A systematic review and meta-analysis of tenderness metrics in control groups used in comparative nutrition experiments[1,2]

M. J. Powell,* C. Yuan,† R. S. Dzikamunhenga,‡ R. Tarté,* E. Huff-Lonergan,* S. M. Lonergan,* and A. M. O'Connor‡[3]

*Department of Animal Science, Iowa State University, Ames 50011; †Department of Statistics, Iowa State University, Ames 50011; and ‡Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University, Ames 50011

ABSTRACT: This review reports the pork quality attributes, Warner-Bratzler Shear Force, Slice Shear Force, Star Probe, pH, marbling, color (Minolta L*/L or Hunter L*/L), and sensory tenderness evaluation, in control groups used in comparative nutrition experiments over the past 20 yr. The original aim of this study was to evaluate if changes in pork quality based on the above metrics occurred over time. To address this question, it was anticipated that data may come from 3 sources with decreasing relevance: representative retail pork surveys, representative post-harvest carcass surveys, and control groups from comparative nutrition experiments. To identify the study population, a review of studies reported in Centre for Agricultural Biosciences International Abstracts (Web of Knowledge; 1994–2014) was conducted. Two national level surveys of retail pork and 146 relevant nutritional experiments studies, with 228 control groups, were identified by the search. It was not possible to conduct a meta-analysis of the retail pork surveys based on only 2 time points. For the comparative studies, a random effects meta-analysis was conducted with year as a covariate to assess the impact of time on the outcome. In the absence of modifiers, there was no evidence of meaningful change in the mean Warner-Bratzler Shear Force, pH, color, marbling, or sensory scores over the study period. There was evidence of substantial between-study heterogeneity in the characteristics of control pigs used over the years for Warner-Bratzler Shear Force and measures of color. The absence of publicly-available representative surveys of pork quality meant the changes in pork quality over time were not clear. If changes in pork quality have occurred, the data suggest that pigs used as controls in experiments may have become less representative of commercial pigs over time and the translatability of study findings from nutrition experiments might be reduced over time. Alternately, if commercial pigs have not changed, then control pigs reflect this. The study does not address if control groups in other experimental intervention studies had similar tenderness patterns as reported here for nutritional interventions. A large amount of potentially available data was excluded from the analysis due to incomplete reporting in the original study reports.

Key words: meta-analysis, pork, quality, systematic review, tenderness

Transl. Anim. Sci. 2017.1:261–276
doi:10.2527/tas2017.0031

[1]This research was funded by a grant from the National Pork Board (NBP Project # 14 -233).

[2]Authorship declarations: MJP prepared the protocol, assisted with the conduct of the search, screening and data extraction, prepared the interpreted the results, and prepared the report for publication. CY assisted with the analysis of the data including all figures. RSD conducted data extraction and assisted with report preparation. EHL and SML provided expertise for designing protocol, addressing eligibility issues and model design during the conduct of the review, interpretation of the results. RT provided comments on the interpretation of the results. AMOC conceived and coordinated the project conduct, coordinated development of the protocol, coordinated the search, supervised the screening, data extract and analysis, interpreted the data, and assisted in report preparation. All authors read the final draft and approved the final manuscript for the review. The authors have no conflict of interests relevant to this topic.

[3]Correspondence email: oconnor@iastate.edu

Received April 12, 2017.

Accepted June 6, 2017.

INTRODUCTION

The degree to which consumers appreciate fresh pork quality affects demand for pork, and therefore, value of fresh pork. Several metrics of retail fresh pork are measured as indicators of tenderness and overall quality including: Warner-Bratzler Shear Force (WBSF), pH, marbling scores, subjective color, instrumental color (conducted using Minolta or Hunter technology), and sensory tenderness evaluation (Bray, 1966; Huff-Lonergan et al., 2002; Choe et al., 2016). How these factors interact to provide the consumer with a satisfying dining experience is a topic of constant research. As consumers are increasingly provided with a wide variety of protein sources, it is critical that pork is a consistently high quality product. The importance of quality metrics to the consumer and the absence of industry wide surveys of pork quality metrics has caused some to hypothesize that pork tenderness may have decreased over the years with the focus on lean production. Therefore, it was of interest to determine if changes in pork quality metrics occurred over the past 20 yr.

The objective of this review was to gather information regarding tenderness metrics of pork over the past few decades. To address this objective, a review of primary literature published between 1994 and 2014 was conducted to determine if there was evidence of change in these metrics during that period. The goal was to find representative retail pork surveys, representative post-harvest surveys, and control group pigs used in comparative nutrition studies to evaluate trends in tenderness metrics over time. The latter group was selected based on the concept that such experiments are designed to translate interventions to commercial production, and, therefore, the control group "should" represent the commercial pigs over time.

The relevant population of interest was US and Canadian produced pork loin. The null hypothesis tested was, "There is no evidence that pork quality metrics have decreased or increased between 1994 and 2014". The goal of this review was to assess the following tenderness metrics: WBSF, pH, marbling, instrumental color, and sensory tenderness evaluation.

MATERIALS AND METHODS

No animals where used in this study , as we used secondary data, therefore institutional animal care and use approval was not required.

Protocol

A review protocol documenting the intended eligibility criteria, information sources, search strategy, study selection, and data collection process (including a draft of the data extraction form) was developed prior to the conduct of the review (Supplemental Material 1).

Studies of Interest

The population of interest was pork cuts from the Longissimus muscle described in publications as follows: Longissimus dorsi, Longissimus lumborum, Longissimus thoracis, loin, loin chops, chop, Longissimus muscle (LM), and not ground, minced, or sausage, from gilts or barrows, not sows or boars.

Relevant surveys were limited to those that described a random selection process for sample selection and obtained samples from retail stores or conducted in abattoirs on carcasses. Comparative nutrition experiments, in which gilts or barrows, were raised in the United States or Canada under conventional production systems, not niche, organic, or outdoors, and slaughtered at or above 100 kg body weight were also considered relevant. It was assumed carcasses were hung by the digital flexor and extensor tendons, unless otherwise stated. Diet was considered a control diet if it met or exceeded recommended nutrient requirements and/or it did not include any additional ingredients not found in a typical commercial diet. Pigs fed a diet containing ractopamine and/or ethanol co-products were excluded because of the variation in use during the study period. The pigs were considered a conventional breed as long as they were not a heritage breed, for example, Mangalitsa, or a foreign breed, for example, Polish Landrace. The housing of the pigs was to follow a conventional indoor system, either in research or industry locations, providing a similar environment to what is used in industry. As such information is often not reported, it was decided instead that studies that specially referenced outdoor, free range, or any other niche form of housing were excluded from the review. Studies that did not mention these factors where otherwise considered likely to be conventional and therefore relevant.

Outcome Measures

The outcomes of interest were measurements related to quality and tenderness: WBSF, Slice Shear Force (SSF), Star Probe (SP), pH, marbling scores, color (Minolta L*/L or Hunter L*/L), and trained sensory tenderness evaluation.

Information Sources

The electronic database interface Web of Science was used to identify articles used in this review. The search was limited to publications cited in the Centre for Agricultural Biosciences International: Centre for Agricultural Biosciences Abstracts and Global Health, which was expected to have almost full coverage of published literature available for the topic of interest.

Search

The search string used to find the literature was {TS = [Pork AND (Tenderness OR Intramuscular fat OR quality OR Warner-Bratzler OR "Shear Force" OR "Sensory panel" OR marbling*)]}. The search was limited from 1994 to 2014 and conducted in January 2015. The initial protocol did not include pH or color, however, after review of the initial results, it was concluded that these were metrics of interest. To address this, the search was repeated on June 29, 2015, adding "color OR pH" to the search string. Search results were downloaded into bibliographic management software (Endnote X7.5.1, Thomas Reuters, 1988 to 2016). Endnote group screening was used to remove non-English language abstracts, duplicates, and abstracts that described assessing ham, sausages, patties, and/or salami. When the results of a study were available in more than 1 source, all sources were used to provide the most complete data for extraction. Final citations were uploaded into DistillerSR systematic review management software (DistillerSR, Evidence Partners, Ottowa, Canada) for relevance screening and data extraction.

Approach to Selecting Studies Relevant to the Review Question

Two levels of relevance screening were used: the first level was based only on the title and abstract and the second level was based on the full text. The screening questions are in Supplemental Material 2. When authors did not report the location of the study, it was assumed the study was conducted in the USA or Canada if all the authors reported were USA or Canada based. Funding sources from outside the USA and Canada, and country specific journals were considered evidence of non-relevance. During the search for surveys, the funding agency made us aware of 2 retail pork surveys they considered potentially relevant (Wright et al., 2005; Newman, 2014).

