. and T. C. Schroeder, et al. (2010). "Consumer Willingness to Pay for Cue Attribute: The Value Beyond Its Own." Journal of International Food & Agribusiness Marketing 22 (1-2): 108-124.

Gao Z., L. House and X. Yu (2010): Using Choice Experiment to Estimate Consumer Valuation: the Role of Experiment Design and Attribute Information Loads, Agricultural Economics, Vol. 41 (6):555-565.

Grolleau, G. and J. Caswell (2007). "Interaction Between Food Attributes in Markets: The Case of Environmental Labeling." Journal of Agricultural and Resource Economics 31 (3): 471-484.

Grunert, K. G. (2005). "Food Quality and Safety: Consumer Perception and Demand." European Review of Agricultural Economics 32 (3): 369 -391.

Hensher, D. A. (2004). "Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings." Journal of Transport Economics and Policy 38 (3): 425-446.

Hensher, D. A. (2006a). "How do Respondents Process Stated Choice Experiments? Attribute Consideration under Varying Information Load." Journal of Applied Econometrics 21 (6): 861-878.

Hensher, D. A. (2006b). "Revealing Differences in Willingness to Pay due to the Dimensionality of Stated Choice Designs: An Initial Assessment." Environmental and Resource Economics 34 (1): 7-44.

Hensher, D. A. (2007). Attribute Processing in Choice Experiments and Implications on Willingness to Pay. Valuing Environmental Amenities Using Stated Choice Studies. B. Kanninen, Springer Netherlands. 8: 135-157.

Hensher, D. A. (2010). "Attribute Processing, Heuristics and Preference Construction in Choice Analysis." State-of Art and State-of Practice in Choice Modelling, Emerald Press, UK: 35-70.

Hensher, D. A. and J. M. Rose, et al. (2005). Applied Choice Analysis: A Primer, Cambridge University Press.

Hess, S. and D. A. Hensher (2010). "Using Conditioning on Observed Choices to Retrieve Individual-Specific Attribute Processing Strategies." Transportation Research Part B: Methodological 44 (6): 781-790.

Hess, S. and D. A. Hensher (2013). "Making Use of Respondent Reported Processing Information to Understand Attribute Importance: a Latent Variable Scaling Approach." Transportation 40 (2): 397-412.

Hiligsmann, M. and C. van Durme, et al. (2013). "Nominal Group Technique to Select Attributes for Discrete Choice Experiments: an Example for Drug Treatment Choice in Osteoporosis." Patient Prefer Adherence 1 (7): 133-139.

Islam, T. and J. J. Louviere, et al. (2007). "Modeling the Effects of Including/Excluding Attributes in Choice Experiments on Systematic and Random Components." International Journal of Research in Marketing 24 (4): 289-300.

Kragt, M. E. (2013). "Stated and Inferred Attribute Attendance Models: A Comparison with Environmental Choice Experiments." Journal of Agricultural Economics 64 (3): 719-736.

Krinsky, I. and A. L. Robb (1986). "On Approximating the Statistical Properties of Elasticities." The Review of Economics and Statistics 68 (4): 715-719.

Loureiro, M. L. and W. J. Umberger (2003). "Estimating Consumer Willingness to Pay for Country-of-Origin Labeling." Journal of Agricultural and Resource Economics 28 (2): 287-301.

Louviere, J. J. and D. A. Hensher, et al. (2000). Stated Choice Methods: Analysis and Applications. UK, Cambridge University Press.

Louviere, J. J. and T. Islam, et al. (2008). "Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price?" Journal of Consumer Research 35 (2): 360-375.

Louviere, J. J. and T. N. Flynn, et al. (2010). "Discrete Choice Experiments Are Not Conjoint Analysis." Journal of Choice Modelling 3 (3): 57-72.

Louviere, J. and K. Train, et al. (2005). "Recent Progress on Endogeneity in Choice

Modeling." Marketing Letters 16 (3-4): 255-265.

Lusk, J. L. and F. B. Norwood (2005). "Effect of Experimental Design on Choice-Based Conjoint Valuation Estimates." American Journal of Agricultural Economics 87 (3): 771 -785.

Lusk, J. L. and J. Roosen, et al. (2003). "Demand for Beef from Cattle Administered Growth Hormones or Fed Genetically Modified Corn: A Comparison of Consumers in France, Germany, the United Kingdom, and the United States." American Journal of Agricultural Economics 85 (1): 16 -29.

Lusk, J. L. and T. C. Schroeder (2004). "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." American Journal of Agricultural Economics 86 (2): 467-482.

Martinsson, P. and F. Carlsson, et al. (2001). Using Choice Experiments for Non-Market Valuation, Göteborg University, Working Papers in Economics, No. 52.

Mennecke, B. E. and A. M. Townsend, et al. (2007). "A Study of the Factors that Influence Consumer Attitudes toward Beef Products Using the Conjoint Market Analysis Tool." Journal of Animal Science 85 (10): 2639-2659.

Muhammad, B. and Abdulai, A. (2016). Impact of Ex-Ante Hypothetical Bias Mitigation Methods on Attribute Non-Attendance in Choice Experiments" Forthcoming in *American Journal of Agricultural Economics.*

Naspetti, S. and A. Bodini (2008). "Consumer Perception of Local and Organic Products: Substitution or Complementary Goods?" The International Journal of Interdisciplinary Social Sciences 3 (2): 111-122.

