U.S. Consumers' Willingness to Pay for Flavor and Tenderness in Steaks as Determined with an Experimental Auction

Dillon M. Feuz, Wendy J. Umberger, Chris R. Calkins, and Bethany Sitz

In a study of beef quality, consumers tasted steak samples and participated in an experimental auction to determine their willingness to pay. Steaks differed in marbling, tenderness, country of origin, and aging method. Marbling and tenderness had statistically significant impacts on consumers' palatability ratings for steaks. Tenderness significantly impacted consumers' willingness-to-pay values. There appear to be threshold levels of marbling and tenderness, below which consumers discount steaks. Steaks from Australia were rated lower for overall acceptability, and bids were lower than for the U.S. steak samples. Dry-aging methods negatively impacted taste panel ratings and bids.

Key words: beef, country of origin, experimental auctions, marbling, tenderness

Introduction

After facing nearly 20 years of declining demand (i.e., from 1979–1998), the beef industry has recently invested in research to determine the factors impacting consumers' satisfaction with beef (Smith et al., 1995). Several independent taste panel studies have found that tenderness (Boleman et al., 1997; Morgan et al., 1991; Shackelford et al., 2001), marbling, the amount of intra-muscular fat deposits (Savell et al., 1987; Morgan et al., 1998; Neely et al., 1998; Killinger et al., 2001), and production processes (Jeremiah et al., 1998; Umberger et al., 2002) are important determinants of consumers' overall satisfaction with beef. All of these previous studies have focused on only one specific beef attribute (e.g., tenderness, marbling, or production processes) that influences beef palatability. Questions remain as to whether or not a tender, but low marbled beef product, would be preferred over a less tender, but highly marbled beef product. Is tenderness more important than marbling, the primary determinant of USDA beef quality grades? What beef palatability characteristics, in addition to marbling and tenderness, are most important to consumers' overall satisfaction with beef?
Perhaps of equal or even greater importance to identifying what beef palatability characteristics consumers desire is determining their willingness to pay (WTP) for specific beef palatability attributes. Most meat science research on beef palatability has focused only on determining consumers' preferences for palatability attributes. However, agricultural economists and meat scientists have conducted a few joint studies where consumers' WTP for beef palatability attributes have been elicited. These studies have used experimental auctions to determine WTP after consumers had actually sampled the product (Lusk et al., 2001b; Melton et al., 1996; Umberger et al., 2002).

This study extends prior research, by having consumers taste and rate various beef samples which vary in tenderness, marbling (percent fat in lean muscle tissue), production processes, and country of origin (U.S. corn-fed, Australian grass-fed, Canadian barley-fed), and aging method (dry-aged, wet-aged). In addition to tasting and rating steak samples, consumer panelists were asked in this study to participate in a random nth-price sealed bid auction. The auction bid values revealed the consumers' WTP for their preferred palatability attributes in beef steaks. The primary objective of this analysis was to determine the impact that each of these palatability attributes had on consumer preferences and on consumer WTP. The impact of selected consumer demographic variables and beef purchasing characteristics on WTP for beef steaks was also assessed.

**Experimental Procedures**

In June and July of 2002, a sample of consumers in Denver, Colorado, and Chicago, Illinois, were randomly screened and selected by telephone to participate in a study of beef quality. Qualifying individuals (those willing to eat beef) were told they would have the opportunity to taste and to purchase New York strip beef steaks, and would be paid $50 for two hours of their time. Individuals agreeing to participate were scheduled for one of 12 panels conducted in each city. After arriving at the research facility, consumers were paid the $50 promised to them and were asked to complete surveys describing their meat-purchasing behavior, eating preferences, knowledge about beef, and their socio-demographic characteristics.

After completing the survey, panelists were informed they would be asked to taste and rate four pairs of steak samples for flavor, juiciness, tenderness, and overall acceptability. Ratings were established using an eight-point hedonic scale (where 1 = extremely dry, extremely tough, extremely undesirable for flavor, extremely undesirable for overall acceptability; and 8 = extremely juicy, extremely tender, extremely desirable for flavor, extremely desirable for overall acceptability). Panelists were then informed they would have an opportunity to participate in an auction and to submit sealed bids (in $/pound of steak) for each steak sample. Bidding would take place after each pair of steaks had been tasted and rated.

The procedures for a variant of the random nth-price auction (Shogren et al., 2001) were explained to participants, and they participated in three non-binding trial auctions on visually evaluated New York strip steaks to familiarize them with the auction process. The auctions were randomly chosen to be a second-, third-, or fourth-price auction, and this procedure was explained to the panelists prior to the auction. They were informed that only one randomly selected auction in each pair of auctions would be binding. Participants were encouraged to bid exactly the amount they believed the
product was worth to them, and were reminded that if they “won” a binding auction, they would be obligated to purchase the one-pound package of steaks at the auction market price.

