

23 **Introduction**

24 Acupuncture is a complementary or alternative medical procedure that involves the
25 stimulation of points on or below the skin to achieve therapeutic effects (Xie and Preast, 2007;
26 Chan et al., 2001; Kaptchuk, 2002). Acupuncture methods vary, and include traditional dry
27 needle acupuncture, electroacupuncture, acupressure, gold bead acupuncture, and other
28 techniques. Traditional acupuncture is the use of needles alone, electroacupuncture employs
29 electricity to stimulate the needles, acupressure uses pressure at specific points in the absence of
30 needles, and gold bead acupuncture is the insertion of small gold beads under the skin. The
31 theory behind acupuncture is that illness can arise due to changes in the flow of an energy,
32 known as “Qi”, and restoration of health can be achieved by stimulating specific points on the
33 body known as acupoints. By stimulating the correct points, Qi is said to be brought into balance
34 and the disease condition is resolved (Kaptchuk, 2002). Acupuncture’s popularity has increased
35 in human medicine (Barnes et al., 2004; Tindle et al., 2005) and owners may seek out this
36 therapy for their pets.

37

38 Randomized controlled trials (RCTs) provide the best evidentiary value for investigating
39 the efficacy of an intervention under real world conditions (Sargeant et al., 2014). When properly
40 conducted, RCTs reduce the likelihood of introducing many types of bias (Pandis, 2011).
41 Systematic reviews provide a scientifically defensible method for evaluating the efficacy of a
42 treatment (Roudebush et al., 2004). These reviews provide a transparent method for collecting
43 evidence from multiple RCTs (or other study designs), evaluating the quality of that evidence,
44 and synthesizing the results across studies. In 2006, the authors of a systematic review of

45 veterinary acupuncture were unable to recommend or reject acupuncture for any condition in
46 domestic animals (Habacher et al., 2006) due to the low quality of evidence and the small
47 number of controlled trials available. That review took place 10 years ago and there is a need to
48 update the evidence on veterinary acupuncture's efficacy. Previously, we conducted a scoping
49 review of veterinary acupuncture that identified a large body of literature on the subject (Rose,
50 2016). In that review, controlled studies represented 21 percent of the published literature, and
51 there appeared to be a body of evidence on acupuncture for musculoskeletal conditions in dogs
52 sufficient to synthesize.

53

54 The objective of this systematic review was to investigate the evidence for efficacy of
55 acupuncture for musculoskeletal conditions in dogs.

56

57 The following research question was developed using the PICO method for systematic
58 reviews on interventions (EFSA, 2010):

59 "What is the evidence for efficacy of acupuncture for any outcome measure related to
60 musculoskeletal conditions in dogs?"

61

62 P (population) – dogs of any breed, sex, or age

63 I (intervention) – any form of acupuncture used for therapeutic or preventive purposes

64 C (comparison) – any concurrent control (excluding another form of acupuncture)

65 O (outcome) – all reported outcomes related to musculoskeletal conditions measured in a live
66 dog

67

68 **Methods**

69 This review is reported following the guidelines of the *Preferred Reporting Items for*
70 *Systematic Reviews and Meta-Analyses* (PRISMA) (Moher et al., 2015). Decision rules for each
71 step were discussed a priori unless otherwise indicated, but an explicit protocol was not created.

72

73 *Eligibility criteria*

74 Eligible Study Types: RCTs or controlled trials (CTs). Controlled trials were defined as
75 any experimental trial with investigator control of allocation to treatment group, that did not
76 randomized subjects but still included a control group. Deliberate disease induction studies were
77 also eligible.

78

79 Eligible Participants: Dogs of any breed, age, or sex, and from any geographical region.
80 All dogs must have been used to investigate the effects of acupuncture on musculoskeletal
81 conditions. Eligible conditions were: conditions of muscles, joints, or the skeletal system
82 including injury, disease, and congenital defects. Studies investigating neoplasia of bone or
83 muscle were not eligible. Mechanistic studies were not eligible; these were studies in which no
84 disease condition was investigated and that were, instead, focused on how acupuncture might
85 work.

86

87 Eligible Interventions: Acupuncture techniques that included the insertion of needles at
88 acupoints or the insertion of materials at acupoints. Eligible techniques included traditional
89 acupuncture (needles at acupoints only), electroacupuncture, acupressure, moxibustion
90 acupuncture, laser acupuncture, injection acupuncture, and implant acupuncture. These
91 treatments could be administered alone or in combination with other treatments.

92

93 Eligible Control Groups: Sham acupuncture (e.g. stimulation of non-acupuncture points),
94 another treatment, or no treatment as control. Comparisons of one form of acupuncture to a
95 different form of acupuncture were excluded.

96

97 Eligible Outcomes: All outcomes related to musculoskeletal conditions, measured in live
98 animals.

99

100 *Search strategy:*

101 Studies were identified in a previously conducted scoping review of veterinary
102 acupuncture (Rose, 2016). In that scoping review, the following databases were searched without
103 restrictions on language, date published, or region: MEDLINE, CAB Direct, AGRICOLA,
104 CINAHL, TOXNET, Science.gov, and Web of Science. Searches were conducted from Jan 26 to
105 Jan 29, 2014 with an update search conducted on June 6, 2015. Search terms were developed to
106 identify all studies investigating the use of acupuncture in dogs, cats, and horses. Each electronic
107 database was searched using seven individual search inputs. Two of the search inputs were
108 exclusive to dogs. The first was: Dog AND (Acupuncture OR acupressure OR
109 electroacupuncture). The second was: Canine AND (Acupuncture OR acupressure OR

110 electroacupuncture). The other 5 searches substituted cat, feline, horse, equine, or veterinary in
111 the place of dog or canine. No contact was made with authors for missing information.

112

113 The output from each of the searched databases was uploaded into the RefWorks
114 reference management tool. Duplicate records that were removed using RefWorks reference
115 management tool (RefWork reference management tool, ProQuest LLC).

116 Results were then exported into Microsoft Excel (Microsoft Office Professional Plus
117 2013, Version 15.0.4779.100) for relevance screening.

118

119 Study selection: Relevance screening was conducted in two stages, the first based on title,
120 abstract, and journal, and the second based on the full article. Two questions were developed to
121 determine eligibility: “Does the article address acupuncture as a therapeutic or preventive
122 intervention?” and “Does the article mention a dog, cat, or horse with the outcome measured in
123 live animals?”. Mechanistic studies that did not address efficacy were excluded based on the first
124 question. At each stage, two independent reviewers examined each citation to determine
125 relevance. Results from the two reviewers were compared and any differences were resolved
126 through consensus.