Rose, J. M. and D. A. Hensher, et al. (2009). "Identifying Differences in Willingness to Pay Due to Dimensionality in Stated Choice Experiments: A Cross Country Analysis." Journal of Transport Geography 17 (1): 21-29.

Scarpa, R. and R. Zanoli, et al. (2013). "Inferred and Stated Attribute Non-attendance in Food Choice Experiments." American Journal of Agricultural Economics 95 (1): 165-180.

Swait, J. and J. Louviere (1993). "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." Journal of Marketing Research 30 (3): 305-314.

Swait, J. and W. Adamowicz (2001a). "Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice." Organizational Behavior and Human Decision Processes 86 (2): 141-167.

Swait, J. and W. Adamowicz (2001b). "The Influence of Task Complexity on Consumer Choice: a Latent Class Model of Decision Strategy Switching." Journal of Consumer Research 28 (1): 135-148.

Taglioni, C. and A. Cavicchi, et al. (2011). "Influence of Brand Equity on Milk Choice: A Choice Experiment Survey." International Journal on Food System Dynamics 2 (3): 305-325.

Tonsor, G. T. (2011). "Consumer Inferences of Food Safety and Quality." European Review of Agricultural Economics 38 (2): 213-235.

Train, K. E. (2003). Discrete Choice Methods with Simulation, Cambridge University Press.

Trine, K. (2005). "A Review of the Discrete Choice Experiment: with Emphasis on Its Application in Health Care." Health Economics Papers.

Yu. X., Z. Gao and S. Shimokawa (2016) Consumer Preferences for US Beef Products: A Meta-Analysis. Italian Review of Agricultural Economics (REA). Vol. LXXI,(2): 177-195.

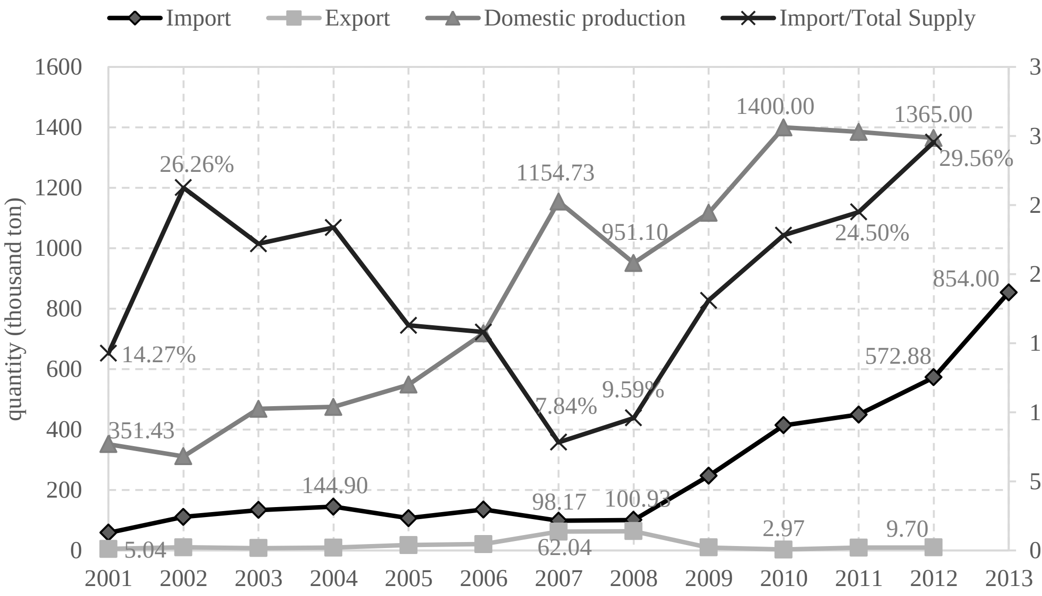

Figure 1 China's Annual Production, Export and Import of Milk formula

Data sources: China's Dairy Industry Yearbook, 2001 to 2013.

Table 1. Attributes used in Choice Experiment

Name of the Attribute	Range of Value
Price	{159, 238, 316, 395}
Imported Food	{Yes, No}
Organic Certificated	{Yes, No}
Probiotics Contained	{Yes, No}
Sugar-free Food	{Yes, No}

Table 2. Attributes in Choice Experiments across Five Groups

	Group1	Group2		Group3		Group4		Group5	
	F1	F2	S2	F3	S3	F4	S4	F5	S5
Price	√	√	√	√	√	√	√	√	√
Imported Food	√	√	√	√	√	√	√	√	
Organic Certificated	√	√	√	√	√	√		√	√
Probiotics Contained	√	√	√	√		√	√	√	√
Sugar-free Food	√	√		√	√	√	√	√	√

