CSIRO PUBLISHING

www.publish.csiro.au/journals/ajea *Australian Journal of Experimental Agriculture*, 2008, **48**, 1425–1433

Meta-analysis of the published effects of HGP use on beef palatability in steers as measured by objective and sensory testing

Ray Watson

Department of Mathematics and Statistics, University of Melbourne, Parkville, Vic. 3010, Australia.
Email: rayw@ms.unimelb.edu.au

Abstract. Evidence is presented that suggests strongly that hormone growth promotant (HGP) implantation has a negative effect on beef palatability. This is based on a meta-analysis of results reported in refereed papers that have appeared in the meat-science literature. To be included in this analysis, a paper must have reported results for control samples (no HGP) and treatment samples (HGP) for either objective testing (Warner-Bratzler shear-force) or consumer preference (tenderness score). The paper must also have reported estimates and standard errors. Further, we consider only the case of steers, and the *M. longissimus dorsi* (striploin). While most of these studies yielded non-significant differences, most gave an estimate indicating that the HGP treatment had a negative effect on beef palatability. When these results are combined using a meta-analysis, they provide significant evidence that the use of HGP implants negatively influences palatability.

Additional keywords: beef palatability, meta-analysis, tenderness score, Warner-Bratzler shear-force.

Introduction

A fundamental component of both scientific research and policy decision making is the review of existing information on the methodology or intervention of interest. This information is commonly quantitative in nature and arises from a variety of sources. Inherent in this information is some variability both in the information itself and in the perceived reliability of the information supplied from each source. Questions arise therefore as to how this information is to be utilised to best aid decision making or further research on the intervention of interest. This quantitative approach, combining the results from each information source, together with the collection and selection of information has come to be widely known as meta-analysis: 'the analysis of analyses' (Glass 1976, p. 4). A meta-analysis can take many forms, corresponding to the field of research in which it is applied. These fields are many, including environmental sciences (Hilborn and Lierman 1998; Gurevitch and Hedges 1999), management (Forza and Dinuzzo 1998; Phillips 1998), education and the behavioural sciences (Dunn *et al.* 1995; Quinn *et al.* 1999) and medical research (Halpem *et al.* 1998; Bent *et al.* 1999). It can also be applied in meat science. The appeal of meta-analysis is that it presents a scientific, objective means of reviewing and analysing existing quantitative information. It is presented as 'a rigorous alternative to the casual, narrative discussions of research studies' (Glass 1976).

The intention of this paper is to assemble a complete summary of the published effects of HGP use on beef palatability as measured by tenderness scores and by Warner-Bratzler shear-force measurements for *M. longissimus dorsi* in steers.

The studies considered in this meta-analysis are papers that have appeared in refereed journals for which estimates and standard errors are available, including the recent Australian studies (Hunter *et al.* 2001; Thompson *et al.* 2008*a*, 2008*b*; Watson *et al.* 2008*b*; McIntyre *et al.* unpubl. data) summarised in the Results section. This resulted in 20 studies, with 30 treatment–control comparisons for the tenderness evaluations; and 16 studies, with 22 treatment–control comparisons for the shear-force measurements.

About 70% of these treatment–control comparisons yielded non-significant differences, but more than 90% gave an estimate indicating that the HGP treatment had a negative effect on beef palatability. When these results are combined using a meta-analysis, they provide significant evidence that the use of HGP implants negatively influences palatability. It is found that the estimate of the HGP effect is to increase shear-force by 0.27 kg, with standard error 0.03; and to reduce tenderness by 5.4 points on a 100-point scale, with standard error 0.8.

Publication bias may exist where there is a tendency for papers with significant results to be published. In most papers considered in this meta-analysis, palatability was a subsidiary result to the other reported results, usually the effect of HGP on animal growth. So the significance or otherwise of the palatability result was not important to the paper's publication. As a consequence publication bias is unlikely to occur in this case. Indeed, in most cases, the HGP effect on palatability was found to be non-significant.

Materials and methods

A meta-analysis was conducted of results reported in refereed papers that have appeared in the meat-science literature. Such papers were found using searches on (beef or steer) and (palatability or shear-force or tenderness) and (HGP or

10.1071/EA07174 0816-1089/08/111425

hormone/hormonal or implants or shear-force). To be included in this meta-analysis, a paper must have appeared in a refereed English-language journal and it must have reported estimates and standard errors for either objective testing (Warner-Bratzler shear-force) or consumer preference (tenderness score). A paper may contain more than one treatment–control comparison. To be included in the meta-analysis, such a comparison must have a specific control sample (no HGP) and a treatment sample or samples (HGP).

To avoid possible sex effects and muscle differences, attention was restricted to papers reporting on steers, and the *M. longissimus dorsi* (striploin). This choice ensured a wide range of results, as most trials for which a comparison has been made between control (no implant) and treated (with HGP implant) include steers, striploins, and shear-force or tenderness scores. This resulted in 20 eligible papers for the tenderness evaluations; and 16 for the shear-force measurements.

In order to make the results of the objective tests and the sensory tests comparable, a 'toughness' measure is used in each case: in the objective case the shear-force, and in the sensory case a negative perceived tenderness. In each case then the HGP effect is the mean increase in toughness when HGP is used, as compared with a control where HGP is not used.

