European Review of Agricultural Economics Vol **38 (1)** (2011) pp. 55–78
doi:10.1093/erae/jbq043
Advance Access Publication 30 November 2010

A meta-analysis of consumer willingness to pay for farm animal welfare

Carl Johan Lagerkvist* and Sebastian Hess

Swedish University of Agricultural Economics, Uppsala, Sweden

Received January 2010; final version accepted September 2010

Review coordinated by Christoph Weiss

Abstract

Farm animal welfare (FAW) concerns have prompted scientific research, development and standard setting (especially in the EU and USA) having ethical, production, economic, social, cultural and trade implications. We meta-analyse the literature on consumer willingness to pay (WTP) for FAW, examining 24 studies reporting 106 estimates of consumer FAW values. Meta-regressions indicate that respondent income and age significantly affect WTP, substantial geographical disparities are unsupported, information provision about farm animal living conditions significantly alters WTP estimates and suggestions that FAW be legislatively required significantly reduce WTP. We conclude that the public good aspect of FAW merits further investigation.

Keywords: consumer, farm animal welfare, meta-analysis, willingness to pay

JEL classification: Q1, Q18, Q19, C49

1. Introduction

Since the mid-1990s, economists have been seeking to quantify farm animal welfare (FAW) in economic terms to provide inputs for analysing the implications of animal production systems for resource use and food cost.

Quantifying animal welfare in economic terms is challenging and ultimately depends on ethical assumptions concerning the value of animal lives and conditions. Anthropocentric welfarism, i.e. the belief that only human utility or wellbeing count and that animal welfare is a subset of human welfare, has been the standard welfare economics assumption in much of the literature in the field to date (McInerney, 1993). In line with such preferences, people might be willing to pay *directly* to improve animal welfare qualities to the extent that their own utility is directly affected by such improvements, or *indirectly* (i.e. instrumentally) out of altruistic concern for other people's behaviour or opportunities, and in consideration of animal

*Corresponding author. E-mail: carl-johan.lagerkvist@ekon.slu.se

conditions. Recent results, however, suggest that people's animal welfare preferences would be more consistent with consequentialistic, or utilitarian, ethics (Johansson-Stenman, 2006). Such a view broadens the scope of intrinsic animal welfare externalities related to animal product production and consumption, as the consequences for the animal *per se* will account for an explicit, weighted part of the utility derived from such activities.

In Europe, legislation has historically been the main vehicle by which FAW has been protected and improved (Bennett, 1997). Animal protection laws are issued by national governments and may exceed the minimum standards set by corresponding EU Directives. Existing European directives specify welfare provisions for each major farm animal species and a specific directive currently covers each farm animal species during rearing and slaughter [see Veissier *et al*. (2008) for an overview of European animal welfare regulatory work]. Bock and Van Huik (2007) found that the UK, Sweden and Norway have imposed stricter FAW measures in their national laws, and noted a remarkable similarity in welfare requirements at the level of minimum standards in other European countries.

Paralleling European legislation has been an increasing emphasis on integrating FAW measures into industry-based quality assurance schemes (QAS) or certification schemes, and the increasing use of these by retailers for quality 'gate keeping'. The use of such QAS covering animal welfare is often related to a production differentiation strategy (Manning *et al*., 2006). The tendency of such schemes also to include animal-focused welfare outcomes and indicators is pronounced (Veissier *et al*., 2008). This line of development is consistent with a utilitarian view of animal welfare. Consumers need, and generally favour, information or assurances on which they can base purchasing decisions to satisfy their animal welfare preferences (Mayfield *et al*., 2007).

The present paper conducts a meta-analysis of studies examining values of consumer willingness to pay (WTP) for FAW and determines the effects of the characteristics of the studied consumers, methods of value elicitation and the FAW attributes on valuation estimates. The goal is to generate a set of findings about consumer WTP that are not conditional on the particulars of a single study, and to provide researchers and policy makers with a concise synthesis of the research results. This line of work should be relevant to the recent European Commission (EC) initiative to develop a labelling scheme concerning animal welfare conditions (EC, 2006). We note that FAW is a contentious issue in US animal husbandry as well (Norwood and Lusk, 2009b). Because of the multidimensional character of animal welfare and its assessment, and the increasing amount of research in the area of FAW economics, we do not claim to provide a comprehensive review of the FAW valuation literature. However, we have collected a reasonably large and representative sample of such studies for analysis. The data set comprises 24 separate studies that collectively provide 106 estimates of consumer FAW values.

In the next section, we discuss our method of selecting studies for analysis and whether and to what extent publication bias is of concern in our sample; we also describe the data collected. Detailed results of a meta-regression analysis are then presented. The final section offers concluding remarks, suggestions for future research and policy implications based on the meta-analysis.

2. Methods

Meta-analytic studies pool information from existing studies, comparing methods and weighting information in the literature data set according to certain quality criteria (e.g. Stanley, 2001).

Our first step was to define the boundaries of the literature to be meta-analysed with respect to the included animal species and welfare items. First, we included six farm animal categories, i.e. beef cattle, veal, dairy cows, hogs, broiler chickens and laying hens. Second, as the literature on animal welfare assessment uses a range of indicators or schemes that are typically animal centred, to facilitate interpreting the empirical results in the literature, we used the main criteria from Botreau *et al*. (2007), who developed a set of four criteria (i.e. good feeding, good housing, good health and appropriate behaviour). However, if consequentialist ethics underlies people's FAW perceptions, such an animal-focused scheme will likely provide an insufficient set of welfare items for considering human–animal welfare assessment. We therefore also included GM feed, hormone treatment, antibiotic use, use of biotechnology in animal production practices and transport of live animals to slaughter as additional FAW classifying criteria. Each criterion could be expected to identify potential animal production externalities. In addition, we further differentiated the criteria from Botreau *et al*. (2007) by using a set of general animal husbandry criteria that in principle apply to all farm animal species (Grauvogl *et al*., 1997).

