



A comparison of two new take-away strategies and their relation to rating and ranking of extrinsic properties of dry cured ham

Tormod Næs^{a,c,*}, Erminio Monteleone^b, Anne Segtnan^a, Margrethe Hersleth^a

^a Nofima, Oslovegen 1, 1430 As, Norway

^b Dipartimento Delle Biotechnologie Agrarie, Via Donizetti, University of Firenze, Italy

^c University of Copenhagen, Dept. Food Science, 1958 Fredriksberg, Copenhagen, Denmark

ARTICLE INFO

Article history:

Received 20 October 2011

Received in revised form 30 May 2012

Accepted 5 June 2012

Available online 23 June 2012

Keywords:

Rating

Ranking

Take-away choice

Individual differences

Cured ham

Extrinsic properties

ABSTRACT

This paper is a study of the relationship between rating, ranking and take-away choices of dry cured ham. Extrinsic attributes related to country of origin, ageing time and price are considered and we study average population results as well as individual differences. Two tests with different sample size and different take-away strategies are explored. The results show that the ranking and rating data provide similar information in both tests. From an average point of view the stated results have also similarities with the take-away results, but with some notable differences. Relatively large individual differences in ranking of the products were observed in the two tests. Another finding in this study is that information about type of meal has a strong effect on the ranks of the products for all responses.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In experimental consumer studies one always has to make a decision about which methodology to use and about which responses to measure. One important approach is to use a so-called rating test based on asking the consumers to rate the products according to their degree of liking or intention to buy (see for instance, Lawless and Heymann, 2010; Næs, Brockhoff, and Tomic, 2010). When concerns statistical analysis of the results, such rating responses have a number of advantages since standard regression analysis and analysis of variance (ANOVA) can be used. Another common approach is to ask consumers to rank the products (Lawless & Heymann, 2010) according to their degree of liking or purchase intention. Arguments that have been put forward in favour of product ranking are that it is easier for the consumer to rank the products than to rate them and that ranking is closer to what happens in real buying situations. A third alternative is to use choice tests in which the consumers choose the most preferred sample from a number of choice sets (Gustafsson, Herrmann, & Huber, 2003). The arguments for this are similar to the arguments for the ranking tests. Usually generalisations of logistic regression, either mixed logit or probit analysis (Jaeger & Rose, 2008; Ortuzar

& de Dios, 2010; Train, 2009) are used for analysing the data. These methods are somewhat less easily available in software and also slightly more difficult to handle and interpret than regular regression and ANOVA. Another aspect is that they are not easily applicable in cases where only one or a few alternatives are present at the same time, which happens in for instance certain context studies (Hersleth, Mevik, Næs, & Guinard, 2003).

There exist a number of approaches that monitor more directly what people actually do when they have a choice. The most well-known tests are probably the various types of experimental auctions (Lange, Martin, Chabanet, Combris, & Issanchou, 2002) and studies based on monitoring real choices made by the consumer in restaurants and canteens (see for instance Caporale, Policastro, Tuorila, & Monteleone, 2009; King, Weber, Meiselman, & Lv, 2004; Rosas-Nexticapa, Angulu, & O'Mahony, 2005). Other approaches based on for instance various take-away strategies (Mustonen, Hissa, Huutilainen, Miettinen, & Tuorila, 2007; Weiss, O'Mahony, & Wichchukit, 2010; Wichchukit & O'Mahony, 2011) have also been suggested. At first glance an even better strategy would be to monitor real behaviour in a store, but such studies are difficult to conduct in practice and they also have some limitations. For instance, they cannot be used in many standard product development situations since product launch has to be done prior to monitoring the behaviour. In such cases, one will therefore have to rely on the information obtained in studies based on stated hedonic liking, purchase intent or actual take-away choice, with the

* Corresponding author at: Nofima, Oslovegen 1, 1430 As, Norway. Tel.: +47 64 97 0165; fax: +47 64 97 0333.

E-mail address: tormod.naes@matforsk.no (T. Næs).

latter being presumably the most realistic alternative. However, for studies concerning more basic factors for food choice, monitoring of real behaviour is of high importance.

For qualitative methodologies such as focus groups and repertory grid (RGM, see for instance Gains, 1994; Thomson & McEwan, 1988), information about the eating situation is frequently used for providing the most reliable data. There exists, however, also quantitatively oriented literature that demonstrates that consumers' planned use of a product may influence their response pattern (Weiss et al., 2010; Wichchukit & O'Mahony, 2011). This indicates that a possible refinement of the traditional experimental tests is to add information about intended use of the product. Hein, Jaeger, Carr, and Delahunty (2008) used an approach where they asked people to think of and write down a typical situation for consumption of the product and to have this in mind during tasting. Another possibility is to provide additional information about an eating situation to the consumer by means of written scenarios and/or pictures. Testing in different real contexts can also be done as described in for instance, Hersleth et al. (2003); Hersleth, Ueland, Allain, and Næs (2005), but this approach is more resource demanding. Home use testing is another way of providing more realistic contexts. The drawback with home use tests, however, is lack of control with the test procedure and thus more noise in the data. One can argue that all these tests implicitly incorporate appropriateness (Schutz & Martens, 2001) into the consumers' scoring pattern and that they therefore can be considered as combinations of tests of appropriateness and liking (or purchase intention).

The main aim of this paper is to propose two new take-away strategies and compare them with information obtained from stated rating and ranking of the same products. The experiment is based on extrinsic properties of dry cured ham related to country of origin (Spain, Italy and Norway), price and ageing. The consumer study is done in Norway where hams from all the three countries are available in the market and consumed in reasonably large quantities. In one of the take-away strategies the consumers are given money that they can use for purchase of products when leaving the test room while in the other case, no money is involved. As a part of the input to the consumer, two different meal descriptions are given. Special emphasis will be put on the comparison between stated rating and stated ranks as predictors of take-away choice and on how the results are influenced by the meal descriptions given.