127

128 Potentially relevant articles were acquired from the library holdings and library journal
129 subscriptions of the Tri-University Group of Libraries (University of Guelph, University of
130 Waterloo, and Wilfred Laurier University), and open-access sources such as Google Scholar.
131 Articles that could not be accessed from these sources were requested via the University of
132 Guelph inter-library loan service (Rose, 2016). If full text was not available through these

133 sources the article was excluded. If full text was not published in English, it was excluded. Some
134 articles included in the scoping review were categorized solely on the bases of review of the
135 abstract when full text was not available. Full text articles were categorized by publication type.
136 Experimental studies, case-control, and case-series studies were further categorized by species
137 studied, acupuncture methods used, and outcome categories (Rose, 2016). Two reviewers
138 independently categorized each article and any conflicts were resolved through consensus.
139 Experimental studies that were conducted on dogs and that investigated acupuncture's effects on
140 musculoskeletal conditions were considered for this systematic review. Studies identified by the
141 scoping review that were categorized by abstract alone were excluded unless the full text
142 subsequently became available.

143

144 *Data extraction and risk of bias assessment:*

145 Two independent reviewers extracted relevant data using a form designed in Microsoft
146 Access (Microsoft Office Professional Plus 2013, Version 15.0.4779.100). Text fields and check
147 boxes were used to collect study details. The data extracted included:

148

149 Study Population: Mean age in years (standard deviation [SD]), breeds of participants,
150 comorbidities, and number of female/male dogs.

151

152 Intervention: Methods used in administering acupuncture (electro, acupressure, gold
153 implant, injection, traditional, and laser), number of dogs in the intervention group, frequency of
154 treatment, size of acupuncture needle, acupuncture needle material, and needle placement
155 location.

156

157 Comparison: Description of the control intervention, number of dogs in the control group,
158 and frequency of treatment. If sham acupuncture was used for comparison, then needle
159 placement location was recorded.

160

161 Outcome: Specific musculoskeletal condition(s) being investigated. All outcomes measures
162 reported in a publication were extracted if they were relevant to a specific musculoskeletal
163 condition. For each outcome measure, the following general description data were extracted: a
164 description of the outcome, follow-up interval length, number of follow-up visits, duration of
165 overall follow-up, number of subjects enrolled, number of subjects in each group, and number of
166 subjects lost to follow-up. Although a general description of every outcome measure was
167 extracted; only outcome measures that met one of three criteria were fully extracted. These
168 criteria were developed after data extraction and were created by the review team:

169

- 170 1. Reported a treatment effect or effect measure (mean, mean difference [MD], relative risk
171 [RR], or odds ratio [OR]) as well as a measure of variability (SD, variance, or confidence
172 interval [CI]),
- 173 2. Reported the raw data such that an effect measure and variability measure could be
174 calculated, or
- 175 3. Reported Likert scale data in sufficient detail to allow for dichotomization into
176 “improvement” or “no improvement” for each treatment group. A RR was then
177 calculated for improvement vs no improvement.

178

179 If the outcome met at least one of these criteria, then effect measure data would be extracted
180 into Microsoft Excel. In the case of the first criteria, the effect measure (mean, MD, RR, or OR)
181 and the measure of variability (SD, Variance, or CI) were extracted for every time point. For the
182 second criteria, all reported raw data were extracted. For the final criteria, number dogs in each
183 of the Likert scale categories were consolidated into improvement and no improvement
184 categories by treatment group. The number of animals in each category would be extracted for
185 each time point.

186

187 For continuous or discrete scale measures such as pain scores, a standard deviation was
188 calculated if not reported (Higgins and Green, 2011).

189

190 Pre-testing of the form was conducted using one published study. Reviewers met regularly
191 throughout the data extraction process to discuss their findings and to come to consensus on any
192 differences.

193

194 Two reviewers independently assessed each of the articles using the Cochrane risk of bias
195 tool for randomized studies (Higgins et al., 2011). Risk of bias assessment was performed at the
196 study level due to the large number of eligible outcomes. Low risk was indicated if the authors
197 adequately addressed the category. Unclear risk was reported if we could not determine if the
198 authors adequately addressed the category and high risk was reported if it was clear that the
199 authors did not adequately address the category. Consensus was reached for any disagreements.
200 “Other” risk of bias was reported if there was any additional areas where reviewers felt bias
201 could have been introduced.

202

203 *Data synthesis and analysis:*

204 To summarize outcome data measured on a continuous scale across studies, we used MD
205 between treatment groups. Continuous scale measures included visual analog scales (of any
206 length); these are visual scales where owners or veterinary assessors indicate their perception of
207 pain or dysfunction for example. In the case of visual analog pain scores, the measure was the
208 MD in the reduction of mean pain score between control and treatment groups. These variables
209 could have been measured at multiple time points. The MD between baseline and the last follow-
210 up period was calculated for both the control and treatment groups. The control and treatment
211 groups were then compared using the difference between the control MD and the treatment MD
212 (treatment group MD– control group MD). A pooled SD was calculated for this overall MD
213 using standard formulae (Thalheimer and Cook, 2009). The overall MDs, pooled SD, and sample
214 size were then entered into Comprehensive Meta-Analysis (CMA) software (Comprehensive
215 Meta-Analysis, version 3.3.070). CMA calculated a confidence interval for each overall MD.

216

217 To summarize data across studies with outcomes measured on a discrete scale, we also
218 used MD. Discrete scales included pain scores that were measured on a point scale. The same
219 methods were used to attain an overall MD as those used for continuous scales.

220

221 To summarize data across studies on outcomes measured on Likert scales or for
222 dichotomous outcomes, we used RR. Likert scale measures that could be dichotomized into
223 “improvement” or “no improvement” were eligible. If a RR or OR was reported directly it would

224 be extracted. If both raw data and summary data were available summary data would be
225 extracted.

226

227 Meta-analyses were divided by disease condition and conducted on seven critical
228 outcomes that were selected post-hoc based on their clinical relevance. These seven outcomes
229 were: improvement in locomotion based on clinical investigator assessment assessed using RR,
230 improvement in locomotion based on owner assessment assessed using RR, improvement in
231 quality of life based on owner assessment assessed using RR, MD in the reduction in mean pain
232 scores assessed by clinical investigator, MD in the reduction in mean pain score assessed by
233 owner, MD in the reduction in mean lameness score assessed by clinical investigator, and MD in
234 the reduction in mean lameness score assessed by owner. A meta-analysis was conducted on
235 each of these outcomes if there were at least 2 studies that reported the outcome within a disease
236 condition. Random effects models were used to calculate a weighted RR or standardized MD.
237 All analyses were conducted using CMA. Forest plots were produced for each meta-analysis.
238 Heterogeneity was quantified using I^2 (Higgins and Green, 2011) and was calculated using CMA
239 software. Funnel plots were not created for each musculoskeletal condition due to the limited
240 number of studies available.