Study Selection and Data Collection Process

Two independent reviewers conducted the study selection process using pretested screening and data extraction forms. After screening a small subset of studies, to ensure agreement and similar understanding of the eligibility criteria, reviewers discussed and resolved any conflicts and screening proceeded until all citations were completed. For the data extraction process, a similar procedure was used and conflicts were resolved by discussion. After data were extracted from a series of papers, all previously excluded abstracts were rescreened as the review team had better familiarity with relevant manuscripts.

Data Items Extracted from Relevant Papers

The data extraction form is provided in Supplemental Material 3. The data items extracted were population characteristics based on the pigs in the control group: weight at slaughter, Halothane (Hal) gene status, Rendement Napole gene (RN) status, and sex. Hal and RN gene statuses were noted as present when authors noted when the pigs where Hal or RN positive. Otherwise it was assumed animals were negative if not specifically stated in the publication, as that was industry standard. Instrumental measures of tenderness of interest were WBSF reported in kilograms, SSF reported in kilograms, or SP reported in kilograms. Outcomes measured as pounds or Newtons were converted where applicable. Final cooking temperature was extracted in °C or converted to °C from F. The pH at 24-h minimum was extracted. Color, reported as a subjective color score, was extracted when following National Pork Producer's Council (NPPC) color standards along with the range scale used. When color was measured with an instrument, Minolta or Hunter (L*/L values representing lightness, 0 = black and 100 = white), light source, observer angle, and aperture size data were extracted when reported. Sensory tenderness evaluation data and scales used were extracted only if trained panelists conducted the evaluation.

Measures of precision were extracted, when reported, for all outcomes, and standard deviation (SD) was converted to standard error of the mean (SEM). Two publications reported Residual Standard Deviation, but it was not clear if this could be translated to a standard error for the control group, so these were not converted, and therefore, excluded from the meta-analysis.

Summary Effect Size

For WBSF, pH, Minolta, or Hunter L*/L values, the control group mean was obtained. For marbling and tenderness sensory analyses, to account for scale differences, the measures were standardized as follows: standardized mean = [(observed mean- scale minimum)/scale range] and standardized SEM = (observed SEM/scale range).

Synthesis of Results

The goal of the analysis was to evaluate the effect of year on the quality metric outcome; therefore, meta-regression with year as a covariate was conducted for each outcome. Year of study was inconsistently reported, and therefore, publication year was used as a proxy for study year.

For all analyses, a random effects model was used with a maximum likelihood estimation method. The outcome (y_i) was the control group mean. The weighting variable was the SEM from each control group. Heterogeneity was assessed using a Q test which has the null hypothesis that the observed variation between studies is that expected to arise from chance. The heterogeneity was also using the I^2 statistic which describes the proportion of the variation in observed effects is due to variation in true effects, then (by definition) the proportion of this variation that would remain if no sampling error occurred. (Higgins and Thompson, 2002; Borenstein et al., 2017). The rationale for both metrics is that was is a hypothesis test while the other is an "effect size" and therefore they provide different information. A random effect for study was used to account for the non-independence of control groups from the same publication. The fixed effects included in each model varied based on the outcome metric and are described below. The analysis was conducted in the R package metaphor (Viechtbauer, 2010).

As the meta-analysis used observational data and was planned without a power analysis, the power to detect a year effect was unknown, therefore, in preference to hypothesis testing the year effect, it was reported and inference made based on the point estimate and precision of the year effect ($\beta_{\text{publication year}}$). For each metric, when the amount of total variation attributed to among study effects was below a set threshold, $I^2 < 50\%$, the observed study effect and the model predicted effect against the year was plotted in a forest plot, and reported the summary effect and its confidence interval. When the amount of total variation was $I^{2\ 3}$ 50%, a predicted effect or the summary effect was not reported. There is debate about heterogeneity of the effect size in meta-analyses; the approach used in this study was not to present summary effect size, when there was evidence of notable heterogeneity, others may reach the conclusion that the heterogeneity was adjusted for by the use of random effects model (Higgins and Thompson, 2002). The rationale for assessing heterogeneity is to determine if a pooled summary effect size is meaningful, if the data suggest that we can reject the hypothesis of homogeneity of effect sizes then pooling the study effects in a single summary number is potentially not meaningful.

For the meta-analysis, a change was made to the protocol to include only Hal and RN gene negative pigs. This change meant that factorial studies that assessed main effects for genetic variables and found them nonsignificant were excluded if the resulting group means combined Hal+ with Hal- animals or RN+ with RN- animals. This represents an additional eligibility criteria not specified in the original protocol.

The dataset used for the analysis of WBSF was limited to control groups that met the following criteria: used a final cooking temperature between 68 °C and 72 °C, reported a mean and SEM for WBSF (or SEM could be calculated), the pH of the sample, and publication year. Fixed effects of interest were pH and publication year.

The outcome of pH dataset for analysis was limited to those control groups that met the following criteria: reported sex of the control group, reported a mean and SEM for the average pH of the group, and publication year. Fixed effects of interest were sex and publication year.

The color outcome, based on Minolta L* or Minolta L system or the Hunter L* or Hunter L system, dataset for analysis was limited to those control groups that met the following criteria: reported sex of the control group, reported mean and SEM for the average color of the group, publication year and pH. Fixed effects of interest were pH and publication year. Ideally, light source, observer angle, and aperture would have been used in the model, but could not be included due to lack of reporting. Therefore, conversions could not be made from L to L* values and separate meta-analyses were conducted for the color outcomes. Meta-analysis was not conducted on subjective color (NPPC color standards). The marbling and sensory scale data analysis was conducted on the standardized data, provided a SEM was available. The meta-regression did not include any covariates other than publication year.

RESULTS

Study Selection

The entire search yielded 9,006 citations. After adjusting for language and select terms that were unlikely to yield relevant citations (Fig. 1), 4,760 citations remained. Of these, 4,254 citations were discarded when abstracts were reviewed. After evaluating the full text, a further 360 citations were discarded for reasons indicated in Fig. 1. Two relevant surveys of retail pork were identified by the National Pork Board (Wright et al., 2005; Newman, 2014) and no others were found. The study by Wright et al. (2005) was found by the search, though not the study by Newman (2014), however, as it was not available as a peer reviewed publication, this was not surprising. No publically available surveys of carcasses were identified. One hundred and forty-six comparative studies met the inclusion criteria, however, not all studies reported all outcomes, and therefore the number of studies used for each analysis varied and is reported in the corresponding section.