To search for FAW studies in English that treated FAW willingness-to-pay estimates, we used electronic databases (i.e. the ISC Web of Knowledge, American Economic Association EconLit and AgEcon Search) and reference lists from identified studies. We also sought unpublished literature using a Google search of key researchers' websites, since publication bias could skew the results of our meta-regression (Stanley, 2005). The literature search used the following keywords, individually and in combination: farm, animal, welfare, economic, valuation, WTP and consumer. Several studies had to be disregarded because: (i) their data sets were identical to those used in other studies; (ii) they presented aggregated WTP estimates for FAW as an overall concept, but so as to preclude referring to our FAW categorisation results; (iii) no explicit WTP estimate was presented, although consumer responses to FAW improvements were investigated and (iv) they met all our selection criteria but missing observations were noted (predominantly for socio-economic variables). The remaining 24 studies provided 106 individual WTP estimates. The included studies and their main characteristics are presented in Table A1. Literature that we might have missed by this procedure includes (i) confidential reports, for example, conducted by multinationals in the food processing industry and (ii) research reports and scientific publications not published in English or not listed in one of the databases mentioned above. With regard to conducting a meaningful meta-analysis, overlooked literature could be crucial, since omission of certain areas of the literature might bias our findings in a manner similar to publication bias. The latter has long been of concern among meta-analysts, and many economic studies have found evidence of publication bias in certain strands of the economic literature (e.g. Stanley, 2005; Rosenberger and Johnston, 2009). At the end of the next section, we explain how we tested for publication bias in our literature sample.

3. Specification of the meta-analysis function and variables used

The meta-analysis function has a panel structure, as some original studies report multiple WTP estimates related to animal welfare attributes and an unbalanced structure exists as the number of reported estimates differs between studies. Following Stanley (2001), let the explicit specification of the meta-regression be

$$\mathrm{WTP}_{mn} = \alpha + \sum_{k=1}^{k} \beta_k x_{k,mn} + e_m + u_n, \tag{1}$$

where m denotes the candidate study from which the WTP estimate comes (i.e. $m = 1, \ldots, M$) and n denotes the WTP estimate reported in the study (i.e. $n = 1, \ldots, N_m$). In the case in which each study (m) provides a single estimate (n), then $N_m = 1$ and e_m collapse into u_n. When each study provides one or more estimates, then we must account for the common error across estimates (u_n) and the group-specific or panel error in a study (e_m). The total number of estimates is $N = \sum_{m=1}^{M} N_m$. The variations in WTP_{mn} are explained by the scope of a vector of explanatory variables, i.e. $k = 1, \ldots, K$, denoted $x_{k,mn}$. The estimates within a study may partly or completely share several explanatory variables, whereas the estimates across studies may differ in many of those exogenous variables. A nested error structure will exist, as the estimates might not be independent in a given study. The decomposed error variance at the study level, e_m, and error at the estimation level, u_n, are assumed to be normally distributed with zero mean and constant variances, σ_e^2 and σ_u^2, respectively (Bijmolt and Pieters, 2001). In equation (1), α is the intercept term and β_k is a vector of slope parameters to be estimated.

Several retrieved studies present multiple WTP estimates or WTP for animal welfare as an aggregate product/production process characteristic and in so doing do not associate the welfare measure with a specific animal type. Multiple estimates are typically available when results are presented for (i) different subgroups of the sample population, (ii) different model specifications and/or (iii) different levels of the FAW attribute in question. Meta-regression can handle these issues if one accounts for heterogeneous variances and correlation among the error terms. Following Van Houtven (2008), non-overlapping subsample estimates were included when they could be accounted for by citing explanatory variables.

The vector $x_{k,mn}$ contains (i) variables capturing the type of FAW change, (ii) socio-economic variables about respondents' characteristics and (iii) the categorical and methodological characteristics of each study (descriptive statistics are available on request). Furthermore, the WTP assessments typically refer to changes in the living conditions of farm animals relative to a given reference scenario or status quo. The reference cases reflect conditions under which a certain type of farm animal is kept in a given country the year the corresponding study was conducted. Each study m then confronts survey respondents with $n \geq 1$ suggested changes in FAW, which must be viewed as occurring relative to a reference scenario. This implies that the magnitude of these m,n changes can be measured if the base levels to which consumers respond are known. In the following, we refer to each of these m,n FAW changes as 'experiments'. In the meta-data set, therefore, experiments were expressed relative to the following base levels: reductions in stocking density were measured as changes in kg/m^2, while changes in other determinants of animal living conditions were expressed in terms of days, hours, etc., and when this was impossible, we used a dummy approach instead. For this purpose, average animal weights and size requirements were obtained from animal science and agronomy sources (e.g. Grauvogl *et al*., 1997). In the case of hens in battery cages, changes were expressed relative to the minimum stipulated cage size given country and year; for indoor pork and poultry production systems, stocking densities were obtained from providers of animal production equipment and genetics.

All WTP values were expressed relative to the base price relative to which respondents made their statements. Thus, the *dependent variable of our analysis is price premium*, defined as the per cent payment increase over some base price attributable to respondent WTP for a given attribute level.

Income values were adjusted for inflation using the consumer price index. In addition to the inflationary adjustments in the WTP values reported in the international studies, purchasing power parity (PPP) indices were used to adjust the values for intercomparability (OECD, 2009).

To test for publication bias, we follow Stanley (2005) and regress $t_{mn} = \beta_0 + \beta_1(1/\text{Standard Error}_{mn}) + \varepsilon$, with t_{mn} referring to the t-values of the reported WTP changes (as calculated from the available information). In addition, following the discussion in Rosenberger and Johnston (2009: 419ff.), we include in this regression alternative fixed effects for (i) first author, (ii) each publication and (iii) whether or not the study was conducted under peer review. For this auxiliary regression without fixed effects, we fail to reject H_0: $\beta_0 = 0$ and reject H_0: $\beta_1 = 0$, both at a 99 per cent significance level (implying that we reject publication bias through this test and fail to reject the existence of a genuine empirical effect). For the alternative auxiliary regression with fixed effects, we find a significant coefficient only for one author and one publication, respectively; however, we find a highly significant coefficient for the dummy of whether or not a study has been conducted under peer review. We therefore do not investigate further sources of potential publication bias and instead include this dummy in our set of explanatory variables.

4. Meta-data set and estimation procedure

We used initial regressions based on the non-parametric Epanechnikov kernel regression estimator (Hayfield and Racine, 2008) to test for appropriate functional relationships between the dependent variable and our set of explanatory variables. Results revealed no evidence that would lead us to reject a linear regression model due to severe misspecification of the functional relationships. Therefore, we followed Stanley (2001) and employed a parametric meta-regression model using the empirical specifications described here.

We included fixed effects for the location of each study, year of data collection and farm animal type. Although we included socio-economic variables, fixed effects are important because they may reflect local or national preferences, rare events such as food scandals, or latent species-specific preferences. However, the fixed effects for time of study were rejected due to severe collinearity with most other explanatory variables. It should be noted that the strong collinear dependence in the meta-data set is only partly due to the relatively small number of studies included; it is also a consequence of the thorough data collection process that *a priori* aimed to incorporate as much information as possible about the literature sample.