2. Statistical methodology

2.1. Linear regression for stated rating data

The most used methods for analysing rating data are linear regression and analysis of variance (ANOVA). Since the focus of this paper is on differences between measurements and methodology rather than on effects of factors involved in the design of the products, main emphasis will be given to ANOVA with the six different products considered as different levels of one single factor. The model used can then be written as

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (1)$$

where the α_i 's are the product effects and the β_j 's are the consumer effects. In the results section we will present average product differences in terms of LS-means (i.e., least square means) and p -values. The ANOVA will here be used for each of the two meals descriptions separately.

One major advantage of this approach is that all results are presented in the same units as the original measurements and are thus easy to interpret.

2.2. Probit analysis of stated rank data

The most used way of analysing rank data is to use rank order logistic regression, its more general rank order probit regression (Gustafsson et al., 2003) or the mixed logit regression (Train, 2009). All these methods are extensions of standard logistic regression models.

In this paper we will use the probit regression model which is (like the other methods mentioned), based on the linear utility (Gustafsson et al., 2003) model

$$U_{ij} = \alpha_i + e_{ij} \quad (2)$$

where the index j represents consumer and i represents product. As above α_i represents the product effects and the e 's the random errors. The main difference between this model and model (1) is that here the dependent variables, the U 's, are not observable. For the probit model the distribution of the e_{ij} is assumed to be multinormal with expectation equal to zero and covariance matrix equal to Σ . As can also be noted, the additive consumer effect used in model (1) is not present. The reason for this is that a constant can be added to a utility model without changing the distribution of the ranking pattern. Since we in our case have independent parameters α_i , it is, however, necessary to fix one of them in order to get a unique solution (the so-called base alternative). The ranking pattern is also independent of the scaling or the variability of the random noise. In order to handle these aspects one can without loss of generality assume that the level for product 1 (the so-called base level) is equal to 0 and that one of the variances in Σ is equal to a constant. For further properties of the method we refer to Gustafsson et al. (2003).

Probit analysis is less developed than standard regression and ANOVA and software is not so easily available as for ANOVA. As can be noted, the regression coefficients in model (2) have no obvious relation to a directly interpretable scale and therefore comparisons with rating data must be done by looking at relative differences and p -values.

2.3. Indices of performance for individual differences among consumers

In the present study, also individual differences related to consistency over rating, ranking and take-away choices will be considered. In the following we describe some indices that will be used for validating and comparing ranking and rating strategies for the different individuals.

2.3.1. Consistency over data collection methods (CoM)

This is a statistic which measures the similarity between stated ranks and the ranks based on rating. This index is defined as the Spearman's rank correlation for the ranking and rating data. The index must be calculated for each consumer and each meal (see experimental section below) separately.

2.3.2. Consistence of first choice (CoF)

This index is defined as the percentage of consumers having the same sample ranked as number one both in the stated rank test and in the take-away test. This means that the consumer group is first split according to which meal they have in mind for the take-away products (see experimental Section 3) and then the product with the highest rank is compared to the stated rank of the same product in the corresponding meal situation. The index may be computed for stated ranks directly and for ranks obtained by the stated rating responses. In order to investigate the effect of the meal description, one can also compute the CoF by switching the meal situation, i.e., by comparing the take-away products with

the rank of the corresponding product in the opposite meal situation (i.e., opposite to the one indicated in the take-away test).

2.3.3. Consistency of ranks (CoRa)

This index is based on the correlations between the ranks in the take-away test and the stated ranks (or the ranks obtained by the stated rating). This is essentially the same as CoM, but refers to relations between take-away results and stated results. As above, the index is calculated for each of the two meal situations as given in the take-away response

Note that while CoF only considers the sample which is selected first, i.e., the sample which is most liked, the CoRa measures the correlation between all responses in the take-away test and the stated ranks, i.e., it gives a measure of the total agreement. Both indices are relevant, but from different perspectives (see discussion).

Histograms will be provided for the first and the last of these indices.

3. Experimental protocols

In both consumer studies considered here, the recruitment criteria, the products tested and the information about meal situations given to the consumer were the same.

Consumer recruitment criteria

Consumers were recruited from local clubs and associations in the community, and the criterion used for participation was that they like dry cured ham, buy and eat dry cured ham regularly and had eaten Norwegian ham plus either Italian or Spanish ham during the last 6 months.

Products used in the study

Six packages of dry cured ham with the following information were evaluated by all the consumers:

Norwegian products:

4 months ageing, 45 NOK (5.8 Euro) per 100 gram (N4)

15 months ageing, 65 NOK (8.4 Euro) per 100 grams (N15)

Spanish products:

9 months ageing, 55 NOK (7.1 Euro) per 100 grams (S9)

18 months ageing, 75 NOK (9.7 Euro) per 100 grams. (S18)

Italian products:

15 months ageing, 65 NOK (8.4 Euro) per 100 grams (I15)

24 months ageing, 85 NOK (11.0 Euro) per 100 grams (I24)

The information was given on white labels, the origin was indicated by flag and the prices are similar to real prices in the Norwegian market. Note that price and ageing are confounded and that the levels are different for each country.

Information about meals

The consumers were asked to rate and rank the products according to purchase intent for two different situations of use (here called meals).

Meal 1: a traditional meals (in Norway) consisting of dry cured hams, scrambled eggs and/or potato salad.

Meal 2: a more novel meals (in Norway) with several small dishes, i.e., “finger food” or “tapas”.

As attention should be given to extrinsic attributes of the packages, the samples inside all six packages consisted of the same (neutral/typical) dry cured ham.

A moderator did the interviews one by one and gave the following instruction to the consumers: “Please imagine that you are in a

shop and that you are going to buy a package of dry cured ham, rank these packages of hams in order of what would be your choice of ham for 1) a meal with dry cured hams and scrambled eggs” and 2) a meal with several small dishes i.e., “finger food” or a tapas-meals”.

3.1. Consumer test 1 (32 consumers)

This test was done as a part of a larger study which focused on the Repertory Grid Methodology (RGM) for exploring consumers’ sensory vocabulary of dry cured ham (Hersleth et al., submitted for publication). Before the RGM-session, the consumers were asked to rank and to rate six samples of dry cured hams according to their probability of buying, based on the extrinsic attributes only.

3.1.1. Test 1 – stated rating and ranking

Consumer group. 32 consumers from Eastern Norway (Oslo region) in the age group between 30 and 60 years old participated in consumer test 1.

