Clark B, Stewart G, Panzone L, Kyriazakis I, Frewer LJ.

Citizens, consumers and farm animal welfare: A meta-analysis of willingness-to-pay studies.

***Food Policy* 2017, 68, 112-127.**

Citizens, consumers and farm animal welfare: a meta-analysis of willingness-to-pay studies

Clark, B*, Stewart, G., Panzone, L., Kyriazakis, I., and Frewer, L.J.

AFRD, Newcastle University, Newcastle upon Tyne, Tyne and Wear, NE1 7RU.

***Author to whom correspondence should be addressed. B.Clark@newcastle.ac.uK**

Abstract

The sustainable intensification of animal production systems is increasing as a consequence of increased demand for foods originating from animals. Production diseases are particularly endemic in intensive production systems, and can negatively impact upon farm animal welfare. There is an increasing need to develop policies regarding animal production diseases, sustainable intensification, and animal welfare which incorporate consumer priorities as well as technical assessments of farm animal welfare. Consumers and/or citizens may have concerns about intensive production systems, and whether animal production disease represent a barrier to consumer acceptance of their increased use. There is a considerable body of research focused on consumer willingness-to-pay (WTP) for improved animal welfare. It is not clear how this relates specifically to a preference for reduced animal production disease incidence in animal production systems. A systematic review and meta-analysis were conducted to establish the publics' WTP for farm animal welfare, with a focus on production diseases which arise in intensive systems. Systematic review methodology combined with data synthesis was applied to integrate existing knowledge regarding consumer WTP for animal welfare, and reduced incidence of animal production diseases. Multiple databases were searched to identify relevant studies. A screening process, using a set of pre-determined inclusion criteria, identified 54 studies, with the strength of evidence and uncertainty for each study being assessed. A random effects meta-analysis was used to explore heterogeneity in relation to a number of factors, with a cumulative meta-analysis conducted to establish changes in WTP over time. The results indicated a small, positive WTP (0.63 standard deviations) for farm animal welfare varying in relation to a number of factors including animal type and region. Socio-demographic characteristics explained the most variation in the data. An evidence gap was highlighted in relation to reduced WTP for specific production diseases associated with the intensification of production, with only 4 of the 54 studies identified being related to this. A combination of market and government based policy solutions appears to be the best solution for improving farm animal welfare standards in the future, enabling the diverse public preferences to be taken into consideration.

Keywords

Systematic review; animal production diseases; willingness-to-pay; farm animal welfare policy; policy options

Highlights

- Overall a small consumer WTP for farm animal welfare was identified
- Only 4 of the 54 studies established WTP for reduced production diseases

- Differences in WTP for welfare existed between animal type, regions and population
- Gaps in evidence for policy development were identified
- The results support the use multiple policy options for improving animal welfare

1. Introduction

The sustainable intensification of animal production represents a potential policy response required to increase the availability foods in relation to growing concerns about food security, and increasing consumer demand for foods derived from animals (Foresight, 2011). However, there is evidence that consumers have very little or no understanding of modern agrifood production systems (Bennett *et al,* 2012). This includes the impact that production diseases can potentially have on animal health and subsequently farm animal welfare (FAW), and the prevalence and nature of occurrence of such diseases in intensive production systems. There is, however, evidence to suggest that FAW is of increasing ethical concern to the European public, with the resulting expectation that foods derived from animals must take due account of welfare issues arising in the production process (Veissier *et al,* 2008; Frewer *et a*l, 2005). Public perceptions of animal health represent an important component within FAW, and represent a potentially important driver of consumption behaviours of European consumers (European Commission, 2007).

The public are an important stakeholder with interests in the food chain, and drive demand for specific foods and commodities (Jensen, 2006). Consideration of their views, needs and preferences regarding the design and operationalisation of animal production systems in FAW policies is essential if they are to be acceptable, and regulatory options reflect public priorities, expectations and requirements. (Farm Animal Welfare Council, 2014; Bennett *et al,* 2002), and a number of aspect of FAW policy have been updated to reflect public concerns (de Jonge & van Trijp, 2014). A number of approaches can be taken by stakeholders to improve FAW (Ingenbleek *et al,* 2012). Government based solutions, in the form of legislation, have traditionally been the main method for ensuring or improving welfare (Bennett, 1997). However, animal production systems which promote higher standards of animal welfare are believed to lead to higher environmental and financial costs (Leinonen *et al,* 2012; World Bank, 2011), which will ultimately be passed onto the consumer unless subsidies or tax breaks are put in place for producers (Bennett, 1997; 1995). In addition, due to the subjective evaluation of animal welfare, individuals may have different opinions as to what counts as a minimally acceptable standard (McInerney, 1994). It is thus difficult to establish a baseline level of animal welfare in production systems that will satisfy all individuals, and which can be used as the initial point for subsequent policy development.

Market based approaches offer an alternative to aligning different approaches to FAW, as different public needs can potentially be met, assuming ethically acceptable *de minimis* welfare standards are applied. They also ensure that producers and consumers are not priced out of the market should any additional costs be passed down the supply chain (McInerey, 1994). Market based solutions are reflected through the increased numbers of private standards being introduced with many businesses adopting welfare friendly stances, including the incorporation of welfare into corporate social responsibility schemes or the adoption of FAW labelling schemes (Marks and Spencer, 2015; McDonalds, 2014). As FAW standards are demand driven, it is important to establish the market potential for these. One approach is to assess consumer/citizen willingness-to-pay (WTP) for FAW. WTP has also been used as a proxy for attitude (Ryan & Spash, 2011) and as an indication of public preferences (Harvey & Hubbard, 2013), and so can be used to assess the acceptability of different FAW practices, to consumers. This evidence can then subsequently be utilised in policy development.

WTP is a measure of value of goods or services to an individual (Hanley *et al*, 2011), and is defined as the price premium or maximum price an individual is willing to sacrifice to obtain a certain benefit or to avoid undesirable characteristics (Breidert *et al*, 2006; Hanley *et al*, 2001). Typically, WTP studies have tried to quantify concerns in relation to the value placed on animal lives, their welfare conditions (Lagerkvist & Hess, 2011) and the higher expected benefits associated with them, including product quality that consumers tend to associate with improved welfare (Verbeke, 2009; European Commission, 2007).

Although previous reviews of the WTP literature have been conducted, these have either not used meta-analysis (Bennett *et al*, 2012), or have not comprehensively explored the grey literature as part of rigorous systematic review methodology combined with meta-analysis (Lagerkvist & Hess, 2011). Furthermore, the issue of consumer WTP for reduced *animal production diseases* has not been a focus of these reviews. Combining systematic review with meta-analysis improves outcome precision and acts to minimise bias in relation to both selection and reporting, taking a comprehensive approach to obtaining and extracting data to ensure that the totality of evidence is considered (Koricheva *et al*, 2013). This will provide more robust evidence on which to base policies. In addition, the increase in intensive production systems in Europe (and indeed internationally) has resulted in attitudes and opinions being potentially influenced by changes in agricultural practices, more intense media reporting of FAW issues, and increased societal discussion of FAW. Precise understanding of consumer attitudes and WTP for FAW interventions specifically designed to address production diseases in intensive systems is required if policy development is to take due account of consumer concerns and priorities.

Both previous reviews have acknowledged the large amount of heterogeneity (variability) in WTP for FAW, for which a number of moderators have had varying explanatory effects. These include different aspects of welfare (Napolitano *et al*, 2008), socio-demographic variables (Bernard & Bernard, 2009; Bennett, 1996) and socio-economic characteristics (Carlsson *et al*, 2007). There is also evidence that WTP for FAW differs between animal types (Cicia & Colantuoni, 2010; Carlsson *et al*, 2007), which may have implications for both producers and FAW policies. In addition, the previous meta-analyses failed to distinguish between consumers of animal products and general citizens. This potential disparity in opinions and attitudes between citizens and consumers is acknowledged in the wider FAW literature (Grunert, 2006; Harper & Henson, 2001), with both known to have favourable attitudes towards higher FAW systems and concerns over more modern or intensive production systems (Blandford *et al*, 2002). However, whereas consumers are able to express these attitudes through the purchasing of animal based products from higher welfare systems such as free range, citizens, including vegetarians and vegans, may not purchase animal products regardless of welfare standards, yet still have an interest in the issues surrounding the implementation of and production of these products (Grunert, 2006). In addition, individuals may behave differently in their dual roles as citizens and consumers, expressing preferences for higher welfare systems when asked (Vanhonacker *et al*, 2007), yet not taking these into consideration when in purchasing situations due to other product attributes taking priority (Blandford *et al*, 2002), or due to a number of perceived barriers to purchasing higher welfare products (Clark *et al*, 2016; Harper & Henson, 2001) These differences are potentially important when developing FAW policies which align with the preferences and priorities of all societal stakeholders.

In light of the increase in published work regarding WTP for FAW since 2011, and in the absence of a review on WTP for reduced animal production diseases specifically, this systematic review and meta-

analysis seeks to extend the work by Lagerkvist & Hess (2011) and aims to establish; 1) what the public are willing-to-pay for FAW, and 2) what the public are willing-to-pay for interventions to reduce production diseases. In addition, heterogeneity within the data will be explored to examine whether certain factors explain the variability in the public's WTP. This will be conducted in relation to; 3) animal type, 4) socio-demographic or socio-economic characteristics, 5) being vegetarian and 6) whether there is a difference in WTP between citizens and consumers.

2. Materials and Methods

2.1. Literature search

The search strategy and meta-analysis protocol were published online prior to starting the review to provide transparency and to enable feedback on the planned research (Clark *et al*, 2014). Relevant publications were identified through searching Scopus, ISI Web of Knowledge, AgEcon Search and Google Scholar using a combination of keywords outlined in table 1, the latter 2 databases enabling the identification of "grey" literature. Search terms were refined after several trial searches to ensure the most effective search terms were used. Both the trialled and final search terms can be obtained by contacting the corresponding author. Face validity of the searches was addressed by checking returned searches for key authors and articles included in both the Lagerkvist & Hess (2011) meta-analysis and Bennett *et al* (2012) review. Animal specific search terms were not used as they frequently returned studies that originated in the natural rather than the social sciences.

Table 1: Keywords considered for search

Type of Study and Outcome	valu* OR intention* OR behav* OR purchas* OR WTP OR willingness to pay OR willingness to buy OR ITP OR buy OR pref*OR economic OR reject* OR consumer OR demand OR choice
	AND
Animal Type	farm animal OR production animal
	AND
Animal Welfare	animal welfare OR health OR disease OR welfare OR production disease

In addition to the database searches, reference lists associated with the studies included were assessed, and key authors in the field were contacted, to identify any additional studies not returned from the searching process. The results were then exported into Endnote for further analysis. The studies which had been identified at this stage were screened in a 2 stage process in order to assess their relevance in relation to the pre-determined inclusion criteria outlined in table 2.

Quantitative empirical studies were included in the review, specifically those that examined the public's WTP for FAW. Measures of welfare were deemed to include anything that was described to participants as altering the lives of animals, ranging from vague descriptors, such as general improvements to overall welfare, to very detailed aspects such as specific stocking densities per m^2, and in relation to specific production diseases. Most studies reported multiple welfare measures and all were extracted for data analysis. All farm animal types were considered for inclusion, including fish. Welfare measures in relation to production diseases were deemed to be anything that specifically mentioned reducing or controlling for diseases. Antibiotic use, including the use of growth promoters (Hughes & Heritage, 2002), was also considered as a proxy for interventions to reduce production diseases, and all studies that measured WTP for animal products produced specifically with or without antibiotics were included.

A broad range of methods used to measure WTP were considered for inclusion in the review. This included, but was not limited to; revealed preference measures (market data, experimental auctions) and stated preference measures (conjoint analysis, contingent valuation studies, choice experiments) as highlighted in table 1. Only studies published in English were included. Studies with duplicate populations (where the same data was presented in 2 or more publications) were removed, with the study with the lowest critical appraisal (see section 2.2) or which reported the fewest WTP or socio-demographic measures being excluded from the current analysis.

Table 2: Eligibility criteria

Study design	English, quantitative empirical; conjoint analysis, auction, dichotomous choice, contingent valuation, choice experiments, additional methods of willingness-to-pay or intention to purchase
Population	Consumers and/ or citizens
Outcome	Willingness-to-pay, intention-to-purchase, price premium

An overview of the search process can be found in the PRISMA flow diagram in figure 1 (Moher *et al* 2009), including the number of studies excluded at each stage. The references of studies excluded at full text stage are provided as a supplementary file (Appendix A).

2.2. Data extraction and critical appraisal

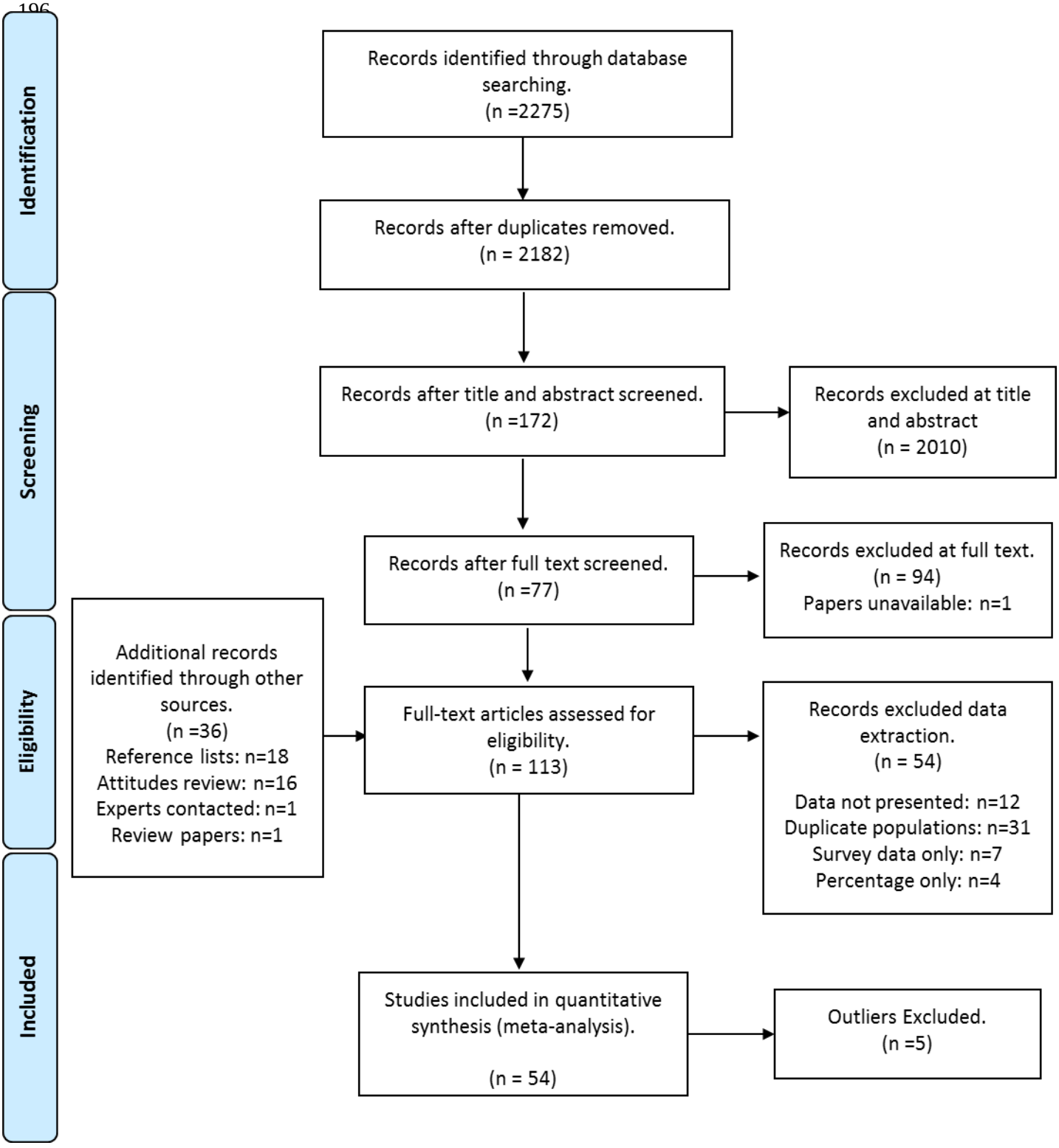
Information was extracted from all papers in relation to the objectives. WTP was extracted as the price premium expressed by participants to purchase products produced to defined FAW standards and was considered as the dependent variable in the analysis. Values were extracted as the mean ± the standard deviation or standard error or with 95% confidence intervals (95% CI) where reported.