Sensory scores are reported on varying scales. To ensure that all sensory scores are comparable, they are converted to a scale from 0 to 100. If x denotes the tenderness score on the scale from a to b, then the following definition is used:

$$\text{Tenderness} = \frac{100(x-a)}{b-a}, \qquad \text{Toughness} = \frac{100(b-x)}{b-a}.$$

In performing the meta-analysis, it is assumed that all HGP treatments have the same palatability effect: i.e. that each is sampled from a random effect distribution with a common mean. This assumption is not contradicted by the published results. Further, in cases where a number of HGP treatments have been compared against a common control, the HGP effect estimates are not independent as each has a common control. In such cases, the HGP treatments are pooled before comparison with the control.

In combining independent estimates, the optimal linear combination of estimates was used, i.e. with weights inversely proportional to the square of the standard error. Given independent estimates, est_i with standard error s.e._i $(i = 1, 2, \ldots, k)$, the combined estimate is given by

$$\text{est} = \frac{\sum_{i=1}^{k} w_i \, \text{est}_i}{\sum_{i=1}^{k} w_i}, \quad \text{where} \quad w_i = 1/\text{s.e.}_i^2;$$

and for this estimate the standard error is such that

$$\text{s.e.}^2 = \frac{1}{\sum_{i=1}^{k} w_i}.$$

The number of HGP effect estimates reported here is determined by the number of independent control groups. The relevant estimates, standard errors and confidence intervals were extracted, combining results as necessary. These results were included in the meta-analysis of published results given below.

Results and discussion

Recent Australian studies

CRC trial: Hunter et al. *(2001), Thompson* et al. *(2008*b*)*

This trial has been reported in Hunter *et al.* (2001) where a preliminary report of the sensory results from 203 animals was given. Thompson *et al.* (2008*b*) presented further results from the same trial for 486 animals, though the results given there did not specifically elaborate on the estimates and standard errors of the HGP effects. These are specified below. In the trial, 509 Brahman and F1 Brahman crossbred steers were used. At weaning, these animals were allocated to three market endpoints (Domestic = 220 kg, Korean = 280 kg, Japanese = 320 kg) by two nutritional finishing strategies (grain-based feedlot, pasture-fed). About half of each group was allocated as control (no implant) and the other half as treated (20 mg estradiol-17 about every 100 days). Fuller details of the experimental procedure can be found in Hunter *et al.* (2001) or Thompson *et al.* (2008*b*). The numbers of animals in the six groups are indicated in Table 1. The completed study contains more than twice the number used in Hunter *et al.* (2001). However, as pointed out in the previous papers, there were some problems with processing some of the carcasses, resulting in cold-shortening: Fig. 1 indicates that there were a number of extremely tough samples. There were also a number of extreme ultimate pH values. To avoid distortion from outliers, meat for which pHu > 5.7 were deleted from the sample. As a result, meat corresponding

Table 1. Number of animals, grouped by nutritional strategy and market endpoint, in the CRC HGP trial

Values in italics are the numbers used in the preliminary report of Hunter *et al.* (2001)

	Domestic		Korean		Japanese		Total
Feedlot	18	18	38	37	53	52	216
	18	*17*	*20*	*17*	*16*	*16*	*104*
Grass-fed	40	39	50	48	58	58	293
	19	*18*	*19*	*14*	*17*	*12*	*99*
Total	58	57	88	85	111	110	509
	37	*35*	*39*	*31*	*33*	*28*	*203*

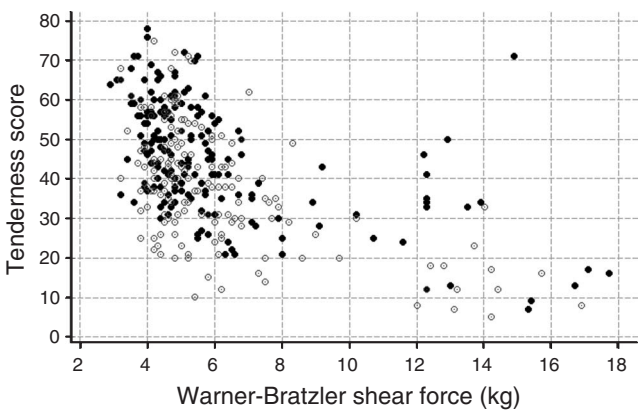

Fig. 1. Relationship between tenderness score and Warner-Bratzler shear-force measurements for the meat samples from HGP implanted (open circles) and control (solid circles) steers from the CRC study.

to the points to the right of Fig. 1 were discarded from the analysis.

The tenderness scores were measured on a 0–100 scale: zero representing very tough and 100 very tender. The scores were obtained as described in Polkinghorne *et al.* (1999) and further described by Watson *et al.* (2008*a*). The shear-force (measured in kg) was measured on 250 g blocks of *M. longissimus* cooked in a water bath for one hour at 70°C. The methodology is described in detail in Perry *et al.* (2001).

Estimates and standard errors of the effect of HGP implants on tenderness and shear-force are presented in Table 2.

Table 2. Effect (estimate ± s.e.) of HGP on the negative tenderness scores and the Warner-Bratzler shear-force (kg) for meat samples from the CRC trial

	Domestic	Korean	Japanese
	Negative tenderness scores		
Feedlot	7.05 ± 3.55	3.74 ± 3.09	3.09 ± 2.52
Grass-fed	12.69 ± 2.92	4.48 ± 2.59	1.28 ± 2.46
	Warner-Bratzler shear-force (kg)		
Feedlot	0.45 ± 0.23	0.22 ± 0.20	0.26 ± 0.15
Grass-fed	0.09 ± 0.29	0.50 ± 0.27	0.33 ± 0.28

WA-1 trial: Thompson et al. *(2008*a*)*

This trial, utilising three cooking methods, covered a wide range of muscles for heifers and steers. The details are covered in full in Thompson *et al.* (2008*a*). Here only the details for the *M. longissimus* for steers are extracted. There were 47 control steers and 44 treated steers with results given in Table 3.