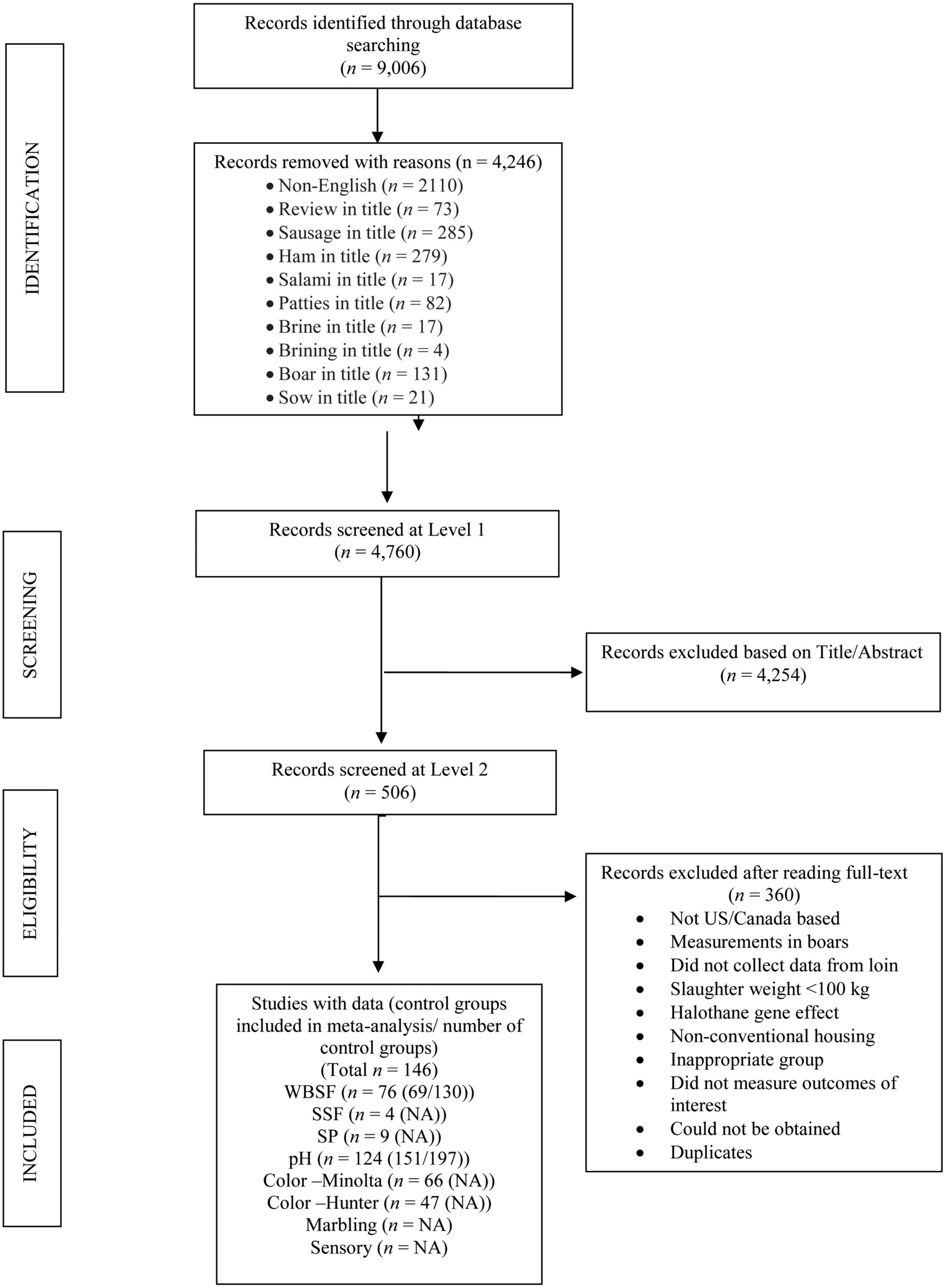

Figure 1. Flow diagram showing study identification and selection process.

Study Characteristics

The 2 retail pork surveys were conducted 10 yr apart in 2002 and 2012. In the 2002 survey, mean WBSF was 2.96 kg (SD = 7.69, *n* = 228), mean pH was 5.64 (SD = 0.26, *n* = 229), mean NPPC color survey was 3.52 (SD = 0.85, *n* = 600), Hunter L* was 48.07 (SD = 5.59, *n* = 599; Wright et al., 2005). In the 2012 survey, mean WBSF was 2.38 kg (SD = 0.68, *n* = 1910), mean pH was 5.87 (SD = 0.3, *n* = 1817), mean NPPC color survey (1 through 6) was 3.12 (SD = 0.85, *n* = 2795), Minolta L* was 55.30 (SD = 3.70, *n* = 1705). Both surveys used enhanced and unenhanced products. Newman (2014) did not make a comparison of the 2 surveys other than for subjective measures, color, and marbling, and concluded, "These results suggest that subjective pork quality attributes observed in the retail meat case are fairly consistent with what was observed previously by Wright et al. (2005)."

The 146 studies yielded 230 control groups. Several published studies reported multiple relevant groups. After reviewing the data, 2 control groups were excluded because these reported color data on the ventral side of the loin and the cut side (Lowe et al., 2014). It was decided by the co-authors that only the cut side data was applicable; thus, the final number of control groups was 228. Twenty-two studies reported the year the study was conducted.

Results of Individual Studies

Because of the number of control groups included (*n* = 228), it was not possible to present the individual results for each control group in a table. The frequency of distribution of characteristics among the studies is shown in Table 1. The different groupings of breeds, crosses, or genetic lines, as well as diets, are shown in Table 2. The manuscripts reporting each outcome are included in Supplemental Material 4. No study explicitly reported the method by which carcasses were hung, therefore, it was assumed that carcasses were hung by the digital flexor and extensor tendons, which is the industry-typical practice.

After review of the analysis, it was confirmed that 2 of the studies, on which authors of this review where co-authors, Patton et al., 2008a; Patton et al. (2008b), had mislabeled precision measures as SEM when they were actually SD, for weight, SP, pH, color, and marbling. These were corrected post-hoc. Nineteen studies (30 control groups) reported the year the study was conducted, hence the decision to use publication year as the measure of year. Eighty-two of the 146 publications (56%), with 105 control groups, reported using random allocation methods to assign animals to treatment group, otherwise it was not documented how animals were assigned.

Table 1. Frequency of reporting characteristics by the control groups (*n* = 228)

Variable	Frequency	Percent
Date of publication		
1994–1995	5	2
1996–2000	24	10
2001–2005	83	36
2006–2010	74	32
2011–2014	41	18
Not discernible	1	0
Year of study		
1996–2000	11	5
2001–2005	5	2
2006–2010	14	6
2011–2014	1	0
Not reported	198	86
Sex		
Barrows only	52	23
Gilts only	31	14
Gilts and barrows	127	56
Not reported	18	8
Setting		
Industry farm	35	15
Research farm	73	32
Industry and research farm	1	0
Not discernible	119	52
Random allocation to group		
Yes	105	46
No, not reported, not discernible	123	54
Halothane status		
Negative, assumed negative	221	97
Carrier, Carrier and Negative	7	3
RN status		
Negative, assumed negative	227	100
Some expressing	1	0
Reported weight		
Slaughter weight/ending weight/ final weight	172	75
Measured WBSF	130	57
Measured Slice Shear Force	4	2
Measured Star Probe	9	4
Measured pH	196	86
Color		
Minolta L*	116	51
Hunter L*	70	31
Other or NR	41	18
Marbling	156	68

Outcomes: Warner-Bratzler Shear Force (WBSF)

One hundred and thirty control groups (from 76 studies) assessed WBSF. Sixty-nine control groups (from 38 studies) were eligible for inclusion in the meta-analysis because they reported relevant covariates needed for the model (Supplemental Material 4). In the final meta-regression model that included publication year and pH, neither moderator was found to be

Table 2. Distribution of breeds, crosses, or genetic lines as well as the different diets of the 228 control groups reported by the studies included in systematic review

Groupings of breeds, crosses, or genetic lines	Diet grouping
Berkshire cross	barley and wheat
Berkshire or Duroc or crossbred	canola, wheat, barley, soybean diet
Crossbred	corn
Duroc	corn and soybean
Duroc cross	corn-wheat middlings-soybean meal
Duroc cross or Large White	corn, peas, barley-based finishing diet
Duroc or Yorkshire or Duroc cross	Ingredients not specified
Duroc × Pietrain	met or exceeded nutrient requirements
Halothane carrier cross	milo-soybean
Halothane carrier hybrid	ractopamine added
Hybrid	sorghum based
Lacombe	wheat, barley, fava beans
Large White	wheat, barley, soybean
Pietrain cross	
Pietrain hybrid	
Yorkshire	
Yorkshire or crossbred	
Yorkshire or Yorkshire cross	

significant [$\beta_{publication\ year}$ = –0.0082 (95% CI = –0.09; 0.07)]; β_{pH} = –1.54 (95% CI = –3.42; 0.35)] with marked residual heterogeneity evident (I^2 = 93.96%, $p < 0.0001$) as shown in Fig. 2. Despite this heterogeneity, the conclusion from this analysis was that publication year was not associated with mean WBSF, and therefore, there was no evidence of changes over time.

Outcomes: Slice Shear Force (SSF) and Star Probe (SP)

Four groups (from 2 studies) reported SSF data (Shackelford et al., 2012; Miar et al., 2014). Nine groups, from 9 studies, reported SP. (Huff-Lonergan et al., 2002; Wiegand et al., 2002; Stoller et al., 2003; Custodio et al., 2006; Lampe et al., 2006; Schwab et al., 2006; Patton et al., 2008a; Patton et al., 2008b; Smith et al., 2011). This was insufficient information to address the research question about year effects, so these data were not included in the analyses.

Outcomes: pH

Of the 197 control groups (from 124 studies) that assessed pH, 151 control groups (from 100 studies) were eligible for inclusion in the meta-analysis (Supplemental Material 4). The final meta-regression model included publication year and sex. The publication year did not appear to have an effect [$\beta_{publication\ year}$ = 0.0042 (95% CI = –0.0018; 0.01)]. None of these moderators were significant (Q = 4.45, p = 0.35). Little residual unexplained heterogeneity was evident (Q = 112.93, p = 0.98) as shown in Fig. 3. The conclusion from this analysis was that year was not associated with mean pH, and therefore, there was no evidence of changes over time.

Outcomes: Minolta L/L Color*

There were 116 control groups (66 studies) that assessed Minolta L*/L color. Only 12 control groups reported light source, observer angle, and aperture, therefore, separate meta-analyses were conducted for Minolta L* and Minolta L. Fourty-five control groups (30 studies) were eligible for inclusion for Minolta L*. In the final meta-regression model that included publication year and pH for Minolta L*, neither moderator was found to be associated with color [$\beta_{publication\ year}$ = 0.09 (95% CI = –0.22; 0.40); β_{pH} = –0.25 (95% CI = –7.12; 6.72)]. Substantial residual among-study variation in effect estimates were still evident [I^2 = 94.13%, Q $p < 0.0001$] as shown in Fig. 4. Fifty-four control groups (33 studies) were eligible for inclusion for Minolta L meta-analysis. In the final meta-regression model for Minolta L, neither moderator was associated with changes in color [$\beta_{publication\ year}$ = 0.25 (95% CI = –0.14; 0.64); β_{pH} = –4.33 (95% CI = –12.18; 3.52)]. Again, substantial among-study variation in effect estimates was seen [I^2 = 95.21%, Q $p < 0.0001$] as shown in Fig. 5. The conclusion from these analyses was that year was not associated with mean color as measured by Minolta L* or Minolta L, and therefore, there was no evidence of changes over time.