WTP values were adjusted for inflation, based on the year of data collection, and when this was not reported it was assumed to be the year of publication (Australian Bureau of Statistics, 2015; inflation.eu, 2015a; 2015b; 2015c; Statistics Sweden, 2015; Bank of Canada, Accessed 26th March 2015; Bank of England, Accessed 26th March 2015; Bureau of Labor Statistics, Accessed 26th March 2015; Eurostat, Accessed 26th March 2015; Statistics Denmark, Accessed 26th March 2015). As the majority of studies were conducted in Europe, the WTP values were converted into Euros to provide a consistent currency across studies (European Central Bank, Accessed 26th March 2015). Additional variables were extracted as moderator variables to help in the explanation of heterogeneity within the data. These were either extracted as mean values (income, age), percentages of the study population (female, vegetarian and college/university education) or coded as categorical variables (animal type, product, welfare aspects, country of data collection, study methodology and economic model used). European countries were grouped according to region based on the United Nations Statistics Division (2013) classification, with the study methods were grouped according to whether they were revealed or stated preference measures as outlined by the Competition Commission (2010).

194 ***Figure 1: Flow Diagram of the search and selection process***

195

196



Sample sizes were recorded in order to provide weights for the meta-analysis. Multiple measures of secondary variables were extracted for each paper reflecting the diverse nature of welfare, and because a large number of studies collected WTP information for a number of animal types and animal products. Summary characteristics for all studies included can be found in appendix B. Due to heterogeneity in the WTP data, effect sizes were calculated for each price premium recorded in order to provide a standardized value and suitable unit for subsequent data analysis. The price premium was recorded as opposed to the percentage price increase (a ratio approach). Although this does not enable a direct determination of the percentage price increase consumers are willing-to-pay as provided by a ratio approach, it does have the advantage of enabling a more direct comparison between effect sizes The effect size, a measure of the magnitude of association between 2 variables, was calculated as the mean WTP divided by the standard deviation (Ferguson, 2009), thus providing a unit of measurement in terms of standard deviations and enabling comparison across studies. When the standard deviation was not available, the standard error and 95% CI were used to calculate this, as *per* Lipsey & Wilson (2001). For studies where standard deviations, standard errors or 95% CI were not reported the variance of the data was imputed using the sample size and mean WTP. This was true for 17 of the 54 studies (31.5%) and these studies will be referred to as 'imputed values' in the analysis. Studies where the effect size was calculated without imputation will be referred to as 'complete case studies'.

The validity and the impact of bias of studies included was addressed by use of a critical appraisal document (appendix C) that examined a number of quality criteria that had the potential to impact on the results of the study; the WTP method used, the economic model used, the sample population and the sampling technique for each study. The document met the guidelines set by the Cochrane Handbook for Systematic Reviews of Interventions (Higgins & Green, 2011), Campbell Collaboration (2001), guidelines and recommendations provided by the Centre for Reviews and Dissemination (2009), and provided a document based in a non-healthcare context. No studies were excluded based on the critical appraisal, with the findings being taken into account during the evidence synthesis when assessing the overall strength of evidence as part the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) analysis (Meader *et al*, 2013). The results of the critical appraisal can be found in Appendix D

2.3. Meta-analysis

Meta-analysis provides a formal and objective way of summarising and interpreting the full range of evidence, with the emphasis of including as much of the literature as possible (Stanley, 2001). A meta-analysis provides increased, power, precision and enables for the formal exploration of the consistency of the variables being examined (Higgins & Green, 2011), which is not provided by a standard literature review alone. This provides more precise and quantifiable evidence from which to develop policies. Meta-analysis was conducted using the open source 'metafor' package (Viechtbauer, 2010) in 'R' (R Core Team, 2013). A positive effect size indicates a WTP a premium in relation to current prices for FAW, and results are reported using the estimate, 95% CI, and I^2, the latter of which examined the amount of heterogeneity remaining within the data. The adaptive GRADE framework (Meader *et al*, 2013) was used to assess the strength of evidence for each study, and was adapted to reflect a non-healthcare setting (c.f. Barański *et al*, 2014). GRADE is based on the following 5 criteria:

1. **Risk of bias** – determined from the critical appraisal of studies.

2. **Imprecision** – the consideration of how big the effect size in relation to zero, as indicated by the lower 95% CI. Values of less than 0.25 are interpreted as being below a minimally important difference (low bias assessment) and values of greater than 0.25 being interpreted as an important difference (moderate bias assessment; Revicki *et al*, 2008; Copay *et al*, 2007).
3. **Indirectness** – due to the focus of the project this focused on whether the majority of studies included (greater than 50%) were conducted in Europe.
4. **Inconsistency** – the variation of effect sizes in relation to the line and spread of the data to establish whether studies were presenting the same picture, and established by visual inspection of forest plots.
5. **Publication bias** – funnel plots were used to establish publication bias, despite their known limitations, with Egger's test also being used to establish funnel plot asymmetry. The results are reported as either undetected or strongly suspected.

A summary of these findings for the analysis can be found in the strength of evidence table in appendix E, with the corresponding weighted mean WTP values for the studies included in each separate analysis to aid with the interpretation of the results.

Random-effects, as opposed to fixed effects meta-analysis was used to calculate the effect size, due to the heterogeneity of studies included in the analysis (Hedges & Vevea, 1998). The random-effects model is a special case of the general linear model, and provides an unconditional inference about a larger set of studies, for which the sample of studies in the meta-analysis is only a random sample of the totality of evidence (Hedges & Vevea, 1998). Random effects meta-analysis helps to deal with the diversity of studies (Borrenstein *et al*, 2009), working on the assumption that the effects from the different studies included in the analysis are not identical but follows some distribution i.e. there is not one true effect size. The centre of this distribution describes the average effect, whereas the width of the distribution details the heterogeneity (Higgins & Green, 2011). Therefore although the model does not solve the problem of heterogeneity of included studies, it does enable analysis of a broad spectrum of studies, indicating the extent of heterogeneity within the data. All analyses were conducted using the restricted maximum likelihood estimator, which is better suited for smaller sample sizes (Viechtbauer, 2010). Effect sizes were calculated for both the complete case studies separately and overall for all included studies (complete case and imputed values). Due to minimal differences in values (appendix E), only the complete case values were used for further analysis. Funnel plots and forest plots were generated for each analysis, with Egger's test also being conducted for each to test for funnel plot asymmetry, which provides an indication of publication bias. Publication bias is defined as "*the tendency toward preparation, submission and publication of research findings based on the nature and direction of the research results*" (Dickersin, 2006), which can lead to the failure to obtain a true representative sample of studies. As this can lead to the overestimation and formation of unreliable conclusions (Dwan *et al*, 2013) and is therefore important to test for.

Forest plots were used to indicate inconsistency in the data and highlight any outliers, which were deemed to be any paper with an effect size more than 10 times the pooled effect size for the complete case studies. Five papers were therefore removed from the analysis; Kehlbacher *et al*, (2012); Chang *et al*, (2010), Pouta *et al* (2010), Glass *et al* (2005) and Dickinson & Bailey (2002), with effect sizes between 18 and 387 times bigger than the pooled effect size. These studies used a variety of different WTP methods, welfare measures and payment vehicles and so shared no obvious similarities.

Sub-group random effects meta-analysis and meta-regression were then used to investigate heterogeneity within the data. Variables included in the sub-group analysis related to both the secondary objectives of the study and exploration of Akaike Information Criterion (AIC) amongst the data, used to minimize over-fitting by establishing the most parsimonious compromise between model fit and model complexity (Koricheva *et al*, 2013). Only findings in relation to animal type, region, socio-demographic characteristics and method are reported in the results section of the report, and additional sub-group analyses can be obtained by contacting the corresponding author. Significant differences were examined between subgroups by establishing whether there was overlap in confidence intervals. A GRADE assessment was also conducted for each subgroup. Finally, the effect size and the weight of effect size values were averaged across studies to provide aggregated values, enabling a cumulative meta-analysis to be conducted, structured by year of publication. A cumulative forest plot was generated to illustrate this.

As per the calculation of the overall effect size, multivariate analysis was conducted using random-effects meta-analysis using the restricted maximum likelihood estimator. A data driven approach was taken with variables selected based on AIC. In addition, due to problems with missing data, and subsequent problems of data not being present across all variable levels, certain variables were excluded from the analysis in order to minimise bias. For example, the socio-demographic characteristics of education and percentage vegetarian were not included, due to too few measures being present (only 74 and 59 respectively of the 227 complete case measures).

Variables included in the analysis were added sequentially based on the number of the measures they contained. For categorical variables (region and animal type) the subgroup with the lowest individual estimate was used as the comparator. In total 5 variables were used (region, animal type, age, gender and income) resulting in 6 models for comparison.

3. Results

3.1. Descriptive statistics

54 studies were included in the final review providing 335 measures of WTP, a summary of which can be found in appendix A. The majority (43) were stated preference studies, 10 were revealed preference studies, and one contained both stated and revealed preference measures. Data came from 17 different countries, with over half of studies being conducted in Europe (56%), 37% being conducted in the USA, and the remaining studies being conducted in Canada, Australia and South Korea. Pigs and laying hens were the 2 most frequently researched animals (14 studies each), with a further 10 studies reporting WTP values for multiple animals, with all but one of these including pigs and either broiler chickens or layer hens. The majority of studies (30) reported WTP for a variety of different welfare measures, with the majority of individual measures relating to overall welfare, free range produce and outdoor access for animals. A range of products were also used, with the most common being eggs and pig meat (pork, ham, salami) reflecting the 2 most common animal types.

In relation to the main objective, only 4 studies specifically examined the public's WTP for FAW related to production diseases; Koistinen *et al* (2013) examined disease and health in pigs and beef cows, Grimsrud *et al* (2013) examined WTP for disease resistance in fish, McVittie *et al* (2006) examined tighter (lower) limits for broiler chickens failing health checks in relation to foot pad lesions

(and associated dermatitis) and Doherty & Campbell (2014) examined WTP for earlier disease detection in broiler chickens. A further 10 studies examined WTP for animal products produced without antibiotics, with the majority of these being conducted in North America (7), and the remainder in Canada, Korea and Denmark. Although other measures which may indirectly affect production diseases were presented, i.e. high stocking density may be associated with respiratory diseases, the relationship between these measures and disease were not described to participants, and it is assumed that study participants were therefore unaware of the link. As a result a WTP value in relation to production diseases cannot be inferred from these results.

Twelve studies did not report any socio-demographic characteristics, with gender, age and income being the 3 most commonly provided variables by 36, 33 and 32 studies respectively. The percentage of vegetarian participants was the least frequently provided piece of information (n=13) with a number of these have the percentage inferred from their study description i.e. all participants described as consumers of meat products. It is possible that vegetarians were excluded from some studies.

3.2. Overall WTP

The results of the GRADE assessment (appendix E) indicate a low strength of evidence for all assessments of general WTP, meaning results should be interpreted with caution. Forests plots and I^2 values highlighted a large amount of variation in the data, for complete case, overall (complete case and imputed) and aggregated values (appendix E). Visual inspection of the funnel plots and results of the Egger's tests confirm the existence of publication bias for the complete case analysis (effect size 0.6302 (95% CI 0.5016, 0.7587), significance $p < 0.0001$), overall data analysis (0.5709 (0.4599, 0.6819), $p < 0.0001$) and aggregated value analysis (0.6135 (0.4106, 0.8524), $p < .0001$), implying that small studies with large effect size estimates appear to be missing from the search process. Both the complete case (effect size 0.6302, (95% CI 0.5016, 0.7587), I^2 99.71) and the overall data (0.5709, (0.4599, 0.6819), 99.76) produced similar estimates, indicating a low price premium for improved FAW. The aggregated WTP estimate was also relatively small price premium (0.4690, (0.2075, 0.7036), 99.72).

The results of the cumulative meta-analysis highlight that measures of WTP have become much more precise over time (figure 3), as indicated by the confidence intervals becoming much narrower, and effect sizes converging as more studies are added chronologically to the analysis. The evolution and convergence of WTP estimates since 2013 reflects the reduction in the number of methods used, with the more recent studies predominantly being choice experiments or auctions. As all 54 studies report WTP for a variety of different animal types and for a number of different welfare measures the converging WTP implies that the positive WTP may not be affected by these individual variables i.e. individuals are willing-to-pay for improved welfare regardless of the individual welfare aspect or animal type.

3.3. Animal Type

Table 3: Summary of results from the animal species subgroup analysis

Animal Type	I^2	Estimate	Lower 95% CI	Upper 95% CI	Number Measures	Number Studies	Weighted Mean WTP (€)	Egger's Test Result
Pig	98.33	0.2843	0.1936	0.3750	90	13	0.54	2.4579, p=0.0140
Layer Hen	99.88	0.7823	0.3594	1.2053	47	10	0.09	1.1088, p=0.2675
Broiler Chicken	97.92	0.4024	0.2653	0.5394	26	8	1.24	4.1308, p<0.0001
Dairy Cow	99.53	1.1176	0.7776	1.4575	27	7	0.50	2.8086, p-0.005
Beef Cow	99.84	1.2022	0.7294	1.6750	24	7	5.00	0.7436, p=0.4571
Multiple	94.78	0.6547	0.4206	0.8888	6	2	11.20	-0.0606, p=0.9516
Fish	99.29	0.3712	-0.0073	0.7497	6	3	3.53	-0.4668, p=0.6406

A summary of findings for the different animal types can be found in table 3 bar calves as results were only obtained from 1 study. Analysis of the subgroups indicated a WTP for all animal types, with the lowest for pigs (0.2843, (0.1936, 0.3750), 98.33), and the largest for beef cows (1.1176, (0.7776, 1.4575), 99.53).

The lower 95% CI for pigs and fish were below the minimally important difference of 0.25 indicating that the WTP for these animals is not significantly different from zero. Overlap of the confidence intervals also indicates that the WTP estimates for laying hens, broiler chickens and pigs are not significantly different from one another, although WTP estimates for both pigs and broilers are significantly different to that from dairy cows, beef cows and multiple animal types.

I^2 values were above 97% for all species, indicating that variation was high within the data, even with the animal type accounted for. Both visual inspection of funnel plots and Egger's test result indicate funnel plot asymmetry and publication bias for 3 of the animal types. Other GRADE criteria also demonstrate bias indicating a low overall strength of evidence.

3.4. Region

Table 4 summarise the results from the region subgroup analysis. Only 1 study was conducted in Asia and so the results are not presented in the table or subsequent discussion. All regions reported a premium as confirmed from the model estimates, with the lowest in Northern Europe (0.1060, (0.0376, 0.1744), 97.84) and the largest in Southern Europe (1.4329, (0.9577, 1.9082), 99.73).