After the practice auctions were completed, panelists were moved into individual taste panel booths. They were then given a warm-up steak sample to taste and rate. Next, panelists were served the first sample of the first pair on a plate, which was identified with a random sample identification number. The panelists wrote the sample identification number on their record sheet and tasted and evaluated (using the eight-point scale) the first sample for each of the four sensory characteristics: juiciness, tenderness, flavor, and overall acceptability. The second sample of the pair was served, and participants again evaluated the sample for each of the four sensory traits. After both samples had been tasted and evaluated, the panelists were given two bid sheets and were asked to simultaneously submit bids ($/pound) for each sample. Participants indicated their WTP for each steak on its individual bid sheet, which was labeled with the sample's unique identification number. The panelists' two bid sheets were collected, and the auction monitors determined the purchase price for each of the samples. The purchase price for each sample was randomly determined and was the second-, third-, or fourth-highest bid for the sample. After the purchase price was determined, participants were given a slip of paper for each steak sample informing them of the steak's purchase price, and whether or not they were a “potential winner.” This entire process was repeated three additional times for the remaining three pairs of steaks.

Participants knew beforehand that only one steak auction in each pair would be binding, so they knew that the most steaks they could potentially purchase would be four (one out of each pair). However, “potential winners” did not know whether they would actually need to purchase the steak until all eight steaks were auctioned and the binding auctions for each pair announced. For example, participants receiving slips of paper stating they were a “potential winner” of only one of the steaks in a pair did not know whether or not they would actually need to purchase the steak until the four binding auctions were announced at the end of the experiment. After all eight auctions were completed, the four binding auctions were randomly determined and the identification numbers of panelists who did not win any of the binding auctions were announced. These individuals were free to leave. The panelists who had won binding auctions stayed to purchase their steaks.

A variant of the random nth-price auction (Shogren et al., 2001) was used in order to elicit consumers’ true valuation of various palatability attributes of beef steaks. The WTP mechanism identified by an auction is less hypothetical than contingent valuation methods (Fox et al., 1995). Non-random, nth-price auctions, such as the frequently used second-price Vickrey auction, have been employed in previous WTP studies (see, for example, Buhr et al., 1993; Lusk et al., 2001a; Melton et al., 1996; Menkhaus et al., 1992; Umberger et al., 2002) because theoretically, this auction mechanism has been shown to be demand-revealing and to induce auction participants to reveal their true WTP for an auctioned good or service (Vickrey, 1961). There is, however, evidence suggesting the second-price auction may not be demand-revealing in practice (e.g., Kagel, Harstad, and Levin, 1987; Knetsch, Tang, and Thaler, 2001). This is particularly true when multiple-round auctions are used (Lusk, 2003; Shogren et al., 1994). As pointed out by both Lusk (2003) and Shogren et al. (2001), subjects who are off-margin may become disengaged when participating in nth-price auctions because they discover that
they either will not win or will not lose the auction, consequently creating poor incentives for these participants to behave as economic theory would predict.\footnote{Off-margin bidders are auction participants whose values are either relatively low or relatively high compared to the market-clearing price (Shogren et al., 2001).}

The random nth-price auction (Shogren et al., 2001) was introduced in order to take advantage of the active market participation feature of the Vickrey auction (demand-revealing and endogenous market-clearing price), to engage off-margin bidders, and to reduce the incentive for strategic bidding in repeated auctions. In a recent study comparing the precision and bias of the second-price versus the random nth-price auction, Parkhurst, Shogren, and Dickinson (2004) found the second-price auction to be precise, but biased; in particular, the highest-positive-value bidders tended to overstate their benefits while the lowest-negative-value bidders tended to understate their costs. On the other hand, aggregate bidding in the random nth-price auction was found to be imprecise, but bidding behavior was demand-revealing regardless of the induced value.

Based on the findings cited above, for this research a variant of the random, nth-price auction was selected over the second-price auction as the auction mechanism to elicit WTP, because of its unique ability to engage all bidders and to reduce the possibility that panelists would attempt to bid strategically. It is more difficult for panelists to determine how their bid values compare to other panelists’ bids if they do not know how many “winners” there are in each auction. For example, if a panelist was notified after the first set of auctions that she was a potential “winner” of one of the auctions, she did not know whether hers was the highest, second-highest, or third-highest bid.\footnote{As one reviewer pointed out, because our auctions were randomly chosen to be second-, third-, or fourth-price auctions, versus allowing the market price of the auction truly to be randomly determined (for example, allowing the auction to be a sixth-price or tenth-price auction), our variant of the random nth-price auction may not be as efficient at engaging off-margin bidders as a true, random nth-price auction.}