Outcomes: Hunter L/L Color*

There were 70 control groups (47 studies) that assessed Hunter L*/L color, 13 control groups reported both characteristics, therefore, again, separate meta-analyses were conducted. Seventeen control groups (from 14 studies) were eligible for inclusion for Hunter L*. In the final meta-regression model, publication year and pH were associated with a negative nonsignificant effect [$\beta_{publication\ year}$ = –0.06 (95% CI = –0.73; 0.61); [β_{pH} = –6.14 (95% CI = –21.76; 9.50)], an increase in year and pH was associated with a decrease in mean Hunter L*. Marked among-study variation in effect estimates was still evident [I^2 = 97.62%, Q p = < 0.0001] as shown in Fig. 6. Twenty control groups (17 studies) were eligible for inclusion in the meta-analysis for Hunter L. In the final meta-regression model that included publication year, year was not found to be significant [$\beta_{publication\ year}$ = 0.40 (95% CI = –0.14; 0.92); however, pH was negatively associated with color [β_{pH} = –9.04 (95% CI = –16.04;

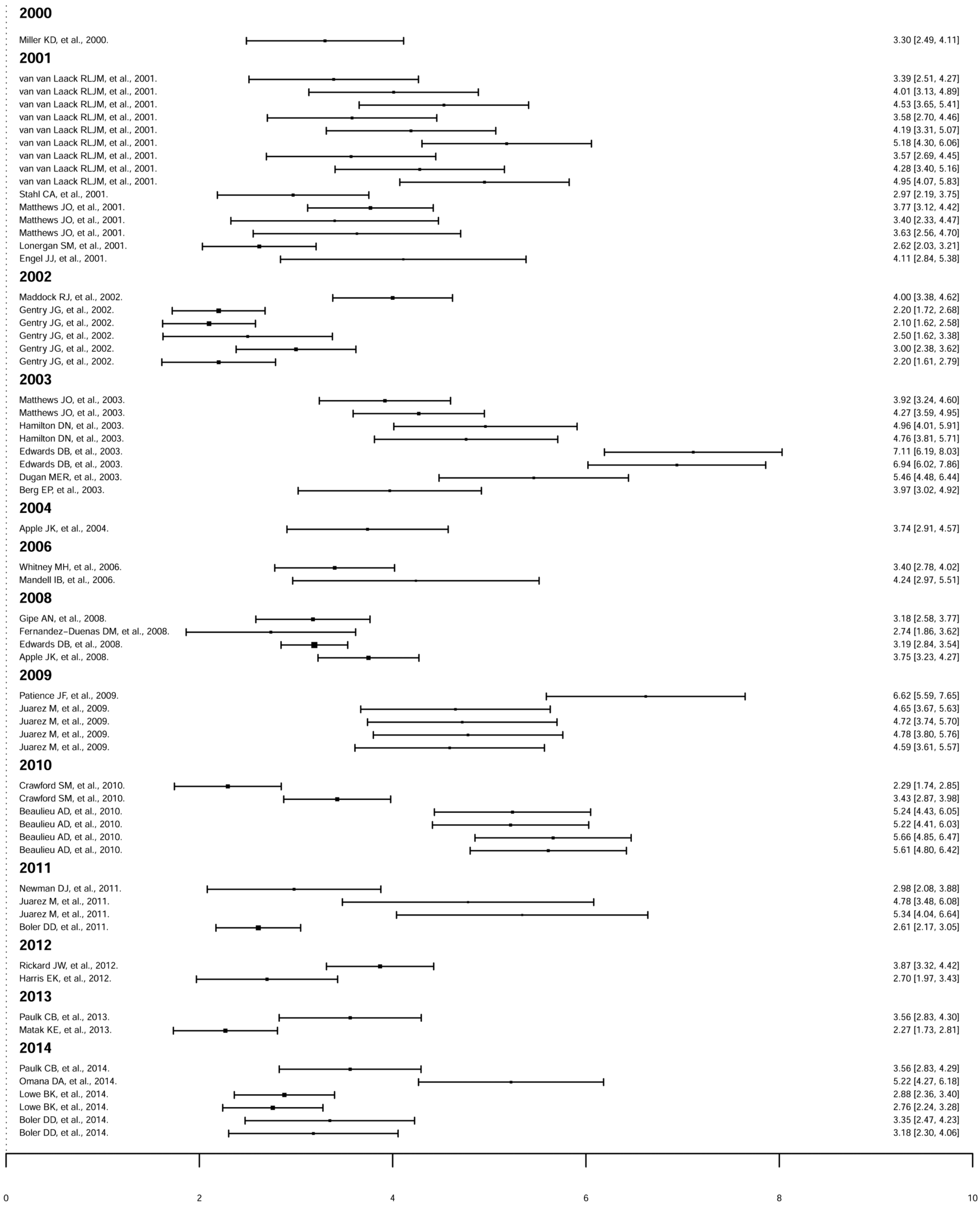

Figure 2. Mean Warner-Bratzler Shear Force (WBSF) for the 69 control groups of gilts and barrows between publication years 2000 and 2014. The black lines represent the study data.