241

242 A GRADE (Grading of Recommendations, Assessment, Development, and Evaluation;
243 Guyatt et al., 2008) assessment was completed for each critical outcome. Evidence was assigned
244 one of four quality of evidence ratings: high, moderate, low, or very low. Quality was
245 determined by examining five domains: risk of bias, indirectness, inconsistency, publication bias,
246 and imprecision. Evidence was initially rated as high quality and was downgraded by one point

247 per domain when concerns were identified in any of the five domains. When a high risk of bias
248 was identified in a category that could have influenced the results, evidence was downgraded one
249 level. If the outcome was a surrogate measure, that outcome was downgraded for indirectness. If
250 CIs between studies did not overlap, or there was a high (>50%) calculated I^2 , the evidence was
251 downgraded for inconsistency. We did not evaluate the potential for publication bias because the
252 small number of studies for each condition precluded the evaluation of possible small study
253 effects using funnel plots or formal statistics (Mavridis and Salanti, 2014). Therefore, no
254 outcomes were downgraded based on this domain. Evidence was downgraded for imprecision
255 when total sample size was less than the estimated sample size needed to detect an appreciable
256 difference in a single study, when CIs crossed both the “no effect” and “appreciable benefit”
257 lines, or when there was only one study examining a given outcome. Estimated sample sizes to
258 detect an appreciable difference were calculated using Epi Tools sample size calculations (Epi
259 tools, 2016). RR sample sizes were calculated using a two proportion sample size calculator with
260 95% confidence and 80% power. MD sample sizes were calculated using a difference between
261 two means sample size calculator with unequal sample sizes and variances. This sample size
262 calculator used 95% confidence and 80% power. In the absence of guidelines for appreciable
263 benefit, we selected a minimum of 1.25 for RR and $\Delta 2$ points for MD.

264
265 A separate summary of findings table was constructed for each disease condition where
266 synthesis was possible. For all RR outcomes, an absolute measure of risk in a control population
267 and an absolute measure of risk in a population treated with acupuncture was calculated. Risk
268 with control was calculated using the incidence rate of improvement in the control group. The
269 overall weighted RR was then multiplied by the risk in the control group to determine the risk

270 with acupuncture. GRADEPro was used to construct the summary of findings table which
271 included: RR and MD, CI for MD, anticipated absolute effects for RR calculations, number of
272 participants in each calculation, quality of evidence (GRADE), and any additional comments
273 (GRADEpro, 2014).

274

275 **Results**

276 There were 5,158 citations identified in the previously conducted scoping review, 843 of
277 which were included in that review after relevance screening (Rose, 2016). Of these, 14
278 investigated acupuncture as a treatment for musculoskeletal conditions in dogs and were
279 therefore eligible for inclusion in this systematic review. One of these was excluded because the
280 full text was not published in English, and two were excluded because the full text could not be
281 retrieved. Eleven studies were therefore included in this review. All included studies involved
282 natural disease exposure.

283

284 Population characteristics can be found in Table 1. Breeds were not reported in 3 studies;
285 the remaining studies each included several different breeds. Average age was reported in 4 of
286 the 11 studies. Six of the studies excluded animals with comorbidities. Four studies did not
287 indicate whether the dogs had any comorbidities, and one study reported that some dogs had a
288 variety of disorders (gastrointestinal problems, urinary retention, paralysis) in addition to the
289 musculoskeletal condition being studied. The sex of the animals was reported in 4 studies.

290

291

292
293
294
295
296
297
298
299
300
301
302

Table 1. Population characteristics of 11 studies examining the efficacy of acupuncture for musculoskeletal conditions in dogs

Study	Average age of dogs in years (SD)	Breeds of dogs	Reported comorbidities	Sex (# Male, # Female)
Balaji et al., 1998	NR	NR	None	NR
Bolliger et al., 2002	NR	MM 5, GS 5, MI 2, LR 3, GSC 1, AH 1, GR 3	None	NR
Hayashi et al., 2007	5.49 (2.04)	DAC 37, CS 5, SS 1, PO 5, PEK 1, MI 1, NR 11	Gastrointestinal issues, urinary retention, paralysis	27 Male, 23 Female
Hielm-Bjorkman et al., 2001	5.1 (NR)	SET 5, GS 5, GR 4, SS 2, CC 2, ROT 2, BR 2, BS 2, MI 2, other 12	None	21 Male, 17 Female

Jaeger et al., 2006	6.3 (NR)	GS 19, GR 7, LR 7, 6 MI, NR 24	None	33 Male, 47 Female
Jaeger et al., 2007	6.3 (NR)	NR	None	NR
Joaquim et al., 2010	NR	NR	NR	NR
Kapatkin et al., 2006	NR	LR 6, GR 1, GS 1, MI 1	NR	NR
Sharifi et al., 2009	NR	MI 10	None	5 Male, 5 Female
Um et al., 2005	NR	MI 8	NR	NR
Vecino et al., 2005	NR	BEA 2	NR	NR

AH, Afghan Hound; BEA, Beagle; BR, Briard; BS, Bernese Sennenhund; CC, Chow Chow; CS, Cocker Spaniel; DAC, Dachshund; GR, Golden Retriever; GS, German Shepherd; GSC, Giant Schnauzer; LR, Labrador Retriever; MI, mixed breed; MM, Malamute; NR, not reported; PEK, Pekingese; PO, Poodle; ROT, Rottweiler; SET, Setter; SS, Springer Spaniel; SD, Standard Deviation;

Methods used in the administration of acupuncture included electroacupuncture, gold bead acupuncture, and traditional dry-needle acupuncture (Table 2). Electroacupuncture administration varied in the methods used to apply current. The beads used in gold bead acupuncture were similar in nature, made of 24 karat gold and typically 1 mm in diameter and 2 mm long. Beads were inserted just below the skin at selected acupoints and remained there for the duration of treatment. In one study examining traditional dry-needle acupuncture, a corticosteroid and an analgesic were also administered when necessary (Hayashi et al., 2007). The specific acupoints stimulated varied (Table 2). A variety of methodologies for locating acupoints on the dog's skin were employed,

including the measurement of conductivity using a galvanometer, measurement of resistance using an ohmmeter, and reference to the published literature.

The characteristics of each control group also varied between studies (Table 2). Controls included no treatment, superficial needle punctures at non-acupoints, corticosteroid and analgesic administration, and decompressive surgery. Two studies provided no information on the nature of the control group (Jaeger et al., 2006; Vecino et al., 2005).

303

304 In the eleven studies included in this review, eight musculoskeletal conditions were

305 investigated (Table 3). The most common condition was hip dysplasia (4 studies).

306 Thoracolumbar intervertebral disk disease [IVDD] was the second most commonly studied (two
307 studies).

308

309

Table 3. Musculoskeletal conditions investigated in each study examining the efficacy of acupuncture for musculoskeletal conditions in dogs

Specific musculoskeletal condition investigated	Study(s)
Achilles tendon injury	Sharifi et al., 2009
Chronic elbow joint osteoarthritis	Kapatkin et al., 2006
Hip Dysplasia	Bolliger et al., 2002; Hielm-Bjorkman et al., 2001; Jaeger et al., 2006; Jaeger et al., 2007
Intervertebral Disk Disease	Hayashi et al., 2007; Joaquim et al., 2010
Osteoporosis	Vecino et al., 2005
Stifle joint chronic arthritis	Um et al., 2005
Stifle ligament injury	Balaji et al., 1998
Thigh muscle injury	Balaji et al., 1998

310

311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329

330 The results of the risk of bias assessments at the study level can be found in Table 4.
331 Random sequence generation was not detailed in two studies resulting in an unclear risk of bias.
332 In three studies, there was no randomization or the randomization protocol was ceased during the
333 study. This resulted in a high risk of bias. No information related to methods to conceal
334 allocation was reported in any trial, therefore, the risk of bias related to allocation was classified
335 as unclear for all studies. There was a unclear risk of bias in both blinding categories for eight of
336 the eleven studies; this was due to not reporting the methods used to blind. Incomplete outcome
337 data was deemed a high risk in four studies as some animals were not included in the analysis of

338 specific outcome measures without explanation. In one case, the risk of other forms of bias was
339 classified as unclear. This was because the study did not report any specific effect measure
340 figures or measures of variability. This study only reported whether the outcome measure was
341 significant, with not additional information (Balaji et al., 1998).