Initially, we considered all elements obtained from our data retrieval process as elements of $x_{k,mn}$, implying that all elements of $x_{k,mn}$ may be of explanatory relevance in the meta-regression. However, regarding the vector of methodological variables, we knew *a priori* that certain combinations of explanatory variables were especially common in the literature data set. Therefore, we grouped eight categorical variables into three broader groups to minimise potential collinearity due to publication-specific effects, as follows:

Group 1 'Context': context of the experiment, for example, type of question or sample;

Group 2 'Payment': type of payment vehicle, opt-out used or not;

Group 3 'Methodology': type of valuation method, type of elicitation format, random utility model, cheap talk script used or not.

This led to an alternative vector of explanatory variables $x_{k,mn}^{\text{pre-grouped}}$.

Furthermore, we weighted all variables in our meta-regression model by the inverse of the standard error of the dependent variable to ensure that potentially more reliable estimates were not confounded by observations subject to a larger standard error.

Finally, we did not assume a separate random effect for study-level effects, due to the small number of studies in the literature data set. Instead, we sought to stabilise data set variance through the weighting procedure. Beyond that, we pre-grouped the data set to test for remaining heteroskedasticity and correlation between the residuals of each study, using 'school of authors' or 'year of data collection' as pre-defined grouping variables. Our meta-regression model is therefore, following Stanley (2001), a weighted OLS model. The explanatory variables for this model were selected according to the following procedure.

4.1 Meta-regression model selection process

The model selection process was divided into two parts, the first based on the full vector, $x_{k,mn}$, and the second on the pre-grouped vector. As long as we are willing to accept that the pre-grouping of variables does not confound the fundamental idea of a meta-analysis, this leads to two alternative regression models. The model selection process based on $x_{k,mn}$ then proceeds as follows (the process analogous to $x_{k,mn}^{\text{pre-grouped}}$):

1. All explanatory variables, $x_{k,mn}$, are included in an initial regression model, with price premium as the dependent variable and 1/(Standard Error of price premium) as the weighting factor.
2. Several variables are immediately dropped due to perfect collinearity. The resulting model is estimated and variance inflation factors (VIFs) are calculated. The VIF of variable j is defined as $\text{VIF}(j) = 1/(1 - R(j)^2)$, where $R(j)^2$ is the multiple coefficient of correlation between variable j and the other independent variables.
3. The variable with the highest VIF is eliminated from the model and the model is re-estimated. This is repeated until all VIFs are significantly below 10 (Kutner *et al*., 2005). If the VIF of several variables remains close to 10, these variables are also dropped until no VIF exceeds 8.
4. The resulting model, Model (0), includes as many elements from our set of explanatory variables as possible, given the constraint that the VIF of each variable must remain significantly below 10.
5. Testing for heteroskedasticity in this model gives the following results: Studentised Breusch–Pagan (BP) test = 56.9502, df = 47, p-value = 0.1517; Goldfeld–Quandt (GQ) test = 0.8829, df1 = 5, df2 = 5, p-value = 0.5527; and Harrison–McCabe (HMC) test (Harrison and McCabe, 1979) = 0.5803, p-value = 0.895. Thus, we have no indication that would force us to reject the assumption of homoskedasticity for this (weighted) Model (0). For the alternative model based on $x_{k,mn}^{\text{pre-grouped}}$ with pre-grouped categories [Model $(0)^{\text{pre-grouped}}$], the test statistics are as follows: BP = 68.2662, df = 48, p-value = 0.02878; GQ = 3.7722, df1 = 4, df2 = 4, p-value = 0.1133; and HMC = 0.2973, p-value = 0.001.

For Model $(0)^{\text{pre-grouped}}$, we thus fail to reject H_0 for homoskedasticity according to the GQ test, but reject it according to the BP and HMC tests. The BP test investigates the potential dependence of the estimated variance of the residuals on the set of independent variables. The GQ test, in contrast, splits the sample and tests for equal variances in both subsamples.

As recommended for meta-analysis (Stanley, 2001), we suspect heteroskedasticity because, by definition, we are pooling observations from various studies; in contrast, weighting the dependent variable, according to the test results for Model (0), seems to remedy the most obvious sources of heteroskedasticity.

To investigate the structure of potential heteroskedasticity still further, we employ the HMC test. This test allows us to order the data set according to selected explanatory variables (i.e. year of data collection, income/month in USD 2005 equivalents, publication ID and observation ID). After ordering, the variances of the residuals on both sides of a sample split (in our case 0.5 of the ordering variable) are tested against H_0 for homoskedasticity.

According to the HMC test, we fail to reject homoskedasticity only for monthly income of respondents. This variable, however, is highly correlated with many other explanatory variables. We therefore conclude that mild heteroskedasticity is potentially present in the data set even after weighting, and that this heteroskedasticity has an unknown structure. Since our model selection process, based on the stepwise elimination of insignificant variables, could be sensitive to even small changes in p-values, we fit two alternative models: Model (1), weighted least squares (WLS) without robust standard errors, and Model (2), the same WLS but with robust standard errors specified by closely approximating the Jackknife procedure suggested by Davidson and MacKinnon (2004).[1] As is common in the literature, we label these robust standard errors HC3. The model selection then proceeds as follows:

1. Stepwise elimination of explanatory variables according to the highest p-value, with a cut-off point at $p = 0.1$.
2. We test for correlation between the residuals by regressing the residuals on the following categorical variable alternatives: 'publication ID', 'school of authors' and 'year of data collection'. 'School of authors' indicates the same category for two publications if these publications share one or more co-authors. Alternately, another major source of potential correlation between the residuals could be unobserved exogenous influences, such as the BSE crisis. We therefore also test for correlation between residuals with fixed effects for the years of data collection.
3. No significant relationship was detected between the explanatory variables in these auxiliary regressions, so, according to adjusted R^2, these regressions have no explanatory power. We therefore conclude that the weighted regressions do not indicate correlation between the residuals according to any typical sources of heterogeneity frequently observed in meta-analysis.

We then replace the explanatory variables in vector $x_{k,mn}$ with the corresponding variables for the pre-grouping (groups 1–3). The model selection process is then repeated until we reach the unrestricted Model $(0)^{\text{pre-grouped}}$ with VIF <10 and stepwise eliminate the insignificant explanatory variables as described above. The reduced regressions based on this procedure are presented in Table 1. In this respect, Model (3) is obtained without robust standard errors. Model (4), in contrast, is derived under robust standard errors (HC3), according to Davidson and MacKinnon (2004).

We perform the J-test of Davidson and MacKinnon (1981) to determine whether the fitted values from any of Models (1)–(4) possess additional explanatory power. If this is the case, we cannot conclude that the model

1 For this HC3 implementation, we use the software developed by Cottrell and Lucchetti (2008). Most of our other computations are executed in R (R Development Core Team, 2009).