It also is important to point out that panelists in this study participated in multiple bidding rounds and had the opportunity to purchase more than one good (up to four steaks).\footnote{Due to the potential issues associated with wealth effects and demand reduction, only one good should have been sold—in other words, only one bidding round out of the eight (rather than multiple auctions) should have been randomly determined to be binding. However, because one of the goals of this analysis was to address specific meat science issues related to consumers’ at-home preparation of steaks, it was important to the meat scientists involved in the research process to be able to sell steaks out of each pair.} Lusk (2003), and Lusk, Feldkamp, and Schroeder (2004) argue that not controlling for demand reduction or wealth effects is a potential flaw in auction designs where subjects participate in multiple bidding rounds and have the opportunity to bid and to purchase multiple goods. Lusk (2003) suggests the easiest way to control for the problem of demand reduction is to randomly determine the binding, bidding round. Therefore, in order to reduce the impact of potential demand reduction or wealth effects on bids in this study, the “winning” auction in each pair was also randomly determined after all auctions were complete. There still may be some wealth effect or demand reduction present, as participants knew they could potentially purchase more than one steak and they also knew if they were potential auction “winners” or “non-winners” in a previous set of auctions. This potential for wealth effects or demand reduction is empirically addressed in subsequent sections of this article.

The four pairs of steak samples were: (a) U.S. corn-fed versus Australian grass-fed, (b) U.S. corn-fed versus Canadian barley-fed, (c) dry-aged USDA Choice versus wet-aged USDA Choice, and (d) dry-aged USDA Prime versus wet-aged USDA Prime. The order in which paired steak samples were presented to participants for taste evaluation was
randomly chosen for each of the 24 taste panels. Furthermore, half of the panelists in each panel tasted one of the samples first, while the other half of the panelists tasted the other sample in each pair first. For example, if a panel first sampled the pair of USDA Choice aged steaks, then half of the panel would have tasted and rated the dry-aged steak first and the other half of the panel would have tasted and rated the wet-aged steak first. Each panelist would then have tasted and rated the other steak in the pair. Then all panelists would have submitted two bids, one for each sample in the pair. Additionally, the taste panels were “blind,” meaning that consumers did not know there were differences between the steaks they tasted.

The primary focus of this experiment was to determine if panelists could detect flavor differences due to country-of-origin/production practices or aging. For each steak sample, objective measures determined the marbling (percent fat in the lean muscle tissue) and tenderness levels. Additionally, each sample was classified as to country-of-origin/production method (U.S. corn-fed, Australian grass-fed, or Canadian barley-fed) and as to aging method (no additional aging, dry-aged, and wet-aged). An objective measure of marbling, the Soxhlet method (Association of Official Analytical Chemists, 1990) was used to determine the percent of fat in the steaks. Tenderness was measured using standard procedures to determine the Warner-Bratzler shear force (WBSF) value. WBSF measures the amount of force necessary for a fixed blade to shear through a cooked sample of meat (Shackelford et al., 2001).

Tenderness levels and marbling were held constant within paired samples to isolate these potential flavor differences. However, the level of tenderness and marbling for each paired sample differed from one panel to the next. In total, there were 192 different steak samples (24 panels x 8 samples) used in the experiment. Therefore, it is possible to econometrically evaluate the relative impact of country-of-origin/production practices, aging, tenderness, and marbling on panelists’ ratings for flavor, juiciness, tenderness, and overall acceptability. Additionally, because panelists bid on each of the steaks, we can estimate the relative impact of these attributes on panelists’ WTP for each steak sample.

**Empirical Specification and Estimation**

The parameters for the following equation were estimated to determine the impact of the different intrinsic beef quality variables on panelists’ palatability ratings:

\[
FLAVOR_{ij}, JUICINESS_{ij}, TENDERNESS_{ij}, OVERALL_{ij} = \alpha_0 + \beta_1 \%FAT_{ij} + \beta_2 WBSF_{ij} + \beta_3 AUSTRALIA_{ij} + \\
\beta_4 CANADA_{ij} + \beta_5 DRYAGE_{ij} + \beta_6 WETAGE_{ij} + \varepsilon_{ij},
\]

where the dependent variables **FLAVOR, JUICINESS, TENDERNESS, and OVERALL** represent the ith (i = 1–273) panelist’s ratings on an eight-point scale for each of the j (j = 1–192) steaks’ flavor, juiciness, tenderness, and overall acceptability, respectively. The independent variables \%FAT and WBSF are real, continuous variables used to indicate the marbling and tenderness levels of the jth steak. AUSTRALIA, CANADA, DRYAGE, and WETAGE are 0/1 dummy variables indicating the jth steak was Australian, Canadian, dry-aged, or wet-aged, respectively.
FLAVOR, JUICINESS, TENDERNESS, and OVERALL are ordinal variables. A high value represents a high degree of satisfaction for an individual steak with that trait, but it would be inappropriate to assume a change in rating between steaks from 1 to 2 is of the same magnitude as a change from 2 to 3. Likewise, no valid comparison can be made between equal values for different variables. Parameter estimates for equation (1) were estimated using a random-effects (based on individual participants) ordered choice probit procedure for panel data in LIMDEP (Greene, 2002, pp. E18-38).