342

343

344

345

346

347

348

349

350

351

352

353

354

355

**Table 4. Risk of bias analysis conducted at the study level for each study
examining the efficacy of acupuncture for musculoskeletal conditions in dogs
(Cochrane Collaboration assessment tool) (Higgins et al., 2011)**

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of personnel (performance bias)	Blinding of owners (detection bias)	Incomplete outcome data (attrition bias)	Selective Reporting (reporting bias)	Other bias
Balaji et al., 1998	Yellow	Yellow	Yellow	Yellow	Red	Green	Yellow
Bolliger et al., 2002	Green	Yellow	Green	Green	Red	Green	Green
Hayashi et al., 2007	Green	Yellow	Yellow	Yellow	Red	Green	Green
Hjelm-Bjorkman et al., 2001	Yellow	Yellow	Green	Green	Red	Yellow	Green
Jaeger et al., 2006	Green	Yellow	Green	Green	Green	Red	Green
Jaeger et al., 2007	Green	Yellow	Red	Red	Green	Green	Green
Joaquim et al., 2010	Red	Yellow	Yellow	Yellow	Green	Green	Green
Kapatkin et al., 2006	Red	Yellow	Yellow	Yellow	Green	Green	Green
Sharifi et al., 2009	Green	Yellow	Yellow	Yellow	Green	Green	Green
Um et al., 2009	Green	Yellow	Yellow	Yellow	Yellow	Green	Green
Vecino et al., 2005	Red	Yellow	Yellow	Yellow	Green	Green	Green

Red: high risk of bias; green: low risk of bias; yellow: unknown risk of bias

356 Five of the eleven trials reported outcomes that met one of our eligibility criteria for full
 357 extraction. These trails investigated two musculoskeletal conditions: hip dysplasia and

358 thoracolumbar intervertebral disk disease. None of the studies that met our eligibility criteria
359 directly reported OR, MD or RR. We calculated RR and MD for eligible outcomes using the raw
360 data provided by each study. Table 5 contains all outcomes eligible for synthesis that examined
361 acupuncture's efficacy in hip dysplasia and Table 6 contains all eligible outcomes that
362 investigated acupuncture's efficacy in thoracolumbar intervertebral disk disease. These outcomes
363 include days to recovery of ambulation, successful treatment, improvement in proprioception,
364 improvement in orthopedic evaluation, improvement in quality of life, pain scores, and lameness
365 scores. The MD or RR is reported for each of these outcomes with corresponding confidence
366 intervals and the numbers of dogs in each intervention group. All these outcomes were eligible
367 for extraction however only outcomes that corresponded to our 7 critical outcomes were used in
368 meta-analysis and GRADE assessment.

369

370

371

372

373

374

375

376

377

378

379

380

Table 6. Outcomes eligible for extraction related to intervertebral disk disease that examined the efficacy of acupuncture for musculoskeletal conditions in dogs

Trial	Outcome description	Effect measure (95% CI)	# Treat	# Control
Hayashi et al., 2007	Mean difference in days to recovery of ambulation for dogs with grade* 3 and 4 dysfunction	MD: -10.7 (-19.3, -2.2)	10	9
Hayashi et al., 2007	Mean difference in days to recovery of ambulation for dogs with grade 5 dysfunction	MD: -3.3 (-7.8,1.1)	6	8
Hayashi et al., 2007	RR of successful** treatment for all grades* of dysfunction	RR: 1.2 (0.9, 1.5)	26	24
Hayashi et al., 2007	RR of successful** treatment rate for grades* 3-4 of dysfunction	RR: 1.5 (0.9, 2.4)	10	9
Hayashi et al., 2007	RR of successful** treatment for grade* 5 of dysfunction	RR: 4 (0.5, 29.6)	6	8
Hayashi et al., 2007	RR of improvement in urinary control	RR: 1.6 (0.8, 3.1)	10	12
Hayashi et al., 2007	RR of improvement in proprioception	RR: 1.7 (1.2, 2.5)	26	24
Hayashi et al., 2007	RR of improvement in partial to full recovery of ambulation.	RR: 1.6 (1.165, 2.332)	26	24
Joaquim et al., 2010	RR of improvement in lack of deep pain perception	RR: 8.3 (0.5, 125.1)	19	11
Joaquim et al., 2010	RR of successful treatment based on a myelopathy*** score of 1 or 2	RR: 2.0 (0.9, 4.4)	19	11

Negative MD denotes a numerical improvement in the acupuncture group compared to control group. RR > 1 favors acupuncture treatment. *Neurologic dysfunction graded 1 to 5; grade 1 = no neurologic signs except pain associated with IVDD, grade 2 = conscious proprioceptive deficit and ambulatory paraparesis, grade 3 = nonambulatory paraparesis and deep pain perception, grade 4 = nonambulatory paraplegia and deep pain perception with or without urinary dysfunction, and grade 5 = nonambulatory paraplegia and no deep pain perception with or without urinary dysfunction. **Success was considered to have occurred when a dog with grade 3, 4, or 5 dysfunction was able to walk without assistance or had return of deep pain perception or a dog with grade 1 or 2 dysfunction had pain control, and improvement in conscious proprioception and ataxia, or both.

***Myelopathy scoring system: grade 1, pain in the vertebral region with no abnormal neurologic signs; grade 2, able to bear weight, deficits of proprioception, and ambulatory paraparesis; grade 3, unable to bear weight, severe incoordination, intact spinal reflexes or hyperreflexia, and deep pain perception; grade 4, nonambulatory paraparesis, deficits of proprioception, and deep pain perception; and grade 5, any of the aforementioned clinical signs plus paraplegia, no deep pain perception, and bladder dysfunction.

381 All outcomes that met our criteria for extraction related to hip dysplasia and thoracolumbar

382 intervertebral disk disease can be found in tables 5 and 6.