Table 1. Meta-regression estimates with/without grouped methodological variables and with/without heteroskedasticity consistent standard errors (HC3)

	Model (1): ungrouped, no HC3[a]			Model (2): ungrouped, HC3[a]			Model (3): grouped, no HC3[a]			Model (4): grouped, HC3[a]		
	Coefficient	Standard error	*p*-Value	Coefficient	Standard error	*p*-Value	Coefficient	Standard error	*p*-Value	Coefficient	Standard error	*p*-Value
Constant	1.589	0.774	0.043**	1.036	0.176	0.000***	1.368	0.826	0.101	0.543	1.071	0.614
Indoor light change (h)	−0.224	0.038	0.000***				−0.224	0.039	0.000***			
Amenity change (days)	0.002	0.000	0.000***									
Fixation change (h)				−0.002	0.001	0.039**						
Isolation removed? $D = 1$ if yes				0.946	0.308	0.003***	0.509	0.144	0.001***	0.687	0.272	0.013**
Amenities mentioned, but no improvement, $D = 1$ if yes				−0.776	0.261	0.004***	−0.747	0.090	0.000***	−0.862	0.258	0.001***
Amenities mentioned and improved, $D = 1$ if yes	1.684	0.195	0.000***				0.864	0.232	0.000***			

(*continued*)

Table 1. (*continued*)

	Model (1): ungrouped, no HC3[a]			Model (2): ungrouped, HC3[a]			Model (3): grouped, no HC3[a]			Model (4): grouped, HC3[a]		
	Coefficient	Standard error	*p*-Value	Coefficient	Standard error	*p*-Value	Coefficient	Standard error	*p*-Value	Coefficient	Standard error	*p*-Value
Rest areas mentioned and improved, $D = 1$ if yes										1.350	0.572	0.020**
Ground texture mentioned but not improved, $D = 1$ if yes	−0.315	0.097	0.002***				−0.292	0.098	0.004***			
Animal type: hens	0.297	0.099	0.003***	0.970	0.315	0.003***	0.686	0.105	0.000***	0.845	0.260	0.002***
Animal type: hogs										0.244	0.127	0.057*
Age of respondents	−0.034	0.015	0.026**				−0.045	0.016	0.007***	−0.036	0.017	0.043**
Income/month (USD 2005 equivalent)	0.209	0.025	0.000***				0.346	0.042	0.000***	0.390	0.055	0.000***
Location: Denmark										−0.517	0.134	0.000***
Location: France				0.950	0.195	0.000***				1.612	0.193	0.000***
Location: Germany										0.979	0.233	0.000***

'FAW changes to be implemented by law?' $D = 1$ if yes	−0.412	0.150	0.007***	−0.478	0.124	0.000***	−0.556	0.158	0.001***	−0.586	0.155	0.000***
'Product labelling for FAW changes?' $D = 1$ if yes							−0.491	0.205	0.018**	−0.529	0.205	0.012**
Valuation method double-bounded dichotomous, $D = 1$ if yes				−0.836	0.193	0.000***						
Payment product price and no opt-out, $D = 1$ if yes							0.408	0.108	0.000***	0.415	0.155	0.009***
Cheap talk used, $D = 1$ if yes				−0.543	0.251	0.033**						
Peer-reviewed publication? $D = 1$ if yes	−0.266	0.101	0.010***	−0.376	0.152	0.015**						
Adjusted R^2	0.94			0.91			0.94			0.91		

[a]HC3 applied according to the procedure and the notation given in Davidson and MacKinnon (2004).
***Statistical significance at the 1 per cent confidence level.
**Statistical significance at the 5 per cent confidence level.
*Statistical significance at the 10 per cent confidence level.

Table 2. *J*-test results with fitted values from Model (*j*) [significance of *p*-values of the fitted values from Model (*i*) in Model (*j*)]

Model (*i*)	Model (1): no pre-grouping, no HC	Model (2): no pre-grouping, HC3	Model (3): pre-grouping, no HC	Model (4): pre-grouping, HC3
Model (1): no pre-grouping, no HC		$<1 \times 10^{-5}$***	4.85×10^{-4}***	$<1 \times 10^{-5}$***
Model (2): no pre-grouping, HC3	1×10^{-5}***		$<1 \times 10^{-5}$***	$<1 \times 10^{-5}$***
Model (3): pre-grouping, no HC3	5.83×10^{-1}	1×10^{-5}***		$<1 \times 10^{-5}$***
Model (4): pre-grouping, HC3	4×10^{-4}***	6.5×10^{-4}***	6.2×10^{-4}***	

Note: Significance implies that Model (*i*) improved due to fitted values from Model (*j*) and thus cannot be considered superior.
***Indicates significance at the 0.01 level.

that improves due to the fitted values is necessarily the best possible model. Table 2 shows the results of the *J*-test.

Results of the *J*-test indicate that Model (3) outperforms Model (1). Both models were selected without robust standard errors. No significantly different results emerge if Models (2) and (3) are selected based on HC3: each model improves because the fitted values form the other model and vice versa. Thus, the *J*-test does not let us identify models that would perform significantly better than others. We therefore conclude that there is weak evidence favouring Model (3), though no other model can be completely rejected either.

Finally, outlier tests do not reveal the dominance of any individual observations, and residuals follow approximately a normal distribution. In this context, the coefficient of determination indicates that well over 90 per cent of the total variance in WTP estimates of the 106 observations in the sample is explainable.

5. Results

Results of Models (1)–(4) are presented in Table 1. Model (0), together with all explanatory variables and the regressions of potentially correlated categorical variables with the corresponding residuals, is available from the authors on request.

Several variables were found to be sensitive to the selection process, and estimates could change if new observations were added to the data set. This applies especially to variables selected, for example, into only one of the four models presented in Table 1, such as fixed effects for the location where a study has been conducted, or variables that address rather broad categories of information (e.g. 'change in amenities'). Caution is therefore warranted in drawing conclusions regarding the WTP for such variables. Table 1 also reveals that several variables appear in two or more models and are therefore more likely to remain if more studies are included in the sample. Changes in hours of light indoors, removal of isolation, mentioning and/or provision of amenities to the animal, ground texture issues, respondent age, legal changes, labelling and peer review are the variables found to exhibit such persistence.