In addition to understanding how these objective beef quality measures impact palatability ratings, it is also important to understand how quality variables translate into value. Thus, the following equation was estimated to determine the impact of these same quality attributes on participants' WTP as measured through their bids for individual steaks:

\[
BID_{ij} = \alpha_0 + \beta_1 %FAT_{ij} + \beta_2 WBSF_{ij} + \beta_3 AUSTRALIA_{ij} + \beta_4 CANADA_{ij} + \beta_5 DRYAGE_{ij} + \beta_6 WETAGE_{ij} + \beta_7 P2_{ij} + \beta_8 P3_{ij} + \beta_9 P4_{ij} + \epsilon_{ij},
\]

where \( BID \) is the \( i \)th panelist's bid in $/pound for the \( j \)th steak sample; \%FAT, WBSF, AUSTRALIA, CANADA, DRYAGE, and WETAGE are as defined in equation (1). Bidding took place in four sequential time periods, and it was possible (due to the potential wealth effects or demand reduction discussed previously) that participants' knowledge about whether they were a "potential winner" or "non-winner" of an auction in one time period could impact their bids in the subsequent time periods. Therefore, in equation (2), where \( BID \) is the dependent variable, three additional 0/1 dummy variables were included to indicate the auction time period: \( P2, P3, \) and \( P4. BID \) is a real, continuous variable, which could be estimated using ordinary least squares (OLS). However, because the data represent a panel, OLS estimation would not be efficient. As a result, equation (2)'s parameters were estimated using the LIMDEP procedure for panel data with random effects based on individual participants. LIMDEP uses a two-step, generalized least squares (GLS) procedure to estimate this type of model (Greene, 2002, pp. E14-15).

Table 1 reports summary statistics for each of the variables used in equations (1) and (2). Our a priori expectations were that the amount of marbling (\%FAT) would have a positive impact on the dependent variables, and WBSF would have a negative impact on the dependent variables (higher WBSF values indicate less tender beef) in equations (1) and (2). The variables AUSTRALIA and CANADA were both expected to have a negative impact on participants' palatability ratings and their WTP (\( BID \)) for steaks from those two countries. Beef from Australian, grass-fed cattle and Canadian, barley-fed cattle will likely differ in flavor and may differ in juiciness from U.S., corn-fed beef. On average, U.S. consumers are likely accustomed to U.S. corn-fed beef and would be expected to prefer the taste of the U.S. product.\(^4\) There were no a priori expectations for the coefficients corresponding to the impact of aging methods (DRYAGE and WETAGE) on the dependent variables denoting overall acceptability (OVERALL) and WTP (\( BID \)).

\(^4\) In addition to the type of diet (corn-fed, grass-fed, or barley-fed), cattle breed and age may also differ between countries (Sitz, 2003). Berry et al. (1988); Davis et al. (1981); and Sitz et al. (2004) all reported significantly higher palatability ratings for beef from corn-fed animals versus beef from grass-fed animals. Jeremiah et al. (1998) found consumers preferred beef from corn-fed versus beef from barley-fed steers.
Table 1. Summary Statistics for all Variables Used in the Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>BID ($/lb.)</td>
<td>3.117</td>
<td>2.667</td>
<td>0</td>
<td>11.13</td>
</tr>
<tr>
<td>OVERALL</td>
<td>5.464</td>
<td>1.454</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>FLAVOR</td>
<td>5.658</td>
<td>1.426</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>JUICINESS</td>
<td>5.338</td>
<td>1.487</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>TENDERNESS</td>
<td>5.425</td>
<td>1.526</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>%FAT</td>
<td>10.482</td>
<td>3.515</td>
<td>3.65</td>
<td>22.14</td>
</tr>
<tr>
<td>WBSF</td>
<td>3.038</td>
<td>0.653</td>
<td>2.01</td>
<td>5.27</td>
</tr>
<tr>
<td>US</td>
<td>0.750</td>
<td>0.433</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AUSTRALIA</td>
<td>0.125</td>
<td>0.331</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CANADA</td>
<td>0.125</td>
<td>0.331</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>DRYAGE</td>
<td>0.250</td>
<td>0.433</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>WETAGE</td>
<td>0.250</td>
<td>0.433</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BEEFEAT</td>
<td>1.974</td>
<td>0.845</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>MALE</td>
<td>0.271</td>
<td>0.445</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>AGE</td>
<td>6.074</td>
<td>1.930</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>INCOME</td>
<td>7.093</td>
<td>2.277</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>NONCAUCASIAN</td>
<td>0.135</td>
<td>0.342</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>KIDS</td>
<td>0.630</td>
<td>0.483</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>LOCATION</td>
<td>0.484</td>
<td>0.500</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PANELSIZE</td>
<td>11.520</td>
<td>1.090</td>
<td>7</td>
<td>12</td>
</tr>
</tbody>
</table>

* N = 2,184 (273 panelists x 8 steak samples); each panelist submitted bids and provided palatability ratings on eight steaks.