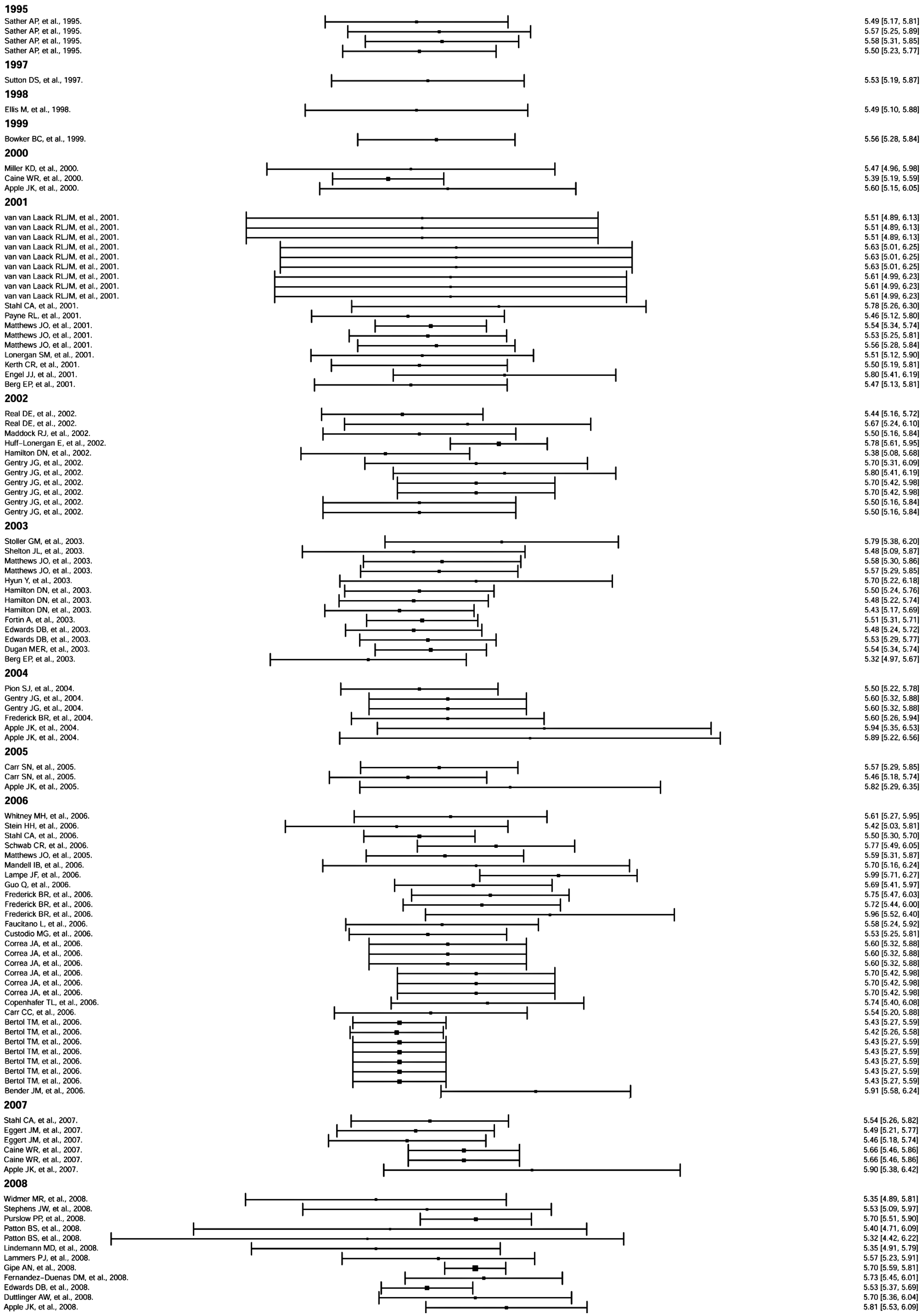

Figure 3. *Continued on next page.*

Figure 3. *Continued.*

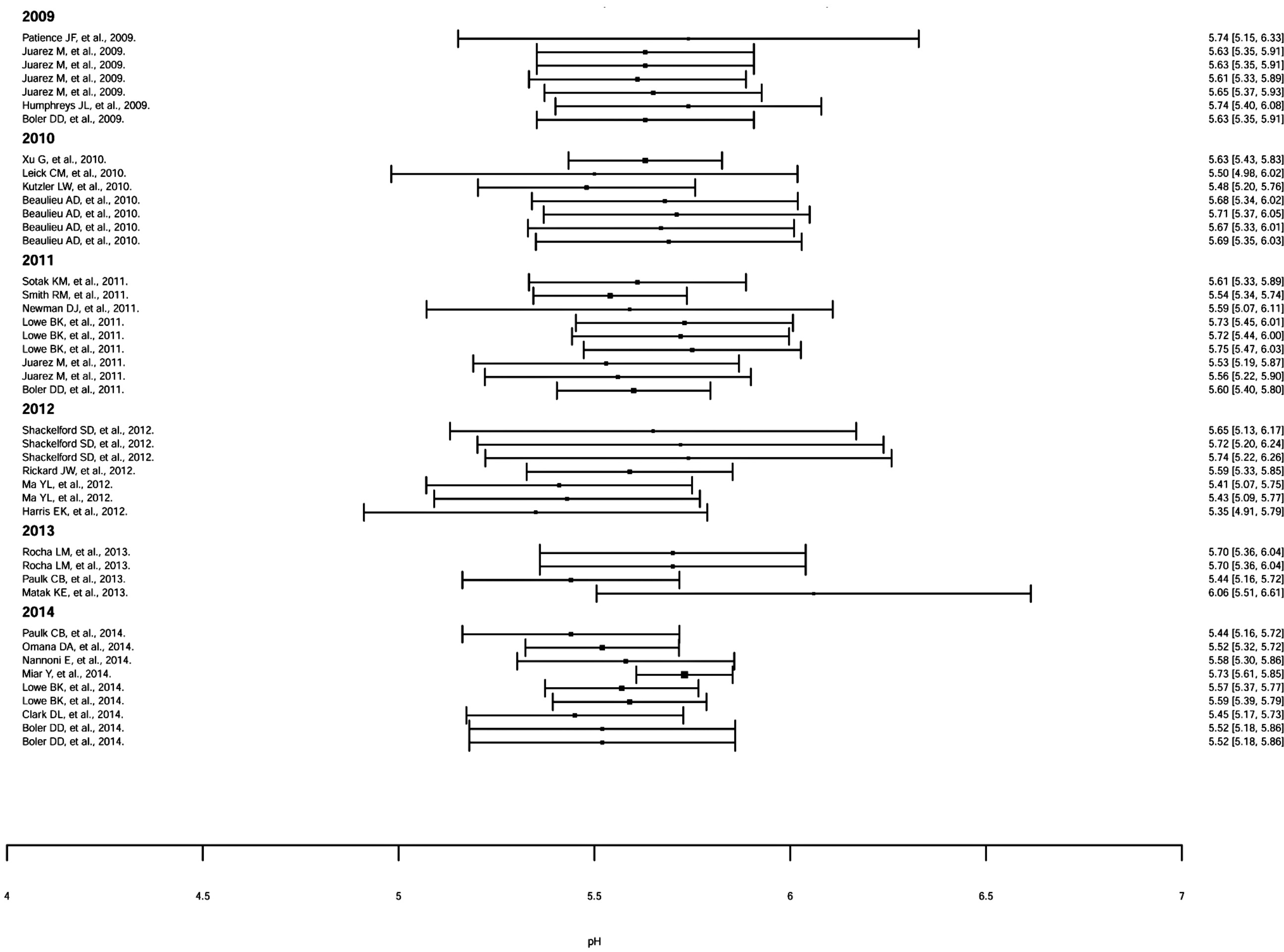

Figure 3. Mean pH for the 151 control groups of gilts and barrows between publication years between 1995 and 2014. The black lines represent the study data and the gray diamonds provide model estimates.

–2.04)], an increase in pH was associated with a decrease in mean Hunter L color score. Marked residual heterogeneity was still evident [I^2 = 95.96%, Q p = < 0.0001] as shown in Fig. 7.

Outcomes: Marbling

One hundred and fifty-seven control groups reported data on marbling scores. However, several different scales were used, for example, Bertol et al. (2006) used 1 = devoid to 10 = abundant, while Mandell et al. (2006) used a scale from 1 = devoid to 5 = abundant. Eighty-nine control groups (from 55 studies) were eligible for inclusion in the meta-analysis based on reporting of marbling scores and scales included. The overall effect was not reported because it was based on a pooled standardized scale that is not meaningful. In the meta-regression model, year was not found to be significant [$\beta_{\text{publication year}}$ = –0.0078 (95% CI = –0.018; 0.003). There was little evidence of substantial heterogeneity [I^2 = 15.67%, p < 0.11], as shown in Fig. 8.

DISCUSSION

The goal of this work was to determine the extent to which pork quality has changed over time. WBSF, pH, marbling scores, subjective color, and instrumental color conducted using Minolta or Hunter technology were the pork quality attributes evaluated. It was anticipated that such information would be obtained from either representative surveys of retail pork, representative surveys of post-harvest carcasses, or the control groups used in comparative nutrition experiments. Based on eligibility criteria, only control groups that identified and reported outcomes of interest were selected for the final review. The findings from that analysis concluded that there was no evidence of changes in pork quality over time and there was too much variation.

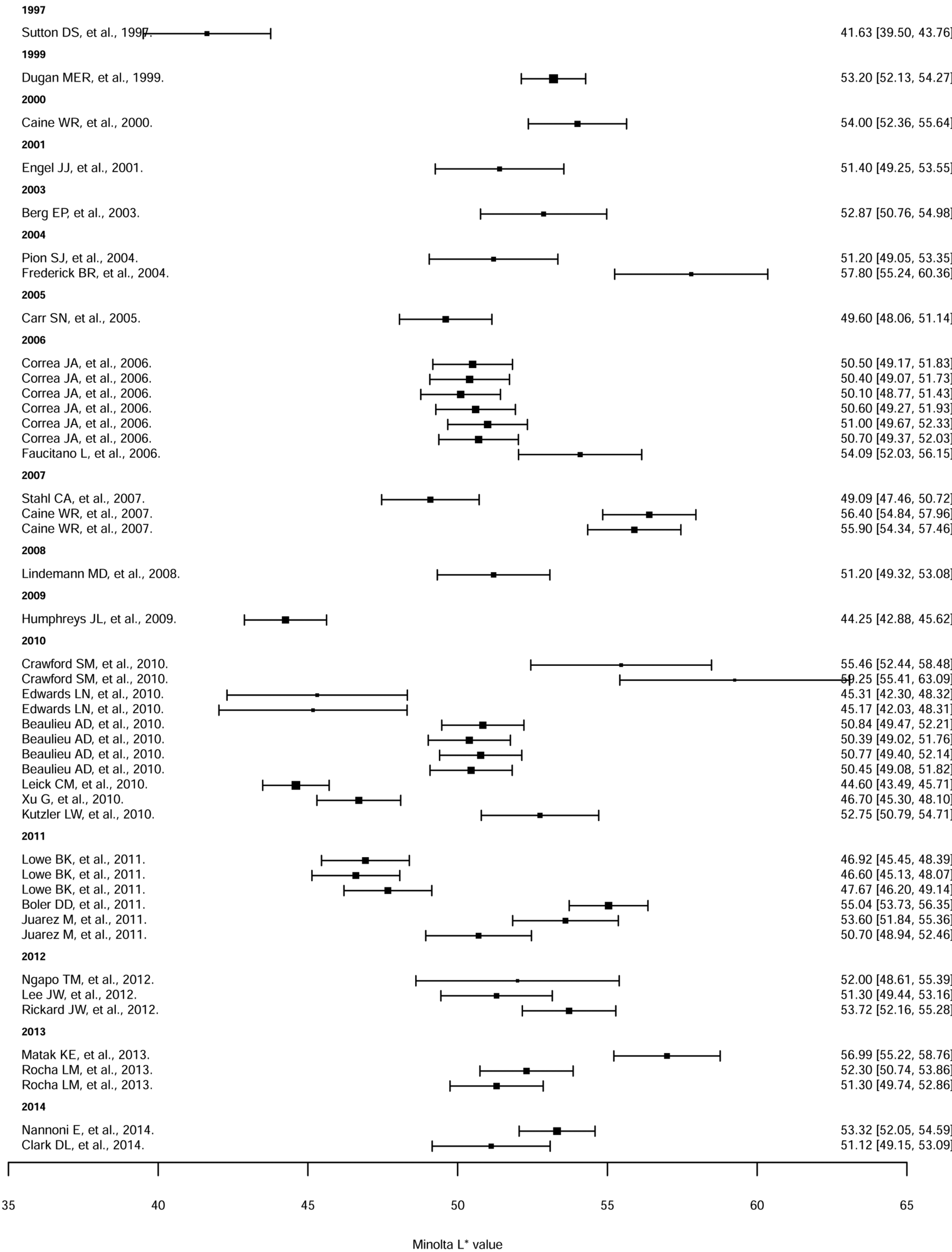

Figure 4. Illustration of heterogeneity for Minolta L* values for the 93 control groups of gilts and barrows for the publication years between 1997 and 2014, mean not meaningful based on lack of reporting. The black lines represent the study data.