Interestingly, lower WTP is associated with a study mentioning that certain amenities are lacking, but where no such amenities were added during the experiment. This coefficient, however, refers to only three observations, all from Norwood and Lusk (2009a), which provides more detailed information to respondents about how animals are kept than does the average sampled study. In contrast, Norwood and Lusk (2009a) use a novel methodology, so this effect might also be due to methodological factors we would not have adequately captured otherwise [note that Norwood and Lusk (2009a) contribute eight observations to the meta-data set].

In contrast, the estimated coefficient is strongly positive if a certain type of amenity is explicitly mentioned as improved during an experiment, indicating that provision of additional amenities is highly appreciated by consumers. Similar results also appear regarding ground texture and removal of isolation. The change in indoor light hours is relevant to poultry production and refers to experiments in which respondents are told about shortened periods of continuous light in indoor facilities, allowing the animals longer sleep and rest periods.

In all four models, the estimates related to hens (in cages) are positive and highly significant. During variable selection, this dummy was subject to collinearity with the strongly related percentage change in stocking density (expressed in kg/m^2). The latter variable captures improvements in the amount of living/moving space per animal. However, although highly significant during initial regression runs, that variable was dropped due to higher VIFs. Banning battery cages is discussed in many studies in the meta-data set, and our result is consistent with that part of the FAW literature.

Regarding the socio-economic variables, the variable about the proportion of children in each sample was dropped due to too few observations, while the variable about the proportions of men and women had to be dropped due to collinearity with income, age and education. This precluded identifying potential stereotypes. If such a gender effect should exist, it is at least statistically severely confounded by related characteristics of the respondents that apply equally to men.

However, we find the interaction effects of these socio-economic variables to be only partly statistically significant while introducing severe collinearity, and so have omitted them from the regression. Of the socio-economic variables, average monthly income exhibited the strongest explanatory power. The extent to which this indicates that FAW is a luxury good is an open question.

The estimates suggest that each additional year of average age reduces WTP for animal welfare by 3–10 per cent. However, we do not know whether older people generally feel less inclined to pay for FAW, or whether a large proportion of European studies captures specific experiences of the post-WWII generation in this respect. In the light of ageing populations in most developed countries, this aspect deserves further empirical investigation.

Notably, WTP is consistently negatively related to the legal regulation of FAW measures. The underlying interactive variable is active if the WTP assessment of a change in farm animal living conditions is accompanied by a remark such as 'and this change is banned'. Several interpretations are possible here. First, respondents could generally express negative preferences for having FAW changes adopted by legislation, reasons being the net of lost option values and public cost related to the negative external effects of removing the attribute under consideration. Then, however, we would expect the estimated coefficient for 'labelling on the product' to be positive if respondents appreciated the opportunity to use labelling to internalise the private disutility of poorer FAW levels. Instead, labelling estimates come out significantly negative as well, and the interaction effects of these dummies with each other and with the socio-economic variables were insignificant throughout, displaying no change in sign that would reverse these findings. Thus, it is unlikely that this coefficient directly describes the fact that respondents are overall against government action regarding FAW issues.

Second, an alternative explanation might be that hypothetical WTP statements generally indicate people's intentions at the moment of being asked, and remarks such as 'legally binding' or 'labelling' may induce strategic choice behaviours. Could this result be linked to the potential public good character of FAW? The negative coefficient may represent each person's free-riding incentive, since it would not make sense to pay more if certain FAW regulations were becoming binding anyway. This view is indirectly supported by the fact that people from Sweden do not appear in the sample as having significantly higher WTP, although Sweden has the longest history of strict legal FAW regulations, including a ban on battery cages long before it was a political issue elsewhere.

Variables controlling for methodological differences between studies display little explanatory power across the meta-regressions, indicating that the observed heterogeneity in WTP across studies is more attributable to the dispersed range of FAW experiments than to the methodological differences. Comparing Models (1) and (2) with the alternative Models (3) and (4), which are based on the joint impact of methodological variables that frequently occur together in studies, demonstrates that, in the latter two models, more methodological variables appear because of the model selection process. This grouping was driven by the aim of creating methodological dummies that would be sufficiently independent from each other. The grouped category 'payment' appears in two models, and the estimated coefficients suggest that results are, all else being equal, higher if respondents react to product price in combination with the absence of an opt-out option. Cheap-talk scripts and double-bounded dichotomous choice are seen to reduce stated WTP.

Furthermore, the dummy variable attributable to peer-review status is significant across all four models, with coefficients ranging from −0.26 to −0.37. Thus, research reports and working papers, possibly not intended for scientific publication, on average produce higher WTP estimates than do models from scientific journals. In contrast, following Stanley (2005), we must consider to what extent publication bias, for example, on the part of referees and editors sceptical about high WTP for FAW, may favour methods leading, all else being equal, to lower WTP estimates, or, in turn, whether reports not subject to peer review may overstate WTP for FAW. Since we have controlled for methodological heterogeneity between studies in the literature sample, our following finding provides a further explanation: The model selection process did not retain the peer-review dummy in those regressions containing pre-grouped methodological variables; in other words, if the explanatory power of methodological variables that typically occur together in a study is bundled, the effect of peer review loses explanatory power relative to these variables in a regression model. This underlines, in turn, that the relationship between methodological pre-selection and the expected WTP outcome of a study is not random, and that the effect of peer review seems to separate those methodological set-ups that lead to rather conservative WTP estimates from those that produce more optimistic outcomes, everything else being equal.

Furthermore, given that those studies in our literature sample that are subject to peer review tend to use survey techniques and empirical methods that have, in different contexts, proved to reduce the hypothetical bias, we interpret our finding regarding the peer-review dummy as a signal of quality control rather than as an indication that peer review in the literature sample might have introduced a publication bias that (e.g. for ideological reasons) would favour only conservative WTP results concerning FAW.

6. Conclusions and discussion

We reviewed the literature on consumer willingness-to-pay (WTP) values for FAW to determine the effects of the characteristics of the studied consumers, methods of value elicitation and FAW attributes on valuation estimates. This meta-analysis does not, as is usual, appear after the publication of most empirical studies of a certain research question, but instead should be seen as an empirical investigation highlighting the need for systematic empirical research into FAW and the most productive directions for such research. Our model specification results do not let us conclude that any of the four proposed models is strictly superior to the others. Therefore, we emphasise that the meta-regressions should be seen as first-order approximations of the general surface of the meta-data obtained from this literature sample.

The results indicate a positive relationship between WTP and respondent income as well as a negative relationship with age. In addition, there is only weak support for cross-country disparities in WTP; in particular, we find that French and German consumers exhibit larger WTP and Danish consumers lower WTP for FAW than do consumers from the other countries represented in the data set. Thus, the WTP for FAW is not generally affected by local or national preferences that could, for example, result from traditions or reflect national policy events such as food scandals.