In general, aged beef is perceived by consumers as being tenderer than beef that is not aged. The dry-aging process (DRYAGE) was expected to have a negative impact on JUICINESS, and could impact FLAVOR. The dry-aging process results in a rich beef flavor which may or may not be desired by some consumers.5 If wealth effects were present, then auction time period would have a negative impact on BID (P2, P3, and P4); however, if participants incorporated market feedback into their valuations, then auction time period could also have a positive impact on BID (Lusk, 2003; Lusk, Feldkamp, and Schroeder, 2004).6

Panelists' WTP in experimental auctions was likely influenced by more than just the palatability characteristics of the beef being auctioned. As shown in a previous experimental WTP study by Umberger and Feuz (2004), consumer demographics, beef purchasing behavior, and experimental design of the auction influence bids.7 To account for these factors, equation (2) was revised to include demographic, beef consumption, and experimental design variables, and was reestimated using OLS to determine the impact on panelists' bids of the different variables:


6 As discussed by Lusk (2003), and Lusk, Feldkamp, and Schroeder (2004), bidder affiliation may also cause subjects to increase their bids in subsequent auctions when posted prices allow participants to receive market feedback. Bidder affiliation can potentially violate the incentive compatibility property of the auction mechanism (Milgrom and Weber, 1982). List and Shogren (1999) found that bidder affiliation from posted prices had only a small effect on median WTP bids and did not affect the bids for familiar market goods.

7 Also see Lusk (2003), and Lusk, Feldkamp, and Schroeder (2004) for a discussion of many of the other experimental auction design issues that may influence auction bids.
where \textit{BEEFEAT} is a categorical variable (1–6) indicating increasing frequency of weekly beef consumption, \textit{NONCAUCASIAN} is a dummy variable (0/1) indicating the panelist is not Caucasian, \textit{MALE} is a 0/1 variable indicating the panelist is male, \textit{AGE} is a categorical variable (1–10) for increasing age categories, \textit{INCOME} is a categorical variable (1–9) for increasing levels of income, \textit{KIDS} is a 0/1 variable indicating the presence of children in the household, \textit{LOCATION} is a 0/1 variable equal to 0 for the Chicago market and equal to 1 for the Denver market, \textit{PANELSIZE} is the number of panelists in each experimental session (7–12). The remaining variables are as previously defined in equation (2). (Table 1 gives summary statistics for each of the additional variables.)

Results

In total, 273 consumers participated in the taste panels and experimental auctions. The majority of the participants were female (73%) and Caucasian (87%). On average, participants were married, were about 40 years of age, had mean household income levels between $50,000 and $60,000 per household, had two children under the age of 18 living in their household, and had some college education. This sample of participants was comparable to the Chicago and Denver 2000 Census populations (U.S. Census Bureau, 2000). However, a higher percentage of female respondents participated than was represented in the general populations of these two cities. This was desirable because females tend to be the primary food shoppers in most households. Slightly more panelists participated in Chicago (141 participants) than in Denver (132 participants).

Estimated parameters for the random-effects ordered probit model for equation (1) and the GLS parameter estimates for equation (2) are presented in table 2. Greene (1997) has cautioned that care be taken in interpreting the coefficients obtained from an ordered probit model. Therefore, the discussion is limited to the signs of the statistically significant variables, while the magnitude of the coefficients is not discussed. However, both magnitudes and signs are relevant for the GLS estimates for the regression on panelists' WTP values (\textit{BID}) from equation (2).

The \%\textit{FAT} variable had a positive and significant impact on the probability of higher consumer palatability ratings for the attributes of juiciness (\textit{JUICINESS}), tenderness (\textit{TENDERNESS}), and overall acceptability (\textit{OVERALL}) (table 2). The \%\textit{FAT} variable did not significantly impact the \textit{FLAVOR} rating, nor did it significantly influence panelists' WTP values (\textit{BID}). The level of tenderness, as measured by the WBSF, had a significant impact on the probability of higher ratings for all palatability ratings: flavor, juiciness, tenderness, and overall acceptability. Tenderness level (\textit{WBSF}) also significantly increased consumers' WTP (\textit{BID}). As a steak became less tender, WBSF increased, and consumers' ratings for steak flavor, juiciness, tenderness, and overall acceptability all tended to decline, as did their WTP for the steak. A one-kilogram increase in WBSF led to a significant decrease in WTP of $0.24/pound for steak. It is interesting to point out that the coefficient for the variable used to measure marbling
The primary determinant of USDA Quality grades, was not significant, while the coefficient for the tenderness variable (WBSF) was significant at the 99% level of confidence in the equation for BID.

The estimated coefficient for the variable AUSTRALIA was significant and negative in all equations for palatability ratings (first four columns of table 2). If a steak was Australian grass-fed beef, then there was a greater probability for lower ratings for flavor, juiciness, tenderness, and overall acceptability. The significance and negative sign of the coefficient on the AUSTRALIA variable in the BID equation appears to indicate that these lower palatability ratings translated into lower WTP values for the Australian steaks than for the domestic steaks. For example, the expected average bid for Australian steaks would be $0.97/pound lower than comparable U.S. steaks. The coefficient for CANADA in the BID equation was significant and negative. The average estimated WTP (BID) for Canadian steaks is $0.28/pound less than for U.S. steaks.