Table 3. HGP effect on toughness [as measured by negative tenderness score and by Warner-Bratzler shear-force (WA-1 trial only)] for (i) the WA-1 trial described by Thompson *et al.* (2008*a*); (ii) the WA-2 trial described by McIntyre *et al.* (unpubl. data); and (iii) the Q-2 trial described in Watson *et al.* (2008*a*, 2008*b*)

Trial	HGP effect	
	Estimate	s.e.
	Toughness (measured by negative tenderness)	
WA-1	12.00	2.97
WA-2	3.18	2.08
Q-2	11.31	3.78
	Toughness [measured by Warner-Bratzler shear-force (kg)]	
WA-1	0.35	0.22

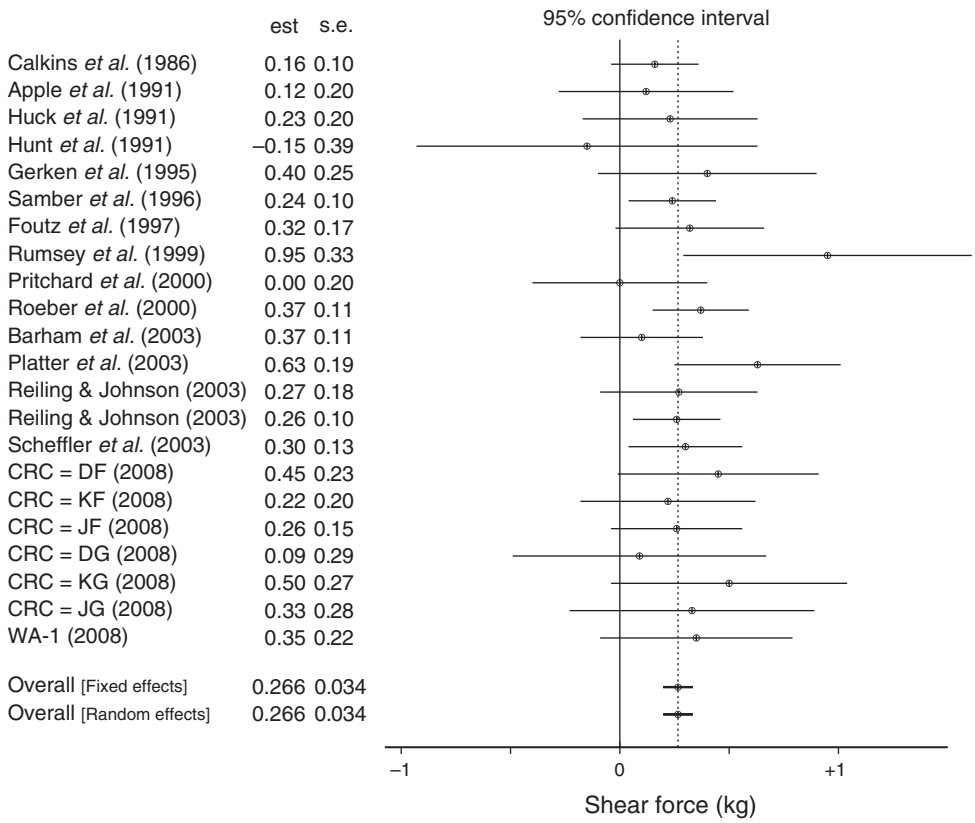

Fig. 2. Published results for HGP effect on shear-force of the *M. longissimus*, and graphical representation of the 95% confidence interval for the HGP effect. Papers that appear more than once have independent control groups. The CRC and WA results relate to studies described earlier in this paper; CRC = XY refers to the CRC experiment with X = (D, K, J) denoting (Domestic, Japanese, Korean) finishing point and Y = (F, G) denoting (Feedlot, Grass-fed) animals.