Table 4: Summary of results from the region subgroup analysis

Region	I^2	Estimate	Lower 95% CI	Upper 95% CI	Number Measures	Number Studies	Weighted Mean WTP (€)	Egger's Test Results
UK	97.07	0.6479	0.5113	0.7845	27	7	1.72	1.9722, p=0.0486
Northern Europe	97.84	0.1060	0.0376	0.1744	76	8	0.41	-2.8201, p=0.0048
Western Europe	94.66	1.0741	0.7720	1.3763	7	3	4.28	2.0213, p=0.4320
North America	99.74	0.7515	0.5026	1.0004	90	16	0.15	1.0488, p=0.2943
Southern Europe	99.73	1.4329	0.9577	1.9082	23	6	0.68	2.3457, p=0.0190

Data collected in Western and Southern Europe reported the 2 highest WTP estimates, although it should be noted that there were only 3 studies in the Western European subgroup, and it was the only region to have a high risk of bias (all others being moderate). Data from the UK indicated the second lowest WTP estimate (0.649, (0.5113, 0.7845), 97.07), which was significantly different from both the Southern (highest) and Northern European (lowest) WTP estimates.

The lower 95% CI for Northern Europe was lower than the minimally important difference and so the WTP estimate cannot be said to be significantly different from zero, however it is significantly different from all the other regions WTP estimates. This was the only region to have an overall moderate strength of evidence, with all others being low, indicating that we can be more confident in interpreting the results as indicating that Northern European consumers would pay a low price premium for higher welfare animal products.

I^2 values for all regions were above 94%, again indicating that a large amount of variability existed across studies. Funnel plots and Egger's test results indicate that publication bias is again strongly suspected for most groups

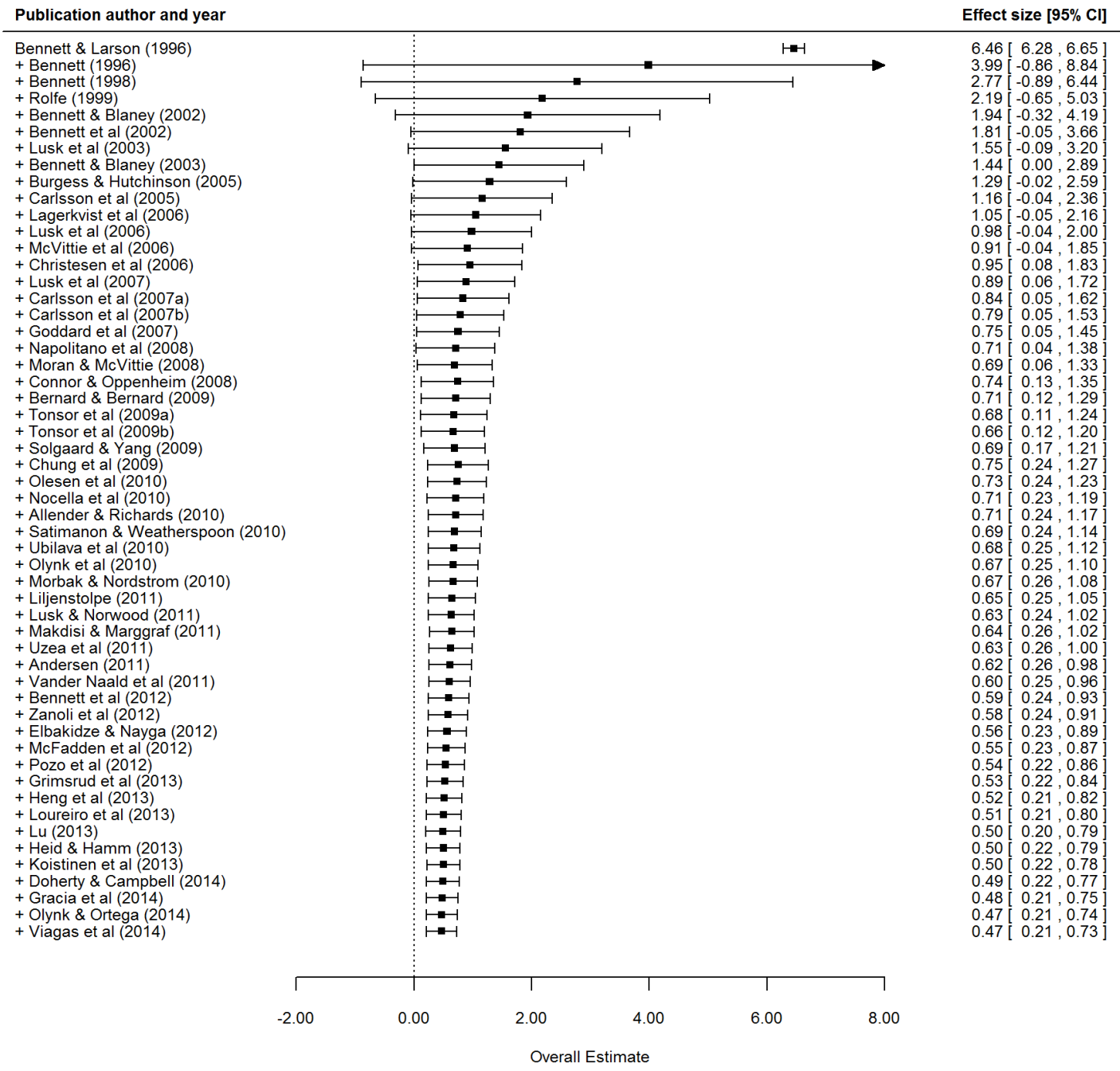


Figure 3: Cumulative meta-analysis of aggregate study values (n=54)

The cumulative meta-analysis of 54 studies demonstrates how WTP has evolved over time, from much larger imprecise estimates, to much smaller, more precise figure of 0.47 standard deviations. This is partly due to methodological advances in cost-benefit analysis.

3.5. Socio-demographic characteristics

Socio-demographic characteristics appear to account for a largest amount of variation within the data, with I^2 values as low as 67.75% for the percentage of vegetarians in the sample. However, these results should be interpreted with caution due to the small sample sizes in the analysis and the lack of reporting of selected variables. This was particularly true for the percentage of vegetarians, which was only reported in 6 studies. In the majority of these the percentage was inferred due to inadequacies in sample descriptions.

The results indicate that WTP decreases with age (-0.0377, (-0.0530, -0.0224), 87.24), increases with income (0.0207, (0.0131, 0.0284), 81.19) and with increased education (0.0086, (-0.0002, 0.0175), 85.69), and is higher for females (0.0246, (0.0113, 0.0379), 98.32) and lower for vegetarians (-0.7024, (-0.9599, -0.4394), 67.75). The WTP estimate for age is significantly different to all the other WTP estimates, highlighting it as an important explanatory variable. Income and level of education are significantly different from one another but are both significantly different to age and the percentage of vegetarians in the sample. Both visual inspection of the funnel plots and the Egger test reveal that publication bias is only detected for education. Imprecision in the results was also low as indicated by visual inspection of the forest plots. Results from the GRADE assessment are therefore more favourable than for previous subgroup analyses with only a low overall strength of evidence for education, with all other variables being either moderate or high (percentage of vegetarians).

Table 5: Summary of results from the socio-demographic subgroup analysis

Socio-Demographic Characteristic	I^2	Estimate	Lower 95% CI	Upper 95% CI	Number Measures	Number Studies	Weighted Mean WTP (€)
Age	87.24	-0.0377	-0.0530	-0.0224	156	24	0.19
Income	81.19	0.0207	0.0131	0.0284	123	19	0.11
Gender	98.32	0.0086	-0.0002	0.0175	157	26	0.19
Vegetarian	67.75	-0.7024	-0.9654	-0.4394	59	6	0.25
Education	85.69	0.0246	0.0113	0.0379	74	15	0.17

3.6. Population

Results from the population subgroup analysis can be found in table 6. Both consumers and citizens reported positive WTP estimates, with citizens reporting a significantly lower WTP estimate (0.5122, (0.3810, 0.6435), 99.65), which was half that of consumers (1.1796, (0.8287, 1.5304), 99.79). Variation in the data remains high with I^2 values of over 99% for both citizens and consumers. Both Egger's test results and visual inspection of the funnel plots indicate that publication bias was strongly suspected, with inspection of the forest plots indicating strong inconsistency in effect size for consumers but only moderate inconsistency in effect size for citizens.

433 ***Table 6: Summary of results from the population subgroup analysis***

Population	I^2	Estimate	Lower 95% CI	Upper 95% CI	Number Measures	Number Studies	Weighted Mean WTP (€)	Egger's Test Results
Citizens	99.65	0.5122	0.3810	0.6435	187	26	0.33	3.7755, p=0.0002
Consumers	99.79	1.1796	0.8287	1.5304	40	11	0.25	1.6097, p=0.1075

434 ## 3.7. Methodological aspects

435 A summary of additional results from the sub-group analysis can be found in table 7. Revealed
436 preference studies reported a significantly higher WTP estimate than stated preference studies, the
437 opposite of what is suggested in the literature e.g. Bateman *et al* (2002). As expected, the use of a
438 cheap talk script (a short statement included in stated preference methods to lower hypothetical bias)
439 reduces consumers stated WTP confirming the importance of its inclusion in stated preference study
440 design. Heterogeneity was extremely high for most other subgroups (I^2 > 90%), indicating that these
441 subgroups do not explain a large amount of variation in the data, as is the case for animal type, region
442 and population type.

443 ***Table 7: Summary of results from additional sub-groups***

Sub-group	I^2	Estimate	Lower 95% CI	Upper 95% CI	Number Measures	Number Studies	Weighted Mean WTP (€)	Egger's Test Results
Revealed Preference	98.72	1.1935	0.9077	1.4793	31	8	€ 0.41	0.3713, p=0.7104
Stated Preference	99.73	0.5416	0.4035	0.6796	196	29	€ 0.25	2.7402, p=0.0061
Cheap talk script used	99.39	0.3595	0.2259	0.4932	83	9	€ 0.11	1.4715, p=0.1412
Cheap talk script not used	99.79	0.6758	0.4600	0.8916	113	20	€ 0.50	1.9413, p=0.0522
Cheap talk script not needed	98.72	1.1935	0.9077	1.4793	31	8	€ 0.41	0.3713, p=0.7104

444 ## 3.8. Multivariate Analysis

445 A summary of the 6 models can be found in table 8. All the models still have high heterogeneity (I^2
446 values of over 98%). Model 6 appears to have the best fit (R^2=55.93%), and AIC score. As per the
447 sub-group analysis (section 3.5), WTP appears to increase with income, decreases with age and is
448 higher for females, although the results are only consistently significant for income.

449 Interaction effects are present (table 8), with certain regions and animal types changing from a positive
450 to a negative WTP with the addition of the socio-demographic characteristics. This is most likely to be
451 due to missing data within the socio-demographic variables, as indicated by the smaller number of
452 measures in the models that containing them (models 4, 5 and 6). The models indicate significant
453 differences between regions, specifically between Asia and North America, and between North

America and Southern Europe. The models also indicate a difference in WTP between animal type, with there being a significant difference between fish and dairy cows in model 6 and between beef cows and layer hens in model 5, and between calves and all other species in models 2 and 3. However, it should be noted that there was only one study in the subgroups for Asia and for calves, both with relatively large WTP values which is likely to account for the consistently significant results.

459 ***Table 8: Multivariate regression models***

	Model 1: Region			Model 2: Animal Type			Model 3: Region + Animal Type		
	Co-efficient	**95% Confidence Intervals**	**p value**	**Co-efficient**	**95% Confidence Intervals**	**p value**	**Co-efficient**	**95% Confidence Intervals**	**p value**
Intercept	0.1049	(-0.0928, 0.3026)	0.297	0.2865	(0.1005, 0.4724)	0.0027 **	0.0162	(-0.1952, 0.2277)	0.8799
Asia[1]	2.2772	(1.3939, 3.1604)	p<0.0001 ***				2.0601	(1.1447, 2.9754)	p<0.0001 ***
North America	0.6461	(0.3769, 0.9153)	p<0.0001 ***				0.4703	(0.1805, 0.7601)	0.0016
Southern Europe	1.3269	(0.9157, 1.7382)	p<0.0001 ***				1.1881	(0.7737, 1.6026)	p<0.0001 ***
UK	0.5478	(0.1611, 0.9344)	0.0057 **				0.6673	(0.1977, 1.1368)	0.0056
Western Europe	0.9776	(0.2931, 1.6621)	0.0053 **				0.9981	(0.2659, 1.7304)	0.0078 **
Beef Cow				0.9152	(0.5100, 1.3204)	p<0.0001 ***	0.3057	(-0.1400, 0.7515)	0.1778
Broiler Chicken				0.1221	(-0.2703, 0.5145)	0.5404	-0.1077	(-0.5878, 0.3725)	0.659
Calves[1]				4.5459	(2.7711, 6.3248)	p<0.0001 ***	4.3479	(2.7091, 5.9867)	p<0.0001 ***
Dairy Cow				0.8311	(0.4434, 1.2187)	p<0.0001 ***	0.354	(0.1627, 0.9058)	0.005 **
Fish				0.0837	(-0.6589, 0.8263)	0.8244	0.354	(-0.3381, 1.0461)	0.3145
Layer Hen				0.4942	(0.1761, 0.8123)	0.0025 **	0.2782	(-0.0380, 0.5944)	0.0844
Multiple				0.3678	(-0.3758, 1.1113)	0.3308	-0.3133	(-1.0788, 0.4521)	0.4206
Age									
Gender									
Income									
n	227			227			227		
AIC	583.1902			573.2498			523.0837		
I^2	99.62			99.64			99.56		
R^2	21.57			18.09			31.3		

	Model 4: Region + Animal Type + Age			Model 5: Region + Animal Type + Age + Gender			Model 6: Region + Animal Type + Age + Gender + Income		
	Co-efficient	**95% Confidence Intervals**	**p value**	**Co-efficient**	**95% Confidence Intervals**	**p value**	**Co-efficient**	**95% Confidence Intervals**	**p value**
Intercept	0.8185	(-0.0925, 1.7294)	0.0779	0.3443	(-0.6885, 1.3772)	0.5108	0.0947	(-0.9242, 1.1137)	0.8541
Asia[1]	1.9448	(1.2902, 2.5995)	p<0.0001 ***	1.7341	(1.0299, 2.4382)	p<0.0001 ***	1.8326	(1.1330, 2.5322)	p<0.0001 ***
North America	0.3963	(0.1678, 0.6249)	0.0008 ***	0.4258	(0.1974, 0.6542)	0.0003 ***	-0.0985	(-0.4092, 0.2121)	0.5309
Southern Europe	1.1824	(0.8729, 1.4919)	p<0.0001 ***	1.0511	(0.7135, 1.3887)	p<0.0001 ***	0.2402	(-0.2231, 0.7036)	0.3064
UK	0.275	(-0.4488, 0.9988)	0.4538	0.2229	(-0.4971, 0.9428)	0.5415			
Western Europe	0.9284	(0.2626, 1.5942)	0.0066) **	0.8626	(0.2001, 1.5251)	0.0111 *			
Beef Cow	0.4021	(0.0198, 0.7845)	0.0394 *	0.4731	(0.0879, 0.8583)	0.0165 *	-0.2595	(-0.7661, 0.2471)	0.3123
Broiler Chicken	-0.1369	(-0.8015, 0.5277)	0.6845	-0.127	(-0.7848, 0.5308)	0.7032	-0.4782	(-1.0650, 0.1087)	0.1092
Calves[1]									
Dairy Cow	0.3798	(-0.0590, 0.8186)	0.0893	0.3208	(-0.1169, 0.7585)	0.14494	0.0996	(-0.2679, 0.4671)	0.5922
Fish	0.017	(-1.0776, 1.1116)	0.9755	0.0444	(-1.0411, 1.1298)	0.9357	-1.5229	(-2.7427, -0.3031)	0.0149 *
Layer Hen	-0.1807	(-0.4393, 0.0779)	0.1693	-0.2558	(-0.5602, 0.0487)	0.099	-0.0679	(-0.3953, 0.2594)	0.6817
Multiple									
Age	-0.0158	(-0.0344, 0.0028)	0.095	-0.0113	(-0.0306, 0.0080)	0.2485	-0.0214	(-0.0415, -0.0013)	0.0371 *
Gender				0.0047	(-0.0040, 0.0133)	0.2875	0.0073	(-0.0025, 0.0170)	0.144
Income							0.0257	(0.0114, 0.4000)	0.0006 ***
n	156			150			121		
AIC	254.8465			242.5592			133.3933		
I^2	98.92			98.93			98.3		
R^2	55.73			48.29			55.93		

Confidence intervals are in parentheses. Significance level: *** 0.01, **0.05, *0.1. [1]Only one study in subgroup.