Test protocol. The ranking was done first for meal 1, and then for meal 2 and the interviewer wrote down the ranks. In the following rating test the same six packages of dry cured ham were presented to the consumers, they were asked to rate the probability of buying each package on a scale between 1 (no probability) and 9 (very high probability). The rating was done first for meal 1 and then for meal 2 with three different orders of samples, i.e., the consumer group was split in three (10 consumers for each order) and each group had its own order.

3.1.2. Test 1 – take-away

In test 1 each consumer was at the point of leaving the test location given 200 NOK (ca. 25 Euro) and asked to buy two packages of ham (with indication of priority 1 and priority 2). They could keep the remaining money after the purchase was done. The consumers were also asked to indicate which kind of meal they had in mind when buying the two packages.

3.2. Consumer test 2 – (120 consumers)

Consumer test 2 was a part of a larger study which consisted of a blind hedonic rating of the same six samples of dry cured ham for the same two meals described above and a ranking and rating test of the same six packages based on the extrinsic attributes only. In this test both ranking and rating was done at the computer and was based on pictures of the packages that were used in test 1. In addition a take-away test was performed (described in Section 3.2.2).

3.2.1. Test 2 – stated rating and ranking

Consumer group. In this case 120 consumers from Eastern Norway (Oslo region) in the age group between 30 and 60 years participated in the consumer test.

Test protocol. The consumer test was run as a Central Location Test (CLT). The test was run over two days and the consumer could arrive any day and time between 3 pm and 7 pm. The test was divided into the following sessions:

- (1) blind testing meal 1, 10 min break,
- (2) blind testing meal 2,
- (3) extrinsic ranking meal 1,
- (4) extrinsic ranking meal 2,
- (5) extrinsic rating meal 1 and finally
- (6) extrinsic rating meal 2.

All samples were presented according to a cyclic design balanced for order and carry-over effects (MacFie, Bratchell, Greenhoff, & Vallis, 1989). The order of the meals was different for the two days; on day 1 meals 1 was presented first, on day 2 meal 2 was presented first. This means that for day 2, test 3 and 4 changed order and test 5 and 6 changed order.

3.2.2. Test 2 – take-away

In this case each consumer received an e-mail two days after the consumer test. They were explained that there were leftovers from the test and that they could come and pick up packages of dry cured ham at specific days and times. They were asked to respond with an order of priority (rank) for choice of all the six samples and an indication of which type of meal they had in mind.

4. Results

In this section we will first compare the stated ratings and stated ranks. This will be done for both consumer tests. In both cases we consider average results as well as individual differences using the proposed indices. The next step is to compare the two stated responses with the take-away ranks. Again, this will be done for both consumer tests and both for the averages as well as the individual differences. The main results will be discussed in a broader context in the discussion section.

4.1. Comparing stated rating and stated ranks of the products

4.1.1. Consumer test 1

Average population effects. The average ratings for probability of buying the six packages of dry cured ham are given in Fig. 1

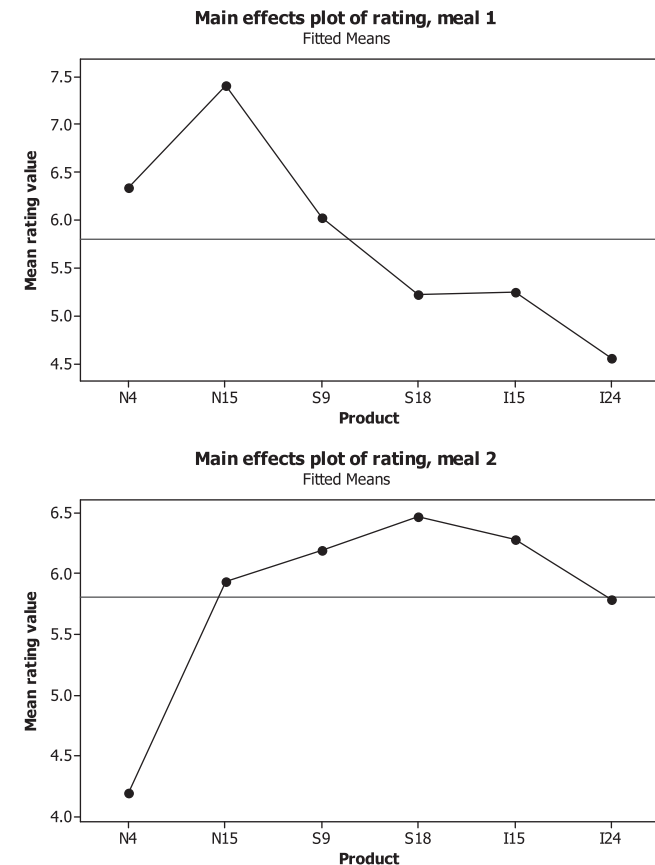


Fig. 1. Results for consumer test 1. Mean rating (LS-means) for the two meals. The abbreviations for the six samples can be found in Section 3. The standard error of the means is equal to 0.37 in both cases.

Table 1

Multiple comparison ANOVA for both consumer tests. The means and significant differences are presented for both meals alternatives and both tests. The meal abbreviations can be found in Section 3.

Meal 1			Meal 2		
Product	Mean	Grouping	Product	Mean	Grouping
<i>Consumer test 1. 32 consumers</i>					
N15	7.4	A	S18	6.5	A
N4	6.3	AB	I15	6.3	A
S9	6.0	ABC	S9	6.2	A
I15	5.3	BC	N15	5.9	A
S18	5.2	BC	I24	5.8	A
I24	4.6	C	N4	4.2	B
<i>Consumer test 2. 120 consumers</i>					
N15	7.5	A	N15	6.5	A
N4	6.1	B	S18	6.2	AB
S9	5.7	BC	S9	6.1	AB
I15	5.1	C	I15	5.9	AB
S18	5.0	C	I24	5.5	B
I24	4.2	D	N4	4.6	C

(standard error equal to 0.37 in both cases) and the multiple comparisons based on ANOVA are shown in Table 1. As can be seen from Fig. 1, the general structure of the average ratings is different for meal 1 and meal 2. For meal 1 (traditional meal), the Norwegian hams received higher ratings for probability of buying than the other hams, while for meal 2 (novel meal), the opposite is the case. The most striking difference between meals is found for N4 which has the clearly lowest rating in meal 2, but has the second highest rating for meal 1. The residuals from the ANOVA indicate no clear deviation from normality.