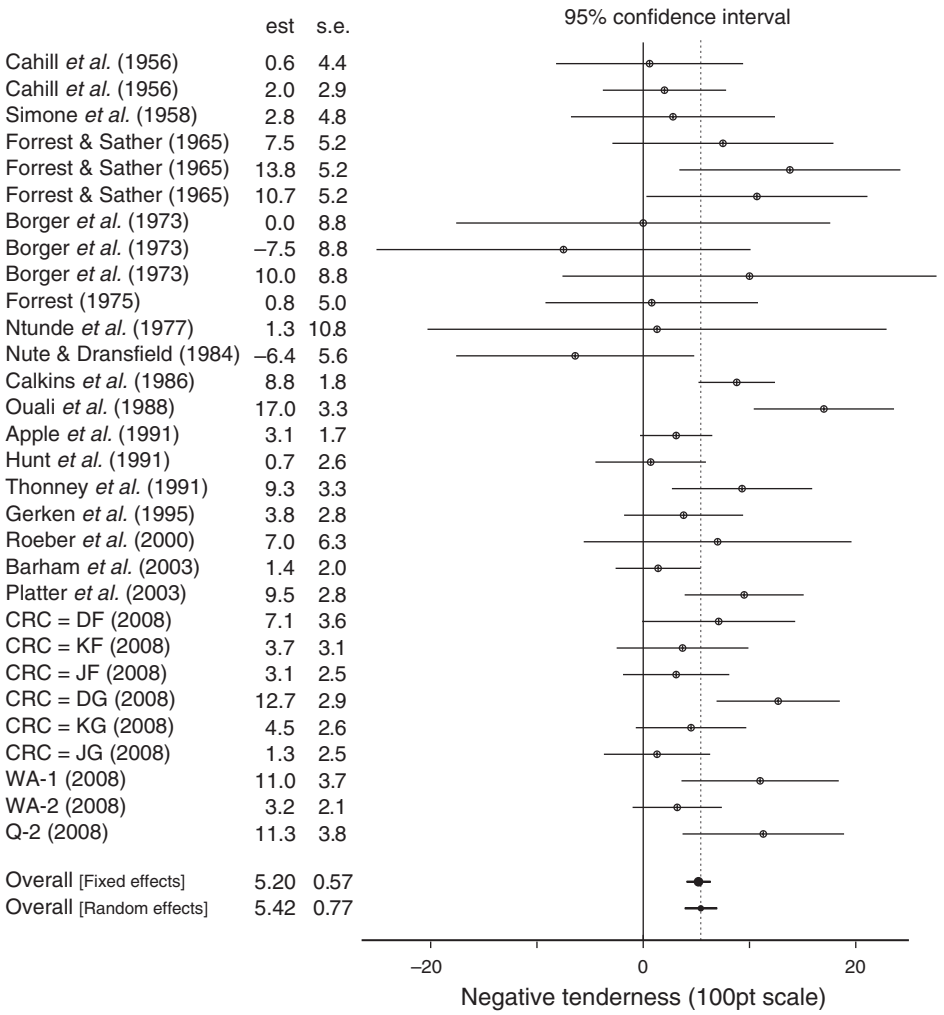

Fig. 3. Published results for HGP effect on reported negative tenderness of the *M. longissimus*, and graphical representation of the 95% confidence interval for the HGP effect. Papers that appear more than once have independent control groups. The CRC and WA results relate to studies described earlier in this paper; CRC = XY refers to the CRC experiment with X = (D, K, J) denoting (Domestic, Japanese, Korean) finishing point and Y = (F, G) denoting (Feedlot, Grass-fed) animals.

WA-2 trial: McIntyre et al. *(unpubl. data)*

This trial, on steers only, and using only the *M. longissimus*, was primarily concerned with the timing of implants: it compared three HGP treatment modes in which the timing of the implant differed. Again, our concern here is to simply compare HGP treatment with control. In this case, only tenderness scores were obtained, with results given in Table 3.

Q-2 trial: Watson et al. *(2008*b*)*

This trial, on steers only, tested a variety of HGP implants on three muscle portions: *Mm. psoas major*, *longisimuss dorsi thoraics* and *lumborum* portions. Details of the procedure and the results can be found in Watson *et al.* (2008*b*). In this trial, only the sensory (tenderness) scores were available. Considering only the *M. longisimuss dorsi lumborum*, and combining all the HGP treatments, yields the result given in Table 3.

Meta-analysis

The papers used in the meta-analysis are listed in Figs 2 and 3. The results of the meta-analysis are summarised in Table 4.

Table 4. The meta-analysis summary of the HGP effect on palatability

Warner-Bratzler shear-force:
Homogeneity test: Q = 17.59 (d.f. = 21); $P = 0.675$
Inter-study variance estimate: $\hat{\tau}^2 = 0$
Fixed-effects model: estimate = 0.266, s.e. = 0.034; 95%CI = (0.199, 0.334)
Random-effects model: estimate = 0.266, s.e. = 0.034; 95%CI = (0.199, 0.334)

Negative tenderness:
Homogeneity test: Q = 38.13 (d.f. = 29); $P = 0.119$
Inter-study variance estimate: $\hat{\tau}^2 = 5.62$
Fixed-effects model: estimate = 5.20, s.e. = 0.57; 95%CI = (4.09, 6.32)
Random-effects model: estimate = 5.42, s.e. = 0.77; 95%CI = (3.91, 6.93)

The details of the meta-analysis procedure are summarised in the Appendix.

In neither case does the homogeneity test show any significant difference between the studies. In the case of shear-force, the best estimate of the inter-study variance is zero, so that the random-effects model reduces to the fixed-effects model. In the case of the sensory data, there is some indication of a difference between studies, although this is not significant ($P = 0.119$). Despite this, the random effects model seems preferable, so as to allow for the possible differences between studies. It is a more general model, and it includes the fixed effects model as a special case. Both fixed-effects and random-effects model results are reported: the random-effects confidence interval is slightly wider, to allow for the possible variation between studies.

The meta-analysis is illustrated in Figs 2 and 3. In these figures, each line represents a treatment–control comparison, from the specified paper. The (reported or calculated) estimate and standard error are given and, on the same line, the estimate and the 95% confidence interval are shown: the dot representing the estimate and the line the confidence interval, with respect to the specified scale. The overall estimates and confidence intervals (for both the fixed effects and random effects models) are given at the bottom of the diagram.

The confidence intervals for the separate experiments are usually quite wide, and mostly include zero, indicating non-significance. But when the results are accumulated the result is a narrow confidence interval which clearly excludes zero, and is therefore highly significant.

There is significant evidence here that HGP affects palatability negatively. Initially it was felt that some time limitation should be put on the studies considered. However, although there is some (non-significant) indication of a variation between studies, the estimates show no trend with year of publication (see Fig. 4). In fact the results for tenderness after 1980 yield a slightly larger estimate of HGP effect. There seemed to be no good reason for excluding any of the published studies.

There was no *a priori* indication for a publication bias, and most results were non-significant in any case. A funnel plot for each of the sets of estimates is show in Fig. 5. There is no indication of asymmetry, which would suggest a publication bias.