4. Discussion

4.1. Consumers' WTP for farm animal welfare

The results from the meta-analysis indicate that consumers are willing-to-pay a small price premium for FAW, equivalent to just over half a standard deviation, which is lower than the WTP identified by Lagerkvist & Hess (2011), who obtained premiums of between 50 and 150% in their analysis (Harvey & Hubbard, 2013). This is higher than some existing premiums in the market (Baltzer, 2004). This large disparity may in part be due to the 28 studies included since 2010, which are likely to have used more refined methods, therefore reducing the premiums. As any increases in FAW standards are likely to incur increases in production costs, consumers will need to be prepared to absorb some of these in order to allow the chain to compete effectively (Harvey & Hubbard, 2013), an accurate assessment of WTP is important. The small additional price consumers are WTP for improved welfare may result in consumers changing their behaviour and purchasing higher welfare products, assuming an appropriate and trustworthy identification and certification policy can be implemented to facilitate consumer recognition of such products.

Publication bias was strongly suspected. Lagerkvist & Hess (2011) found that peer reviewed studies had lower WTP estimates typical of publication bias, although they suggest that this indicates that peer review acts as a form of quality control, rather than traditional publication bias. As a consequence of this bias and the other GRADE assessment criteria such as the high heterogeneity, varied populations and low critical appraisal score, an overall low strength of evidence has been identified in the reviewed literature indicating that the results of the analysis should be treated with caution. As the goal of meta-analysis is to study patterns of answers, heterogeneity within the data is expected (Borenstein *et al,* 2009), especially given the range of assessment of WTP methods, animal types and welfare measures available for analysis. Although extensive variation can be problematic, it can be explored by examining moderator variables or by using regression analysis (Stanley, 2001) to provide a better understanding of underlying variation, as discussed in section 4.2. Further primary research is therefore needed to confirm the small price premium found, and given the heterogeneity within the data and significant differences between sub-groups, it would be more relevant to look at this in relation to specific animal products or animal types. In light of how current legislation is currently structured, investigating public WTP in this manner will be more relevant to policy makers. Given the strength of evidence assessment, this is likely to produce much more robust evidence upon which policy can be formulated.

The results of the cumulative meta-analysis indicate that WTP estimates have evolved over time from large imprecise estimates, to much smaller, precise values. This could reflect the developments in WTP methodologies over the past 20 years (Lagerkvist & Hess, 2011), such as advances in modelling and the introduction of cheap talk scripts, the latter of which is known to reduce hypothetical bias (Carlsson *et al,* 2005a). This also reflects the shift towards the use of 2 main types of study; auctions and choice experiments which have been predominantly used since 2012, and are typically viewed as the more preferable measures of WTP (Competition Commission, 2010; Bateman *et al,* 2002). The convergence in WTP estimates from the cumulative meta-analysis also indicate that a positive WTP exists regardless of the animal type or measure of welfare since the studies all reported different combinations of these. This corresponds with the high I^2 values for the different animal types indicating that they do not explain variance data, especially when compared to variables such as socio-demographic characteristics (see section 4.2). The consistently positive WTP reinforces the negative

perceptions the public have of modern farming (Clark *et al*, 2016) and also demonstrates that consumers are concerned about all aspects of welfare therefore a holistic approach to animal wellbeing needs to be considered in policy, which takes into consideration all aspects of welfare such as housing, environment and transport.

A policy evidence gap was demonstrated in relation to consumer WTP for reduced production diseases in intensive farming systems, with only 4 of the 54 studies specifically mentioning these, reflecting findings from a similar review into public attitudes towards production diseases associated with FAW (Clark *et al*, 2016). Two of these studies addressed production diseases in relation to human health (i.e. food safety), whereas the other studies addressed production diseases from an animal perspective, with the study by Grimsrud *et al* (2013) examining WTP to reduce production diseases in farmed salmon and the study by McVittie *et al* (2006) examining WTP to minimise the incidence of foot pad lesions (and associated dermatitis) in broiler chickens, by reducing the threshold for the acceptable number of cases detected. Considering the current trend towards implementation of intensive production systems in Europe, and the role these have to play in future food security policies (e.g. Foresight, 2011), there is a need to address this research gap, as it is not yet clear whether the incidence of production diseases, or the interventions and processes used to treat these will be acceptable to the public, or whether they will represent a barrier to their increased use. Zingg & Siegrist (2012) found that although consumers seem accepting of vaccination programmes for animals for both epidemic and zoonotic diseases, they were not as willing to consume meat from animals that had been vaccinated, which could have serious market implications This also has implications for further development of policies designed to promote FAW though reduced incidence of animal diseases and associated improvements in animal health. In addition, it is important to ensure that the public's views are taken into account in future policy decisions concerning the management of these systems, including any future legislation surrounding animal health and welfare standards to ensure that factors associated with product diseases are effectively incorporated. A better understanding of consumer preferences in relation to aspects associates with production diseases and their interventions may also help to realise any value-added potential from alternative production systems, such as free-range or organic, which are perceived by consumers to be less risky although this may not necessarily be the case (Norwood & Lusk, 2013).

Ten studies addressed consumer WTP for antibiotic free meat and dairy products, with the majority (80%) of these based in the United States of America where antibiotic use as a growth promoter is still permitted (Hughes & Heritage, 2002), although as indicated by results of the WTP studies, is not favourably viewed. This has implications for how much longer they will continue to be used. The lack of research in Europe highlights a need to examine more relevant aspects related to intensive production, such as whether antibiotic use for disease treatment or disease prevention, is considered acceptable to the European public, especially as previous reviews into consumer attitudes have been identified as a concern (Clark *et al*, 2016) being viewed as unnatural, unnecessary and raising concerns in relation to product safety and human health in addition to FAW. At a time when agricultural antibiotic use for economic gain is coming under increased scrutiny due to antibiotic resistance (Mateus *et al*, 2016), and is being widely contested (Chang *et al*, 2015; Hughes & Heritage, 2002), there is a need to consider the evaluation of alternatives including the public acceptability of these.

Despite individual studies considering production and welfare attributes associated with the occurrence of production diseases, they were not presented to participants in the studies included in the systematic review. Hence, future research needs to explicitly describe the link between the welfare attributes presented and the role they play in disease reduction, and animal health, in intensive production diseases, thereby providing further insights on how consumers and citizens value specific measures to reduce production diseases, as consumers do not appear to have the necessary knowledge to make the link themselves, with almost 1 in 3 European consumers (28%) reporting to know nothing at all about the conditions of farm animals in their respective countries (European Commission, 2007).

4.2. Willingness-to-pay by subcategory

4.2.1. Animal type

The public reported a positive WTP estimate for FAW in all animal types indicating that regardless of the animal, consumers are willing-to-pay a premium for improved FAW. Again the GRADE assessment indicated a low overall strength of evidence for all types, with the exception of layer hens. The lowest average WTP estimate obtained was for pigs and the highest for beef and dairy cows. The low WTP estimate for pig welfare is surprising due the large number of studies in this field and attention paid to ethical issues in modern pig production (Lassen *et al*, 2006). However, as the majority of studies were conducted in Sweden (71%), and with Northern European consumers having a statistically lower WTP estimate than other regions, this could have influenced the results. Pig meat may also be considered as a low-value meat suitable for everyday consumption and hence other attributes may be more important than welfare, such as price, origin or various intrinsic product characteristics (Grunert, 2006), rather than FAW.

WTP estimates for the welfare of pigs and broilers are significantly different from dairy cows, beef cows and multiple animal types, but not from laying hens. The higher WTP estimates for both dairy cows and layer hens could be because there are few or no substitutes for these animal products (Kjærnes *et al,* 2007), with dairy alternatives, such as soya, only becoming more readily available over the past few years, after the majority of the studies concerning dairy cows were conducted. Individuals are also familiar with paying price premiums for products such as free range eggs, and for organic dairy products, a production method which is often associated with higher welfare standards (Bernard & Bernard, 2009). In addition, beef is considered as a premium meat (Verbeke & Viaene, 1999) and so consumers may be more willing to pay a premium for this. Alternatively, chicken and pork can quite easily be substituted by other meats, resulting in consumers having a having a lower price premium for these products as they may buy whichever is cheapest, especially if they are price sensitive consumers.

Results of a systematic review towards public attitudes towards FAW also found that consumers are more concerned about chickens (both layers and broilers), although dairy cow welfare was viewed more favourably by the public (Clark *et al,* 2016). From a policy perspective, this implies that although the public appear to have priorities in relation to certain animals, having acceptable *di minimus* standards for all animals is important. The preferences exhibited in both this review and Clark *et al* (2016) also highlights the role of information provision and the media within as playing an important role within this, with the animal types generating the highest WTP estimates being those that have received greatest press attention in recent years. Therefore, ensuring that the public are

provided with, and have access to, accurate information should also be an important policy consideration (European Union, 2011).

4.2.2. Region

WTP estimated were found to be significantly different across regions. The Northern European region had a significantly lower WTP compared to all the other regions. This compares to findings by Lagerkvist & Hess (2011), whose results indicated that the Swedish public did not have a significantly higher WTP. Sweden provided the most measures of WTP for Northern Europe in this analysis, and had the lowest estimate in subgroup analysis for individual countries, although not significantly different from the other Northern European countries included in the analysis.

Even within Europe, significant differences were detected between regions with Southern European countries likely to pay a higher premium than countries within Northern Europe and the UK. This may reflect that Northern Europe and the UK put more emphasis on regulations and less on market-based solutions than Southern Europe (Veissier *et al*, 2008; Kjærnes *et al*, 2007). These regions also have stricter minimum welfare requirements compared to the legislative minimum and so the rest of Europe (Spoolder *et al*, 2011; Bock and van Huik, 2007). Swedish and Norwegian consumers have also been found to be less worried about, and have more trust in, national animal production systems (Kjærnes *et al*, 2007), placing responsibility for ensuring welfare standards with the government rather than themselves (Veissier *et al*, 2008). As a result, individuals from these countries may not be as receptive to, or may not perceive that, further incremental improvements in welfare are required, which may explain the significant difference in WTP estimates compared to Southern Europe. This reflects findings from the 2006 Eurobarometer survey where only 68% Swedes and 67% Finns responded that further improvements to national animal production systems were needed, compared to 90% of Portuguese and 91% of Cypriots (European Commission, 2007). The differences in WTP between European regions implies that future European policy decisions in relation to animal welfare will need to respect the cultural diversity whilst ensuring that the common ethically acceptable *di minimus* standard of welfare enforced by EU legislation does not restrict consumer's freedom of choice by out-pricing them from the market by imposing too large an increase in welfare. Legislation also needs to ensure that European producers are protected from imports from outside the European Economic Area (McGlone, 2001) which may not have been produced to as high a welfare standard and so could be sold at a much lower price. This would also apply to European regions with higher than *di minimus* standards of welfare in their national legislations, where lower priced animal products from other European countries could be imported. If policy cannot protect producers from this trade issue, then products should be regulated so that they are labelled accordingly, so that consumers can make informed purchase decisions.

4.2.3. Socio-demographic variables

Socio-demographic variables were found to account for the largest proportion of heterogeneity within the data compared to the other subgroups analysed, and all had a greater strength of evidence compared to other moderator variables, apart from education. However, due to problems with missing data, especially for the percentage of vegetarians in the study populations, the results should again be interpreted with caution. Age was significantly different to all the other socio-demographic variables, with an increase in age leading to a decrease in WTP, implying different preferences between older and younger individuals. This finding is in line with the previous meta-analysis by Lagerkvist & Hess

(2011) who report a similar decrease in WTP in relation to age. As the authors state, this finding warrants further research to understand the reasoning behind this, due to the aging population in Europe, although it is likely that as people get older their preferences for FAW will not change. Income and education were not significantly different to each other or the percentage of females in the sample, due to non-overlapping confidence intervals. It is not surprising that these variables are linked to a higher WTP, as more educated individuals are more likely to have higher disposable incomes which will provide an individual with a greater ability to reflect their attitudes in their purchasing behaviour.

The heterogeneity within the socio-demographic and socio-economic characteristics confirms that niche markets for products produced to higher FAW standards do exist (Wathes *et al*, 2013), and that segments of the market with different FAW preferences need to be catered for (de Jonge & van Trijp, 2014). This indicates that a blanket policy that maintains acceptable *di minimus* welfare standards for the majority, can be supplemented with market based options, thus ensuring that not all are priced out of the market by high *di minimus* standards of welfare, whilst also maintaining freedom of choice. The lower heterogeneity in WTP estimates with socio-demographics implies that these traditional market segmentation variables are important for the identification of niche markets, especially as there is evidence to suggest that FAW improvements could pay for themselves if products are suitably presented in a way so that consumers can differentiate them from products produced to minimal welfare standards (Compassion in World Farming, 2014; Verbeke, 2009). This also supports findings than market based solutions are a key solution in improving FAW standards (Ingenbleek *et al*, 2012). Ensuring that consumers have enough information to make informed decisions (Mayfield *et al*, 2007) highlights the importance of labelling of higher welfare products as part of market based solutions, providing an authenticity cue that delivers additional assurance to consumers who are prepared to pay more for these products. The authenticity of food is important to consumers enabling them to believe products are what they claim to be (van Rijswijk *et al*, 2008), with consumers willing-to-pay more for authenticity labels (Cicia & Colantuoni, 2010), including on–farm traceability and animal welfare. Suitable independent regulation and monitoring should be in place to ensure that labelling scheme(s) are maintaining the claimed higher standard so as to obtain public trust.

As the results are confounded and affected by bias, further exploration of the population in relation to these characteristics, or additional attitudinal variables, may help to better understand the variability and rationale behind choices, and add further explanation to these findings so that a stronger evidence-base is created. Therefore, future market based research should seek to further explore the socio-demographic characteristics that accounted for the largest amount of heterogeneity. This will enable more definitive recommendations to be made benefitting both policy makers and producers. In addition, consistent reporting standards would ensure that key variables are recorded in future studies, in order to provide greater transparency and aid in subsequent comparison of results and synthesis. These include but are not limited to; socio-demographic characteristics, standard errors, standard deviations or 95% CI, and key methodological details such as year of data collection.

4.2.4. Population

The literature acknowledges the difference between citizens and consumers in relation to animal welfare (Grunert, 2006; Harper & Henson, 2001), recognising that both are likely to express positive attitudes towards improving FAW standards. However, consumers are able to express these attitudes through purchasing products whereas citizens, including vegetarians and vegans, may not purchase

(some) animal products regardless of the welfare standards, yet still have an interest in the issues surrounding the implementation of and production of these products (Vanhonacker *et al,* 2007). The results of the meta-analysis confirm that consumers have a WTP estimate more than double that of citizens, implying that consumers have much stronger, positive attitudes towards higher welfare products than citizens, thereby exhibiting a much greater WTP. Consumers are also likely to place more value on the products involved (obtaining a private benefit from a perceived higher quality or healthier product), and so are willing-to-pay more to secure this (Bennett *et al,* 2012; European Commission 2005). It should be noted that only 40% of the options used in non-consumer experiments were suitable for vegetarians (i.e. tax increases, eggs or dairy products), therefore the difference in WTP between citizens and consumers may be exaggerated.