The ranking data were analysed by Box-plots and the probit analysis described above (results not shown). When comparing the medians in the Box-plot with the averages in Fig. 1 the general structure is the same. Also the coefficients and their p-values from the probit model confirm a similar pattern as for the rating data.

We can conclude that from a population average point of view, the two types of responses provide very similar information about the probability of buying for both meal alternatives. The meal description, however, had a strong impact on the results.

Individual differences. In Fig. 2 is shown the histogram for the CoM index. There were some minor differences between the two meals, but the overall pattern in the histograms was similar. The results show that there are several consumers with a relatively low consistency between the two data collection methods. Many consumers had a correlation value higher than 0.8, but many also had a value much lower than 0.8. Significance at 5% level is here at a

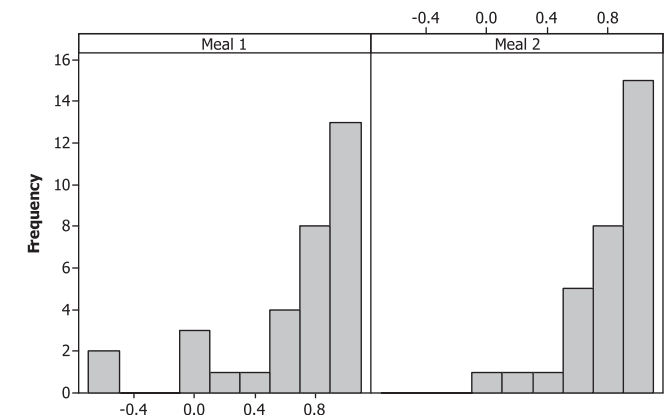


Fig. 2. Results for consumer test 1. Histogram of CoM (see Section 2) is presented for both meal alternatives (meal indicated at the top of the figure).

correlation value equal to 0.89. In other words, on an individual basis consumers are quite inconsistent between ranking and rating, but for the whole group the results for the two tests are very similar.

4.1.2. Consumer test 2

Average population effects. The average values from rating, including the multiple comparisons and the means, are presented in Table 1 (standard error for the means are 0.19 and 0.20 respectively for meal 1 and meal 2). Corresponding results from ranking are shown in Fig. 3. As can be seen, there are still substantial differences between the two meals. The general tendency is very similar to the results obtained in consumer test 1. The results from the probit analysis (results not shown) confirm the tendencies in Fig. 3.

The main conclusion is that for the averages the differences between the ranks and the ranks obtained from rating is moderate or small. In addition, the results show that type of meal affects both types of responses. The average results are comparable to the results in test 1, in particular for meal 1.

Individual differences. In order to study the individual differences between the two data collection methods, the histogram of CoM is considered (Fig. 4). Again the two meals gave very similar histograms (as in Fig. 2) and only the average results taken over the two meals (for each consumer) are shown. As can be seen from Fig. 4, the consistency is also here very different for the different consumers. But again, the tendencies for the two tests are the same when looking at results for the whole group of consumers.

4.2. Comparing stated rank and stated rating responses with the take-away ranks

Since rating and ranking give very similar results on an average level, the comparison on an average level with take-away choices will be more or less the same for both responses. This means that we can concentrate on one of them. On an individual level, however, the results for both the stated responses must be considered.

4.2.1. Consumer test 1

Since a full ranking of all samples in the take-away was not done here, we will here only give attention to the first selected sample in the take-away test. Note that this analysis will also be done for test 2. The main results for the take-away study are presented in Fig. 5 for the 23 respondents who indicated take-away for use in meal 1 (or both meal 1 and meal 2, or none) and for the 22 respondents who indicated use in meal 2 (or both meal 1 and meal 2, or none) for the take-away products. There is some overlap

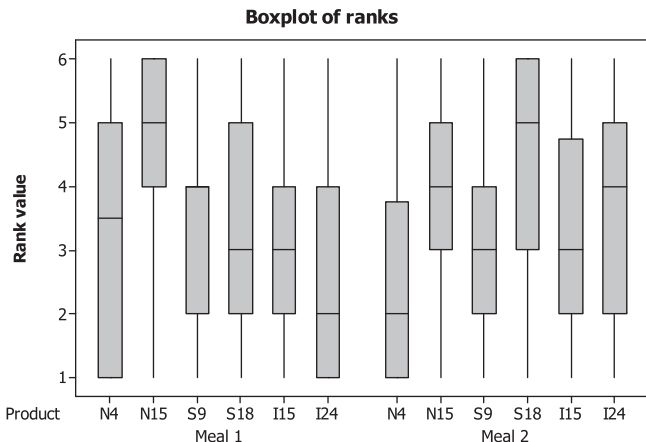


Fig. 3. Results for consumer test 2. Box plot of ranks for both meal alternatives. Sample abbreviations are defined in Section 3.

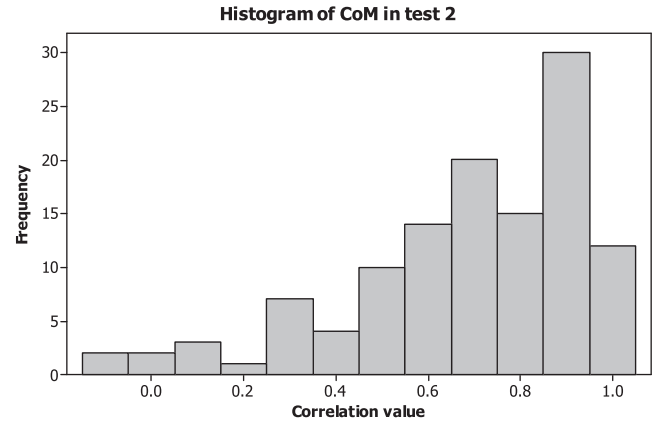


Fig. 4. Results for consumer test 2. Histogram of CoM (see Section 2) averaged over the two meal alternatives.