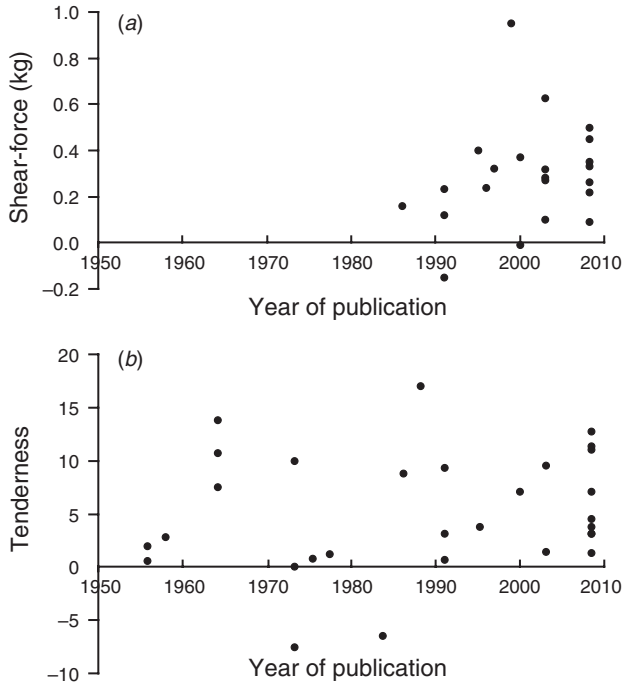

Fig. 4. Published estimates of HGP palatability effects [(*a*) shear-force (kg) and (*b*) tenderness; HGP – control] against year of publication.

Conclusion

The evidence presented here suggests very strongly that HGP has a negative effect on beef palatabilty. Evidence for the sensory results dates back 50 years and little seems to have changed: the results for 1955–1980 are quite similar to the results for 1981–2008.

The question which must now be asked is what might be done to ameliorate it. Or is it just a price to be paid for efficient beef production?

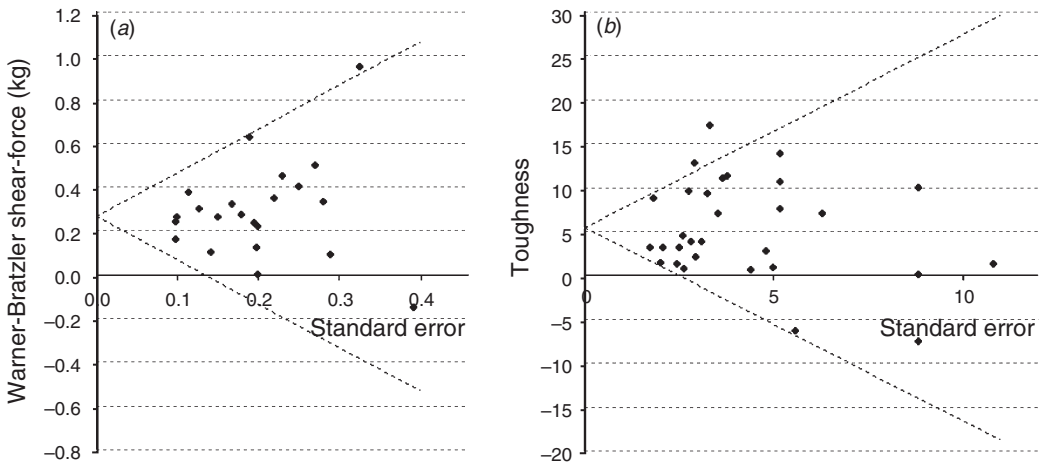

Fig. 5. Funnel plot for estimates of HGP palatability effects [(*a*) Warner-Bratzler shear-force and (*b*) toughness] against standard error.

Acknowledgements

The author is grateful to two anonymous referees for their thorough reviews and helpful criticisms.

References

Apple JK, Dikeman ME, Simms DD, Kuhl G (1991) Effects of synthetic hormone implants, singularly or in combinations, on performance, carcass traits, and longissimus muscle palatability of Holstein steers. *Journal of Animal Science* **69**, 4437–4448.

Barham BL, Brooks JC, Blanton JR Jr, Herring AD, Carr MA, Kerth CR, Miller MF (2003) Effects of growth implants on consumer perceptions of meat tenderness in beef steers. *Journal of Animal Science* **81**, 3052–3056.

Bent S, Saint S, Vittinghoff E, Grady D (1999) Antibiotics in acute brichitis. *American Journal of Medicine* **107**, 62–67.

Borger ML, Wilson LL, Sink JD, Ziegler JH, Davis SL (1973) Zeranol and dietary protein level effects on live performance, carcass merit, certain endocrine factors and blood metabolite levels of steers. *Journal of Animal Science* **36**, 706–711.

Cahill VR, Kunkle LE, Klosterman EW, Deatherage FE, Wierbicki E (1956) Effect of Diethylstilbestrol implantation on carcass composition and the weight of certain endocrine glands of steers and bulls. *Journal of Animal Science* **15**, 701–709.

Calkins CR, Clanton DC, Berg TJ, Kinder JE (1986) Growth, carcass and palatability traits of intact males and steers implanted with zeranol or estradiol early and throughout life. *Journal of Animal Science* **62**, 625–631.

Cochran WG (1937) Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society* **4**(Supp.), 102–118.

Danesh J, Collins R, Appleby P, Peto R (1998) Association of fibrinogen, C-reactive protein, albumin, or leucocyte count with coronary heart disease. *Journal of the American Medical Association* **279**, 1477–1482.

DerSimonian RF, Laird NM (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.

Dunn R, Griggs S, Olson J, Beasley M (1995) A meta-analytic validation of the Dunn and Dunn model of learning-style preferences. *Journal of Educational Research* **88**, 353–362.