Table 3. Summary Statistics of Consumer Demographics by City and Group

Variables	Beijing					Xi'an				
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 1	Group 2	Group 3	Group 4	Group 5
Age	31.432	31.410	31.586	31.261	31.189	34.021	34.629	33.638	33.168	34.132
	(6.729)	(7.090)	(6.423)	(4.425)	(7.365)	(8.891)	(9.611)	(9.976)	(8.699)	(10.171)
Baby age	2.495	1.523	1.974	1.928	1.476	1.602	2.120	1.001	1.589	1.377
	(1.890)	(1.848)	(1.922)	(1.930)	(1.840)	(1.797)	(2.235)	(1.470)	(1.818)	(1.769)
Gender										
Female	72.80%	65.38%	65.71%	60.87%	63.51%	61.82%	54.29%	49.28%	56.84%	48.61%
Male	27.20%	34.62%	34.29%	39.13%	36.49%	38.18%	45.71%	50.72%	43.16%	51.39%
Relationship										
Parents	86.00%	69.87%	82.86%	81.16%	64.86%	69.09%	74.29%	52.17%	67.37%	63.19%
Grandparents	3.60%	3.21%	1.43%	2.90%	2.70%	5.76%	8.57%	5.80%	7.37%	7.64%
Others	10.40%	26.92%	15.71%	15.94%	32.43%	25.15%	17.14%	42.03%	25.26%	29.17%
Education										
1	0.00%	3.21%	1.43%	0.00%	1.35%	5.15%	8.57%	0.00%	12.63%	9.03%
2	5.20%	3.21%	1.43%	4.35%	12.16%	22.42%	24.76%	14.49%	22.11%	38.89%
3	73.20%	66.67%	70.00%	63.77%	64.86%	65.76%	58.10%	23.19%	56.84%	47.92%
4	21.60%	26.92%	27.14%	31.88%	21.62%	6.67%	8.57%	62.32%	8.42%	4.17%
Monthly Income(￥)										
Lower than 4000	3.20%	5.13%	2.86%	10.14%	16.22%	19.70%	27.62%	37.68%	29.47%	27.08%
(4,000, 8,000]	15.20%	27.56%	40.00%	20.29%	35.14%	39.70%	36.19%	39.13%	41.05%	38.89%
(8,000, 15,000]	49.20%	37.82%	34.29%	27.54%	33.78%	27.27%	28.57%	18.84%	23.16%	27.08%
More than 15,000	32.40%	29.49%	22.86%	42.03%	14.86%	13.33%	7.62%	4.35%	6.32%	6.94%
Employment										
Full time	90.80%	89.10%	95.71%	95.65%	85.14%	82.12%	78.10%	82.61%	81.05%	77.78%
Part time	4.00%	4.49%	1.43%	2.90%	8.11%	6.06%	8.57%	4.35%	5.26%	9.72%
Unemployed	2.40%	4.49%	1.43%	1.45%	4.05%	7.27%	5.71%	5.80%	9.47%	4.86%
Retired	2.80%	1.92%	1.43%	0.00%	2.70%	4.55%	7.62%	7.25%	4.21%	7.64%
No. of respondents	250	156	70	69	74	330	105	69	95	144

Notes: Reported statistics of age and baby age are means and standard deviations (in parentheses). For all other variables, reported statistics are frequency of the levels among respondents.

Age and Baby age: Age in years.

Babe age: For families containing more than one babies, we asked the age of the major one for milk consumption. For respondents who purchased milk formula not for their own babies, we skipped the question of baby age.

Education: 1=middle school and below; 2=high school graduate or college's degree; 3=Bachelor's degree; 4=Master's or Ph.D. degree.

Table 4. Estimates of RPL Model for Beijing Consumers Segmented by Group

Variables	Group 1	Group 2		Group 3		Group 4		Group 5	
	F1	F2	S2	F3	S3	F4	S4	F5	S5
Means									
Price	-0.570**	-0.408**	-0.550**	-0.207*	-1.913**	-0.234*	-1.106**	-0.488**	-1.183**
	(0.070)	(0.079)	(0.100)	(0.091)	(0.521)	(0.097)	(0.250)	(0.123)	(0.250)
Constant for None	0.546*	0.348	1.300**	0.157	3.026	0.345	0.0952	0.713	1.054
	(0.217)	(0.248)	(0.364)	(0.319)	(2.046)	(0.346)	(0.683)	(0.373)	(0.813)
Imported	2.139**	1.734**	2.629**	0.799**	7.683**	0.777**	4.734**	0.607**	
	(0.131)	(0.148)	(0.185)	(0.164)	(1.039)	(0.198)	(0.657)	(0.179)	
Organic	1.276**	1.109**	1.382**	0.682**	6.153**	0.151		2.092**	4.183**
	(0.128)	(0.128)	(0.165)	(0.193)	(1.010)	(0.190)		(0.273)	(0.475)
Probiotics	1.145**	0.690**	0.857**	0.119		0.539**	3.376**	0.974**	2.347**
	(0.115)	(0.130)	(0.244)	(0.150)		(0.169)	(0.609)	(0.209)	(0.389)
Sugar-free	0.312**	0.219*		0.273	2.129**	0.509**	0.569	0.287	0.976*
	(0.083)	(0.109)		(0.148)	(0.700)	(0.153)	(0.338)	(0.192)	(0.391)
Standard Deviations									
Imported	1.146**	1.173**	1.308**	0.836**	2.947**	1.226**	4.018**	0.691**	
	(0.122)	(0.152)	(0.180)	(0.190)	(0.597)	(0.217)	(0.687)	(0.192)	
Organic	1.346**	0.752**	1.158**	0.946**	0.391	0.937**		1.297**	2.591**
	(0.144)	(0.172)	(0.199)	(0.212)	(0.606)	(0.216)		(0.283)	(0.486)
Probiotics	0.775**	0.699**	2.348**	0.219		0.495*	4.083**	0.999**	2.248**
	(0.140)	(0.168)	(0.266)	(0.223)		(0.215)	(0.752)	(0.256)	(0.561)
Sugar-free	0.319*	0.68**		0.629**	2.048**	0.607**	1.598**	0.849**	2.534**
	(0.177)	(0.157)		(0.215)	(0.603)	(0.214)	(0.367)	(0.212)	(0.527)
Lower Covariance Matrix									
Org: Imp	-1.233**	-0.826**	-1.463**	-0.496*	-1.023	0.464		-0.385	
	(0.264)	(0.255)	(0.429)	(0.271)	(1.785)	(0.316)		(0.235)	
Pro: Imp	0.156	0.139	2.254**	-0.134		0.176	14.100**	0.288	
	(0.188)	(0.210)	(0.553)	(0.189)		(0.249)	(4.844)	(0.193)	
Sug: Imp	-0.246	0.160		-0.047	0.410	0.098	0.366	0.117	
	(0.141)	(0.192)		(0.188)	(1.976)	(0.248)	(1.718)	(0.154)	
Pro: Org	0.018	-0.024	-1.471**	0.035		-0.203		0.37	5.557**
	(0.203)	(0.150)	(0.561)	(0.197)		(0.224)		(0.410)	(1.937)
Sug: Org	0.280	-0.041		-0.356	0.314	-0.178		0.506	4.680**
	(0.151)	(0.150)		(0.257)	(1.177)	(0.209)		(0.417)	(1.820)
Sug: Pro	-0.073	-0.084		0.093		0.209	3.112	0.042	3.360*
	(0.117)	(0.141)		(0.144)		(0.134)	(1.694)	(0.306)	(1.409)
No. of Respondents	250	156	156	70	70	69	69	74	74
Observations	6000	3744	3744	1680	1680	1656	1656	1776	1776
Log likelihood	-1611.3	-1065.9	-893.1	-558.8	-193.8	-539.9	-269	-507.9	-287.6