Dry-aged steaks (DRYAGE) had a greater probability of being rated lower for flavor and overall acceptability than steaks that were not dry-aged. Additionally, BID was
$0.33/pound lower, on average, than for a similar steak that had not been dry-aged. The negative relationship between dry-aging and palatability ratings and the fact that panelists' WTP for dry-aged steaks were on average less than WTP for steaks that had not been aged is interesting considering the substantial, additional costs related to dry-aging beef compared to wet-aging or no aging. Steaks that were wet-aged did not significantly impact panelists' palatability ratings. Furthermore, bids for wet-aged (WETAGE) steaks were not significantly different from steaks that had no additional aging.

The variables $P_2$, $P_3$, and $P_4$ were used to determine if panelists' bids in the respective periods were significantly different than the first bidding period. As noted previously, significant and negative signs on these variables may indicate possible demand reduction, as panelists learned they were "potential winners" of steaks in previous auctions and became satiated. Alternatively, positive signs on the bidding period coefficients may indicate that participants incorporated market feedback into their bid values when simultaneously submitting bids for pairs of steak samples. The coefficients for all three variables are positive in the BID equation; however, only the coefficients on $P_3$ and $P_4$ were significant. Panelists' bid values were significantly higher for the third and fourth pair of steaks sampled and auctioned compared to the first pair of steaks auctioned. Therefore, one interpretation of this significant and positive result may be that demand reduction did not have a significant impact on bids. Another potential explanation for the increase in bid values in bidding periods three and four is that certain participants in a panel may have become more aggressive with their bids if they had not previously been a "potential winner" of an auction and if they were desirous of winning. However, because these coefficient values are the net effect of bidding period on bid values, we cannot actually determine if demand reduction occurred. This inconclusive result is because potential decreases in bids associated with demand reduction may have been offset by potential increases in bids due to bidder affiliation caused by market feedback information.

Equation (3) was estimated by using OLS to determine how both beef quality and production attributes and panelist demographics impact WTP values. Table 3 presents the estimated coefficients of equation (3). The signs and significance of the coefficients on the beef attribute variables %FAT, WBSF, AUSTRALIA, DRYAGE, and WETAGE are similar to those estimated in equation (2). As in the estimation of equation (2), the sign of the coefficient on the CANADA variable was also negative; however, it became significant at the 95% level of confidence in explaining BID when demographic variables were added into the estimation.

In addition to the beef attribute variables, a number of demographic variables were significant in predicting participants' WTP measured by BID (table 3). The coefficients on the demographic variables BEEFEAT and MALE were significant and positive. Male panelists who were heavy beef eaters (BEEFEAT) had higher bid values than panelists who ate beef less frequently and who were female. The coefficients on the variables AGE and KIDS were negative and significant. Thus, older panelists and panelists with children in their household tended to bid less for steaks than did younger panelists and panelists with no children present in the household.

Similar to equation (2), experimental design variables were included in the estimation of equation (3) to determine the impact of location, panel size, and auction order on WTP values. The coefficient on the LOCATION variable was not significant, but the number
Table 3. OLS Regression Results Using Demographic Variables to Predict BID: Equation (3)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.307</td>
<td>0.812</td>
</tr>
<tr>
<td>%FAT</td>
<td>-0.027</td>
<td>0.132</td>
</tr>
<tr>
<td>WBSF</td>
<td>-0.440**</td>
<td>0.103</td>
</tr>
<tr>
<td>AUSTRALIA</td>
<td>-1.086***</td>
<td>0.206</td>
</tr>
<tr>
<td>CANADA</td>
<td>-0.385**</td>
<td>0.197</td>
</tr>
<tr>
<td>DRYAGE</td>
<td>-0.352**</td>
<td>0.166</td>
</tr>
<tr>
<td>WETAGE</td>
<td>-0.032</td>
<td>0.188</td>
</tr>
<tr>
<td>BEEFEAT</td>
<td>0.162**</td>
<td>0.067</td>
</tr>
<tr>
<td>NONCAUCASIAN</td>
<td>0.190</td>
<td>0.164</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Coefficient</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>0.271**</td>
<td>0.132</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.003***</td>
<td>0.001</td>
</tr>
<tr>
<td>INCOME</td>
<td>-0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>KIDS</td>
<td>-0.411***</td>
<td>0.119</td>
</tr>
<tr>
<td>LOCATION</td>
<td>0.067</td>
<td>0.120</td>
</tr>
<tr>
<td>PANELSIZE</td>
<td>0.290***</td>
<td>0.053</td>
</tr>
<tr>
<td>P2</td>
<td>-0.013</td>
<td>0.158</td>
</tr>
<tr>
<td>P3</td>
<td>0.302</td>
<td>0.157</td>
</tr>
<tr>
<td>P4</td>
<td>0.397**</td>
<td>0.159</td>
</tr>
</tbody>
</table>

Adjusted $R^2 = 0.530$

Notes: Double and triple asterisks (*) denote statistical significance at least at the $\alpha = 0.05$ and 0.01 levels, respectively; $N = 2,184$.