Forrest RJ (1975) Effects of castration, sire and hormone treatments on the quality of rib roasts from Holstein-Friesian males. *Canadian Journal of Animal Science* **55**, 287–290.

Forrest RJ, Sather LA (1965) The effect of hormone on the rate of gain feed consumption of Holstein-Friesian steers slaughtered at 340, 522, and 703 kilograms body weight. *Canadian Journal of Animal Science* **45**, 173–179.

Forza C, Dinuzzo F (1998) Meta-analysis applied to operations management – summarizing the results of empirical research. *International Journal of Production Research* **38**, 837–861.

Foutz CP, Dolezal HG, Gardner TL, Gill DR, Hensley JL, Morgan JB (1997) Anabolic implant effects on steer performance, carcass traits, subprimal yields, and longissimus muscle properties. *Journal Animal Science* **75**, 1256–1265.

Gerken CL, Tatum JD, Morgan JB, Smith GC (1995) Use of genetically identical (clone) steers to determine the effects of estrogenic and androgenic implants on beef quality and palatability characteristics. *Journal of Animal Science* **73**, 3317–3324.

Glass G (1976) Primary, secondary and meta-analysis of research. *Educational Researcher* **5**, 3–8.

Gurevitch J, Hedges L (1999) Statistical issues in ecological meta-analysis. *Ecology* **80**, 1142–1149.

Halpem S, Leighton B, Ohlsson A, Barrett J, Rice A (1998) Effect of epidural vs parenteral analgesia on the progress of labor—a meta-analysis. *Journal of the American Medical Association* **280**, 2105–2110.

Hardy RJ, Thompson SG (1998) Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* **17**, 841–856.

Hilborn R, Liermann M (1998) Standing on the shoulders of giants: learning from experience in fisheries. *Reviews in Fish Biology and Fisheries* **8**, 273–283.

Huck GL, Brandt RT Jr, Dikeman ME, Simms DD, Kuhl GL (1991) Frequency and timing of trenbolone acetate implantation on steer performance, carcass characteristics, and beef quality. *Journal of Animal Science* **69**(S1), 560.

Hunt DW, Henricks DM, Skelley GC, Grimes LW (1991) Use of trenbolone acetate and estradiol in intact and castrate male cattle: effects on growth, serum hormones, and carcass characteristics. *Journal of Animal Science* **69**, 2452–2462.

Hunter RA, Burrow HM, McCrabb GJ (2001) Sustained growth promotion, carcass and meat quality of steers slaughtered at three liveweights. *Australian Journal of Experimental Agriculture* **41**, 1033–1040.

Ntunde BN, Usborne WR, Ashton GC (1977) Responses in meat characteristics of Holstein–Friesian males to castration and diet. *Canadian Journal of Animal Science* **57**, 449–458.

Nute GR, Dransfield E (1984) The quality of sirloin from zeranol implanted steers. *Journal of Food Technology* **19**, 21–28.

Ouali A, Zabari M, Renou JP, Touraille C, Kopp J, Bonnet M, Valin C (1988) Anabolic agents in beef production: effects on muscle traits and meat quality. *Meat Science* **24**, 151–161.

Perry D, Shorthose WR, Ferguson DM, Thompson JM (2001). Methods used in the CRC program for the determination of carcass yield and beef quality. *Australian Journal of Experimental Agriculture* **41**, 953–958.

Phillips J (1998) Effects of realistic job previews on multiple organizational outcomes: a meta-analysis. *Academy of Management Journal* **41**, 673–698.

Platter WJ, Tatum JD, Belk KE, Scanga JA, Smith GC (2003) Effects of repetitive use of hormonal implants on beef carcass quality, tenderness, and consumer ratings of beef palatability. *Journal of Animal Science* **81**, 984–996.

Polkinghorne R, Watson R, Porter M, Gee A, Scott J, Thompson J (1999) Meat Standards Australia, A 'PACCP' based beef grading scheme for consumers. The use of consumer scores to set grade standards. In 'Proceedings of the 45th International Congress of Meat Science and Technology', pp. 14–15. (1–6 August 1999, Yokohama, Japan).

Pritchard RH, Bruns KW, Birkelo CP (2000) A comparison of lifetime implant strategies on production and carcass characteristics in steers. *Journal of Animal Science* **78**(S1), 47.

Quinn M, Kavale K, Mathur S, Rutherford R, Forness S (1999) A meta-analysis of social skill interventions for students with emotional or behavioral disorders. *Journal of Emotional and Behavioural Disorders* **7**, 54–64.

Reiling BA, Johnson DD (2003) Effects of implant regimens (Trenbolone acetate-estradiol administered alone or in combination with zeranol) and Vitamin D3 on fresh beef colour and quality. *Journal of Animal Science* **81**, 135–142.

Roeber DL, Cannell RC, Belk KE, Miller RK, Tatum JD, Smith GC (2000) Implant strategies during feeding: impact on carcass grades and consumer acceptability. *Journal of Animal Science* **78**, 1867–1874.

Rumsey TS, Elsasser TH, Kahl S, Solomon MB (1999) The effect of roasted soybeans in the diet of feedlot steers and Synovex-S Ear implants on carcass characteristics and estimated composition. *Journal of Animal Science* **77**, 1726–1734.

Samber JA, Tatum JD, Wray MI, Nichols WT, Morgan JB, Smith GC (1996) Implant program effects on performance and carcass quality of steer calves finished for 212 days. *Journal of Animal Science* **74**, 1470–1476.