Notes: Numbers in parentheses are standard errors. One and two asterisks respectively indicate statistical significance at the 0.05 and 0.01 level.

Table 5. Estimates of RPL Model for Xi'an Consumers Segmented by Group

Variables	Group 1	Group 2		Group 3		Group 4		Group 5	
	F1	F2	S2	F3	S3	F4	S4	F5	S5
					Means				
Price	-0.619**	-0.779**	-3.586**	-0.801**	-3.846**	-0.522**	-2.739**	-0.502**	-2.071**
	(0.061)	(0.143)	(0.676)	(0.163)	(0.865)	(0.095)	(0.460)	(0.079)	(0.273)
Constant for None	0.249	0.136	0.232	-0.146	-1.788	0.223	-0.543	-0.111	0.32
	(0.180)	(0.371)	(1.774)	(0.446)	(2.033)	(0.340)	(1.255)	(0.265)	(0.803)
Imported	1.939**	1.764**	7.801**	2.147**	8.888**	1.753**	7.639**	0.425**	
	(0.103)	(0.180)	(1.076)	(0.267)	(1.431)	(0.179)	(0.931)	(0.163)	
Organic	1.634**	2.470**	9.032**	1.580**	7.321**	0.338		1.112**	5.233**
	(0.121)	(0.255)	(1.261)	(0.266)	(1.395)	(0.194)		(0.123)	(0.478)
Probiotics	1.015**	1.964**	4.451**	0.419		1.452**	5.142**	1.339**	4.549**
	(0.091)	(0.180)	(0.875)	(0.296)		(0.190)	(0.724)	(0.121)	(0.414)
Sugar-free	0.277**	0.531**		1.296**	3.878**	0.614**	3.870**	0.988**	2.199**
	(0.076)	(0.186)		(0.215)	(1.070)	(0.138)	(0.808)	(0.105)	(0.434)
					Standard Deviations				
Imported	0.984**	0.654**	4.440**	1.224**	4.519**	0.894**	5.571**	1.533**	
	(0.103)	(0.209)	(0.762)	(0.256)	(0.955)	(0.178)	(0.812)	(0.163)	
Organic	1.469**	0.948**	2.140**	1.225**	2.250**	1.239**		0.276	4.307**
	(0.124)	(0.237)	(0.660)	(0.298)	(0.812)	(0.200)		(0.225)	(0.597)
Probiotics	0.698**	0.726**	3.498**	1.751**		0.837**	1.078	0.436*	1.591**
	(0.122)	(0.204)	(0.780)	(0.341)		(0.208)	(0.661)	(0.204)	(0.521)
Sugar-free	0.495**	1.052**		0.994**	4.208**	0.571**	2.676**	0.152	4.039**
	(0.120)	(0.204)		(0.294)	(0.838)	(0.193)	(0.622)	(0.140)	(0.570)
					Lower Covariance Matrix				
Org: Imp	-0.802**	-0.219	-9.241*	-1.349*	-9.080	-0.319		0.251	
	(0.190)	(0.215)	(3.974)	(0.561)	(4.716)	(0.265)		(0.216)	
Pro: Imp	0.439**	0.099	2.88	-0.235		0.697**	5.339	0.0889	
	(0.141)	(0.199)	(3.799)	(0.506)		(0.259)	(3.990)	(0.210)	
Sug: Imp	-0.020	-0.139		0.102	2.372	0.091	10.730*	0.157	
	(0.106)	(0.287)		(0.373)	(5.582)	(0.177)	(4.886)	(0.175)	
Pro: Org	0.171	0.573	0.363	0.239		-0.636		-0.009	6.554**
	(0.177)	(0.302)	(2.709)	(0.564)		(0.326)		(0.071)	(2.295)
Sug: Org	0.091	-0.367		0.395	1.964	-0.286		0.014	15.95**
	(0.150)	(0.295)		(0.490)	(3.215)	(0.242)		(0.046)	(4.412)
Sug: Pro	-0.040	-0.315		-0.418		0.129	1.019	0.052	6.321**
	(0.091)	(0.260)		(0.534)		(0.196)	(1.241)	(0.074)	(2.139)
No. of Respondents	330	105	105	69	69	95	95	144	144
Observations	7920	2520	2520	1656	1656	2280	2280	3456	3456
Log likelihood	-2146.3	-586.6	-273.9	-421.6	-184.2	-654	-278.4	-974.6	-455.1