Table 4. Average Bids for Steaks of Various USDA Quality Grades and Tenderness Levels

<table>
<thead>
<tr>
<th>Tenderness Category</th>
<th>Select</th>
<th>Low Choice</th>
<th>Upper 2/3 Choice</th>
<th>Prime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tough</td>
<td>$2.24</td>
<td>$2.23</td>
<td>$2.49</td>
<td>$2.58</td>
</tr>
<tr>
<td>Moderate</td>
<td>$1.98</td>
<td>$3.09</td>
<td>$3.03</td>
<td>$3.09</td>
</tr>
<tr>
<td>Tender</td>
<td>NA</td>
<td>$3.19</td>
<td>$3.39</td>
<td>$3.17</td>
</tr>
</tbody>
</table>

Notes: Means in the "Select" column and in the Tough row are significantly lower than the means in the other columns and rows, respectively, based on the Duncan test at the $\alpha = 0.05$ level; $N = 192$ steaks. Values appearing in boldface/italics are statistically larger than the other values in the table.

of panelists in each panel (PANELSIZE) did impact the bid levels. Increasing panel size by one participant increased bid values by $0.29/pound. Umberger and Feuz (2004) also found a positive relationship between panel size and bid in their experiment. As observed from table 3, when demographic variables were added, the auction time period had only a minimal impact on bid values. Only the fourth, or last, auction time period (P4) had a statistically significant and positive impact on bid values.

Further analysis of the impact of marbling and of tenderness on bids was conducted, and results are reported in table 4. All 192 steak samples used in the previous analysis were classified by USDA quality grade (based on marbling level as measured by %FAT) and by tenderness level (based on WBSF values). Three levels of tenderness were defined: (a) Tough, WBSF > 4 kilograms; (b) Moderate, 3 kilograms ≤ WBSF ≤ 4 kilograms; and (c) Tender, WBSF < 3 kilograms. Huffman et al. (1996) established a threshold value for WBSF of 4.1 kilograms or less for an acceptable level of tenderness, and Wheeler, Shackelford, and Koohmaraie (1997) found that consumers rated steaks with WBSF values of 3.0 kilograms to 7.9 kilograms as unacceptable.

Mean bids were calculated for each classification and tests for differences among them were conducted using the Duncan test in SAS Proc GLM (SAS Institute, Inc., 1999). The average bids for the Tough category of steak, regardless of USDA quality grade, were significantly lower than for the other categories of tenderness. This would
tend to substantiate the threshold WBSF value of 4.1 kilograms or less proposed by Huffman et al. (1996). Bids for steak samples with sufficient marbling to only qualify for the USDA Select grade, regardless of tenderness, were significantly less than bids for the other USDA quality grades. There were no statistical differences at the 5% level in the mean bids for steaks in the Moderate and Tender categories with sufficient marbling to grade USDA low Choice or above.

**Summary and Implications**

In June and July of 2002, consumers from Denver and Chicago were randomly screened and selected by telephone interview to participate in a study of beef quality. Consumer panelists rated steaks for juiciness, tenderness, flavor, and overall acceptability. Panelists also participated in a random nth-price, sealed-bid auction for each steak sample to determine their willingness to pay. For each steak sample, objective measures were taken to determine the steak's marbling and tenderness, and each steak was classified as to country of origin (United States, Australia, or Canada) and aging method (no aging, dry-aging, wet-aging).

The research project was undertaken to determine the beef palatability characteristics, in addition to marbling and tenderness, that are most important to consumers' overall satisfaction and WTP for beef steaks. The marbling level of a steak, as measured by the percent fat in a steak, had a positive and significant impact on consumers' palatability ratings for juiciness, tenderness, and overall acceptability. The level of tenderness, as measured by WBSF, had a significant impact on all palatability ratings (flavor, juiciness, tenderness, and overall acceptability), with a more tender steak receiving higher ratings. Marbling, the primary determinant of USDA beef quality grades, did not significantly impact panelists' WTP values. However, tenderness, a beef quality attribute not currently used as a measure to determine USDA quality grades, did significantly impact panelists' WTP. More tender steaks with lower WBSF values significantly increased panelists' auction bids.