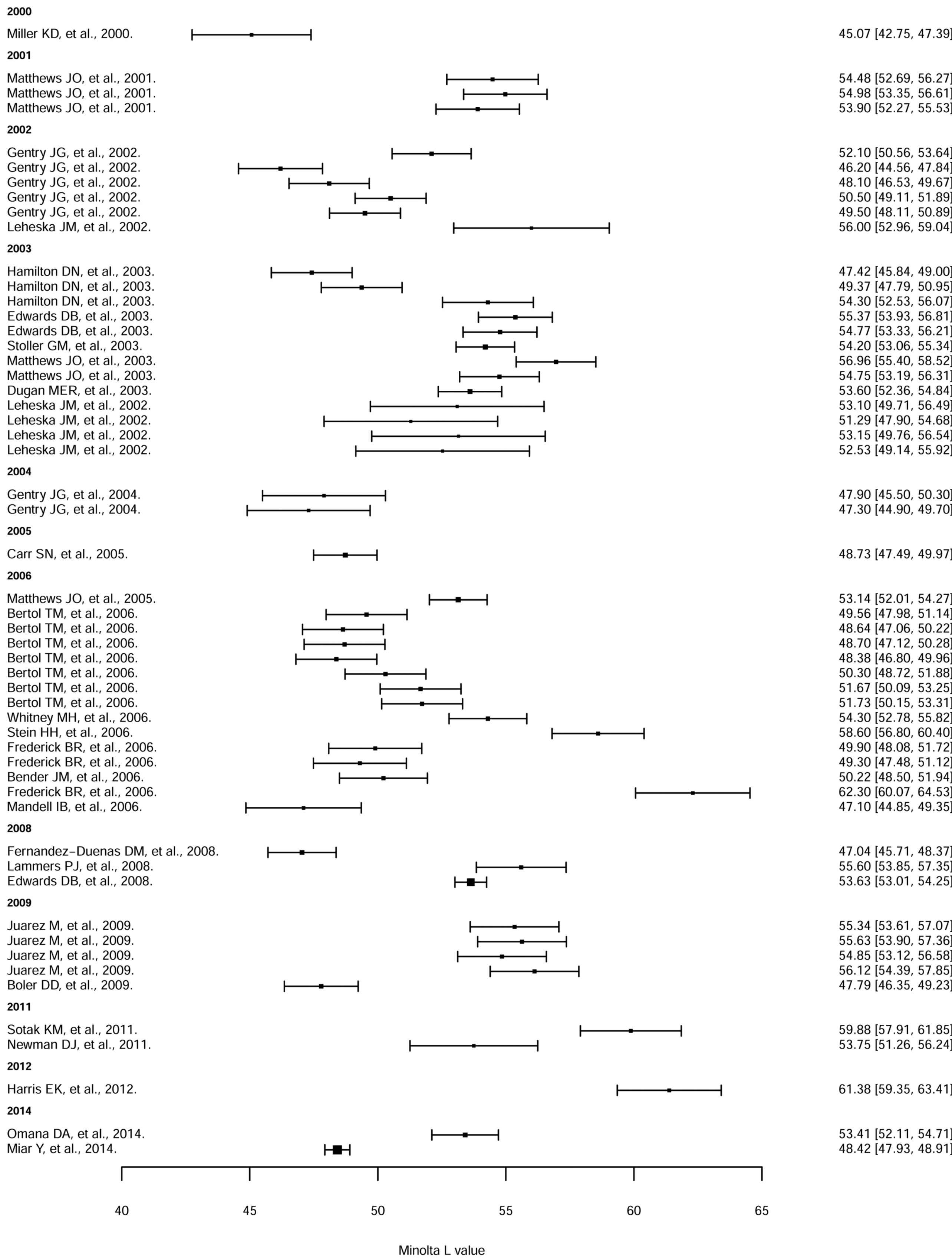

Figure 5. Illustration of heterogeneity for Minolta L values for the 54 control groups from publication years between 2000 and 2014, mean not meaningful based on lack of reporting. The black lines represent the study data.

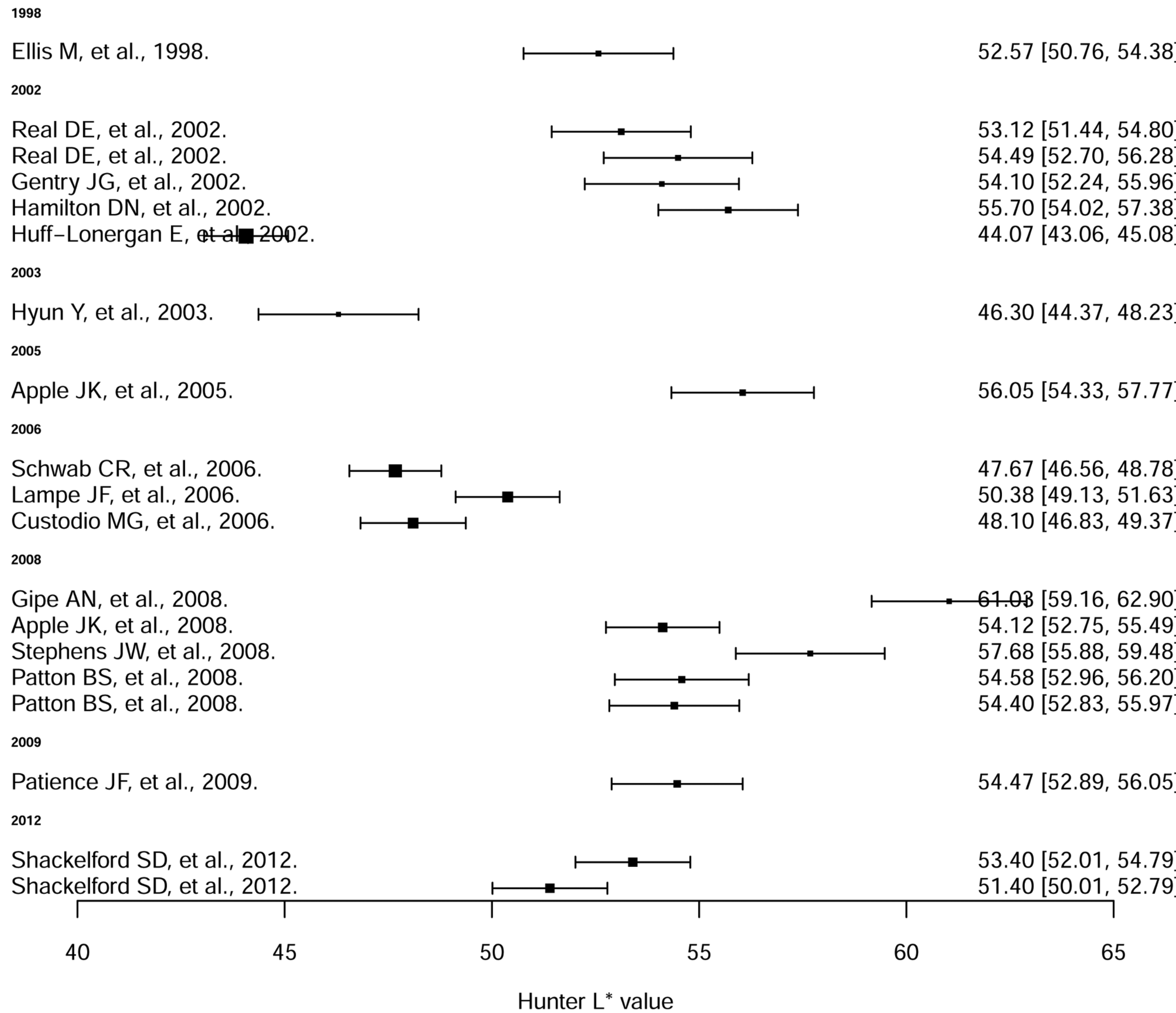

Figure 6. Illustration of heterogeneity for Hunter L* values for the 17 control groups from the publication years between 1999 and 2012, mean not meaningful based on lack of reporting. The black lines represent the study data.