Notes: Numbers in parentheses are standard errors. One and two asterisks respectively indicate statistical significance at the 0.05 and 0.01 level.

Table 6. Statistics of Simulated WTP Estimates Segmented by Group

City & Group	WTP for	F set			S set			Unpaired *t*-test		
		Mean	Std. Dev.	Percent	Mean	Std. Dev.	Percent	Diff.	Std. Err.	P-value
BJ,G2	Imported	4.301	2.931	46.99%	4.822	2.424	54.96%	-0.522	0.085	0.000
BJ,G2	Organic	2.663	1.891	29.10%	2.456	2.159	28.00%	0.207	0.064	0.001
BJ,G2	Probiotics	1.653	1.670	18.06%	1.495	4.218	17.04%	0.159	0.101	0.118
BJ,G2	Sugar-free	0.535	1.724	5.85%						
BJ,G2	*Total*	9.153	3.009		8.774	4.821		0.379	0.127	0.003
BJ,G3	Imported	3.926	4.110	43.66%	4.042	1.570	48.57%	-0.116	0.098	0.237
BJ,G3	Organic	3.104	4.648	34.52%	3.209	0.209	38.56%	-0.105	0.104	0.315
BJ,G3	Probiotics	0.568	1.070	6.32%						
BJ,G3	Sugar-free	1.393	3.037	15.49%	1.071	1.058	12.87%	0.322	0.072	0.000
BJ,G3	*Total*	8.992	2.244		8.323	1.830		0.669	0.065	0.000
BJ,G4	Imported	3.405	5.334	40.14%	4.346	3.706	55.42%	-0.942	0.145	0.000
BJ,G4	Organic	0.532	3.948	6.28%						
BJ,G4	Probiotics	2.339	2.124	27.58%	3.037	3.695	38.72%	-0.697	0.095	0.000
BJ,G4	Sugar-free	2.206	2.592	26.01%	0.459	1.412	5.86%	1.747	0.066	0.000
BJ,G4	*Total*	8.483	8.687		7.842	7.610		0.641	0.258	0.013
BJ,G5	Imported	1.269	1.445	15.93%						
BJ,G5	Organic	4.178	2.685	52.45%	3.573	2.232	55.62%	0.605	0.078	0.000
BJ,G5	Probiotics	1.951	1.991	24.50%	1.993	1.917	31.03%	-0.042	0.062	0.499
BJ,G5	Sugar-free	0.568	1.787	7.13%	0.858	2.160	13.36%	-0.290	0.063	0.000
BJ,G5	*Total*	7.966	4.856		6.424	5.775		1.542	0.169	0.000
XA,G2	Imported	2.280	0.856	26.57%	2.197	1.263	37.19%	0.083	0.034	0.015
XA,G2	Organic	3.120	1.226	36.35%	2.503	0.611	42.37%	0.617	0.031	0.000
XA,G2	Probiotics	2.489	0.918	29.00%	1.207	0.965	20.44%	1.282	0.030	0.000
XA,G2	Sugar-free	0.693	1.349	8.07%						
XA,G2	*Total*	8.581	1.879		5.907	1.367		2.674	0.052	0.000
XA,G3	Imported	2.706	1.557	40.41%	2.331	1.198	44.95%	0.375	0.044	0.000
XA,G3	Organic	1.922	1.569	28.70%	1.884	0.600	36.33%	0.038	0.038	0.311
XA,G3	Probiotics	0.488	2.159	7.28%						
XA,G3	Sugar-free	1.581	1.250	23.61%	0.971	1.064	18.72%	0.610	0.037	0.000
XA,G3	*Total*	6.697	2.588		5.186	1.471		1.511	0.067	0.000
XA,G4	Imported	3.390	1.748	42.48%	2.824	2.074	45.96%	0.565	0.061	0.000
XA,G4	Organic	0.549	2.385	6.88%						
XA,G4	Probiotics	2.831	1.645	35.48%	1.876	0.395	30.53%	0.955	0.038	0.000
XA,G4	Sugar-free	1.210	1.091	15.17%	1.445	0.998	23.51%	-0.234	0.033	0.000
XA,G4	*Total*	7.980	3.192		6.145	3.198		1.835	0.101	0.000
XA,G5	Imported	0.899	3.112	11.63%						
XA,G5	Organic	2.203	0.543	28.48%	2.563	2.120	44.00%	-0.360	0.049	0.000
XA,G5	Probiotics	2.663	0.860	34.43%	2.200	0.775	37.78%	0.462	0.026	0.000
XA,G5	Sugar-free	1.969	0.300	25.46%	1.061	1.954	18.22%	0.908	0.044	0.000
XA,G5	*Total*	7.735	3.895		5.825	4.761		1.910	0.138	0.000