Scheffler JM, Buskirk DD, Rust SR, Cowley JD, Doumit ME (2003) Effect of repeated administration of combination trenbolone acetate and estradiol implants on growth, carcass traits, and beef quality of long-fed Holstein steers. *Journal of Animal Science* **81**, 2395–2400.

Simone M, Clegg MT, Carroll F (1958) Effect of methods of Silbestrol administration on quality factors of beef. *Journal of Animal Science* **17**, 834–840.

Thompson JM, McIntyre BM, Tudor GD, Pethick DW, Polkinghorne R, Watson R (2008*a*) Effects of hormonal growth promotants (HGP) on growth, carcass characteristics, the palatability of different muscles in the beef carcass and their interaction with aging. *Australian Journal of Experimental Agriculture* **48**, 1405–1414.

Thompson JM, Polkinghorne R, Porter M, Burrow HM, Hunter RA, McCrabb GJ, Watson R (2008*b*) Effect of repeated implants of oestradiol-17β on beef palatability in Brahman and Brahman cross steers finished to different market end points. *Australian Journal of Experimental Agriculture* **48**, 1434–1441.

Thompson S, Pocock S (1991) Can meta-analysis be trusted? (editorial). *Lancet* **338**, 1127–1130.

Thonney ML, Perry TC, Armbruster G, Berrman DH, Fox DG (1991) Comparison of steaks from Holstein and Simmental × Angus Steers. *Journal of Animal Science* **69**, 4866–4870.

Touloumi GF, Katsouyanni K, Zmirou D, Schwartz J, Spix C (1997) Short term effects of ambient oxidant exposure on mortality: a combined analysis within the APHEA project. *American Journal of Epidemiology*, **146**, 177–185.

Watson R, Gee AM, Polkinghorne R, Porter M (2008*a*) Consumer assessment of eating quality – development of protocols for Meat Standards Australia (MSA) testing. *Australian Journal of Experimental Agriculture* **48**, 1360–1367.

Watson R, Polkinghorne R, Gee A, Porter M, Thompson JM, Ferguson D, Pethick D, McIntyre B (2008*b*) Effect of hormonal growth promotants on palatability and carcass traits of various muscles from steer and heifer carcasses from a *Bos indicus–Bos taurus* composite cross. *Australian Journal of Experimental Agriculture* **48**, 1415–1424.

Manuscript received 12 June 2007, accepted 20 June 2008

Appendix: Meta-analysis

Consider a collection of k studies, each of which measures the effect of a particular intervention. It can be assumed that from each of these studies a point estimate of the effect of intervention and a standard error of that estimate are available. This Appendix considers two standard methods by which these studies may be combined in order to obtain an estimate and confidence interval for the overall effect of the intervention of interest.

For each of the k studies let $\hat{\theta}_i$ $(i = 1, 2, \ldots, k)$ denote the estimated effect of intervention and θ_i the true effect of intervention. A general model is then specified by

$$\hat{\theta}_i = \theta_i + e_i \quad \text{where} \quad e_i \sim \text{N}(0, \sigma_i^2), \quad i = 1, 2, \ldots, k,$$

and the e_i are assumed to be independent. The estimated effect size $\hat{\theta}_i$ can be any measure of effect provided the assumption of normality is (at least approximately) appropriate.

In practice the σ_i^2 are unknown and so estimated variances are used. It is widely assumed, however, that the individual studies provide reasonable estimates of the σ_i^2. This assumption is almost universally made when the studies are analysed individually and should be the case provided the studies are at least moderate in size. In order to emphasize that the within study variances are estimated, the notation $\hat{\sigma}_i^2$ is used, where $\hat{\sigma}_i$ is the standard error of $\hat{\theta}_i$. The general model is therefore rewritten as

$$\hat{\theta}_i = \theta_i + e_i \quad \text{where} \quad e_i \sim \text{N}(0, \hat{\sigma}_i^2), \quad i = 1, 2, \ldots, k.$$

For the meta-analyses considered here, the parameter of interest is the overall effect of intervention, denoted μ. In the remainder of this Appendix outlines both the fixed and random effects models and describes how each relates μ to the θ_i. In both cases the estimates $\hat{\theta}_i$ are point estimates of μ. Mention is also made of a test of homogeneity, widely used to select either the fixed or random effects model.

The fixed effects approach

The fixed effects model for meta-analysis assumes that the k studies are homogeneous—each having the same true effect of intervention. Therefore $\theta_i = \mu$ for all $i = 1, 2, \ldots, k$, giving the model

$$\hat{\theta}_i = \mu + e_i \quad \text{where} \quad e_i \sim \text{N}(0, \hat{\sigma}_i^2), \quad i = 1, 2, \ldots, k.$$

This model only allows for within-study variation.

The overall effect μ is commonly estimated using a weighted average. Ignoring the sampling error in the $\hat{\sigma}_i^2$, it is optimal to use weights proportional to $1/\hat{\sigma}_i^2$ giving

$$\hat{\mu} = \frac{\sum \hat{w}_i \hat{\theta}_i}{\sum \hat{w}_i} \quad \text{where} \quad \hat{w}_i = \frac{1}{\hat{\sigma}_i^2}.$$

The notation $\hat{w}_i$ is used to indicate that the weights use estimates $\hat{\sigma}_i^2$, although the variation associated with these estimates is ignored in practice. Under the assumptions of independence and normality,

$$\widehat{\text{var}}(\hat{\mu}) = \frac{1}{\sum \hat{w}_i},$$

and an approximate 95% confidence interval for μ is given by

$$\hat{\mu} \pm \frac{1.96}{\sqrt{\sum \hat{w}_i}}.$$

The random effects method and testing for homogeneity

Under the random effects model a distribution is assumed for the θ_i giving the two-stage model

$$\hat{\theta}_i = \theta_i + e_i \quad \text{where} \quad e_i \sim \text{N}(0, \hat{\sigma}_i^2), \quad i = 1, 2, \ldots, k,$$

$$\theta_i = \mu + \varepsilon_i \quad \text{where} \quad \varepsilon_i \sim \text{N}(0, \tau^2), \quad i = 1, 2, \ldots, k;$$

and where the e_i and ε_i are assumed to be independent. In this model, the true effect for study i is centred around the overall effect μ, allowing the individual studies to vary in both estimated and true effect. The between-study variance parameter, τ^2, is a measure of the heterogeneity between studies and clearly this model permits both within and between-study variation.