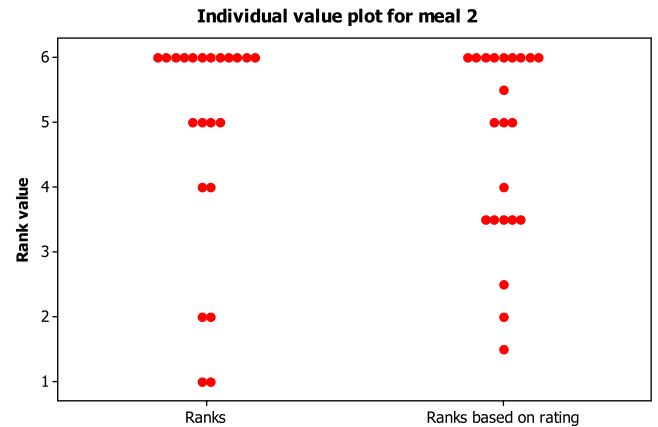
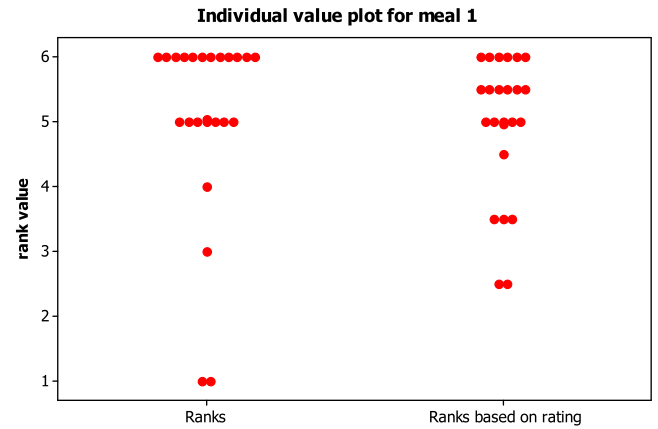


Fig. 5. Results for consumer test 1. Plots of the relation between the first selected product in the take-away test and the corresponding rank in the stated test. The plots show results for both the stated ranks and the ranks based on the stated ratings. The meal 1 results (defined by the take-away response) are presented in Fig. 5a (upper panel) and the meal 2 results are presented in Fig. 5b (lower panel).

between the two groups since those who respond with “both occasions” or “no special occasion” are incorporated in both groups. The plots show the stated rank for the product which is chosen first in the take-away test. For both meals, both stated ranks and the ranking (with ties) based on the stated rating are presented. As can be seen from the number of samples present at the different rank values, the results are comparable for stated rank and ranks based on stated rating, with a slightly better correspondence with

the take-away data for the stated rank data. But there were many ties in the rating data which means that the visual interpretation is slightly more complex.

Despite this similarity, the CoF is as low as 52% for stated ranks and ranks based on stated rating for meal 1 (i.e., stated meal 1 in the take-away test) and 55% and 45% respectively for meal 2 (i.e., stated meal 2 in the take-away test). For both meals the products with rank values of 6 and 5.5 (tie) were taken as the most preferred for the rating data. This means that when considering exact correspondence between take-away choice and maximum rank, the correspondence is quite poor.

If we, however, extend perspective and consider not only the sample with the highest rank, but also samples with stated rank 5, the correspondence is much better. For the stated ranks for meal 1, only 4 of the consumers stated a rank of less than 5 for the first choice in the take-away study. For meals 2 the corresponding number was 6.

For comparison we also looked at the frequency of first choice samples selected. For meal 1 (or both meal 1 and meal 2, or none, see above), the number of the choices for the 6 samples (N4, N15, S9, S18, I15, I24) were 2, 9, 3, 5, 0, 4 respectively. For meal 2 the corresponding numbers were 1, 7, 3, 8, 1, 2. As can be seen, the majority vote is for N15 and S18 in both cases (the values 9 and 5 for meal 1 and the values 7 and 8 for meal 2). For meal 1, N15 got the highest vote and for meal 2 the S18 got the highest. The results correspond well to the average stated rank results.

4.2.2. Consumer test 2

Average results. The averages of the ranks in the take away choice test as well as their 95% confidence intervals (based on the standard method of the mean \pm (approx) 2 times the standard deviation) are presented in Fig. 6. The split between the meal in Fig. 6 is done according to which meal was reported as situation of planned use given by consumers in their responses to the e-mail with the take-away invitation. When concerns the extremes and the relative ranking of countries, these results are in reasonable agreement with results from ranking and rating. The most striking difference is, however, the fact that in the take-away, the most aged (and with the highest price) hams were ranked systematically higher than the less aged for all three countries. This was not the case for probability of buying in the central location test.

If we as a contrast look at all the results without taking the meal situation into account, the most striking tendency is that Italian and Spanish hams are ranked higher than hams from Norway. Still there is a systematic difference between short aged and long aged

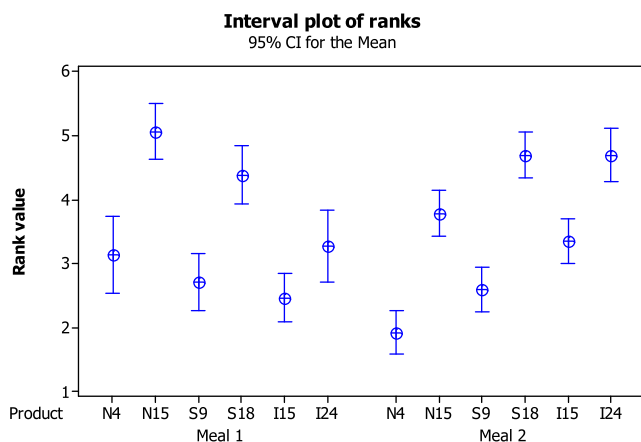


Fig. 6. Results for consumer test 2. Means and confidence intervals for the ranks in the take away choice test. There were 37 consumers in the meal 1 group and 58 consumers in the meal 2 group. Sample abbreviations are defined in Section 3.

as above. Comparing these results and the results in Fig. 6 with the stated results, we see that specification of a concrete meal gives an improved predictability with respect to differences between the countries of origin.