Notes: BJ denotes Beijing, XA denotes Xi'an, Gn denotes group n. The unit for WTP is 100¥/900g. Reported means and standard deviations are obtained from 2,000 simulated WTP estimates utilizing Krinsky and Robb (1986) bootstrapping method. All WTP mean values are significantly not equal to zero at 0.01 level. Total WTP is the sum of all WTPs for individual attributes. Percentage value is the ratio of mean WTP for an individual attribute to the mean of total WTP. Difference value is the mean of WTP differences between F and S.

Table 7 Statistics of Individual-level WTP Estimates Segmented by Group

City & Group	WTP for	F set			S set			Paired *t*-test		
		Mean	Std. Dev.	Percent	Mean	Std. Dev.	Percent	Diff.	Std. Err.	P-value
BJ,G2	Imported	4.290	2.219	46.45%	4.782	1.859	53.93%	-0.491	0.150	0.001
BJ,G2	Organic	2.707	1.384	29.30%	2.516	1.552	28.37%	0.191	0.101	0.061
BJ,G2	Probiotics	1.672	0.969	18.10%	1.570	3.579	17.70%	0.103	0.266	0.700
BJ,G2	Sugar-free	0.568	0.957	6.14%						
BJ,G2	*Total*	9.237	2.045		8.867	4.088		0.370	0.305	0.227
BJ,G3	Imported	3.884	2.859	42.93%	3.959	1.326	47.96%	-0.075	0.374	0.842
BJ,G3	Organic	3.244	3.229	35.86%	3.220	0.157	39.01%	0.024	0.386	0.951
BJ,G3	Probiotics	0.575	0.639	6.36%						
BJ,G3	Sugar-free	1.344	1.815	14.86%	1.076	0.739	13.04%	0.268	0.226	0.240
BJ,G3	*Total*	9.047	0.913		8.255	1.588		0.792	0.214	0.000
BJ,G4	Imported	3.381	4.273	39.28%	4.422	3.190	53.25%	-1.042	0.610	0.092
BJ,G4	Organic	0.712	2.804	8.28%						
BJ,G4	Probiotics	2.292	1.285	26.63%	3.270	3.199	39.38%	-0.978	0.419	0.023
BJ,G4	Sugar-free	2.221	1.521	25.81%	0.613	1.067	7.38%	1.608	0.209	0.000
BJ,G4	*Total*	8.606	7.284		8.305	6.677		0.301	1.088	0.783
BJ,G5	Imported	1.243	0.975	15.82%						
BJ,G5	Organic	4.157	2.125	52.92%	3.541	1.997	55.77%	0.615	0.326	0.063
BJ,G5	Probiotics	1.874	1.380	23.86%	1.997	1.674	31.45%	-0.123	0.190	0.520
BJ,G5	Sugar-free	0.581	1.360	7.39%	0.812	1.786	12.78%	-0.231	0.225	0.308
BJ,G5	*Total*	7.854	4.143		6.350	5.260		1.505	0.581	0.012
XA,G2	Imported	2.264	0.463	26.26%	2.150	1.086	36.35%	0.114	0.086	0.191
XA,G2	Organic	3.170	0.828	36.76%	2.530	0.510	42.77%	0.640	0.072	0.000
XA,G2	Probiotics	2.518	0.604	29.21%	1.236	0.765	20.89%	1.283	0.069	0.000
XA,G2	Sugar-free	0.670	0.939	7.77%						
XA,G2	*Total*	8.622	1.220		5.916	1.167		2.706	0.150	0.000
XA,G3	Imported	2.604	1.165	38.55%	2.249	1.042	43.29%	0.355	0.195	0.073
XA,G3	Organic	2.062	1.173	30.53%	1.932	0.494	37.19%	0.130	0.159	0.416
XA,G3	Probiotics	0.423	1.748	6.26%						
XA,G3	Sugar-free	1.665	0.930	24.65%	1.015	0.964	19.53%	0.650	0.161	0.000
XA,G3	*Total*	6.754	1.911		5.196	1.361		1.558	0.245	0.000
XA,G4	Imported	3.312	1.282	41.69%	2.831	1.909	46.01%	0.481	0.230	0.039
XA,G4	Organic	0.712	1.716	8.96%						
XA,G4	Probiotics	2.730	1.201	34.36%	1.882	0.335	30.59%	0.848	0.126	0.000
XA,G4	Sugar-free	1.191	0.579	14.99%	1.440	0.790	23.40%	-0.250	0.104	0.018
XA,G4	*Total*	7.944	2.312		6.153	2.936		1.791	0.373	0.000
XA,G5	Imported	0.890	2.557	11.48%						
XA,G5	Organic	2.221	0.287	28.64%	2.485	1.928	43.75%	-0.264	0.162	0.104
XA,G5	Probiotics	2.670	0.372	34.44%	2.178	0.716	38.35%	0.492	0.066	0.000
XA,G5	Sugar-free	1.972	0.200	25.43%	1.017	1.801	17.90%	0.955	0.150	0.000
XA,G5	*Total*	7.753	3.175		5.680	4.423		2.073	0.448	0.000