From a methodological perspective, the literature on WTP for FAW has moved from early contingent valuation studies or conjoint analyses (mostly choice experiments) to experimental auctions. Methodological developments to handle choice behaviour and choice heuristics leading to, for example, problems of hypothetical biases have been introduced and refined in the literature. Our results suggest that the nature of FAW change experiments causes more variation in WTP estimates than does the choice of methodology, even though some methods typically used to alleviate hypothetical bias problems actually relate to lower WTPs. It might still be too early for a meta-analysis to distinguish the impact on WTP of the recent use of real choice methods.

From an FAW policy perspective, it is interesting to see that WTP can be viewed as relatively non-species specific. The major exception relates to caged hens, for which there is a consistently positive WTP related to housing conditions. The latter finding suggests that the public regards FAW more as a general ethical issue (i.e. a public good) than a product attribute, but that they, as respondents, when provided information on FAW, seem to form distinct preferences for animal welfare improvement. As consumers in a real purchase situation typically lack such complete information as is given in an experimental setting, we interpret this as evidence that a general FAW coding scheme is potentially capable of condensing the complex set of information related to species-specific aspects of FAW. This could be important, because it is questionable whether future consumers would be willing to digest even more species-specific information about a certain FAW problem. However, this potential need to condense information about FAW changes for consumers (and policy makers) has to be distinguished from the scientific need to increasingly generate objective information regarding those factors that really influence animal wellbeing, as well as to take into account that the differences among producers in terms of the environment provided for animals have more impact on FAW than does a single FAW attribute by itself (e.g. Dawkins *et al.*, 2004). Thus, while the latter type of scientific knowledge likely has to address issues highly specific to species and production systems, our results suggest potential for a consistent condensation of this information at the point of sale, for example, related to labelling schemes or legislative standard setting. Such a general coding scheme for FAW-related issues may be less costly to administer and therefore could be economically more efficient, as the public does not seem to make very fine distinctions between individual animals and generally lacks knowledge of how farm animals are kept.

Moreover, several public policy considerations arise, as follows.

(a) FAW is likely perceived as a credence good to most consumers, and the public's lack of knowledge of FAW can by itself generate market failure. Higher FAW standards risk being under-provided if the information provided to consumers does not convincingly present the conditions under which the animals were raised, because there is little reason for farmers to assume higher costs and improve standards if consumers, due to any combination of lack of prior knowledge, awareness or label information, do not change their purchasing decisions in favour of FAW-friendly products. Mandatory labelling or legislative initiatives as ways to alleviate such problems may result in welfare losses for consumers, for example, due to higher prices for FAW-friendly products due to the costs of segregating such products and to consumer aversion towards non-FAW-friendly products. Labelling does let the consumer internalise the private disutility derived from consuming non-FAW-friendly products, but the negative external effect (i.e. the disutility of knowing that others might consume non-FAW-friendly products) potentially causing the market failure may still persist, especially if the disutility component is large enough. This point has, despite early discussions of this distinction by McInerney (1993) and Bennett (1995), received scant attention. Any public intervention touching on FAW features must thoroughly distinguish between the values of the private and public aspects of FAW to maximise the welfare gains achieved by public intervention.

(b) We find a significant relationship between respondent income and WTP for FAW (e.g. the value of a scarce resource). This finding is relevant in an analysis of the ability of the market to coordinate itself when FAW is considered with respect to negative externalities or as a potential market failure. Any regulation or market solution that stipulates and enforces stricter FAW standards could generate a loss of choice for certain members of society (e.g. restricting access to less expensive food) as well as cause social injustice, with wealthier people more likely to have the opportunity to exercise their FAW preferences, while less wealthy people with equally strong FAW preferences might be forced to reduce their purchases or substitute away from the product. Direct referendums regarding FAW legislation/regulations might be more socially acceptable, as they are neutral regarding the division of consumer surplus changes across consumer segments. It can also be argued that a direct referendum would efficiently resolve the market failure originating from the incentives for consumers to free-ride off other consumers' purchases of FAW-friendly products. From an international perspective, direct votes concerning FAW laws would also help distinguish regional voter preferences regarding FAW measures that governments may introduce, for example, for reasons of product discrimination against imports.

(c) Most studies report WTP for positive changes relative to a base level. It is known that people are more loss averse than gain averse, and it is far from certain that FAW measures taken by chain actors are always positive. One might also argue that improvements of single FAW attributes should not be seen in isolation from changes at other attribute levels. Focusing on single measures, such as stocking density, may not improve animal well-being if other criteria are not simultaneously altered. The existing literature has not systematically addressed this; in particular, the design of choice experiments has not considered interdependencies among included attributes. One might ask how such considerations could affect WTP estimates and the need for public intervention.

(d) In a human-centric approach, FAW clearly includes both a human and a human–animal element. A clearer separation of these elements is important in work attempting to derive FAW preferences. From which perspective do we seek values? How do respondents trade off consequences for themselves versus consequences for the animal? The answers concern the weights assigned in addressing the social benefits of policy improvements. They also concern altruism: how much am I willing to give up to improve FAW conditions, and why? Even though the meta-data set includes work directed to these questions, there is a substantial lack of rigour in developing and presenting these aspects in general.

(e) Another aspect concerns preferences and value formation with respect to FAW: how sensitive are these constructs to the distinction between personal feelings and rationality, bearing in mind that, while FAW is an emotional issue for many people, it is simultaneously something 'forgotten' when shopping? This is a very obvious aspect of the literature related to consumer WTP for FAW.

(f) Finally, there is a link between FAW and food safety, from the perspective of both animals (e.g. antibiotics) and humans (e.g. antibiotics, GM and growth hormones). For example, antibiotic provision should make the animal better off, while consumers may want to see antibiotic use in animal production limited, for example, due to the presence of residues in food and public health concerns, or because antibiotic use in animal production signals poorer FAW causing animal health problems. In this respect, our meta-model selection process includes none of these variables in the final regressions, implying that the current literature contains no evidence in this regard statistically strong enough to distinguish respondent WTP for healthy food for themselves from WTP for animal wellbeing. For this reason, studies explicitly focusing on this aspect of the human–animal relationship seem justified; again, meta-analysis would provide a way to compare these studies empirically and estimate demand systems across publications.

In closing, we emphasise that the economics literature on WTP for FAW, according to our literature sample, appears much smaller than public attention to this topic would suggest.

Acknowledgements

We would like to thank Dianne Burgess, Alistair McVittie, Morten Mørkbak and John Rolfe for providing additional data, and Jayson Lusk and Glynn Tonsor for inputs during the AAEA Organized Symposium in 2009, where a first version of our analysis was presented. We are especially grateful for comments and suggestion from three anonymous reviewers and Christoph Weiss.