Differences in WTP for different population segments further supports a portfolio policy response for improving welfare s, depending on the context and the target audience (Ingenbleek *et al,* 2012). A higher consumer WTP again supports the adoption of market based solutions for improving FAW. Companies are increasingly looking to use FAW standards as a point of differentiation as a driver for company codes of practice (Broom, 2010) and to protect brand capital (Blandford *et al,* 2002). Competitive pressures also thought to encourage the adoption of best practice (Harvey & Hubbard, 2013), as demonstrated by number of retailers have already adopted welfare friendly strategies as part of their differentiation strategies, including to entire product categories (White, 2016), a stance which can appeal to non-consumers too. As aforementioned in section 4.2.3, suitable regulation and monitoring of private sector schemes is essential, and as they continue to grow in popularity, it may be the case that a more formalised legislation is required to ensure the transparency, consistency and authenticity of these. The ready identification of these consumers (perhaps by the variables discussed in 4.2.3) along with improved access the higher welfare products, either through transparent labelling, education programs or private assurance schemes (Kehlbacher *et al,* 2012; Ingenbleek *et al,* 2012) are an integral part of this for ensuring success. From a producer perspective, the identification of target markets and indication of whether certain changes to production process will be financially viable is important (Compassion in World Farming, 2014), with transparent regulation of these independent standards being important from a consumer perspective.

Although citizens were WTP less than consumers the positive WTP highlights that the wider public (i.e. non-users) still place a value on higher welfare production systems. It is important to ensure that the growing proportion of those who do not consumer products still have their views taken into consideration in policy formulation., In addition, as the majority of those who consume animal products may adopt dissonance strategies (Clark *et al,* 2016) therefore market based solutions cannot be the only route forward. Policy therefore needs to take into account individuals who still take value from animals being in higher welfare systems yet may not be able to reveal their preferences through market choices, reflecting the social, as opposed to just private benefits, accrued from improvements to animal welfare (Bennett *et al,* 2012). These social benefits support government and farmer based policy solutions, such as legislation and subsidies which would act to guarantee minimum standards or incentivise higher ones respectively. This is supported by the findings of the review with WTP estimates decreasing as the proportion of vegetarians in the sample increased. Although it would be expected that vegetarians could still obtain utility from higher FAW, especially if this was a motivation for their choice, it could be explained by all the products in the analysis being meat (pork, ham and beef) and so would not be directly consumed by them. Therefore, a multi-faceted policy

approach is necessary to support both citizen and consumer preferences and to provide a feasibly acceptable *di minimus* standards of welfare supplemented with more stringent independent standards.

4.2.5. Methodological aspects

Revealed preference studies provided higher WTP estimates than stated preference studies, which is the opposite of what is currently suggested in the literature suggests (Bateman *et al*, 2002). Half of the revealed preference studies were auctions, which also contributed more measures to the analysis. These typically were associated with much higher WTP estimates than either the hedonic analysis or real choice experiments which may explain this difference. Half of these studies were also conducted in Southern European countries, (where the highest WTP estimates were reported), which may also have contributed towards the higher WTP estimates for revealed preference studies. The addition of a cheap talk script, used to reduce hypothetical bias in stated preference methods, to a study decreased the WTP highlighting it as an important and recommended factor in WTP study design (Competition Commission, 2010; Carlsson *et al*, 2005a; 2005b). However, as cheap talk scripts were included only 10 of the 27 stated preference studies published since its first occurrence in the FAW literature in 2005, it appears that this recommendation is still not common practice despite its apparent benefits.

4.2.6. Multivariate analysis

The results from the multivariate models support those from the subgroup analysis, confirming the importance of socio-demographic characteristics and that WTP values appear to differ slightly (although not largely significantly) between animal species and regions. The results for the socio-demographic characteristics confirm that WTP decreases with age, increases with income and is higher for females, with both age and income being significant. This again highlights the potential market based solutions have in improving animal welfare standards (Ingenbleek *et al*, 2012), and confirms these as important marketing segmentation variables in the targeting of niche markets. As per the subgroup analysis, both animal type and regional differences are apparent in the model, although significant differences were not consistent across models. Again this reinforces the conclusions from the respective subgroup analyses in that legislation needs to take into account regional and cultural differences and should also account for concerns towards specific farm animals. The negative WTP estimates for animal type appear to indicate that individuals wish to be compensated for higher welfare products, although they could be due to interaction effects in the data. This negative WTP is unlikely to be in monetary terms but in terms of the additional credence attributes often associated with higher FAW, such as quality, health and taste (European Commission, 2007; Harper & Henson, 2001).

As per the subgroup analysis, there were problems with missing data. Not all animal types were surveyed in all regions, and some regions had very few studies which reported socio-demographic characteristics (mainly those from the UK and Western Europe). This is a potential source of the interaction effects present across animal types and regions. However, despite this limitation it is likely that the regional and species differences as well as the trends observed for the socio-demographic variables are likely to be apparent, due to the consistency with results observed in the subgroup analysis.

4.3. Limitations

The heterogeneity of the data, in relation to a number of the study variables, makes it difficult to draw firm conclusions from the findings especially in relation to overall WTP values. Missing data introduced bias into the results and subsequently, due to the small sample sizes and heterogeneity for certain variables, the sub-group analysis should be interpreted cautiously. It is also important to note that the multivariate analysis was exploratory, due to missing data problems and cofounding between variables, as indicated by AIC and the interaction effects apparent in the analysis. Multivariate analysis also assumed linearity of the variables, which may not be the case. It is also possible the results may be confounded by ecological bias (Stewart *et al,* 2012). However, despite the heterogeneity in the data, and small sample sizes in some sub-group analyses, meta-analysis was still a useful method for formal exploration of the data, providing a more transparent, formalised and robust assessment of the consistency of the effect (Higgins & Green, 2011; Pettigrew & Roberts, 2008) compared to a simple *ad hoc* summary of the literature (Borenstein *et al,* 2009). This transparent framework can also be updated as more evidence becomes available, and has had a transformative effect in other disciplines (Koricheva *et al*, 2013; Higgin & Green, 2011). In addition, as only a small amount of heterogeneity was explained by study characteristics (table 7), it is likely that variation in WTP estimates are more likely due to other factors, such as socio-demographic characteristics and other potential influential drivers not measured or reported in primary studies. This large variability, combined with the public preferring a holistic approach to welfare (Spooner *et al,* 2014), question the usefulness of economic measures such as WTP on highly specific welfare aspects, in the first instance, when there is very little replication of methodologies to enable direct comparison.

4.4. Policy implications summary

- An evidence gap was highlighted in relation to the public's WTP for improved FAW through improved health from a reduction in production diseases. Policies and industry strategies currently focus on increased implementation of intensive production systems in relation to food security, which may not be acceptable to (some) consumers and citizens. It is therefore important that research is conducted in this area to explicitly examine public acceptability of the systems and interventions proposed, as some of the latter may be more controversial than others, especially in relation to antibiotic use.

- Although there is a group of highly concerned consumers, the majority of those who consume animal products are unlikely to consider FAW at the point of purchase, with there also being a growing proportion of the population who do not consume animal products altogether. Market based solutions can therefore only be part of the strategy for improving FAW, with legislation also required to reflect the concerns of non-purchasers. Consequently, a multifaceted response is needed to provide a feasibly acceptable *di minimus* standards of welfare supplemented with more stringent independent standards.

- There is a need to ensure that FAW standards and regulations are continually meeting public preferences, and as these are likely to continue to change (Mann, 2005) and be of great public concern (Mason, 2016), more formalised links between legislation and public perceptions should be developed to foster more acceptable *di minimus* standards. This will enable public views to be taken into account in future policy decisions, especially those concerning the management and handling of animals, on which the majority of the studies included in the review were based.

- A portfolio policy response/ industry response with higher again standards of welfare should be considered, providing a means for consumers with the opportunity to purchase products produced to higher than average welfare standards.

- Given the small sample sizes and heterogeneity within the data, it would seem pertinent to repeat analyses to test the replicability of results and an indication of how stable public preferences are and also further explore animal types and regions where there is little research. Given the problems with missing data in analysis, future research should also focus on being as transparent and as high a standard as possible to ensure that policy and can be based upon as sound an evidence base as possible.

- Due to the low strength of evidence identified in the WTP literature, and as economic valuation methodology (and WTP in particular) is a widely used in a number of fields including policy, more systematic and transparent review processes should be adopted. This includes the more formalised assessment of publication bias, and the robustness and the rigour of the methods used.

5. Conclusion

This study sought to establish public WTP for FAW with a specific focus on interventions to reduce production diseases in intensive production systems. The results indicated that consumers report a small premium for higher FAW products, although this should be treated cautiously due to the high heterogeneity and low strength of evidence presented. Although there are attributes which relevant for reducing production diseases, a research gap was highlighted in relation to the primary objective of the study with only four of the 54 studies specifically examining production diseases.

Further research is therefore required to explore this research gap and to better understand the heterogeneity in WTP in relation to socio-demographic characteristics, which were found to account for the greatest proportion of heterogeneity within the data, and could be used to segment consumers to better facilitate market based solutions for improving FAW.

6. References

- Australian Bureau of Statistics (2015). *Consumer price index inflation calculator.* Available at: http://www.abs.gov.au/websitedbs/d3310114.nsf/home/Consumer+Price+Index+Inflation+Calculator (Accessed: 26th March 2015).

- Baltzer, K. (2004). Consumers' willingness to pay for food quality–the case of eggs. *Food Economics-Acta Agriculturae Scandinavica, Section C*, *1*(2), 78-90.

- Bank of Canada (no date). *Inflation calculator.* Available at: http://www.bankofcanada.ca/rates/related/inflation-calculator/ (Accessed: 26th March 2015).

- Bank of England (no date). *Inflation calculator.* Available at: http://www.bankofengland.co.uk/education/Pages/resources/inflationtools/calculator/flash/default.aspx (Accessed: 26th March 2015).

- Barański, M., Średnicka-Tober, D., Volakakis, N., Seal, C., Sanderson, R., Stewart, G. B., ... & Leifert, C. (2014). Higher antioxidant and lower cadmium concentrations and lower incidence of pesticide residues in organically grown crops: a systematic literature review and meta-analyses. *British Journal of Nutrition*, 112(5): 94-811.

- Bateman, I., Carson, R., Day, B., Hanemann, N., Hanley, N., Jones-Lee, M., Loomes, G., Mourato, S. & Ozdemiroglu, E. (2002). *Economic Valuation with Stated Preference Techniques: A Manual.* Edward Elgar, Cheltenham.

- Bennett, R.M., Butterworth, A., Jones, P.J., Kehlbacher, A. & Tranter, R.B. (2012) *Valuation of animal welfare benefits. A report to Defra.* University of Reading.

- Bennett, R.M. (1997). Farm animal welfare and food policy. *Food Policy*, 22(4), 281-288.

- Bennett, R.M. (1995). The value of farm animal welfare. *Journal of Agricultural Economics*, 46(1), 46-60.

- Bennett, R.M. (1996). People's willingness to pay for farm animal welfare. *Animal Welfare*, 5(1): 3-11.
- Bernard, J.C. & Bernard, D.J. (2009). What is it about organic milk? An experimental analysis. *American Journal of Agricultural Economics*, 91(3): 826-836.

- Blandford, D., Bureau, J.C., Fulponi, L. and Henson, S. (2002). 'Potential implications of animal welfare concerns and public policies in industrialized countries for international trade' in Krissoff, B., Bohman, M. & Caswell, J.A. (eds). *Global food trade and consumer demand for quality*, New York: Springer, 77-99.

- Bock, B.B. & van Huik, M.M. (2007). Animal welfare: the attitudes and behavior of European pig farmers. *British Food Journal*, 109(11): 931-944.

- Borenstein, M., Hedges, L.V., Higgins, J. and Rothstein, H.R. (2009). 'When Does it Make Sense to Perform a Meta-Analysis?' in Borenstein, M., Hedges, L.V., Higgins, J.P.T. & Rothstein, H.R. (eds). *Introduction to Meta-analysis*, Chichester: John Wiley & Sons, 357-364.

- Breidert, C., Hahsler, M. & Reutterer, T. (2006) A review of methods of willingness-to-pay. *Innovative Marketing*, 2(4): 8-32.

- Broom, D.M. (2010). Animal welfare: an aspect of care, sustainability, and food quality required by the public. *Journal of Veterinary Medical Education*, 37(1): 83-88.

- Bureau of Labor Statistics (no date). *CPI Inflation calculator.* Available at: http://www.bls.gov/data/inflation_calculator.htm (Accessed: 26th March 2015).

- Campbell Collaboration (2001) Guidelines for preparation of review protocols [Online]. *Available at: http://www.campbellcollaboration.org/artman2/uploads/1/C2_Protocols_guidelines_v1.pdf* (accessed: 17th October 2014).

- Carlsson, F., Frykblom, P. & Lagerkvist, C.J. (2007). Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics*, 34(3): 321-344.

- Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2005a). Using cheap talk as a test of validity in choice experiments. *Economics Letters*, 89(2): 147-152.

- Carlsson, F., Frykblom, P. & Lagerkvist, C.J. (2005b). Consumer preferences for food product quality attributes from Swedish agriculture, *AMBIO: A Journal of the Human Environment*, 34(4): 366-370.

- Centre for Reviews and Dissemination (2009) *Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care.* York, University of York.

- Chang, Q., Wang, W., Regev-Yochay, G., Lipsitch, M., & Hanage, W. P. (2015). Antibiotics in agriculture and the risk to human health: how worried should we be? *Evolutionary Applications*, 8(3), 240-247.

- Cicia, G. & Colantuoni, F. (2010). Willingness to pay for traceable meat attributes: a meta-analysis. *International Journal on Food System Dynamics*, 1(3): 252-263.

- Clark B, Stewart G.B, Panzone L.A, Kyriazakis I, Frewer .LJ. (2016). A systematic review of public attitudes, perceptions and behaviours towards production diseases associated with farm animal welfare. *Journal of Agricultural & Environmental Ethics*, 29(3), 455-478.

- Clark, B., Stewart, G.S., Panzone, L.A. & Frewer, L.J. (2014). A protocol for a meta-analysis of consumers' and citizens' willingness-to-pay for farm animal welfare and disease prevention. *PeerJ PrePrints* 2:e675v1 http://dx.doi.org/10.7287/peerj.preprints.675v1

- Compassion in World Farming (2014). Economic implications for moving to improved standards of animal welfare. Available at: http://www.ciwf.org.uk/media/5885638/economic-implications-of-moving-to-improved-standards-of-animal-welfare-2014.pdf (Accessed: 9th October 2014).

- Competition Commission (2010). *Review of stated preference and willingness to pay methods.* Available at: http://webarchive.nationalarchives.gov.uk/+/http:/www.competition-commission.org.uk/our_role/analysis/summary_and_report_combined.pdf (Accessed: 6th February 2015).

- Copay, A.G., Subach, B.R., Glassman, S.D., Polly Jr., D.W. & Schuler, T.C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *The Spine Journal*, 7(5), 541-546.

- de Jonge, J. & van Trijp, H. (2014). Heterogeneity in consumer perceptions of the animal friendliness of broiler production systems. *Food Policy*, 49, 174-185.

- DG SANCO (2012). *EU Animal Welfare Strategy: 2012-2015.* Available at: http://ec.europa.eu/food/animal/welfare/docs/brochure_aw_strategy.pdf (Accessed: 1st June 2015).

- Dickersin, K. (2006). 'Publication bias: recognising the problem, understanding its origins and scope and preventing harm'. in: Rothstein, H.R., Sutton, A.J. & Borenstein, M. (eds). *Publication bias in meta-analysis: Prevention, assessment and adjustments*, Chichester: John Wiley & Sons; 12-33.

- Dwan, K., Gamble, C., Williamson, P.R. & Kirkham, J.J. (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. *PloS one*. 8(7), e66844.