383

384 Meta-analyses were possible for hip dysplasia and intervertebral disk disease. Among studies
385 examining hip dysplasia, a meta-analysis was possible for 6 of the 7 clinically relevant outcomes
386 the results of which can be found in Table 7. For RR of improvement in locomotion based on
387 clinical investigator assessment three studies reported outcomes that could be used in this
388 analysis: Bolliger et al., 2002, Jaeger et al., 2006, and Jaeger et al., 2007. The outcome used from
389 Bolliger et al., 2002 was: RR of improvement of orthopaedic evaluation conducted by veterinarian at 3
390 months. The outcome used from Jaeger et al., 2006 was: RR of improvement in lameness after 6 months
391 assessed by clinical investigator. The outcome used from Jaeger et al., 2007 was: RR of improvement in
392 lameness assessed by clinical investigator. For RR of improvement in locomotion based on owner
393 assessment three studies reported outcomes that could be used in this analysis: Bolliger et al., 2002,
394 Jaeger et al., 2006, and Jaeger et al., 2007. The outcome used from Bolliger et al., 2002 was: RR of
395 improvement in owner questionnaire at three months Category 2: Difficulties in climbing stairs, jumping
396 onto furniture or into the car. The outcome used from Jaeger et al., 2006 was: RR of improvement in the
397 signs of hip dysplasia according to owner's general impression of their dog's behavior in its daily life
398 after 6 months. The outcome used from Jaeger et al., 2007 was: RR of improvement of owner's
399 impression of dog's behaviour. For RR of improvement in quality of life based on owner assessment two
400 studies reported outcomes that could be used in this analysis: Jaeger et al., 2006, and Jaeger et al., 2006.
401 The outcome used from Jaeger et al., 2006 was: RR of improvement in owner's assessment in change of
402 quality of life at 6 months. The outcome used from Jaeger et al., 2007 was: RR of improvement of
403 owner's impression in the change of dog's quality of life. For MD in the reduction of mean pain scores
404 assessed by clinical investigator two studies reported outcomes that could be used in this analysis: Jaeger
405 et al., 2006 and Jaeger et al., 2007. The outcome used from Jaeger et al., 2006 was: MD in reduction of
406 total hip pain score assessed by clinical investigator from time 0-6 months between treatment and control.
407 The outcome used from Jaeger et al., 2007 was: MD in the reduction of mean hip pain score evaluated by
408 clinical investigator from 0 to 24 months between acupuncture and control. For MD in the reduction of

409 mean pain scores assessed by owners two studies reported outcomes that could be used in this analysis:
410 Jaeger et al., 2006 and Jaeger et al., 2007. The outcome used from Jaeger et al., 2006 was: MD in
411 reduction of mean pain score assessed by owner from time 0-3 months between treatment and control.
412 The outcome used from Jaeger et al., 2007 was: MD in the reduction of mean hip pain signs score
413 evaluated by owners from 0 to 24 months between acupuncture and control. No meta-analysis was
414 possible for MD in the reduction of mean lameness score assessed by clinical investigator as only one
415 study reported this outcome. For MD in the reduction of mean lameness score assessed by owners two
416 studies reported outcomes that could be used in this analysis: Jaeger et al., 2006 and Jaeger et al., 2007.
417 The outcome used from Jaeger et al., 2006 was: MD in reduction of owner assessed dysfunction score
418 from time 0-6 months between treatment and control. The outcome used from Jaeger et al., 2007 was:
419 MD in the reduction of mean dysfunction score evaluated by owner from 0 to 24 months between
420 acupuncture and control. The heterogeneity as measured by I-squared can be found in each meta-analyses
421 forest plot (Figure 1-5). None of the analyses had CI that indicated appreciable benefit. All grade
422 assessments indicated a low or very low confidence in the effect estimate. A summary of findings table
423 for hip dysplasia can be found in Table 7.

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

Figure 1: RR of improvement in locomotion for dogs with hip dysplasia based on clinical investigator assessment

443

Study name

444

Risk ratio and 95% CI

445

Risk ratio **Lower limit** **Upper limit**

446

Bolliger et al., 2002 0.500 0.055 4.583

447

Jaeger et al., 2007 5.657 0.389 82.254

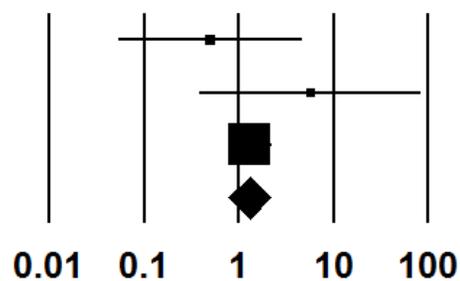
448

Jaeger et al., 2006 1.300 0.767 2.203

449

1.303 0.788 2.157

450



451

I-squared: 0.00

452

453

454

452

453

454

455

456

457

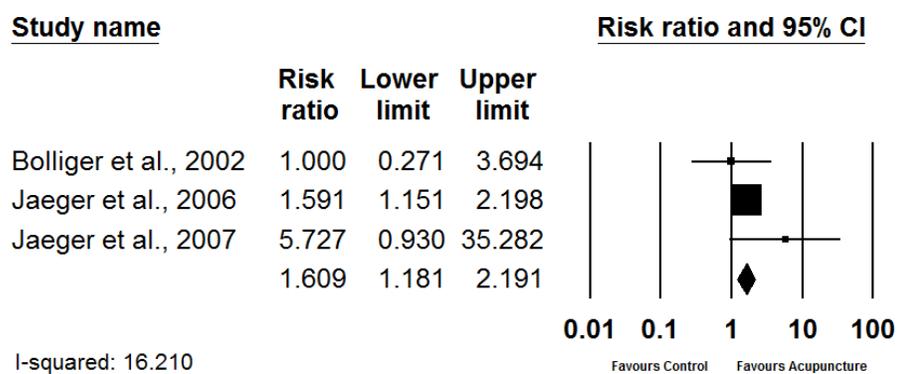
458

459

460

461

Figure 2: RR of improvement in locomotion for dogs with hip dysplasia based on owner assessment



462

463

464

465

466

467

468

469

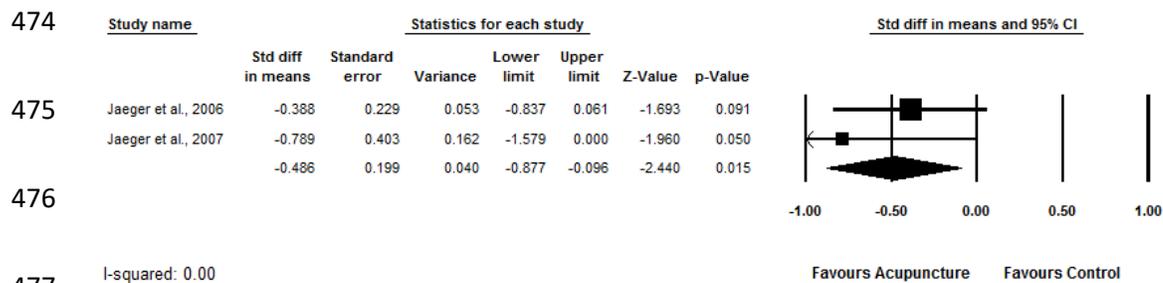
470

471

472

473

Figure 3: MD in the reduction of mean pain scores for dogs with hip dysplasia assessed by clinical investigator