References

Andersson, J. and Frykblom, P. (1999). Exploring nonmarket values for the social impact of farm animal legislation. Working Paper Series 1999:2. Uppsala: Department of Economics, Swedish University of Agricultural Sciences.

Andersson, C., Lagerkvist, C. J., Carlsson, F., Hannerz, N., Lindgren, K. and Frykblom, P. (2004). Valuation of pork on a local market: a market research. JTI Report 325. Uppsala: Swedish Institute of Agricultural and Environmental Engineering.

Bennett, R. (1995). The value of farm animal welfare. *Journal of Agricultural Economics* 46: 46–60.

Bennett, R. (1997). Farm animal welfare and food policy. *Food Policy* 22: 281–288.

Bennett, R. and Blaney, R. (2002). Social consensus, moral intensity and willingness to pay to address a farm animal welfare issue. *Journal of Economic Psychology* 23: 501–520.

Bennett, R. and Larson, D. (1996). Contingent valuation of the perceived benefits of farm animal welfare legislation: an explanatory survey. *Journal of Agricultural Economics* 47: 224–235.

Bennett, R., Anderson, J. and Blaney, R. J. (2002). Moral intensity and willingness to pay concerning farm animal welfare issues and the implications for agricultural policy. *Journal of Agricultural and Environmental Ethics* 15: 187–202.

Bijmolt, T. H. A. and Pieters, R. G. M. (2001). Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters* 12: 157–169.

Bock, B. B. and van Huik, M. M. (2007). Animal welfare: the attitudes and behaviour of European pig farmers. *British Food Journal* 109: 931–944.

Botreau, R., Veissier, I., Butterworth, A., Bracke, M. B. M. and Keeling, L. J. (2007). Definition of criteria for overall assessment of animal welfare. *Animal Welfare* 16: 225–228.

Burgess, D., Hutchinson, G., McCallion, T. and Scarpa, R. (2003). Investigating choice rationality in stated preference methods for enhanced farm animal welfare. CSERGE Working Paper ECM 03-02. Norwich: University of East Anglia.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2005a). Using cheap talk as a test of validity in choice experiments. *Economic Letters* 89: 147–152.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2005b). Consumer preferences for food product quality attributes from Swedish agriculture. *Ambio* 34: 366–370.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2007a). Farm animal welfare: testing for market failure. *Journal of Agricultural and Applied Economics* 39: 61–73.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2007b). Consumer benefits of labels and bans on GM foods: choice experiments with Swedish consumers. *American Journal of Agricultural Economics* 89: 152–161.

Carlsson, F., Frykblom, P. and Lagerkvist, C. J. (2007c). Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics* 34: 321–344.

Cottrell, A. and Lucchetti, R. (2008). Gnu regression, econometrics and time-series. Available at http://gretl.sourceforge.net/.

Davidson, R. and MacKinnon, J. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49: 781–793.

Davidson, R. and MacKinnon, J. (2004). *Econometric Theory and Methods*. Oxford: Oxford University Press.

Dawkins, M. S., Donelly, C. A. and Jones, T.A. (2004). Chicken welfare is influenced more by housing conditions than by stocking density. *Nature* 427: 342–344.

EC (2006). Labelling: competitiveness, consumer information and better regulation for the EU. Available at http://ec.europa.eu/food/food/labellingnutrition/betterregulation/competitiveness_consumer_info.pdf.

Glass, C. A., Hutchinson, W. G. and Beattie, V. E. (2005). Measuring the value to the public of pig welfare improvements: a contingent valuation approach. *Animal Welfare* 14: 61–69.

Grauvogl, A., Pirkelmann, H., Rosenberger, G. and von Zerboni di Sposetti, H.-N. (1997). *Artgemäße und Rentable Nutztierhaltung*. München, Germany: BLV Verlagsgesellschaft.

Harrison, M. J. and McCabe, B. P. M. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *Journal of the American Statistical Association* 74: 494–499.

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: the np package. *Journal of Statistical Software* 27(5): 1–32. Available at http://www.jstatsoft.org/v27/i05/.

Johansson-Stenman, O. (2006). Should animal welfare count? Working Papers in Economics 197. Department of Economics, School of Business, Economics and Law, Göteborg University. Göteborg, Sweden: Göteborg University.

Kutner, M., Nachtsheim, C., Neter, J. and Li, W. (2005). *Applied Linear Statistical Models,* International Edition. New York: McGraw-Hill.

Lagerkvist, C. J., Carlsson, F. and Viske, D. (2006). Swedish consumer preferences for animal welfare and biotech: a choice experiment. *AgBioForum* 9: 51–58.

Liljenstolpe, C. (2008). Evaluating animal welfare with choice experiments: an application to Swedish pig production. *Agribusiness* 24: 67–84.

Lusk, J. L. and Fox, J. A. (2002). Consumer demand for mandatory labeling of beef from cattle administered growth hormones or fed genetically modified corn. *Journal of Agricultural and Applied Economics* 34: 27–38.

Lusk, J. L., Norwood, F. B. and Pruitt, J. R. (2006). Consumer demand for a ban on antibiotic drug use in pork production. *American Journal of Agricultural Economics* 88: 1015–1033.

Lusk, J. L., Roosen, J. and Fox, J. (2003). Demand for beef from cattle administered growth hormones or fed genetically modified corn: a comparison of consumers in France, Germany, the United Kingdom, and the United States. *American Journal of Agricultural Economics* 85: 16–29.

Manning, L., Baines, R. N. and Chadd, S. A. (2006). Quality assurance models in the food supply chain. *British Food Journal* 108: 91–104.

Mayfield, L. E., Bennett, R. M., Tranter, R. B. and Wooldridge, M. J. (2007). Consumption of welfare-friendly food products in Great Britain, Italy and Sweden, and how it may be influenced by consumer attitudes to, and behaviour towards, animal welfare attributes. *International Journal of Sociology of Food and Agriculture* 15: 59–73.

McInerney, J. P. (1993). Animal welfare: an economic perspective. Paper delivered to the Annual Conference of the Agricultural Economics Society. Oxford.

McVittie, A., Moran, D. and Nevison, I. (2006). Public preferences for broiler chicken welfare: evidence from stated preference studies. Land Economy Working Paper Series 3. Edinburgh: Scottish Agricultural College.

Moran, D. and McVittie, A. 2008. Estimation of the value the public places on regulations to improve broiler welfare. *Animal Welfare* 17: 43–52.