For comparison we also calculated the frequencies of the different samples in the take-away test. These were 1, 20, 1, 9, 0, 6 and 2, 9, 1, 19, 1, 26 for meal 1 and meal 2 respectively. As can be seen, these results correspond quite well to the average results, although with an even more extreme bias towards the expensive and long aged samples vs. the rest. The order of the countries is the same.

Individual differences related to first choice in the take-away test. The individual differences in consistency between take-away and stated ranking were first considered by simple plotting (Fig. 7) and by using the CoF. Fig. 7 shows the relation between the stated first rank and the first take-away choice (as in Fig. 5). This is done for two groups of consumers separately, for those who reported meal 2 and for those who reported meal 1 in take-away choice. For both groups, the comparison was first made with the same meal specification. As a contrast and for assessing the effect of the meal, the comparison was also made with the opposite meal specification. As can be seen, in both cases, in particular for meal 2 the change of meal had an effect. The correspondence is clearly better when

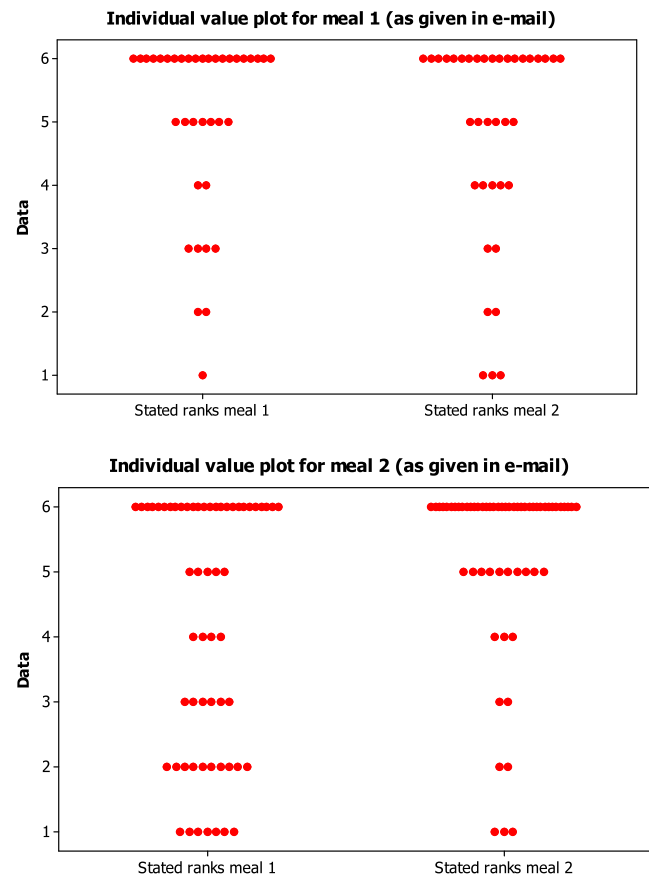


Fig. 7. a,b. Results for consumer test 2. Plot of the relation between first selected product in the take-away test and the corresponding rank given in the stated rank test. The two panels correspond to meal 1 and meal 2 as indicated in the e-mail response. The two columns in each panel correspond to comparisons with ranks in the same and opposite meals situation. For the upper plot, the number of points corresponding to the rank 6 is 21 and 19 respectively (left and right) and the number of points in total is 37 in each column. For the lower plot, the number of points corresponding to the rank 6 is 26 and 38 (left and right) and the total number of points is equal to 58 in each column.

comparing within the same meal, showing that specification of meal improved predictability.

For meal 1 the CoF was equal to 57% and 51% and for meal 2 the CoF was equal to 66% and 45% for the corresponding and opposite meals respectively. The consistency between choice and stated rank is better for meals 2. One can see that even though CoF is quite low, the correspondence is better if we extend attention also to stated rank = 5 (as done also for test 1). In both cases the number of products with lower rank than 4 was equal to 7.

Individual differences related to all hams. Histograms for CoRa for the two take-away groups are presented in Fig. 8. The results are comparable to the comparison between stated rating and stated ranking done above (Fig. 4). A large number of consumers have values higher than 0.8, but a substantial number have lower values and some even have negative values. For these latter consumers there is no relation at all between what they say and “what they do”

In order to study the difference in predictability of the take-away choices from the stated rating values, the same plot was made for ranks based on the rating values (not shown here). The results are quite comparable to the results for the ranks themselves and with large individual differences. But the similarity between take-away choice and stated rating is for meal 1 slightly better than for the stated ranks themselves, but for meal 2 the results are better for the true ranks themselves.

The relation between take-away choice and variables in the design of the samples. As could be seen from above, there was a discrepancy between the stated results and the real take-away choices. The next and final step is then to investigate this discrepancy in relation to the variables in the design of the samples. This is done by splitting the consumer group in two according to the value of the correlation between stated ranks and take-away ranks.

An interval plot for only the consumers with a significant value of CoRa (>0.89) is presented in Fig. 9. As can be seen, the plot is quite similar to the plot obtained for all consumers (Fig. 6).

For the consumers with a lower correlation value than 0.89 we computed the number of times the choice ranks were higher and lower than the stated rank. Special focus was given to the differences between the two groups of samples (N4, S9, I15) on one side and (N15, S18, I24) (i.e., short vs. long ageing, or low vs. high price). The number of responses with a higher rank in the e-mail (take-away) was equal to 13 and 37 respectively for the two groups of samples in meal 1 and 25 and 46 for meal 2. The corresponding average rank difference values were 1.58 and 1.81 for meal 1 and

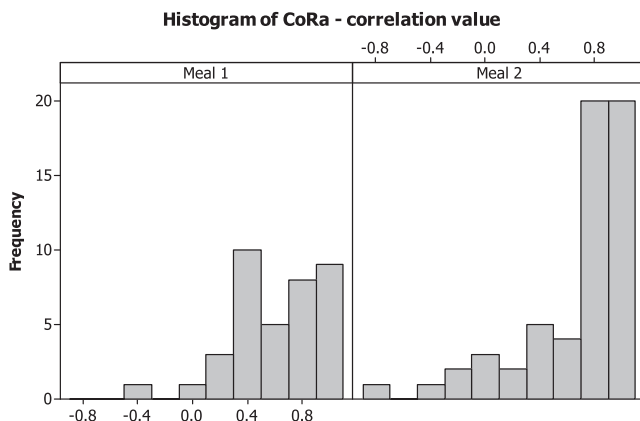


Fig. 8. Results for consumer test 2. Plots of CoRa (see Section 2) for both meals alternatives.