Notes: BJ denotes Beijing, XA denotes Xi’an, Gn denotes group n. The unit for WTP is 100¥/900g. Reported means and standard deviations are obtained from individual-level WTP estimates in each CE. All WTP mean values are significantly not equal to zero at 0.01 level, except that Organic in BJ,G4 and Probiotics in XA,G3 significant at 0.05 level. Total WTP is the sum of all WTPs for individual attributes in each CE. Percentage value is the ratio of mean WTP for an individual attribute to the mean of total WTP. Difference value is the mean of WTP differences between F and S.

Table 8 Comparisons of Changes in WTP across Attributes within Groups

City & Group	Difference of WTP Changes between	Simulated WTP						Individual-level WTP					
		Mean Diff.	Diff. of Absolute Mean	Paired *t*-test		Unpaired *t*-test		Mean Diff.	Diff. of Absolute Mean	Paired *t*-test		Unpaired *t*-test	
				Std. Err.	P-value	Std. Err.	P-value			Std. Err.	P-value	Std. Err.	P-value
Beijing, group2	Imported & Organic	-0.729	0.315	0.005	0.000	0.014	0.000	-0.682	0.300	0.243	0.006	0.180	0.000
Beijing, group2	Imported & Probiotics	-0.680	0.363	0.092	0.000	0.084	0.000	-0.594	0.384	0.218	0.007	0.305	0.053
Beijing, group2	Organic & Probiotics	0.049	0.049	0.090	0.588	0.083	0.560	0.088	0.088	0.337	0.794	0.284	0.756
Beijing, group3	Imported & Organic	-0.012	0.011	0.143	0.934	0.115	0.918	-0.099	0.051	0.703	0.888	0.537	0.854
Beijing, group3	Imported & Sugar-free	-0.438	-0.206	0.108	0.000	0.104	0.000	-0.343	-0.193	0.416	0.412	0.437	0.434
Beijing, group3	Organic & Sugar-free	-0.427	-0.217	0.171	0.013	0.133	0.001	-0.244	-0.244	0.552	0.660	0.447	0.586
Beijing, group4	Imported & Probiotics	-0.244	0.244	0.123	0.046	0.104	0.019	-0.064	0.064	0.514	0.902	0.740	0.932
Beijing, group4	Imported & Sugar-free	-2.689	-0.805	0.073	0.000	0.076	0.000	-2.650	-0.566	0.616	0.000	0.645	0.000
Beijing, group4	Probiotics & Sugar-free	-2.444	-1.050	0.083	0.000	0.119	0.000	-2.587	-0.630	0.366	0.000	0.468	0.000
Beijing, group5	Organic & Probiotics	0.646	0.563	0.078	0.000	0.103	0.000	0.738	0.492	0.245	0.004	0.377	0.053
Beijing, group5	Organic & Sugar-free	0.895	0.315	0.075	0.000	0.119	0.000	0.846	0.384	0.238	0.001	0.396	0.035
Beijing, group5	Probiotics & Sugar-free	0.248	-0.248	0.091	0.006	0.084	0.003	0.108	-0.108	0.253	0.670	0.294	0.714
Xi’an, group2	Organic & Imported	0.533	0.533	0.025	0.000	0.024	0.000	0.526	0.526	0.129	0.000	0.112	0.000
Xi’an, group2	Probiotics & Imported	1.199	1.199	0.010	0.000	0.010	0.000	1.169	1.169	0.087	0.000	0.111	0.000
Xi’an, group2	Organic & Probiotics	-0.665	-0.665	0.024	0.000	0.023	0.000	-0.643	-0.643	0.086	0.000	0.100	0.000
Xi’an, group3	Imported & Organic	0.337	0.337	0.029	0.000	0.023	0.000	0.225	0.225	0.345	0.517	0.251	0.372
Xi’an, group3	Imported & Sugar-free	-0.236	-0.236	0.028	0.000	0.028	0.000	-0.296	-0.296	0.236	0.214	0.252	0.243
Xi’an, group3	Organic & Sugar-free	-0.572	-0.572	0.032	0.000	0.035	0.000	-0.521	-0.521	0.206	0.014	0.226	0.023
Xi’an, group4	Imported & Probiotics	-0.390	-0.390	0.038	0.000	0.032	0.000	-0.367	-0.367	0.167	0.031	0.262	0.164
Xi’an, group4	Imported & Sugar-free	0.800	0.331	0.029	0.000	0.032	0.000	0.731	0.231	0.176	0.000	0.252	0.004
Xi’an, group4	Probiotics & Sugar-free	1.190	0.721	0.053	0.000	0.044	0.000	1.097	0.598	0.129	0.000	0.163	0.000
Xi’an, group5	Probiotics & Organic	-0.822	0.102	0.037	0.000	0.048	0.000	-0.756	0.282	0.108	0.000	0.175	0.000
Xi’an, group5	Organic & Sugar-free	-1.268	-0.548	0.027	0.000	0.057	0.000	-1.219	-0.691	0.039	0.000	0.221	0.000
Xi’an, group5	Probiotics & Sugar-free	-0.446	-0.446	0.031	0.000	0.047	0.000	-0.464	-0.464	0.094	0.000	0.165	0.005