- European Central Bank (no date). *Currency convertor.* Available at: http://sdw.ecb.europa.eu/curConverter.do?sourceAmount=1.0&sourceCurrency=AUD&targetCurrency=EUR&inputDate=02-07-2014&submitConvert.x=263&submitConvert.y=12 (Accessed: 26th March 2015).

- European Commission (2012). *EU Animal Welfare Strategy: 2012-2015.* Available at: http://ec.europa.eu/food/animals/docs/aw_brochure_strategy_en.pdf (accessed: 13th June 2016).

- European Commission (2007). "Attitudes of EU citizens towards Animal Welfare". *Special Eurobarometer 270*. Wave 66.1.TNS Opinion and Social. IR *Available at: http://ec.europa.eu/public_opinion/archives/ebs/ebs_270_en.pdf* (accessed: 20th October 2014).

- Eurostat (no date). *HICP – inflation rate.* Available at: http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tec00118&plugin=1 (Accessed: 26th March 2015).

- Farm Animal Welfare Committee (2014). *Evidence and the welfare of farmed animals. Part 1: The evidence base.* Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/324480/FAWC_report_on_evidence_and_the_welfare_of_farmed_animals_part_1_the_evidence_base.pdf (Accessed 1st June 2015).

- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5): 532-538.

- Foresight (2011). *The Future of Food and Farming. The Final Project Report.* London: The Government Office for Science.

- Fraser, D. (2008). Towards a global perspective on farm animal welfare. *Applied Animal Behaviour Science*, 113(4), 330-339.

- Frewer, L.J., Kole, A., Van De Kroon, S.M.A. & De Lauwere, C. (2005). Consumer attitudes towards the development of animal-friendly husbandry systems. *Journal of Agricultural and Environmental Ethics*, 18(4), 345-367.

- Grunert, K. G. (2006). Future trends and consumer lifestyles with regard to meat consumption. *Meat Science*, *74*(1), 149-160.

- Hanley, N., Shogren, J. F. & White, B. (2001). *Introduction to Environmental Economics*. Oxford: Oxford University Press.

- Harper, G., & Henson, S. (2001). *Consumer concerns about animal welfare and the impact on food choice, Final Report. EU FAIR CT98-3678*, Centre for Food Economics Research, The University of Reading.

- Harvey, D. and Hubbard, C. (2013). Reconsidering the political economy of farm animal welfare: An anatomy of market failure. *Food Policy*, *38*, 105-114.

- Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods*, *3*(4), 486.

- Higgins, J.P.T. & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions. [Online]. *Available at: http://www.cochrane-handbook.org* (accessed: 9th October 2014).

- Hughes, P. & Heritage, J. (2002). *Antibiotics as animal growth promoters.* Available at: http://www.fao.org/docrep/article/agrippa/555_en.htm (Accessed: 3rd July 2015).

- Inflation.eu (2015a). *Historic inflation France – CPI inflation.* Available at: http://www.inflation.eu/inflation-rates/france/historic-inflation/cpi-inflation-france.aspx (Accessed: 26th March 2015).

- Inflation.eu (2015b). *Historic inflation Germany – CPI inflation.* Available at: http://www.inflation.eu/inflation-rates/germany/historic-inflation/cpi-inflation-germany.aspx (Accessed: 26th March 2015).

- Inflation.eu (2015c). *Historic inflation South Korea – CPI inflation* Available at: http://www.inflation.eu/inflation-rates/south-korea/historic-inflation/cpi-inflation-south-korea.aspx (Accessed: 26th March 2015).

- Ingenbleek, P.T.M., Immink, V.M., Spoolder, H.A.M., Bokma, M.H. & Keeling, L.J. (2012). EU animal welfare policy: developing a comprehensive policy framework. *Food Policy*, 37(6), 690-699.

- Jensen, H.H. (2006). *Consumer issues and demand*. Available at: http://www.choicesmagazine.org/2006-3/animal/2006-3-09.htm (Accessed: 30th June 2015).

- 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.

- Kehlbacher, A., Bennett, R. & Balcombe, K. (2012). Measuring the consumer benefits of improving farm animal welfare to inform welfare labelling, *Food Policy*, 37, 627-633.

- Kjærnes, U., Miele, M. & Roex, J. (2007). *Attitudes of consumers, retailers and producers to farm animal welfare. Welfare Quality Reports no.2*. Cardiff: Cardiff University, School of City and Regional Planning.

- Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). *Handbook of meta-analysis in ecology and evolution*. New Jersey: Princeton University Press.

- Lagerkvist, C.J. & Hess, S. (2011). A meta-analysis of consumer willingness to pay for farm animal welfare. *European Review of Agricultural Economics*, 38(1): 55-78.

- Lassen, J., Sandøe, P., & Forkman, B. (2006). Happy pigs are dirty!–conflicting perspectives on animal welfare. *Livestock Science*, 103(3), 221-230.

- Lipsey, M.W. & Wilson, D.B. (2001). *Practical Meta-Analysis. Applied Social Research Methods Series Volume 49.* London: Sage Publications.

- Marks and Spencer (2015). *Animal welfare.* Available at: http://corporate.marksandspencer.com/plan-a/policies/our-policies/food/animal-welfare (Accessed: 1st June 2015).

- Mason, R. (2016). *Ministers abandon plans to scrap farm animal welfare codes.* Available at: http://www.theguardian.com/environment/2016/apr/07/ministers-abandon-plan-to-scrap-farm-animal-welfare-codes-chicken-farming (Accessed: 13th June 2016).

- Mateus, A., Takahashi, E.A., Abdelrazeq Elkholly, D., Ekiri, A., Wylie, C., Guinat, C., Patricio, M., Marshall, L., Ferreira, J.P., Stärk, K. & Guitian, J. (2016). *A systematic review to assess the significance of the food chain in the context of antimicrobial resistance (AMR) with particular reference to pork and poultry meat, dairy products, seafood and fresh produce on retail sale in the UK.* Available at: https://www.food.gov.uk/sites/default/files/amr-systematic-review-final-report-2016.pdf (Accessed: 11th January 2017).

- Mayfield, L.E., Bennett, R.M., Tranter, R.B. & 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 Agriculture and Food*, 15(3), 59-73.

- McClone, J.J. (2001). Farm animal welfare in the context of other society issues: toward sustainable systems. *Livestock Production Science*, 72(1-2), 75-81.

- McDonalds (2014). *Animal welfare.* Available at: http://www.mcdonalds.co.uk/ukhome/whatmakesmcdonalds/questions/food/animal-welfare/what-are-mcdonalds-animal-welfare-standards-and-policies.html (Accessed: 1st June 2015).

- Meader, N., King, K., Llewellyn, A., Norman, G., Brown, J., Rodgers, M., Moe-Byrne, T., Higgins, J.P.T., Sowden, A. & Stewart, G. (2014). A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic reviews*, *3*(1): 1-9.

- Moher, D., Liberati, A., Tetzlaff, J. & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4): 264-269.

- Napolitano, F., Pacelli, C., Girolami, A. & Braghieri, A. (2008). Effect of information about animal welfare on consumer willingness to pay for yoghurt. *Journal of Dairy Science*, 91(3): 910-917. Available at: http://www.foodsafetymagazine.com/magazine-archive1/februarymarch-2013/animal-welfare-and-food-safety/ (accessed: 13th June 2016).

- Norwood, F.B. & Lusk, J.L. (2013). Animal welfare and food safety. *Food Safety Magazine*.

- Pettigrew, M. & Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide.* Chichester: John Wiley & Sons.

- R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at: http://www.R-Project.org/. (Accessed: 1st March 2015).

- Revicki, D., Hays, R.D., Cella, D. & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102-109.

- Ryan, A.M. & Spash, C.L. (2011). Is WTP an attitudinal measure? Empirical analysis of the psychological explanation for contingent values. *Journal of Economic Psychology*, 32(5), 674-687.

- Spoolder, H., Bokma, M., Harvey, D., Keeling, L., Majewsky, E., de Roest, K. & Schmid, O. (2011). *EconWelfare findings, conclusions and recommendations concerning effective policy instruments in the route towards higher animal welfare in the EU*. Deliverable D0.5. Available at: http://www.econwelfare.eu/publications/EconWelfareD0.5_Findings_conclusions_and_recommendations.pdf (Accessed: 20th April 2015).

- Spooner, J.M., Schuppli, C.A. and Fraser, D. (2014). Attitudes of Canadian citizens toward farm animal welfare: A qualitative study. *Livestock Science, 163*, 150-158.

- Stanley, T.D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *The Journal of Economic Perspectives*, *15*(3), 131-150.

- Statistics Denmark (no date). *PRIS9: Consumer price index, annual rate of change (inflation) (1900=100) by type.* Available at: http://www.statbank.dk/PRIS9 (Accessed: 26th March 2015).

- Statistics Sweden (2015). *Inflation in Sweden 1831-2014.* Available at: http://www.scb.se/en_/Finding-statistics/Statistics-by-subject-area/Prices-and-Consumption/Consumer-Price-Index/Consumer-Price-Index-CPI/Aktuell-Pong/33779/Consumer-Price-Index-CPI/33831/ (Accessed: 26th March 2015).

- Steinfeld, H. (2004). The livestock revolution – a global veterinary mission. *Veterinary Parasitology*, 125(1): 19-41.

- Stewart, G.B., Altman, D.G.,, Askie, L.M.,, Duley, L., Simmonds, M.C. & Stewart, L.A. (2012). Statistical Analysis of Individual Participant Data Meta-Analyses: A Comparison of Methods and Recommendations for Practice. *PLoS ONE* 7(10): e46042. doi:10.1371/journal.pone.0046042United Nations Statistics Division (2013). *Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings.* Available at: http://unstats.un.org/unsd/methods/m49/m49regin.htm#europe (Accessed: 16th April 2015).

- Vanhonacker, F., Verbeke, W., Van Poucke, E. and Tuyttens, F. (2007). Segmentation based on consumers' perceived importance and attitude toward farm animal welfare. *International Journal of Sociology of Agriculture and Food, 15*(3), 91-107.

- van Rijswijk, W., Frewer, L.J., Menozzi, D. & Faioli, G. (2008). Consumer perceptions of traceability: a cross-national comparison of the associated benefits. *Food Quality and Preference*, 452-464.

- Verbeke, W. (2009).Stakeholder, citizen and consumer interests in farm animal welfare. *Animal Welfare*, 18(4): 325-333.

- Verbeke, W. & Viaene, J. (1999). Beliefs, attitudes and behaviour towards fresh meat consumption in Belgium: empirical evidence from a consumer survey. *Food Quality and Preference*, 10(6), 437-445.

- Veissier, I, Butterworth, A., Bock, B. & Roe, E. (2008). European approaches to ensure good animal welfare. Applied Animal Behavioural Science, 113(4): 279-297

- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), pp. 1-48. Available at: http://www.jstatsoft.org/v36/i03/ (Accessed: 24th March 2015).

- Wathes, C.M., Buller, H., Maggs, H. & Campbell, M.L. (2013). Livestock Production in the UK in the 21st Century: A Perfect Storm Averted? *Animals*, 3(3): 574-583.

- White, R. (2016). *Waitrose milk cows to be grass fed for at least 100 days a year.* Available at: http://www.thegrocer.co.uk/buying-and-supplying/categories/dairy/waitrose-milk-cows-to-be-grass-fed-for-at-least-100-days-a-year/531153.article (Accessed: 13th June 2016).

- Zingg, A. & Siegrist, M. (2012). People's willingness to eat meat from animals vaccinated against epidemics. *Food Policy*, 37, 226-331.

7. Appendices

Appendix A: Studies included in the systematic review

- Allender, W.J. & Richards, T.J. (2010). Consumer Impact of Animal Welfare Regulation in the California Poultry Industry, *Journal of Agricultural and Resource Economics*, 35(3), pp. 424-442.

- Andersen, L.M. (2011). Animal Welfare and Eggs - Cheap Talk or Money on the Counter?, *Journal of Agricultural Economics*, 62(3), pp. 565-584.

- Bennett, R. (1998). *Measuring public support for animal welfare legislation: A case study of cage egg production,* Animal Welfare, 7(1), pp. 1-10.

- Bennett, R. & Blaney, R. (2002). Social consensus, moral intensity and willingness to pay to address a farm animal welfare issue, *Journal of Economic Psychology*, 23(4), pp. 501-520.

- Bennett, R., Kehlbacher, A. & Balcombe, K. (2012). *A method for the economic valuation of animal welfare benefits using a single welfare score,* Animal Welfare, 21, pp. 125-130.

- Bennett, R. & Larson, D. (1996) Contingent valuation of the perceived benefits of farm animal welfare legislation: an exploratory survey, *Journal of Agricultural Economics*, 47(1-4), pp. 224-235.

- Bennett, R.M. (1996) People's willingness to pay for farm animal welfare, *Animal Welfare*, 5(1), pp. 3-11.

- Bennett, R.M., Anderson, J. & Blaney, R.J.P. (2002). Moral intensity and willingness to pay concerning farm animal welfare issues and the implications for agricultural policy, *Journal of Agricultural & Environmental Ethics*, 15(2), pp. 187-202.

- Bennett, R.M. & Blaney, R.J.P. (2003). Estimating the benefits of farm animal welfare legislation using the contingent valuation method, *Agricultural Economics*, 29(1), pp. 85-98.

- Bernard, J.C. & Bernard, D.J. (2009).What is it about organic milk? An experimental analysis, *American Journal of Agricultural Economics*, 91(3), pp. 826-836.

- Burgess, D., Hutchinson, W.G., McCallion, T. and Scarpa, R. (2003). Investigating choice rationality in stated preference methods for enhanced farm animal welfare, *Working Paper - Centre for Social and Economic Research on the Global Environment*. pp. 1-36. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-55849094325&partnerID=40&md5=5a49c0ff3c953c806ff1b40c5c60935f.

- Carlsson, F., Frykblom, P. & Lagerkvist, C.J. (2005). Consumer preferences for food product quality attributes from Swedish agriculture, *AMBIO: A Journal of the Human Environment*, 34(4), pp. 366-370.

- Carlsson, F., Frykblom, P. & Lagerkvist, C.J. (2007a). Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter, *European Review of Agricultural Economics*, 34(3), pp. 321-344.

- Carlsson, F., Frykblom, P. & Lagerkvist, C.J. (2007b). Farm animal welfare-testing for market failure, *Journal of Agricultural and Applied Economics*, 39(1), p. 61.

- Christensen, T., Mørkbak, M., Hasler, B., Lundhede, T., Porsbo, L.J. & Christoffersen, L.B. (2006). Information, risk perception and consumer behaviour: a choice experiment on food safety and animal welfare. Copenhagen. Fødevareøkonomisk Institut. (Rapport / Fødevareøkonomisk Institut; No. 180).

- Chung, C., Boyer, T. & Han, S. (2009). Valuing quality attributes and country of origin in the Korean beef market, *Journal of Agricultural Economics*, 60(3), pp. 682-698.

- Conner, D.S. & Oppenheim, D. (2008). Demand for pasture-raised livestock products: results from Michigan retail surveys, *Journal of Agribusiness*, 26(1), pp. 1-20.

- Doherty, E. & Campbell, D. (2014). Demand for safety and regional certification of food Results from Great Britain and the Republic of Ireland, *British Food Journal*, 116(4), pp. 676-689.

- Elbakidze, L. & Nayga, R.M., Jr. (2012). The effects of information on willingness to pay for animal welfare in dairy production: Application of nonhypothetical valuation mechanisms, *Journal of Dairy Science*, 95(3), pp. 1099-1107.

- Goddard, E.W., Boxall, P., Emunu, J.P., Boyd, C., Asselin, A. & Neall, A. (2007). *Consumer attitudes, willingness to pay and revealed preferences for different egg production attributes: analysis of Canadian egg consumers*. Department of Rural Economy, University of Alberta.