Mørkbak, M., Christensen, T. and Gyrd-Hansen, D. (2009). Consumer preferences for food safety characteristics in pork. Working Paper. Copenhagen: Institute of Food and Resource Economics, University of Copenhagen.

Norwood, F. B. and Lusk, J. L. (2009a). A calibrated auction-conjoint valuation method: valuing pork and eggs produced under differing animal welfare conditions. Working Paper. Stillwater, OK: Oklahoma State University.

Norwood, F. B. and Lusk, J. L. (2009b). Speciesism, altruism, and the economics of animal welfare. Working Paper. Stillwater, OK: Oklahoma State University.

OECD (2009). Purchasing power parity data. Available at http://www.oecd.org/document/47/0,3343,en_2649_34357_36202863_1_1_1_1,00.html.

R Development Core Team (2009). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org.

Rolfe, J. (1999). Ethical rules and the demand for free range eggs. *Economic Analysis and Policy* 29: 187–206.

Rosenberger, R. and Johnston, R. (2009). Selection effects in meta-analysis and benefit transfer: avoiding unintended consequences. *Land Economics* 85(3): 410–428.

Stanley, T. D. (2001). Wheat from chaff: meta-analysis as quantitative literature review. *Journal of Economic Perspectives* 15: 131–150.

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys* 19(3): 309–345.

Tonsor, G. T., Olynk, N. and Wolf, C. (2009). Consumer preferences for animal welfare attributes: the case of gestation crates. *Journal of Agricultural and Applied Economics* 41: 713–730.

Van Houtven, G. (2008). Methods for the meta-analysis of willingness-to-pay data. *Pharmaeconomics* 26: 901–910.

Veissier, I., Butterworth, A., Bock, B. and Roe, E. (2008). European approaches to ensure good animal welfare. *Applied Animal Behaviour Science* 113: 279–297.

Appendix A

Table A1. Summary of FAW studies selected for analysis

Study	Number of observations in data set	Year of data collection	Location of study	Valuation method	Sample description	Elicitation format	Nature of elicitation question	Animal type
Bennett and Larson (1996) (*n* = 110)	3	1995	USA	Double-bounded dichotomous choice (DC)	Students	In-person survey	Hypothetical	Hens (laying)
Bennett (1997) (*n* = 2,000)	3	1996[a]	UK	Double-bounded DC	Random sample	Mail survey	Hypothetical	Hens (laying)
Rolfe (1999) (*n* = 400, r.r. = 0.26)	1	1996	Queensland, Australia	Open-ended contingent valuation (CV)	Random sample	Mail survey	Hypothetical	Hens (laying)
Andersson and Frykblom (1999) (*n* = 294, r.r. = 0.52)	2	1997	Sweden	DC	Random sample	Mail survey	Hypothetical	Hens (laying)
Bennett *et al*. (2002) (*n* = 120)	1	1998	UK	Open-ended CV	Students	In-person survey	Hypothetical	Hens (laying)
Bennett and Blaney (2002) (*n* = 164)	1	1999[a]	UK	Single-bounded DC	Students	In-person interview	Hypothetical	Hogs
Lusk *et al*. (2003) (*n* = 1,000–2,500, r.r. = 0.045–0.23)	8	2000	France, UK, Germany, and USA	Choice experiment (CE)	Random sample	Mail survey	Hypothetical	Cattle
Lusk and Fox (2002) (*n* = 2,500, r.r. = 0.26)	2	2000	USA	One-and-one-half-bounded DC	Random sample	Mail survey	Hypothetical	Cattle

Carlsson *et al.* (2007a) (n = 800, r.r. = 0.58)	2	2002	Sweden	CE	Random sample	Mail survey	Hypothetical	Hens (laying)
Liljenstolpe (2008) (n = 3,000, r.r. = 0.43)	5	2002	Sweden	CE	Random sample	Mail survey	Hypothetical	Hogs
Carlsson *et al.* (2005a) (n = 800; 1,600, r.r. = 0.47–0.52)	18	2003	Sweden	CE	Random sample	Mail survey	Hypothetical	Broilers, cattle
Carlsson *et al.* (2005b) (n = 1,600, r.r. = 0.47)	7	2003	Sweden	CE	Random sample	Mail survey	Hypothetical	Cattle, hens (laying), hogs
Burgess *et al.* (2003) (n = 191)	4	2003	Northern Ireland	Double-bounded DC	Random sample	In-person interview	Hypothetical	Broilers, cattle, hens (laying), hogs
Glass *et al.* (2005) (n = 1876, r.r. = NA)	4	2003	Northern Ireland	Double-bounded DC	Random sample	Mail survey	Hypothetical	Hogs
Andersson *et al.* (2004) (n = 600, r.r. = 0.52)	3	2004	Locally Sweden	CE	Random sample	Mail survey	Hypothetical	Hogs
Carlsson *et al.* (2007b) (n = 800, r.r. = 0.48–0.52)	8	2004	Sweden	CE	Random sample	Mail survey	Hypothetical	Broilers, cattle

(*continued*)

Table A1. (*continued*)

Study	Number of observations in data set	Year of data collection	Location of study	Valuation method	Sample description	Elicitation format	Nature of elicitation question	Animal type
Lusk *et al.* (2006) ($n = 291$)	1	2004[a]	Locally USA	Conjoint choice	Grocery shoppers	In-person interview	Experiment	Hogs
Carlsson *et al.* (2007c) ($n = 800$, r.r. = 0.45–0.49)	4	2004	Sweden	CE	Random sample	Mail survey	Hypothetical	Broilers, cattle
Lagerkvist *et al.* (2006) ($n = 700$, r.r. = 0.41)	8	2005	Sweden	CE	Random sample	Mail survey	Hypothetical	Hogs
McVittie *et al.* (2006) ($n = 336$)	8	2005	UK	CE	Random sample	In-person interview	Hypothetical	Broilers
Moran and McVittie (2008) ($n = 318$)	1	2005	UK	Double-bounded DC	Random sample	In-person interview	Hypothetical	Broilers
Mørkbak *et al.* (2009) ($n = 1322$, r.r. = 0.43)	2	2006	Denmark	CE	Random sample	Internet survey	Hypothetical	Hogs
Tonsor *et al.* (2009) ($n = 1000$, r.r. = 0.26)	2	2007	Locally USA	CE	Random sample	Mail survey	Hypothetical	Hogs
Norwood and Lusk (2009a)	8	2008	USA	Calibrated auction – conjoint	Random sample	Experiments		Hogs, hens

Note: n refers to the sample size; r.r. is response rate.
[a]Specific information on the year of data was not provided in the paper but had to be 'inferred' from the text.