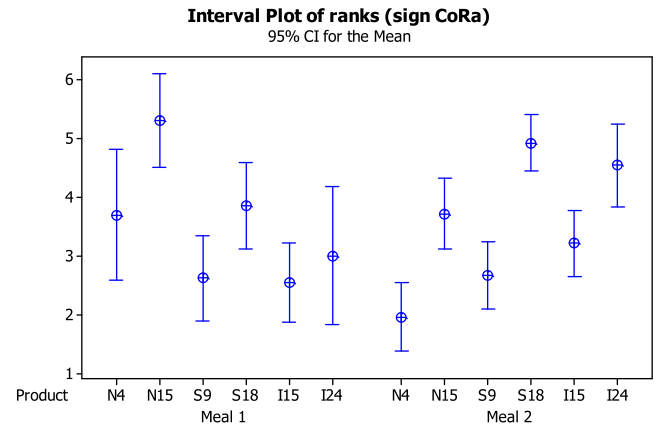


Fig. 9. Results for consumer test 2. Interval plots for those consumers with a significant relation between ranks from take-away and stated ranks. Both meal alternatives are presented. Sample abbreviations are defined in Section 3.

1.70 and 1.97 for meal 2. When looking at the negative differences, the average rank differences were -2.18 and -1.33 for meal 1 and -2.00 and -1.69 for meal 2. In other words, a large number of consumers increase their ranking for the most exclusive products (long ageing and high price) in the take-away test.

5. Discussion

5.1. Rating vs. ranking

The average rating and ranking results were quite similar for both meals. This indicates that at least for this type of products, it doesn't matter in which way the data are collected. Even though it may seem easier for the consumer to rank the products, the same information of the relative differences is obtained also for the stated rating data. It should, however, be mentioned that this may differ from case to case depending on for instance familiarity with the product, the size of the product differences and the type of consumer group in the study.

It should be mentioned that the stated rating and stated ranking tests were done in the same session in this study. It is therefore possible that this could have made the results more similar than they would have been in totally independent studies.

When it comes to individual differences, however, the differences between the two data collection methods are quite large for a large number of consumers, showing that the noise in the data may be considerable. An interesting finding is that the tendencies in both tests are the same when averages over the whole group are considered.

5.2. Comparison of the two take-away strategies

In this paper we focused on the relation between stated ranks and take-away choices. For test 1, only the first choice is considered, while in test 2 both the first choice and all take-away ranks were considered. Both average results as well as individual differences were considered (see also, Wichchukit and O'Mahony, 2011).

In test 2, the results from the stated ranks and the take-away ranks were comparable for the countries of origin, but with a higher rank for products with the longest ageing (and higher price) in the take-away test. For test 1 where an economic incentive was introduced to the consumers in the take-away test, this effect is not seen. This may indicate that for further development of take-away tests, one should consider methods which take the real economic

consequences into account as was the case for test 1. This aspect may deserve further study using also other types of products.

The CoF was, however, not larger than slightly above 50% in all cases. This shows that exact correspondence of first choices is not the same as a reasonable agreement of average rank values. This means that it is probably better to look at the overall pattern than just looking at one single number focusing only on exact correspondence. The plots associated with CoF also indicated that it may be useful to look at approximate correspondence, i.e., not only at the samples with the highest, but also those with the next to highest rank.

Specification of meal situation clearly improved prediction ability of the take-away choice.

5.3. The effect of conceptual meal descriptions

It can be noted that the differences in purchase probability for the two meal situations are quite large. The results also show that improved predictability of the take-away choice was obtained when information about the situation of use was given. This shows clearly that information about the eating situation is important in consumer studies and that it can be used to reduce noise in experiments of this type. This may also have implications for how products should be promoted and advertised.

The specification of the meal situation as a basis for the reported purchase probability corresponds to the introduction of appropriateness as a cognitive-contextual measure of food attitudes suggested by Cardello and Schutz (1996) and Schutz and Martens (2001). If the context for consumption of a product is removed in a test, the consumer may have less involvement with the product and accurate ratings may not be obtained. However, the importance of appropriateness is dependent of the type of food. For a product like dry cured ham (in Norway) this study showed that the context strongly affected the results.

5.4. Small vs. large consumer test

It is interesting to note that on an average level, the results from both tests are quite similar even though the smallest test is based on only 32 consumers. This may indicate that for some cases where the differences between the products are quite large, as in this case with clearly different extrinsic characteristics, it may be enough to use a smaller consumer sample than what is normally used. Therefore, even though there seems to be a general consensus in the literature that preferably more than 100 respondents are required, this aspect may warrant some further investigations. Computer simulations based on extracting subsets of consumers from larger tests can be used to provide useful information. Criteria used for recruitment is an aspect that should also be brought into such a discussion. The more specific such criteria are, the lower can probably the number of consumers in the study be.

5.5. Extrinsic vs. intrinsic attributes

It should be stressed that in this paper only extrinsic attributes are considered. There is reason to believe that the demonstrated effect of meals situation is also important for evaluation of intrinsic attributes in the products. A publication discussing similar aspects is Mueller and Szolnoki (2010).

5.6. The effect of ageing and country of origin

Since the main focus here has been on differences between measurement methods, the main focus in the presentation has been on using the six samples as individual levels of one experimental factor. This is also in harmony with the general methodo-

logical profile of the paper. Treating the origin and ageing/price as independent factors could have been done, but the results would have been somewhat difficult to interpret since ageing time and price are confounded and also since the levels are different for the different countries. Information of the relative contributions of the two factors is, however, implicitly present in the plots and in the multiple comparison tests provided. The effect of ageing/price as a separate factor is also given some attention when investigating reasons for inconsistency.