In addition to marbling and tenderness, differences in production practices between countries and aging methods also impacted participants' palatability ratings and WTP. Australian grass-fed beef steaks had a higher probability of receiving lower palatability ratings for flavor, juiciness, tenderness, and overall acceptability. Furthermore, the bids for Australian grass-fed steaks were substantially lower ($0.97/pound) on average than for comparable U.S. steaks. Ratings for Canadian barley-fed steaks were not significantly different than for U.S. steaks, but bid values for Canadian barley-fed steaks were $0.28 lower on average than for comparable U.S. steaks. Steaks that were dry-aged had a higher probability of being rated lower for flavor and overall acceptability than steaks that had not been aged, and the average bid value was $0.33/pound lower. This result does not support dry-aging beef when one considers the additional costs associated with this aging method. However, these are average values, and it is likely that some consumers prefer the taste and are willing to pay a premium for dry-aged beef. Steaks that were wet-aged did not have significantly different palatability ratings or bids than steaks that were not wet-aged.

The findings regarding differences between WTP for U.S. versus Australian and Canadian beef may raise questions related to the ongoing debate on country-of-origin
It is important to point out that this research was not undertaken to address issues related to COOL for beef. However, the results show that American consumers can taste a difference between steaks originating either from the United States, Australia, or Canada. On average, panelists in the experiment preferred the flavor of U.S. steaks to the flavor of Australian steaks. Furthermore, they were willing to pay more for their preferred flavor preference. These findings do not imply all panelists preferred the U.S. steaks and were willing to pay a higher price for the domestic steaks than for the imported steaks. Sitz et al. (2004) reported that 19% and 29% of panelists in a taste experiment preferred the taste of Australian grass-fed steaks and Canadian barley-fed steaks to U.S. corn-fed steaks, respectively. These consumer segments were willing to pay more for their preferred imported steak. Given that the United States imports about 10% of its total beef for consumption, there are likely viable niche markets for imported steaks, particularly if marketers of imported steaks emphasize potentially desirable attributes, such as being "grass-fed."

In addition to beef quality and production attributes, demographic variables such as gender, age, and the presence of children in the panelist’s household all had significant impacts on WTP. Panelists who consumed beef more frequently had higher WTP levels than panelists who consumed beef less frequently. Similar to other studies, experimental auction variables also had a significant impact on bids. The actual number of panelists present in any given panel influenced panelists' bids, and the bidding period also influenced the bids.

All steaks evaluated by consumers were classified by USDA quality grade and by tenderness level. The average bids for the Tough category of steak, regardless of quality grade, were significantly lower than for the other categories of tenderness. Bids for steak samples with only sufficient marbling to qualify for the USDA Select grade, regardless of tenderness, were significantly less than bids for the other higher USDA quality grades. On average, bids for steak samples in the Moderate and Tender categories of tenderness, but with sufficient marbling to grade USDA low Choice or above, were all equal.

The results of this analysis concur with previous research findings that tenderness is a critical beef steak palatability factor impacting consumers' satisfaction and WTP for beef steaks. However, consumers in this study equally valued steaks that had a WBSF of four kilograms or less and which were USDA low Choice or higher. Participants were willing to pay more, on average, for steaks that were graded USDA Choice or higher compared to steaks graded USDA Select, but marbling did not significantly impact bids in the regression analysis.

---

8 Title X, Section 10816 of the Farm Security and Rural Investment Act of 2002 (the 2002 Farm Bill) amended the Agricultural Marketing Act of 1946 and required retailers to inform consumers of the country of origin of agricultural commodities such as ground meat and muscle cuts from beef, lamb, and pork. According to the 2002 Farm Bill's COOL provision guidelines, for a beef product to be labeled as a "Product of U.S.A.," the beef animal must be born, raised, and processed in the United States (USDA/Agricultural Marketing Service, 2003). The COOL program is currently a voluntary program. In January 2004, President Bush signed Public Law 208-199 postponing implementation of mandatory COOL for all commodities except wild and farm-raised fish and shellfish from September 30, 2004 until September 30, 2006. Proponents of COOL are continually lobbying for reinstatement of mandatory COOL. See Umberger et al. (2002) and Loureiro and Umberger (2003) for a further discussion of the debate on COOL.

9 Almost all of the beef imported by the United States from Australia is "grass fed." Most of the Australian beef exported as "grain fed" (mostly barley and wheat) is tied to Japan (Japanese contracts). Also, much of the initial investment in Australian feedlots came from Japan, but cattle feedlots are still a relatively small part of the Australia beef industry. Therefore, the bulk of the beef that the U.S. imports from Australia is a "grass-fed," boneless product which is used for hamburger, manufactured products, and processed meats (Robb, 2004).
In addition to being affected by tenderness, consumer palatability ratings and WTP were influenced by country of origin (United States, Australia, and Canada) and by aging methods (no aging, dry-aging, wet-aging). As the beef industry continues to mature, additional branding of beef with specific flavor attributes, and with a critical level of tenderness, may be a viable strategy to further segment the market and to increase consumer demand for beef.

[Received November 2003; final revision received August 2004.]

References


Robb, J. Director, Livestock Marketing Information Center, Lakewood, CO. Personal communication, May 25, 2004.