- Gracia, A., Barreiro-Hurle, J. & Lopez-Galan, B. (2014). Are Local and Organic Claims Complements or Substitutes? A Consumer Preferences Study for Eggs, *Journal of Agricultural Economics*, 65(1), pp. 49-67.

- Grimsrud, K.M., Nielsen, H.M., Navrud, S. & Olesen, I. (2013). Households' willingness-to-pay for improved fish welfare in breeding programs for farmed Atlantic salmon, *Aquaculture*, 372, pp. 19-27.

- Heid, A. & Hamm, U. (2013). Animal welfare versus food quality: Factors influencing organic consumers' preferences for alternatives to piglet castration without anaesthesia, *Meat Science*, 95(2), pp. 203-211.

- Heng, Y., Peterson, H.H. & Li, X. (2013). Consumer attitudes toward farm-animal welfare: The case of laying hens, *Journal of Agricultural and Resource Economics*, 38(3), pp. 418-434.

- Koistinen, L., Pouta, E., Heikkilä, J., Forsman-Hugg, S., Kotro, J., Mäkelä, J. & Niva, M. (2013). The impact of fat content, production methods and carbon footprint information on consumer preferences for minced meat, *Food Quality and Preference*, 29(2), pp. 126-136.

- Lagerkvist, C.J., Carlsson, F. & Viske, D. (2006). Swedish consumer preferences for animal welfare and biotech: a choice experiment. *The Journal of Agrobiotechnology Management & Economics*, 9(1).

- Liljenstolpe, C. (2011). Demand for Value-Added Pork in Sweden: A Latent Class Model Approach, *Agribusiness*, 27(2), pp. 129-146.

- Loureiro, M.L., Gracia, A. & Nayga, R.M., Jr. (2013). Do experimental auction estimates pass the scope test?, *Journal of Economic Psychology*, 37, pp. 7-17.

- Lu, Y. (2013). *Consumer Preference for Eggs from Enhanced Animal Welfare Production System: A Stated Choice Analysis.* PhD Thesis, University of Guelph.

- Lusk, J.L., Nilsson, T. & Foster, K. (2007). Public preferences and private choices: effect of altruism and free riding on demand for environmentally certified pork, *Environmental and Resource Economics*, 36(4), pp. 499-521.

- Lusk, J.L. & Norwood, F.B. (2012). Speciesism, altruism and the economics of animal welfare, *European Review of Agricultural Economics*, 39(2), pp. 189-212.

- Lusk, J.L., Norwood, F.B. & Pruitt, J.R. (2006). Consumer demand for a ban on antibiotic drug use in pork production, *American Journal of Agricultural Economics*, 88(4), pp. 1015-1033.

- Lusk, J.L., Roosen, J. & Fox, J.A. (2003). Demand for beef from cattle administered growth hormones or fed genetically modified corn: a comparison of consumers in France, Germany, the United Kingdom, and the United States, *American Journal of Agricultural Economics*, 85(1), pp. 16-29.

- Makdisi, F. & Marggraf, R. (2011). *Consumer willingness to pay for farm animal welfare in Germany – the case of broiler chickens. 30th* September, Halle, Germany. Proceedings of the 51st Annual Conference.

- McFadden, B.R., Lusk, J.L., Crespi, J.M., Cherry, J.B.C., Martin, L.E. & Bruce, A.S. (2012). *Effect of advocacy information on consumer preferences for cage free eggs: a neuroeconomic analysis.* 12-14th August, Seattle, Washington. Agricultural and Applied Economics Association Annual Meeting.

- McVittie, A., Moran, D. & Nevison, I. (2006). Public preferences for broiler chicken welfare: evidence from stated preference studies, Land Economy Paper Series, Land Economy Research Group, Scottish Agricultural College. Available at: http://www.sac.ac.uk/research/publications/lee/workingpapers.

- Moran, D. & McVittie, A. (2008). Estimation of the value the public places on regulations to improve broiler welfare, *Animal Welfare*, 17(1), pp. 43-52.

- Morten Raun, M., Tove, C. & Dorte, G.H. (2010). Consumer preferences for safety characteristics in pork, *British Food Journal*, 112(7), pp. 775-791.

- Napolitano, F., Pacelli, C., Girolami, A. & Braghieri, A. (2008). Effect of information about animal welfare on consumer willingness to pay for yogurt, *Journal of Dairy Science*, 91(3), pp. 910-917.
- Nocella, G., Hubbard, L. & Scarpa, R. (2010). Farm animal welfare, consumer willingness to pay, and trust: Results of a cross-national survey, *Applied Economic Perspectives and Policy*, 32(2), pp. 275-297.
- Olesen, I., Alfnes, F., Rora, M.B. & Kolstad, K. (2010). Eliciting consumers' willingness to pay for organic and welfare-labelled salmon in a non-hypothetical choice experiment, *Livestock Science*, 127(2-3), pp. 218-226.
- Olynk, N.J., Tonsor, G.T. & Wolf, C.A. (2010). Consumer Willingness to Pay for Livestock Credence Attribute Claim Verification, *Journal of Agricultural and Resource Economics*, 35(2), pp. 261-280.
- Pozo, V.F., Tonsor, G.T. & Schroeder, T.C. (2012). How Choice Experiment Design Affects Estimated Valuation of Use of Gestation Crates, *Journal of Agricultural Economics*, 63(3), pp. 639-655.
- Rolfe, J. (1999). Ethical rules and the demand for free range eggs, *Economic Analysis and Policy*, 29(2), pp. 187-206.
- Satimanon, T. & Weatherspoon, D.D. (2010). Hedonic Analysis of Sustainable Food Products, *International Food and Agribusiness Management Review*, 13(4), pp. 57-74.
- Solgaard, H.S. & Yang, Y. (2011). Consumers' perception of farmed fish and willingness to pay for fish welfare, *British Food Journal*, 113(8-9), pp. 997-1010.
- Tonsor, G.T., Olynk, N. & Wolf, C. (2009a). Consumer preferences for animal welfare attributes: The case of gestation crates, *Journal of Agricultural and Applied Economics*, 41(3), pp. 713-730.
- Tonsor, G.T., Wolf, C. & Olynk, N. (2009b). Consumer voting and demand behavior regarding swine gestation crates, *Food Policy*, 34(6), pp. 492-498.
- Ubilava, D., Foster, K.A., Lusk, J.L. & Nilsson, T. (2010). Effects of income and social awareness on consumer WTP for social product attributes, *Technological Forecasting and Social Change*, 77(4), pp. 587-593.
- Uzea, A.D., Hobbs, J.E. & Zhang, J. (2011). Activists and Animal Welfare: Quality Verifications in the Canadian Pork Sector, *Journal of Agricultural Economics*, 62(2), pp. 281-304.
- Vander Naald, B. & Cameron, T.A. (2011). Willingness to pay for other species' well-being', *Ecological Economics*, 70(7), pp. 1325-1335.

- Viegas, I., Nunes, L.C., Madureira, L., Fontes, M.A. & Santos, J.L. (2014). Beef Credence Attributes: Implications of Substitution Effects on Consumers' WTP', *Journal of Agricultural Economics*, 65(3), pp. 600-615.

- Widmar, O., Nicole, J., Ortega, D.L., McKenzie, A. & Paudel, K. (2014). Comparing Consumer Preferences for Livestock Production Process Attributes Across Products, Species, and Modeling Methods, *Journal of Agricultural and Applied Economics*, 46(03).

- Zanoli, R., Scarpa, R., Napolitano, F., Piasentier, E., Naspetti, S. & Bruschi, V. (2013). Organic label as an identifier of environmentally related quality: A consumer choice experiment on beef in Italy, *Renewable Agriculture and Food Systems*, 28(1), pp. 70-79.

Appendix B: Summary information of included studies

Paper	Year data collection	Number of measures	Country	Sample size	Population	Method	Animal species
Bennett *et al* (2012)	2012	3	UK	300	Citizens	Choice Experiment	Beef cows, pigs, broiler chickens
Bennett & Blaney (2002)	2002	1	UK	164	Students	Contingent Valuation	Pig
Bennett *et al* (2002)	1998	2	UK	119	Students	Contingent Valuation	Animals in general and layer hens
Bennett (1998)	1996	1	UK	591	Citizens	Contingent Valuation	Layer hens
Bennett (1996b)	1996	2	USA	140	Students	Contingent Valuation	Layer hens and calves
Napolitano *et al* (2008)	2008	6	Italy	104	Consumers	Auction	Dairy cows
Moran & McVittie (2008)	2005	1	England	318	Citizens	Contingent Valuation	Broiler chickens
Liljenstolpe (2011)	2002	36	Sweden	1250	Citizens	Choice Experiment	Pig

Burgess & Hutchinson (2005)	2003	5	Northern Ireland	192	Citizens	Contingent Valuation	Layer hens, dairy cows, broiler chickens, pigs
Zanoli *et al* (2012)	2008	4	Italy	145	Consumers	Choice Experiment	Beef cows
Bernard & Bernard (2009)	2005	2	USA	154	Citizens	Auction	Dairy cows
Doherty & Campbell (2014)	2010	6	UK and Republic of Ireland	1173, 400	Citizens	Choice Experiment	Broiler chickens
Elbakidze & Nayga (2012)	2009	8	USA	215	Citizens	Auction	Dairy cows
Lopez-Galan *et al* (2013)	2013	1	Spain	803	Consumers	Choice Experiment	Layer hens
Heid & Hamm (2013)	2009	4	Germany	89	Consumers	Auction	Pig
Grimsrud *et al* (2013)	2010	4	Norway	737	Citizens	Choice Experiment	Salmon
Rolfe (1999)	1997	2	Australia	100, 224	Citizens	Contingent Valuation	Layer hens
Heng *et al* (2013)	2012	4	USA	449	Consumers	Choice Experiment	Layer hens

Uzea *et al* (2011)	2008	20	Canada	541	Consumers	Choice Experiment	Pigs
Loureiro *et al* (2013)	2010	6	Spain	92, 62	Consumers	Auction	Pigs
Olesen *et al* (2010)	2004	1	Norway	84	Consumers	Real choice experiment	Salmon
Lusk & Norwood (2011)	2011	6	USA	126, 134	Citizens	Auction	Layer hens and pigs
Nocella *et al* (2010)	2007	5	France, Germany, Italy, Spain, UK	133, 155, 792, 160, 182	Citizens	Contingent Valuation	General
Lusk *et al* (2007)	2004	2	USA	594	Citizens	Choice Experiment	Pigs
Makdisi & Marggraf (2011)	2007	1	Germany	300	Citizens	Contingent valuation	Broiler chickens
McFadden *et al* (2012)	2012	12	USA	29	Citizens	Contingent Valuation and Conjoint Analysis	Layer hens
Lusk *et al* (2003)	2000	8	France, Germany, UK , USA	93, 45, 109, 566	Citizens	Choice Experiment	Beef cows
Lagerkvist *et al* (2006)	2005	8	Sweden	285	Consumers	Choice	Pigs

						Experiment	
Koistinen *et al* (2013)	2010	28	Finland	1623	Consumers	Choice Experiment	Beef cows
Lusk *et al* (2006)	2004	2	USA	291, 432	Citizens	Conjoint Analysis and Contingent Valuation	Pigs
Connor & Oppenheim (2008)	2007	2	USA	253	Consumers	Contingent Valuation	Beef cows and dairy cows
Bennett & Blaney (2003)	1996	1	UK	446	Citizens	Contingent Valuation	Layer hens
Bennett & Larson (1996)	1996	3	USA	137	Students	Contingent Valuation	Layer hens and calves
Carlsson *et al* (2007a)	2002	2	Sweden	450	Citizens	Choice Experiment	Layer hens
Allender & Richards (2010)	2008	1	USA (California)	993	Consumers	Hedonic Analysis	Layer hens
Carlsson *et al* (2005a)	2003	30	Sweden	710	Citizens	Choice Experiment	Broiler chickens, beef cows, pigs, dairy cows, layer hens
Carlsson *et al* (2007B)	2004	12	Sweden	395, 362	Consumers	Choice Experiment	Broiler chickens and beef cows

Goddard *et al* (2007)	2005	3	Canada (Alberta and Ontario)	292, 248	Citizens	Hedonic Analysis	Layer Hen
Tonsor *et al* (2009a)	2007	10	USA	205	Citizens	Choice Experiment	Pigs
Pozo *et al* (2012)	2012	10	USA	1312	Citizens	Choice Experiment	Pigs
Tonsor *et al* (2009b)	2008	1	USA	768	Citizens	Contingent Valuation	Pigs
Solgaard & Yang (2009)	2009	1	Denmark	1000	Citizens	Contingent Valuation	Fish
Ubilava *et al* (2010)	2004	2	USA	197	Citizens	Choice Experiment	Pigs
Olynk & Ortega (2014)	2011	12	USA	500, 399	Citizens	Choice Experiment	Dairy cows, pigs
Andersen (2011)	2000	2	Denmark	844	Consumers	Choice model	Layer hens
McVittie *et al* (2006)	2005	9	England	336, 318	Citizens	Choice Experiment and Contingent Valuation	Broiler chickens
Lu (2013)	2013	14	Canada	518	Citizens	Choice	Layer hens

						Experiment	
Viagas *et al* (2014)	2013	1	Portugal	613	Consumers	Choice Experiment	Beef cows
Vander Naald *et al* (2011)	2006	2	USA	240	Citizens	Conjoint Analysis	Broiler chickens
Satimanon & Weatherspoon (2010)	2007	1	USA	207	Consumers	Hedonic pricing	Layer hens
Olynk *et al* (2010a)	2008	32	USA	669	Citizens	Choice Experiment	Pigs
Morbak *et al* (2010)	2006	2	Denmark	1322	Consumers	Choice Experiment	Pigs
Chung *et al* (2009)	2007	4	Korea	1000	Consumers	Choice Experiment	Beef cows
Christensen *et al* (2006)	2005	2	Denmark	2301	Consumers	Choice Experiment	Broiler chickens

1237

1238

Appendix C: Critical appraisal tool

Considerations				
Type of Study		**Standard**	**Criteria**	**Risk of Bias***
Revealed Preference Measures	Hedonic	Normal	Linear models, criteria based on previous research	Moderate
		Good	Panel data or scanner data, criteria based on previous research	Low
	Real Choice Experiment	Normal	If two or more of the below are not mentioned	Moderate
		Good	Reminded budget constraints, provided with information,	Low
	Auctions	Normal	Reminded budget constraints, practice in the method beforehand	Moderate
		Good	BDM Lottery/ Vickrey auctions, participants trained/ practiced in the method, made clear that participants have a commitment to buy, reminded budget constraints	Low
Stated Preference Measures	Conjoint Analysis	Normal	If two or more of the below are not mentioned	Moderate
		Good	Opt out provided, reminded budget constraints, prior qualitative research, cheap talk script used	Low
	Choice Experiment	Normal	Cheap talk script not used, non-orthogonal design, prior qualitative research	Moderate
		Good	Use of a cheap talk script, orthogonal/ main effects design, cyclical or full/ fractional factorial design maintaining orthogonality, D-optimal design, prior qualitative research	Low
	Contingent Valuation Methods	Normal	Open-ended choice, if more than two of the below are not mentioned	Moderate
		Good	Single bounded, one and one half bounded, double bounded dichotomous choice, reminded budget constraints, realistic choices, cheap talk script , payment card or payment scale	Low
Modelling Used				
Revealed Preference Measures	Hedonic	Normal	Linear models	Moderate
		Good	Semi log model, log-log models, non-linear models	Low
	Real Choice Experiment	Normal	Multinomial and mixed multinomial logit models, probit model, conditional logit model, descriptive statistics	Moderate
		Good	Random parameter logit (mixed logit model), latent class model, nested logit model, random co-efficient logit model	Low
	Auctions	Normal	Descriptive or multivariate statistics, ordinary least squares	Moderate
		Good	Tobit model, random effects tobit model, log-linear model	Low
Stated Preference	Conjoint	Normal	Probit model, ordinary least squares, non-linear least squares	Moderate