477

478

479

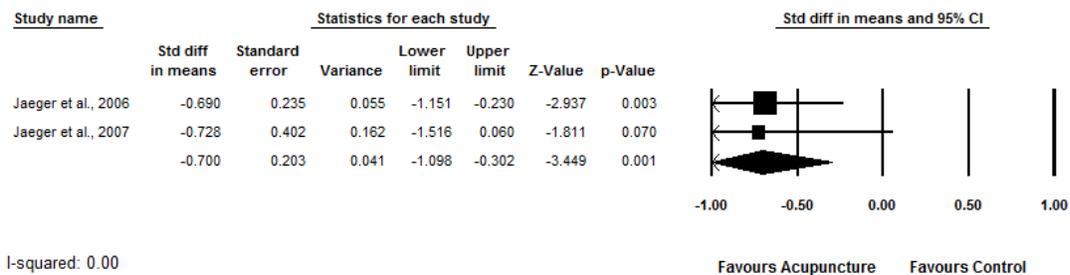
480

481

482

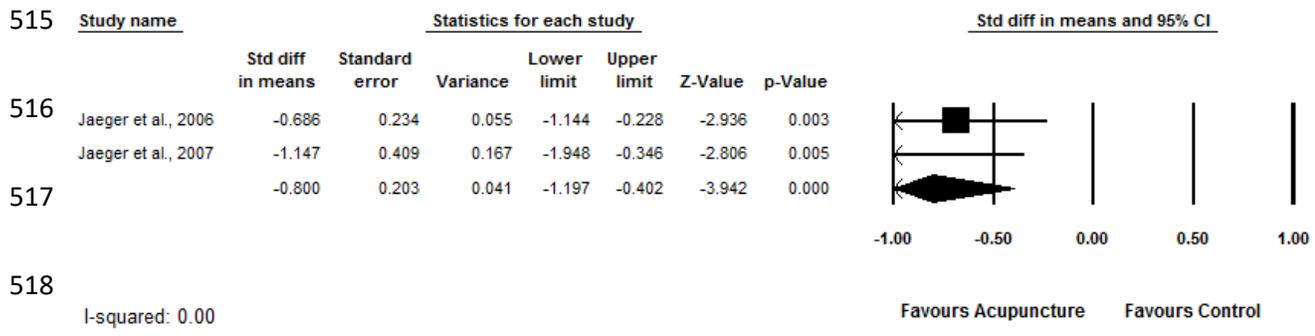
483
 484
 485
 486
 487
 488
 489
 490
 491

Figure 4: MD in the reduction of mean pain score for dogs with hip dysplasia assessed by owners



503
504
505
506
507
508
509
510
511
512
513
514

Figure 5: MD in the reduction of mean lameness score for dogs with hip dysplasia assessed by owners



515
516
517
518
519
520
521
522

523

524

525

526

527

528

529

530

531

532

533 Among studies examining thoracolumbar intervertebral disk disease one critical outcome had sufficient
534 data for meta-analysis. This outcome was RR of improvement in locomotion based on clinical
535 investigator assessment. Two studies reported outcomes for this analysis: Hayashi et al., 2007 and
536 Joaquim et al., 2010. The outcome used from Hayashi et al., 2007 was: RR of successful treatment for
537 all grades of dysfunction. The outcome used from Joaquim et al., 2010 was: RR of successful
538 treatment based on a myelopathy score of 1 or 2. Heterogeneity as measured by I-squared can be
539 found in Figure 6. The 95% CI does not indicate appreciable benefit and the grade assessment indicates
540 that we have very low confidence in the effect estimate. A summary of findings table can be found in
541 Table 8.

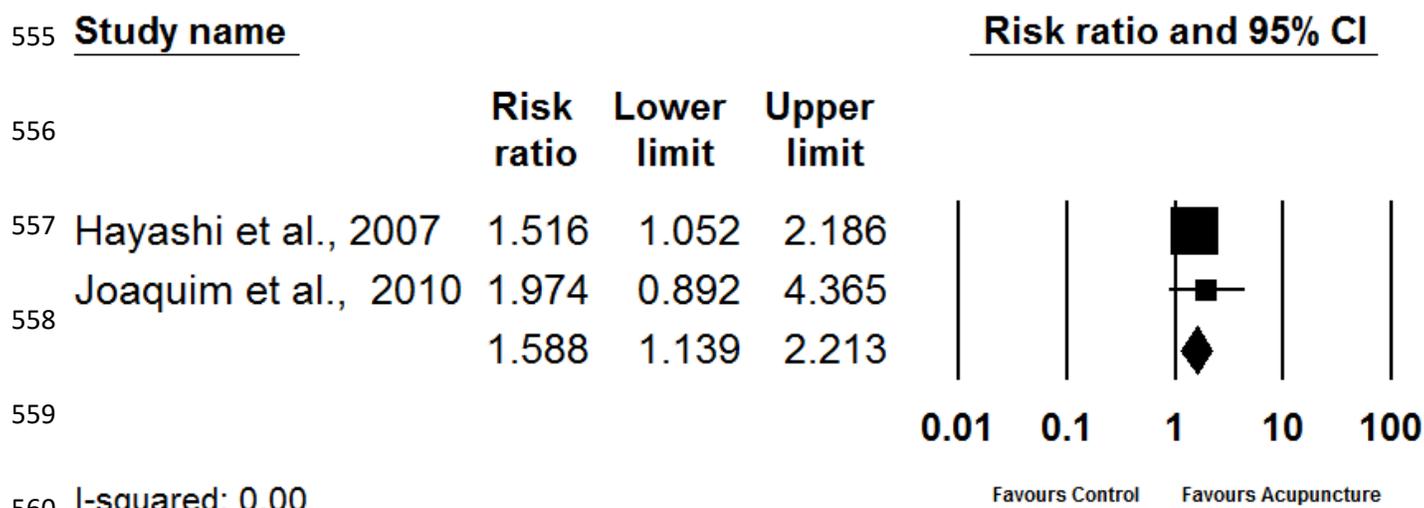
542

543

544

545
546
547
548
549
550
551
552
553

554 **Figure 6: RR of improvement in locomotion for dogs with thoracolumbar intervertebral disc disease based on clinical investigator assessment**



561
562

563
564
565
566
567
568
569
570
571

Table 8: Summary of Finding for Acupuncture compared to control for thoracolumbar intervertebral disc disease in dogs

Patient or population: thoracolumbar intervertebral disc disease in dogs

Setting:

Intervention: acupuncture

Comparison: control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control	Risk with acupuncture				
Relative risk of improvement in locomotion based on clinical investigator assessment assessed with: Likert scale follow up: mean 3.25 months	647 per 1,000	763 per 1,000 (582 to 1,000)	RR 1.179 (0.900 to 2.200)	79 (2 RCTs)	⊕○○○ VERY LOW ^{a-b,c}	

Table 8: Summary of Finding for Acupuncture compared to control for thoracolumbar intervertebral disc disease in dogs

Patient or population: thoracolumbar intervertebral disc disease in dogs

Setting:

Intervention: acupuncture

Comparison: control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control	Risk with acupuncture				

*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

- 572 a. High risk of bias for blinding of outcome assessors
 573 b. High risk of bias for random sequence generation
 574 c. Inadequate sample size to detect a significant difference

575
576

577

578

579 Discussion

580 The purpose of this systematic review was to examine the efficacy of acupuncture for
 581 musculoskeletal conditions in dogs. A previous scoping review (Rose, 2016) identified 5,158
 582 citations pertaining to acupuncture for a broad range of conditions in dogs, cats, and horses. One-
 583 fifth of those publications described controlled studies. Outcomes investigated among those
 584 studies varied widely; some of the most common outcomes were cardiovascular parameters,
 585 anesthesia, changes in blood parameters, and musculoskeletal conditions. We focused on

586 musculoskeletal conditions since, anecdotally, they represent the most common clinical
587 application of veterinary acupuncture (Chan et al., 2001). All acupuncture methodologies,
588 control protocols, and outcomes related to musculoskeletal conditions were included in this
589 review. There were too few published studies to allow the inclusion criteria to be narrowed
590 further.