Notes: Mean difference values are the differences between mean WTP for two attributes, for example, difference value -0.729 between Imported & Organic in Beijing, group2 is the difference of mean WTP change values of Imported and Organic, respectively -0.522 and 0.207 in Table 5, and the corresponding absolute mean difference value is 0.315. P-values are two-tailed significance values for alternative hypothesis of unequal means. For simulated WTP estimates, observations are randomly paired to conduct the paired *t*-test. Reported unpaired *t*-test assumes unequal variances, whose result is similar to the test assuming equal variances presented in supplementary tables.

Table 9 Comparisons of Changes in WTP across Groups

City	Difference of WTP Changes between	Attributes	Simulated WTP			Individual-level WTP		
			Diff.	Std. Err.	P-value	Diff.	Std. Err.	P-value
Beijing	Group2 & Group3	Imported	-0.405	0.058	0.000	-0.416	0.403	0.304
Beijing	Group2 & Group3	Organic	0.312	0.100	0.002	0.167	0.399	0.677
Beijing	Group2 & Group3	*Total*	-0.290	0.084	0.001	-0.422	0.373	0.259
Beijing	Group2 & Group4	Imported	0.420	0.038	0.000	0.550	0.628	0.384
Beijing	Group2 & Group4	Probiotics	0.856	0.128	0.000	1.081	0.496	0.031
Beijing	Group2 & Group4	*Total*	-0.261	0.100	0.009	0.069	1.130	0.952
Beijing	Group2 & Group5	Organic	-0.398	0.094	0.000	-0.424	0.341	0.217
Beijing	Group2 & Group5	Probiotics	0.200	0.093	0.031	0.225	0.327	0.491
Beijing	Group2 & Group5	*Total*	-1.162	0.156	0.000	-1.135	0.656	0.087
Beijing	Group3 & Group4	Imported	0.825	0.067	0.000	0.967	0.716	0.180
Beijing	Group3 & Group4	Sugar-free	-1.425	0.110	0.000	-1.340	0.308	0.000
Beijing	Group3 & Group4	*Total*	0.029	0.096	0.766	0.491	1.109	0.659
Beijing	Group3 & Group5	Organic	-0.709	0.137	0.000	-0.591	0.505	0.244
Beijing	Group3 & Group5	Sugar-free	0.612	0.114	0.000	0.499	0.319	0.120
Beijing	Group3 & Group5	*Total*	-0.872	0.154	0.000	-0.713	0.620	0.253
Beijing	Group4 & Group5	Probiotics	-0.656	0.106	0.000	-0.855	0.460	0.066
Beijing	Group4 & Group5	Sugar-free	2.037	0.099	0.000	1.839	0.307	0.000
Beijing	Group4 & Group5	*Total*	-0.901	0.163	0.000	-1.204	1.233	0.331
Xi'an	Group2 & Group3	Imported	-0.292	0.012	0.000	-0.241	0.213	0.261
Xi'an	Group2 & Group3	Organic	0.579	0.031	0.000	0.510	0.174	0.004
Xi'an	Group2 & Group3	*Total*	1.163	0.047	0.000	1.149	0.288	0.000
Xi'an	Group2 & Group4	Imported	-0.482	0.012	0.000	-0.367	0.245	0.137
Xi'an	Group2 & Group4	Probiotics	0.327	0.032	0.000	0.435	0.144	0.003
Xi'an	Group2 & Group4	*Total*	0.839	0.058	0.000	0.916	0.402	0.024
Xi'an	Group2 & Group5	Organic	0.976	0.047	0.000	0.904	0.177	0.000
Xi'an	Group2 & Group5	Probiotics	0.819	0.025	0.000	0.791	0.096	0.000
Xi'an	Group2 & Group5	*Total*	0.764	0.044	0.000	0.633	0.473	0.182
Xi'an	Group3 & Group4	Imported	-0.190	0.011	0.000	-0.126	0.301	0.675
Xi'an	Group3 & Group4	Sugar-free	0.845	0.041	0.000	0.900	0.191	0.000
Xi'an	Group3 & Group4	*Total*	-0.324	0.060	0.000	-0.233	0.446	0.602
Xi'an	Group3 & Group5	Organic	0.398	0.047	0.000	0.394	0.226	0.084
Xi'an	Group3 & Group5	Sugar-free	-0.298	0.048	0.000	-0.305	0.220	0.168
Xi'an	Group3 & Group5	*Total*	-0.399	0.047	0.000	-0.516	0.511	0.314
Xi'an	Group4 & Group5	Probiotics	0.493	0.040	0.000	0.356	0.142	0.014
Xi'an	Group4 & Group5	Sugar-free	-1.142	0.050	0.000	-1.205	0.183	0.000
Xi'an	Group4 & Group5	*Total*	-0.075	0.057	0.191	-0.283	0.583	0.628

Notes: Difference values are the difference of mean WTP between two groups. P-values are two-tailed significance values for unpaired *t*-tests with alternative hypothesis of unequal means and assuming unequal variances across groups.