This model can be written

$$\hat{\theta}_i = \mu + e_i + \varepsilon_i \quad \text{where} \quad e_i \sim \text{N}(0, \hat{\sigma}_i^2) \quad \text{and} \quad \varepsilon_i \sim \text{N}(0, \tau^2), \quad i = 1, 2, \ldots, k,$$

giving

$$\hat{\theta}_i \sim \mathrm{N}(\mu, \hat{\sigma}_i^2 + \tau^2),$$

under the assumptions of normality and independence. The variance of $\hat{\theta}_i$ includes both within and between-study variance components. The fixed effects model is a special case of the random effects model with $\tau^2 = 0$. Selection of either the fixed effects or random effects models can therefore be carried out by testing the hypothesis $\tau^2 = 0$ against a one-sided alternative. If $\tau^2 = 0$ the studies are considered to be homogeneous. If the σ_i^2 are known, a test of the homogeneity between studies can be carried out using a statistic defined by Cochran (1937):

$$Q_w = \sum \frac{(\hat{\theta}_i - \hat{\mu})^2}{\sigma_i^2},$$

which has a χ^2_{k-1} distribution under the hypothesis $\tau^2 = 0$. In practice however the σ_i^2 are not known and the statistic

$$Q_{\hat{w}} = \sum \hat{w}_i(\hat{\theta}_i - \hat{\mu})^2 \quad \text{where} \quad \hat{w}_i = \frac{1}{\hat{\sigma}_i^2}$$

is used. If $\tau^2 = 0$ then $Q_{\hat{w}} \sim \chi^2_{k-1}$ approximately, and the hypothesis of homogeneity is rejected if $q_{\hat{w}} > c_{0.95}(\chi^2_{k-1})$.

This test is frequently used to determine whether the fixed or random effects model should be adopted (for example Touloumi *et al*. 1997; Danesh *et al.* 1998). It has been suggested, however, that the power of this test can be low (Thompson and Pocock 1991; Hardy and Thompson 1998).

In order to estimate μ under the random effects model an estimate of τ^2, the between-study variance parameter, is required.

Under the random effects model it is assumed that $\hat{\theta}_i \sim \mathrm{N}(\mu, \hat{\sigma}_i^2 + \tau^2)$. As for the fixed effects method a weighted average is generally used to estimate μ where the weights are derived from the variances of the $\hat{\theta}_i$, $i = 1, 2, \ldots, k$. Therefore

$$\hat{\mu}_\tau = \frac{\sum \hat{w}_i(\tau)\hat{\theta}_i}{\sum \hat{w}_i(\tau)} \quad \text{and} \quad \mathrm{var}(\hat{\mu}_\tau) = \frac{1}{\sum \hat{w}_i(\tau)},$$

where

$$\hat{w}_i(\tau) = \frac{1}{\hat{\sigma}_i^2 + \tau^2}.$$

Here the sampling error in the $\hat{\sigma}_i^2$ is ignored and it is assumed that τ^2 is known. Under these assumptions

$$\hat{\mu}_\tau \sim \mathrm{N}\left(\mu, \frac{1}{\sum \hat{w}_i(\tau)}\right).$$

In practice however, τ^2 must be estimated. The most widely-used estimate of τ^2 is one proposed by DerSimonian and Laird (1986):

$$\hat{\tau}^2 = \max\left(0, \frac{Q_{\hat{w}} - k + 1}{W}\right),$$

where $Q_{\hat{w}}$ denotes the observed value of the homogeneity statistic and

$$W = \sum \hat{w}_i - \frac{\sum \hat{w}_i^2}{\sum \hat{w}_i}.$$

This estimate is obtained by equating the expected value of $Q_{\hat{w}}$ with the observed value, and truncating to ensure that $\hat{\tau}^2 \geq 0$.

The estimate of τ^2 is then directly incorporated into the random effects weights, giving

$$\hat{w}_i(\hat{\tau}) = \frac{1}{\hat{\sigma}_i^2 + \hat{\tau}^2}.$$

This yields

$$\hat{\mu}_{\hat{\tau}} = \frac{\sum \hat{w}_i(\hat{\tau})\hat{\theta}_i}{\sum \hat{w}_i(\hat{\tau})} \quad \text{and} \quad \widehat{\mathrm{var}}(\hat{\mu}_{\hat{\tau}}) = \frac{1}{\sum \hat{w}_i(\hat{\tau})}.$$

Confidence intervals for μ are calculated under the assumption of normality; thus a 95% confidence interval for μ is given by

$$\hat{\mu}_{\hat{\tau}} \pm 1.96\sqrt{\widehat{\mathrm{var}}(\hat{\mu}_{\hat{\tau}})}.$$

In cases where $\hat{\tau}^2 = 0$ the random effects method estimate and interval for μ are the same as those for the fixed effects method.