591

592 Reporting of population characteristics was not consistent between studies. Many studies
593 did not report average age, breeds, sex, or comorbidities. Incomplete reporting of these
594 characteristics can limit the reader's ability to judge the generalizability of the results. Future
595 studies should report these characteristics to improve the ability to evaluate external validity.
596 CONSORT reporting guidelines can be followed to improve reporting (Schulz et al., 2010).
597 Comorbidities should always be reported, and if there are none an explicit statement to that effect
598 should be made. When comorbidities are not mentioned, it is impossible to determine whether
599 they were present or not and if there were any significant differences between the control and
600 treatment groups with regard to their health. This further complicates the synthesis of study
601 results.

602

603 Methods used in administering acupuncture varied between studies. Electroacupuncture,
604 gold bead acupuncture, and traditional dry-needle acupuncture were the three methods used in
605 the studies included in this review. Electroacupuncture studies applied different amounts of
606 current to the needles. The acupoints selected also varied between studies, to such an extent that
607 each study used a different set of acupoints with very little commonality with any other study.

608 Several studies employed a baseline set of acupoints and then located additional points by
609 searching for areas of skin with higher local conductivity. Two studies described the use of an
610 ohmmeter to find points with lower resistance, indicating higher conductivity. There was,
611 therefore, little consistency in needle placement on a dog's skin, however this is to be expected
612 as practitioners often sought to individualize treatment to the patient. It may however, introduce
613 potential heterogeneity when attempting to summarize the evidence across studies.

614

615 Differences in the control groups present another challenge in meaningfully summarizing
616 evidence across studies. Depending on the study, controls consisted of no treatment, placement
617 of needles at non-acupoints, drug regimens, or decompressive surgery, and in the case of two
618 studies no details were given about the control group (Jaeger et al., 2006; Vecino et al., 2005).
619 Thorough reporting of all protocol details in future studies will allow for more reliable evidence
620 synthesis. Knowing the nature of the control protocol is necessary to determine if summarizing
621 evidence is appropriate.

622

623 Acupuncture can be used as a combination therapy or alone. In one of the studies
624 (Hayashi et al., 2007) evaluating the use of traditional dry needle acupuncture, analgesic and
625 anti-inflammatory drugs were administered concurrently when required. Such ad hoc
626 combination therapy makes the interpretation of results more challenging. If acupuncture is used
627 as a combination therapy, expectations of efficacy would differ compared to studies in which
628 acupuncture is used alone to treat disease. If acupuncture is considered to be an effective adjunct,
629 it may be more informative to compare conventional therapy with and without acupuncture. If

630 acupuncture is thought to be capable of replacing certain interventions, then a simple comparison
631 would be most appropriate.

632

633 Risk of bias assessment revealed deficiencies in many aspects of study design and the
634 reported information often made it difficult to judge the potential for bias. There was the
635 potential for selection bias in many of the included studies due to deficiencies in random
636 sequence generation and allocation concealment categories. Non-randomized studies can
637 overestimate effect size, illustrating the need for proper randomization (Schulz et al., 1995).
638 Blinding, which was not reported in many studies, plays a pivotal role in reducing the risk of
639 bias, especially where outcome measures are inherently relatively subjective, as is the case with
640 pain scores. Decreasing the number of subjective outcome measures would decrease the risk of
641 bias in future studies, as would blinding of owners to treatment, which should be achievable by
642 using non-acupoint needle placement as the control. Selective outcome reporting was difficult to
643 assess because there is no pre-study registry listing the outcomes to be evaluated for veterinary
644 clinical studies. A pre-study registry would decrease the likelihood of reporting only favorable
645 outcomes, which may be spurious (Toews, 2011). Adherence to reporting guidelines, such as
646 those detailed in the CONSORT and REFLECT statements, can help improve the reporting of
647 studies (Moher et al., 2010; O'Connor et al., 2010). Future studies of veterinary acupuncture
648 should follow such guidelines that facilitate interpretation of results by the reader and inclusion
649 of results in future SRs.

650

651 Many different outcomes were investigated among the included trials and there was little
652 consistency even between trials investigating the same specific musculoskeletal condition. Of the seven
653 clinically relevant outcomes chosen for meta-analysis and GRADE evaluation, only 2 of the 8 specific
654 musculoskeletal conditions included these outcomes. None of the trials reporting pain, lameness scores, or
655 improvement in locomotion appear to have used a validated tool. Due to the subjective nature of such
656 scores, and the lack of validated assessment tools, the reported differences in pain, lameness, and
657 locomotion scores may not be interpretable (Epstein et al., 2015; Wojciechowska et al., 2005; Yeates
658 and Main, 2009). These limitations also affect the quality of life measures reported in some studies.

659
660 None of the MD meta-analyses had 95% CI that indicated an appreciable benefit. We selected a 2 point
661 change as our point of appreciable benefit because of the relatively subjective nature of pain scores and
662 lameness scores, they have been shown to be reliable only in detecting large improvements or
663 deteriorations (Kunz and Oxman, 1998; Quinn et al., 2007). The small improvements in pain or lameness
664 scores may not show clinically meaningful benefits.

665
666
667
668 This review was limited by the small number of published studies, the differences
669 between studies, and the exclusion of three studies due to publication language or inability to
670 locate the full text. There were only 14 studies identified investigating acupuncture's efficacy for
671 musculoskeletal conditions in dogs, and among the 11 included in this review eight different
672 conditions were investigated. This makes evidence synthesis difficult since results are often not
673 comparable across different disease conditions. There were also multiple methods used when

674 administering acupuncture. Additional complexity was introduced by the varied methods
675 employed when conducting electroacupuncture. The heterogeneous nature of treatment methods
676 adds another limitation to evidence synthesis. Outcomes measured also varied widely between
677 studies, resulting in few comparable studies.

678

679 This review revealed a body of evidence that was heterogeneous in treatment methods,
680 control protocols, and outcome measures between studies, severely limiting the potential for data
681 synthesis and therefore the strength of the available evidence. Risk of bias in the available
682 studies was high in multiple categories, further decreasing the strength of evidence. All of the
683 GRADE quality of evidence grades show that we are not confident in the calculated effect
684 estimates. There is a need for several well-designed and reasonably similar studies focusing on
685 the most common clinical uses of veterinary acupuncture in order to establish its efficacy. These
686 studies should use validated and clinically relevant outcome measures and adhere to
687 requirements intended to minimize the risk of bias.