The inference reached above is predicated on the value of control groups. Are these groups representative sources of what is found in the commercial pork industry; It has been concluded yes. An alternative argument might be that these controls are not representative sources of changes that have occurred in the commercial pork industry and the reality is that tenderness metrics of commercial pork have changed over the past 20 yr. If this is the truth, the study findings still have implications for the industry, but rather than relating the changes in tenderness metrics the implications relate to the validity of interventions assessed by comparative experiments. For example, the results of comparative experiments are based on invalid controls. Approaches to modifying and improving these metrics are a constant topic of comparative nutrition experiments. Experimentation creates conditions that differ from commercial production and translation to commercial production will be imperfect because the controlled conditions in experiments lead to larger effect sizes. However, as discussed by Bedford et al. (2016), researchers should seek to "ensure that when an experiment is conducted, the data generated are both accurate and relevant to the intended application." One of the major factors to ensure the data are relevant is to minimize the difference between the baseline characteristics of the study population, the control pigs, and the population the study result aims to translate to, the commercial pigs. For studies of tenderness metrics, if the comparison group being used by researchers is not changing over time to match changes occurring in commercial production, then the comparisons, and the effect sizes, reflect the expected effect size in commercial industry standard less each year.

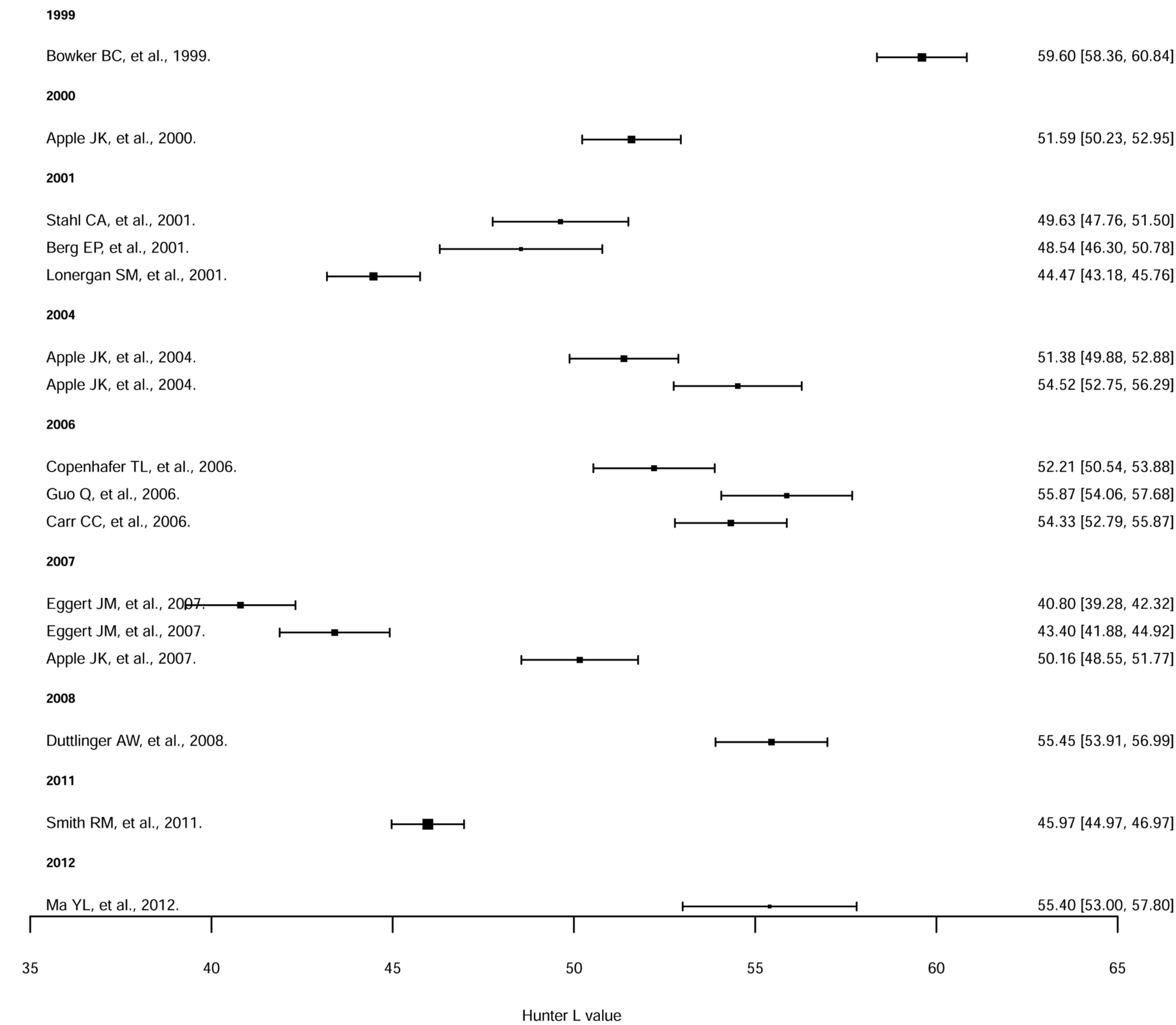

Figure 7. Illustration of heterogeneity for Hunter L values for the 21 control groups from the publication years between 1998 and 2012, mean not meaningful based on lack of reporting. The black lines represent the study data.

Ancillary Findings

An ancillary finding of the review was that many studies had to be excluded from the final review due to missing data. A major reason for exclusion included lack of reporting information on variation of precision estimates, for example, standard errors or standard deviations. The impact of this reporting issue was markedly reduced size and sensitivity of the analysis. For example, the number of studies that assessed WBSF was 76, and 130 control groups were assessed. However, due to incomplete reporting, the number of studies included in the meta-analysis was limited to 32 and the number of control groups assessed was 69, approximately 47% data loss. In the WBSF model, 4 studies did not report the outcome for weight and were excluded from the analysis. Another study did not include the precision outcome of weight and was excluded from the analysis, as well. Although this is just one example, missing data was found in all the parameters of interest. As discussed previously, the marked heterogeneity observed for the color metrics may be associated with the inability to adjust the meta-analyses for covariates: light source, observer angle, and aperture. The inclusion of published data in these analyses requires uniformity in sample handling and method application.

Limitations

This review had several limitations. The first relates to the inference are control groups relevant to the question? Yearly surveys of pork quality metrics in commercial product would be preferable, however, data from comparative study publications is reported here. First, despite including as many citations as possible, for some outcomes, such as color and WBSF, there were marked differences among the studies. These differences were likely related