The two meals gave very different ranking and rating results. The general tendency is that Norwegian products are preferred for meal 1 and the Italian and Spanish products for meal 2.

For more information about importance of ageing time and origin of dry cured ham we refer to Hersleth, Lengard, Guerrero, Verbeke, and Næs (2011), which is more focused on the hams themselves rather than methodology

6. Conclusions

This paper has discussed results from two consumer studies done in Norway exploring effects of extrinsic properties of cured ham from three countries. Stated ranks, stated ratings as well as take-away ranks have been collected for all products. The two studies have different sample sizes and two different take-away strategies are tested. Except for that, the two studies are the same. One of the main findings of this paper is that stated ranking and stated rating gave very similar results when average results are considered. There was, however, a strong inconsistency at an individual level when the two measurements were compared. Another finding is that there is a reasonable correspondence between the average take-away ranks and stated ranks for origin of the products. For the traditional meals, the Norwegian hams got the highest score and rank, while for the more novel meals, the Spanish and Italian hams (plus the long aged Norwegian ham) tend to get the higher scores both in the stated tests and in the take-away tests. However, when consumers are able to make a choice without paying for products in a take-away test, they tend to choose the products with the highest price and the longest ageing time (highest quality image). A third finding is that information about meal situation has a strong impact on the ranks of the products and then on the relation between the different responses measured. Thus, appropriateness measures may give important information in consumer studies.

Acknowledgements

This study was funded by Foundation of Research Levy on Agricultural Products (Food Choice and Consumer Check projects). We would also like to thank Consorzio del Prosciutto crudo di Parma and Nortura, Norway for useful input at an early stage of the project. We thank the referees for useful comments leading to improvements of the paper.

References

- Caporale, G., Policastro, S., Tuorila, H., & Monteleone, E. (2009). Hedonic ratings and consumption of school lunch among preschool children. *Food quality and preference*, 20, 482–489.
- Cardello, A., & Schutz, H. G. (1996). Food appropriateness measures as an adjunct to consumer preference/acceptability evaluation. *Food quality and preference*, 7, 239–249.
- Gains, N. (1994). The repertory grid approach. In H. MacFie & D. M. H. Thomson (Eds.), *Measurement of food preference* (pp. 51–76). Gaithersburg: Aspen.
- Gustafsson, A., Herrmann, A., & Huber, F. (2003). *Conjoint measurement: Methods and applications*. Berlin: Springer.
- Hein, K. A., Jaeger, S. R., Carr, B. T., & Delahunty, C. M. (2008). Comparison of five common acceptance and preference methods. *Food quality and preference*, 19, 651–661.

- Hersleth, M., Mevik, B.-H., Næs, T., & Guinard, X. (2003). The use of robust design methods for analysing context effects. *Food quality and preference*, 14, 615–622.
- Hersleth, M., Ueland Allain, H., & Næs, T. (2005). Consumer acceptance of cheese, influence of different testing conditions. *Food quality and preference*, 16, 103–110.
- Hersleth, M., Lengard, V., Guerrero, L., Verbeke, W., & Næs, T. (2011). Norwegian consumers' acceptance of innovations for dry cured ham: Reduced salt content and new origin. *Food quality and preference*, 22, 31–41.
- Hersleth, M., Næs, T., Guerrero, L., Claret, A., Recchia, A., Dinnella, C. and Monteleone, E. Consumers' perception of dry cured ham – a cross cultural study based on repertory grid methodology in Italy, Norway and Spain. (submitted for publication).
- Jaeger, S. R., & Rose, J. M. (2008). Stated choice experimentation, contextual influences and food choice. A case study. *Food quality and preference*, 19, 539–564.
- King, S. C., Weber, A. J., Meiselman, H. L., & Lv, N. (2004). The effect of meal situation, social interaction, physical environment and choice on food acceptability. *Food quality and preference*, 15(7–8), 645–653.
- Lange, C., Martin, C., Chabanet, C., Combris, P., & Issanchou, S. (2002). Impact of the information provided to consumers on their willingness to pay for champagne: Comparison with hedonic scores. *Food quality and preference*, 13, 597–608.
- Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food principles and practices*. NY: Springer.
- Mustonen, S., Hissa, I., Huotilainen, A., Miettinen, S.-M., & Tuorila, H. (2007). Hedonic responses as predictors of food choice. flexibility and self-prediction. *Appetite*, 49, 159–168.
- MacFie, H. J. H., Bratchell, N., Greenhoff, K., & Vallis, L. V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of sensory studies*, 4, 129–148.
- Mueller, S., & Szolnoki, G. (2010). The relative influence of packaging, labelling, branding and sensory attributes on liking and purchase intent. Consumers different in their responsiveness. *Food quality and preference*, 21, 774–783.
- Næs, T., Brockhoff, P., & Tomic, O. (2010). *Statistics for sensory and consumer science*. Chichester, UK: John Wiley and sons.
- Ortuzar, J., & de Dios (2010). Estimating individual preferences with flexible discrete-choice-models. *Food quality and preference*, 21, 262–269.
- Rosas-Nexticapa, M., Angulu, O., & O'Mahony, M. (2005). How well does the 9-point scale predict purchase frequency. *Journal of sensory studies*, 20, 313–331.
- Schutz, H. G., & Martens, M. (2001). Appropriateness as a cognitive-contextual measure of food attitudes. In L. Frewer, E. Risvik, & H. Schifferstein (Eds.), *Food, People and Society* (pp. 247–266). Berlin Heidelberg: Springer-Verlag.
- Thomson, D. M. H., & McEwan, J. A. (1988). An application of the repertory grid method to investigate consumer perception of food. *Appetite*, 8, 1–14.
- Train, K. E. (2009). *Discrete choice methods with simulation*. NY, USA: Cambridge University press.
- Wichchukit, S., & O'Mahony, M. M. (2011). Liking, buying, choosing and take away preference testes for varying degrees of hedonic disparity. *Food quality and preference*, 22, 60–65.
- Weiss, B. H., O'Mahony, M., & Wichchukit, S. (2010). Various paired preference tests: Experimenter effect on "take home" choice. *Journal of sensory studies*, 25, 778–790.