Measures	Analysis	Good	Logit model	Low
	Choice Experiment	Normal	Multinomial logit model, conditional logit model	Moderate
		Good	Mixed logit model (random parameter logit, WTP-Space model, random co-efficient logit model, mixed multinomial logit model), latent class model, multinomial probit model	Low
	Contingent Valuation Methods	Poor	Descriptive statistics	High
		Normal	Binary logit model (binomial logit model), probit model	Moderate
		Good	Ordered probit model, tobit model	Low
Directness				
Population	General citizen or consumer	Good	Looking at consumers of specified product(s) or wider citizens in general	Low
	Specific population	Normal	Appropriateness of the sample for the product in question (relevance and representativeness)	Moderate
Representativeness	Quota or stratified sample	Good	The sample used is representative of the chosen population in most aspects, national sales data, clustered sampling technique	Low
		Normal	The sample used is representative of the chosen population in only one aspect, regional sales data	Moderate
	Simple random or systematic sample	Normal	A randomly sample of the chosen population, convenience sample, supermarket sample	Moderate

*Where no information is provided, or it is unclear, the risk of bias will be rated as high for that given aspect

NB) Novel or modified methods will be assessed based on the criteria of similar traditional methods e.g. Calibrated Auction Conjoint Method and Matching Method

Appendix D: Critical appraisal of included studies

Study	Causation						Directness			Overall Risk of Bias
	Type	Rating	Risk of Bias	Model	Rating	Risk of Bias	Population	Representative	Risk of Bias	
Bennett *et al* (2012)	Choice Experiment	Normal	Moderate	Mixed Logit Model	Good	Low	Citizens	Stratified, quota sample contacted randomly	Low	Low
Bennett & Blaney (2002)	Contingent valuation	Normal	Moderate	Binary Logit Model	Normal	Low	Students	Random and convenience	Moderate	Moderate
Bennett *et al* (2002)	Contingent valuation	Normal	Moderate	Binary Logit Model	Normal	Moderate	Students	Unclear how chosen	Moderate	Moderate
Bennett (1998)	Contingent valuation	Normal	Moderate	Descriptive Statistics	Poor	High	Citizens	Random stratified sample by socio-economic characteristics	Low	High
Bennett (1996)	Contingent valuation	Normal	Moderate	Probit Model	Normal	Moderate	Students	Described as *a sample*	Moderate	Moderate
Napolitano *et al* (2008)	Auction	Good	Low	Multivariate Statistics	Normal	Moderate	Consumers	Screening criteria, mostly students	Moderate	Moderate
Moran & McVittie (2008)	Contingent valuation	Normal	Moderate	Binary Logit Model	Normal	Moderate	Citizens	Stratified for age and gender, no sampling method	Moderate	Moderate

Liljenstolpe (2011)	Choice Experiment	Normal	Moderate	Latent Class model	Good	Low	Citizens	Unclear due reporting, sample obtained from database	Moderate	Moderate
Burgess & Hutchinson (2005)	Contingent valuation	Normal	Moderate	Binary Logit Model	Normal	Moderate	Citizens	random sample using electoral register	Low	Moderate
Zanoli *et al* (2012)	Choice Experiment	Good	Low	Multinomial Logit Model, Multiple Variations of Random Parameter Logit Model (inc. WTP-Space Model)	Good	Low	Consumers	Quota sample of beef consumers, and occassional organic purchasers	Moderate	Low
Bernard & Bernard (2009)	Auction	Normal	Moderate	Descriptive Statistics	Normal	Moderate	Citizens	Random sample using ads, local organizations	Moderate	Moderate
Doherty & Campbell (2014)	Choice Experiment	Normal	Moderate	Latent Class Model	Good	Low	Citizens	Representative sample collected by research agency	Low	Low
Elbakidze & Nayga (2012)	Real Choice Experiment	Normal	Moderate	Descriptive Statistics	Normal	Moderate	Citizens	Mostly students, no real information on	Moderate	Moderate

								sampling procedure		
Elbakidze & Nayga (2012)	Auction	Normal	Moderate	Descriptive Statistics	Normal	Moderate	Citizens	Mostly students, no real information on sampling procedure	Moderate	Moderate
Gracia *et al* (2014)	Choice Experiment	Good	Low	Latent Class model	Good	Low	Consumers	Stratified, random sample of consumers	Low	Low
Heid & Hamm (2013)	Auction	Normal	Moderate	Descriptive statistics	Poor	High	Consumers	Quota sample, recruitment strategy unclear	Moderate	High
Grimsrud *et al* (2013)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Citizens	Stratified and random sampling to be representative	Low	Low
Rolfe (1999)	Contingent valuation	Normal	Moderate	Binomial Logit Model	Normal	Moderate	Citizens	A random sample of households	Moderate	Moderate
Heng *et al* (2013)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Consumers	Nationally representitive, stratified sample. Use screening questionnaire	Low	Low
Uzea *et al* (2011)	Choice Experiment	Normal	Moderate	Latent Class Model	Good	Low	Consumers	Representative sample collected by	Low	Low

								research agency		
Loureiro *et al* (2013)	Auction	Normal	Moderate	Random Effects Tobit Model	Good	Low	Consumers	Random stratified procedure by age	Moderate	Moderate
Olesen *et al* (2010)	Real Choice Experiment	Good	Low	Random Parameter Logit Model	Good	Low	Consumers	Participants recruited locally is only information	High	Moderate
Lusk & Norwood (2011)	Auction	Good	Low	Descriptive Statistics	Normal	Moderate	Citizens	Randomly recruited by market research companies	Moderate	Moderate
Nocella *et al* (2010)	Contingent valuation	Good	Low	Binary Logit Model	Normal	Moderate	Citizens	Randomly recruited online	High	High
Lusk *et al* (2007)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Citizens	Random sample matched to be representative	Low	Low
Makdisi & Marggraf (2011)	Contingent valuation	Normal	Moderate	Binomial Logit Model	Normal	Moderate	Citizens	Unclear	High	High
McFadden *et al* (2012)	Contingent valuation	Normal	Moderate	Descriptive Statistics	Poor	High	Citizens	A quota sample	Moderate	High
McFadden et al (2012)	Conjoint analysis	Unclear	High	Non-Linear Least Squares	Normal	Moderate	Citizens	A quota sample	Moderate	High

Lusk *et al* (2003)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Citizens	Representative sample accessed through private companies	Low	Low
Lagerkvist *et al* (2006)	Choice Experiment	Good	Low	Binary Logit Model	Normal	Moderate	Consumers	Random sample of the census registry	Low	Low
Koistinen *et al* (2013)	Choice Experiment	Normal	Moderate	Conditional Logit Model and Latent Class Model	Good	Low	Consumers	Representative of internet users	Moderate	Moderate
Lusk *et al* (2006)	Choice based conjoint analysis	Normal	Moderate	Multinomial Logit Model	Normal	Moderate	Citizens	Participants recruited in a grocery store	Moderate	Moderate
Connor & Oppenheim (2008)	Contingent valuation	Unclear	High	Tobit Model	Good	Low	Consumers	Sample of consumers from outside supermarket	Moderate	High
Bennett & Blaney (2003)	Contingent valuation	Good	Low	Binary Logit Model	Normal	Moderate	Citizens	Random stratified sample by socio-economic characteristics	Low	Low
Bennett & Larson (1996)	Contingent valuation	Normal	Moderate	Binary Logit Model	Normal	Moderate	Students	Convenience sample of students	Moderate	Moderate
Carlsson *et al* (2007b)	Choice Experiment	Good	Low	Random Parameter	Good	Low	Citizens	Sample from census date	Moderate	Low

				Logit Model						
Allender & Richards (2010)	Hedonic Pricing	Good	Low	Random Parameter Logit Model	Good	Low	Consumers	From homescan database	Low	Low
Carlsson *et al* (2005)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Consumers	Random sample from census registry	Low	Low
Carlsson et al (2007a)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Consumers	Random sample from census registry	Low	Low
Goddard *et al* (2007)	Hedonic Pricing	Normal	Moderate	Conditional Logit Model	Normal	Moderate	Consumers	From homescan database	High	High
Goddard *et al* (2007)	Choice Experiment	Unclear	High	Conditional Logit Model	Normal	Low	Citizens	Unclear - a market research agency was used	High	High
Tonsor *et al* (2009a)	Choice Experiment	Good	Low	Random Parameter Model and Latent Class Model	Good	Low	Citizens	Unclear, used a market research agency	Moderate	Low
Pozo *et al* (2012)	Choice Experiment	Good	Low	Multinomial Logit Model and Random Parameter Model	Good	Low	Citizens	Representative sample using market research agency	Low	Low

Tonsor et al (2009)	Contingent valuation	Normal	Moderate	Binary Logit Model	Normal	Moderate	Citizens	Representative sample using panel data	Low	Moderate
Solgaard & Yang (2009)	Contingent valuation	Unclear	High	Binomial Logit Model	Normal	Moderate	Citizens	Representative sample using panel data	Low	High
Ubilava *et al* (2010)	Choice Experiment	Unclear	High	Random Parameter Logit Model	Good	Low	Citizens	Representative sample	Moderate	High
Olynk & Ortega (2014)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Citizens	Representative sample using panel data	Low	Low
Andersen (2011)	Choice Model	Good	Low	Mixed Multinomial Logit Model	Normal	Moderate	Consumers	Retail purchase data	Low	Low
McVittie *et al* (2004)	Choice Experiment	Moderate	Normal	Multinomial Logit Model	Normal	Moderate	Citizens	Stratified sample	Moderate	Moderate
McVittie *et al* (2004)	Contingent valuation	Moderate	Normal	Binary Logit Model	Normal	Moderate	citizens	Stratified sample	Moderate	Moderate
Lu (2013)	Choice Experiment	Moderate	Normal	Conditional Logit Model	Normal	Moderate	Citizens	Unclear	High	High
Viagas *et al* (2014)	Choice Experiment	Good	Low	Random Parameter Logit Model	Good	Low	Consumers	Stratified random sample	Low	Low

Vander Naald *et al* (2011)	Conjoint Analysis	Normal	Moderate	Conditional Logit Model	Normal	Low	Citizens	Chosen from jury register	Moderate	Moderate
Satimanon & Weatherspoon (2010)	Hedonic Analysis	Good	Low	Hedonic Model	Unclear	High	Consumers	Sales data	Low	Moderate
Olynk *et al* (2010a)	Choice Experiment	Normal	Moderate	Random Parameter Logit Model	Good	Low	Citizens	Representative, used a market research agency	Low	Low
Morbak & Norstrom (2009)	Choice Experiment	Good	Low	Random Parameter Logit Model	Good	Low	Consumers	Recruited online through an agency	Moderate	Low
Chung *et al* (2009)	Choice Experiment	Normal	Moderate	Multinomial Logit Model and Mixed Multinomial Logit Model	Normal	Moderate	Consumers	Convenience sample of supermarket shoppers	Moderate	Moderate
Christesen *et al* (2006)	Choice Experiment	Good	Low	Multinomial Probit Model	Good	Moderate	Consumers	Panel data, random quota sample	Low	Low

1246

1247

Appendix E: Summary of evidence

Outcome	Number of Studies (number of measures)	Risk of Bias	Imprecision	Inconsistency	Indirectness	Publication Bias	Overall Strength of Evidence	Effect Magnitude (95% CI)	Adjusted WTP
Complete Case Studies	37	Moderate	Moderate	High	Moderate	Strongly Suspected	Low	0.6302 (0.5016, 0.7587)	€ 0.27
Overall (Complete case and imputed value studies)	54	Moderate	Moderate	High	Moderate	Strongly Suspected	Low	0.5709 (0.4599, 0.6819)	€ 0.34
Aggregated (by paper)	54	Moderate	Low	High	Moderate	Strongly Suspected	Low	0.4690 (0.2075, 0.7036)	€ 0.92
Pigs	13(90)	Low	Moderate	High	Moderate	Strongly Suspected	Low	0.2843 (0.1936, 0.3750)	€ 0.54
Layer Hens	10 (47)	Moderate	Moderate	Moderate	Moderate	Strongly Suspected	Moderate	0.7823 (0.3594, 1.2053	€ 0.09
Broiler Chickens	8 (26)	Moderate	Moderate	High	Low	Strongly Suspected	Low	0.4024 (0.2653, 0.5394)	€ 1.24
Dairy Cows	7 (27)	Moderate	High	High	Moderate	Strongly Suspected	Low	1.2276 (0.7776, 1.4575)	€ 0.50
Beef Cows	7 (24)	Low	High	High	Low	Strongly Suspected	Low	1.022 (0.7294, 1.6750)	€ 5.00
Multiple	2 (6)	High	Moderate	Moderate	Low	Strongly Suspected	Low	0.6547 (0.4206, 0.8888)	€11.20
Calves	1 (1)	Moderate	High		High		Low	4.8344 (4.6526, 5.0162)	€8.69

Fish	3 (6)	Moderate	Low	Moderate	Low	Strongly Suspected	Moderate	0.3712 (-0.0073, 0.7497)	€3.53
UK	7 (27)	Moderate	High	High	Low	Strongly Suspected	Low	0.6479 (0.5113, 0.7845)	€ 1.72
Northern Europe	8 (76)	Moderate	Moderate	Moderate	Low	Strongly Suspected	Moderate	0.1060 (0.0376, 0.1744)	€ 0.41
Western Europe	3 (7)	High	High	Moderate	Low	Strongly Suspected	Low	1.0741 (0.7720, 1.7630)	€ 4.28
Southern Europe	6 (23)	Moderate	High	High	Low	Strongly Suspected	Low	1.4329 (0.9577, 1.9082)	€ 0.68
Asia	1 (4)	Moderate	High		High		Low	2.3820 (2.0842, 2.6799)	€ 5.40
North America	16 (90)	Moderate	High	Moderate	High	Strongly Suspected	Low	0.7515 (0.5026, 1.0004)	€ 0.15
Citizens	26 (187)	Moderate	Moderate	Moderate	Moderate	Strongly Suspected	Moderate	0.5122 (0.3810, 0.6435)	€ 0.33
Consumers	11 (40)	Low	High	High	Moderate	Strongly Suspected	Low	1.1796 (0.8287, 1.5304)	€ 0.25
Age	24 (156)	Moderate	Low	High	Moderate	Undetected	Moderate	-0.0377 (-0.0530, -0.0224)	€ 0.19
Income	19 (123)	Moderate	Low	High	Moderate	Undetected	Moderate	0.0207 (0.0131, 0.0284)	€ 0.11
% Female	26 (157)	Moderate	Low	Moderate	Moderate	Undetected	Moderate	0.0086 (-0.0002, 0.0175)	€ 0.19

% Vegetarian	6 (59)	Low	Low	Moderate	Low	Undetected	High	-0.7024 (-0.9654, -0.4394)	€ 0.25
Education	15 (74)	Moderate	Low	High	Moderate	Strongly Suspected	Low	0.0246 (0.0113, 0.0379)	€ 0.17
Revealed Preference	8 (31)	Moderate	High	High	Moderate	Undetected	Low	1.1935 (0.9077, 1.4793)	€ 0.41
Stated Preferences	29 (196)	Moderate	Moderate	High	Moderate	Strongly Suspected	Low	0.5416 (0.4035, 0.6796)	€ 0.25
Cheap talk script used	9 (83)	Low	Low	Moderate	Low	Strongly Suspected	Moderate	0.3595 (0.2259, 0.4932)	€ 0.11
Cheap talk script not used	20 (113)	Moderate	Moderate	Moderate	Moderate	Strongly Suspected	Low	0.6758 (0.4600, 0.8916)	€ 0.50
Cheap talk script not needed	8 (31)	Moderate	High	High	Moderate	Undetected	Low	1.1935 (0.9077, 1.4793)	€ 0.41