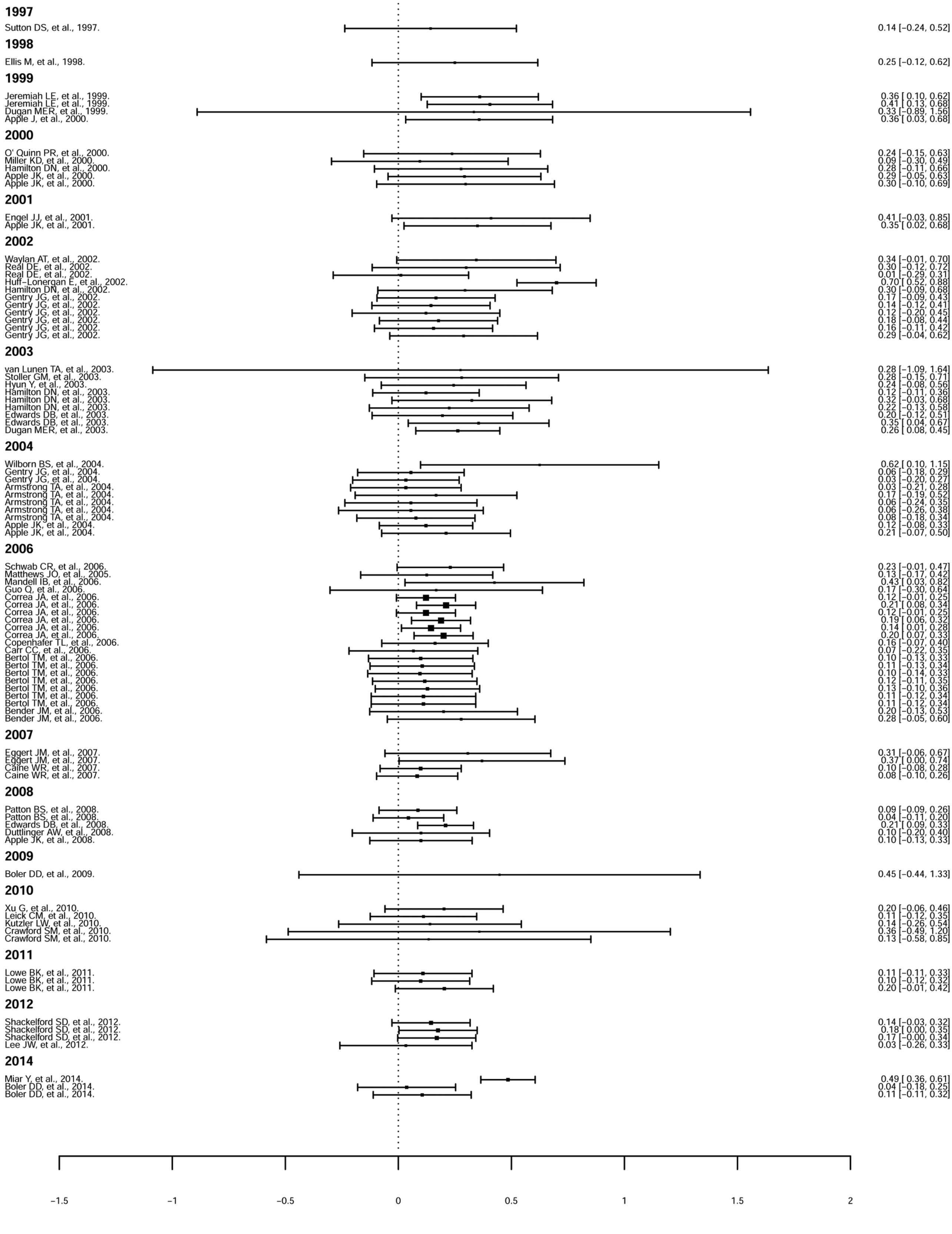

Figure 8. Standardized marbling scores for the control groups from the publication years between 1998 and 2012. The black lines represent the study data and the gray diamonds provide model estimates.

to the individual study variation in characteristics such as diet, weight at slaughter, breed, and setting. The exact influence of these characteristics is beyond the scope of this study. Although missing data reduced the sample size and precision, it seems an unlikely source of systematic bias. For example, it seems unlikely that studies with consistently high or consistently low metrics were more likely to be excluded because of reporting issues.

Conclusions

This systematic review of pork tenderness metrics found that an insufficient number of representative surveys were available to make conclusions about changes in pork tenderness metrics. If control groups used in experiments are indicative of tenderness metrics in commercial pigs, it can be concluded that pork quality traits have not changed over time. Alternatively, if control groups used in experiments are not indicative of tenderness metrics in commercial pigs, then comparative nutrition experiments are using control groups with decreasing relevance to the commercial industry. Further, there is a need to continue to be vigilant about comprehensive reporting of the results of experiments to ensure maximum value is obtained from the synthesis of research results.

LITERATURE CITED

Bedford, M. R., M. Choct, and H. V. M O'Neill. 2016. Nutrition experiments in pigs and poultry: a practical guide. CABI, Boston, MA. p.1

Bertol, T. M., M. Ellis, M. J. Ritter, F. K. McKeith, and D. N. Hamilton. 2006. Variation in glycolytic potential and fresh pork quality traits along the longissimus dorsi of slaughter weight pigs. J. Muscle Foods 17:237–247. doi:10.1111/j.1745-4573.2006.00042.x

Borenstein, M., J. P. T. Higgins, L. V. Hedges, and H. R. Rothstein. 2017. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Res. Synth. Methods 8:5–18. doi:10.1002/jrsm.1230

Bray, R. 1966. Pork quality—definition, characteristics and significance. J. Anim. Sci. 25:839–842. doi:10.2527/jas1966.253839x

Choe, J. H., M. H. Choi, M. S. Rhee, and B. C. Kim. 2016. Estimation of sensory pork loin tenderness using Warner Bratzler Shear Force and texture profile analysis. Asian-australas. J. Anim. Sci. 29:1029–1036. doi:10.5713/ajas.15.0482

Custodio, M. G., W. J. Powers, E. Huff-Lonergan, M. A. Faust, and J. Stein. 2006. Growth, pork quality, and excretion characteristics of pigs fed Bt corn or non-transgenic corn. Can. J. Anim. Sci. 86:462–469. doi:10.4141/A05-082

Higgins, J., and S. G. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21(11):1539–1558. doi:10.1002/sim.1186

Huff-Lonergan, E., T. J. Baas, M. Malek, J. C. M. Dekkers, K. Prusa, and M. F. Rothschild. 2002. Correlations among selected pork quality traits. J. Anim. Sci. 80:617–627. doi:10.2527/2002.803617x

Lampe, J. F., T. J. Baas, and J. W. Mabry. 2006. Comparison of grain sources for swine diets and their effect on meat and fat quality traits. J. Anim. Sci. 84:1022–1029. doi:10.2527/2006.8441022x

Lowe, B. K., G. D. Gerlemann, S. N. Carr, P. J. Rincker, A. L Schroeder, D. B. Petry, F. K. McKeith, G. L. Allee, and A. C. Dilger. 2014. Effects of feeding ractopamine hydrochloride (Paylean) to physical and immunological castrates (Improvest) in a commercial setting on carcass cutting yields and loin quality. J. Anim. Sci. 92:3715–3726. doi:10.2527/jas.2013-7515

Mandell, I. B., C. P. Campbell, and C. F. D. Lange. 2006. Effects of gender, sire line, and penning environment on growth, carcass characteristics, and aspects of pork meat quality at different locations in the loin. Can. J. Anim. Sci. 86:49–61. doi:10.4141/A05-025

Miar, Y., G. S. Plastow, S. S. Moore, G. Manafiazer, P. Charagu, R. A. Kamp, B. van Haandel, A. E. Huisman, C. Y. Zhang, R. M. McKay, H. L. Bruce, and Z. Wang. 2014. Genetic and Phenotypic Correlations between Performance Traits with Meat Quality and Carcass Characteristics in Commercial Crossbred Pigs. PLoS One 9(10):e110105. doi:10.1371/journal.pone.0110105

Newman, D. J. "National Pork Retail Benchmarking Study." Research Report to National Pork Board, 2014, p. 24. National Pork Board. Clive, IA.

Patton, B. S., E. Huff-Lonergan, M. S. Honeymoon, J. D. Crouse, B. J. Kerr, and S. M. Lonergan. 2008a. Effects of deep-bedded finishing system on market pig performance, composition and pork quality. Anim. 2:459-470. doi:10.1017/S1751731107001292

Patton, B. S., E. Huff-Lonergan, M. S. Honeyman, B. J. Kerr, and S. M. Lonergan. 2008b. Effects of space allocation within a deep-bedded finishing system on pig growth performance, fatty acid composition and pork quality. Anim. 2:471-478. doi:10.1017/S1751731107001280

Schwab, C. R., T. J. Baas, K. J. Stalder, and J. W. Mabry. 2006. Effect of long-term selection for increased leanness on meat and eating quality traits in Duroc swine. J. Anim. Sci. 84:1577–1583. doi:10.2527/2006.8461577x

Shackelford, S. D., D. A. King, and T. L. Wheeler. 2012. Chilling rate effects on pork loin tenderness in commercial processing plants. J. Anim. Sci. 90:2842–2849. doi:10.2527/jas.2011-4855

Smith, R. M., N. K. Gabler, J. M. Young, W. Cai, N. J. Boddicker, M. J. Anderson, E. Huff-Longergan, J. C. M. Dekkers, and S. M. Lonergan. 2011. Effects of selection for decreased residual feed intake on composition and quality of fresh pork. J. Anim. Sci. 89:192–200. doi:10.2527/jas.2010-2861

Stoller, G. M., H. N. Zerby, S. J. Moeller, T. J. Baas, C. Johnson, and L. E. Watkins. 2003. The effect of feeding ractopamine (Paylean) on muscle quality and sensory characteristics in three diverse genetic lines of swine. J. Anim. Sci. 81:1508–1516. doi:10.2527/2003.8161508x

Viechtbauer, W. 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36:1–48. doi:10.18637/jss.v036.i03

Wiegand, B. R., J. C. Sparks, F. C. Parrish, and D. R. Zimmerman. 2002. Duration of feeding conjugated linoleic acid influences growth performance, carcass traits, and meat quality of finishing barrows. J. Anim. Sci. 80:637–643. doi:10.2527/2002.803637x

Wright, L. I., J. A. Scanga, K. E. Belk, T. E. Engle, J. D. Tatum, R. C. Person, D. R. McKenna, D. B. Griffin, F. K. McKeith, J. W. Savell, and G. C. Smith. 2005. Benchmarking Value in the Pork Supply Chain: Characterization of Us Pork in the Retail Marketplace. Meat Sci. 71(3):451–463. doi:10.1016/j.meatsci.2005.04.024