688

689 There was no external funding for this review.

690

691

Table 2 Description of treatment group, control groups, and acupoints used for the 11 studies included in this systematic review examining the efficacy of acupuncture for musculoskeletal conditions in dogs

Study	Treatment group intervention(s)	Acupoints used	Control group intervention(s)	Treatment duration (No of sessions)
Balaji et al., 1998	Electroacupuncture, 3.0-4.0 mA and 10-30 Hz	LI 9, TH 9, HT 7, GB 34, St 36, and Sp 6.	No treatment given	5 days (5); evaluation every 48 hours
Bolliger et al., 2002	Gold bead insertion acupuncture, 24k gold beads 1mm in diameter	GB 29, GB 30, BL 54. Additional trigger points were used determined in by measuring galvanic skin response	Superficial needle punctures at non-acupoints, no beads inserted	3 months (N/A gold bead insertion); evaluation at 1 and 3 months
Hayashi et al., 2007	Traditional dry needle acupuncture. Electroacupuncture, 3-100 Hz. Oral administration of prednisone (1 mg/kg [0.45 mg/lb], every 24 hours, for 3 days; followed by 0.5 mg/kg [0.23 mg/lb], every 24 hours, for 5 days; and 0.5 mg/kg, every other day, for 5 days).- If necessary for pain control, tramadol13 (2 mg/kg, q 8 hours, for 7 days)	SI 3, BL 62, BL 20, BL 23, ST 36, KI 3, BL 60, GV 1, lumbar Bai Hui, LI 4, BL25, and GB 30	Oral administration of prednisone (1 mg/kg [0.45 mg/lb], every 24 hours, for 3 days; followed by 0.5 mg/kg [0.23 mg/lb], every 24 hours, for 5 days; and 0.5 mg/kg, every other day, for 5 days).- If necessary, tramadol13 (2 mg/kg, q 8 hours, for 7 days) was prescribed for pain control	At least 2 weeks (3); evaluation at 0, 7, and 14 days
Hjelm-Bjorkman et al., 2001	Gold bead insertion acupuncture, 24k beads, 1mm in diameter and 2mm long	GB 29, GB 30, and BL 54. 1-3 additional acupoints were used at areas with high conductivity	3 small needle holes made at non-acupoints, no beads inserted	24 weeks (N/A gold bead insertion); evaluation at 0, 1, 2, 3, 4, 12 and 24 weeks
Jaeger et al., 2006	Gold bead insertion acupuncture, 24 K gold beads, 1mm in diameter and 2mm in length	5 defined acupoints. Used an ohm meter to find points with higher local conductivity.	Placebo, no additional description	6 months (N/A gold bead insertion); evaluation at 14 days, 3 months, and 6 months

Table 2 (continued) Description of treatment group, control groups, and the acupoints used for the 11 studies included in this systematic review examining the efficacy of acupuncture for musculoskeletal conditions in dogs

Study	Treatment group intervention(s)	Acupoints used	Control group intervention(s)	Treatment duration (No of sessions)
Jaeger et al., 2007	Gold bead insertion acupuncture, 24K gold beads, 1mm in diameter and 2mm in length.	Ohm meter used to find points with higher local conductivity. 5 points used	Skin penetrated at 4 non-acupoints, no beads inserted	24 months (N/A gold bead insertion); evaluation at 24 months
Joaquim et al., 2010	Electroacupuncture, 2-15 Hz. In combination with Decompressive surgery	BL 18, BL 23, BL 40, KI 3, GB 30, GB 34, and ST 36	Decompressive surgery	1-6 months (4-24); evaluation at 6 months
Kapatkin et al., 2006	Electroacupuncture, 2Hz and gradually increased until muscle contraction was observed. Needles 0.22mm in diameter and 24 mm in length	TH 13, TH 10, LU 5, PC 3, PC 6, HT 7, baihui, GB 33, GB 34, BL 10, and GV 14	Placement of a single unstimulated dermal needle at the top of the dog's head at a non-acupoint	3 weeks (3); evaluation at 1, 2, and 3 weeks
Sharifi et al., 2009	Traditional dry needle acupuncture	GB 30, BL 40, and ST 36	No treatment given	15 days (15); evaluation at 15 days
Um et al., 2005	Traditional dry needle acupuncture. 32 gauge 30mm long	BL 40, GB 33, GB 34, LV 8	No treatment given	4 weeks (4); evaluation at 1, 2, 3, and 4 weeks
Vecino et al., 2005	Electroacupuncture. Needles 0.25mm in diameter and 15 mm long	BL 11, BL 23, K 6, GB 30, and	No description provided	15 days (4); evaluation at 5, 10, and 15 days
mA, milliamps; Hz, hertz; LI, large intestine; TH, triple heater; LU, lung; PC, pericardium; HT, heart; GB, gallbladder; BL, bladder; GV, governing vessel; LV, liver; ST, stomach; K, kidney; SI, small intestine; SP, spleen;				

Table 5 All outcomes that were eligible for extraction related to hip dysplasia included in this systematic review that examined the efficacy of acupuncture for musculoskeletal conditions in dogs

Trial	Outcome description	Effect measure (95% CI)	# Treat	# Control
Bolliger et al., 2002	RR of improvement of orthopaedic evaluation conducted by veterinarian at one month	RR: 4 (1.2, 13.9)	9	9
Bolliger et al., 2002	RR of improvement of orthopaedic evaluation conducted by veterinarian at 3 months	RR: 0.5 (0.1, 4.6)	9	9
Bolliger et al., 2002	RR of improvement in owner questionnaire at one month Category 1: Stiffness upon rising, lameness and/or "bunny hopping"	RR: 1 (0.3, 3.7)	9	9
Bolliger et al., 2002	RR of improvement in owner questionnaire at one month Category 2: Difficulties in climbing stairs, jumping onto furniture or into the car	RR: 1.5 (0.324, 6.9)	9	9
Bolliger et al., 2002	RR of improvement in owner questionnaire at one month Category 3: Reduction of activity required	RR: 0.2 (0.01, 2.7)	5	4
Bolliger et al., 2002	RR of improvement in owner questionnaire at three months Category 1: Stiffness upon rising, lameness and/or "bunny hopping"	RR: 1 (0.3, 3.7)	9	9
Bolliger et al., 2002	RR of improvement in owner questionnaire at three months Category 2: Difficulties in climbing stairs, jumping onto furniture or into the car	RR: 1 (0.3, 3.7)	9	9
Bolliger et al., 2002	RR of improvement in owner questionnaire at three months Category 3: Reduction of activity required	RR: 0.4 (0.1, 3.0)	5	4
Bolliger et al., 2002	Mean difference in the change of mean peak vertical force (10^{-2} N/kg) from 0 to 3 months between acupuncture and control	MD: -2.06 (-9.1, 5.0)	9	9
Bolliger et al., 2002	Mean difference in the change of mean vertical impulse (10^{-2} N/kg/s) from 0 to 3 months between acupuncture and control	MD: -0.6 (-1.8, 0.6)	9	9
Bolliger et al., 2002	Mean difference in the change of mean peak braking impulse from 0 to three months between acupuncture and control	MD: -0.04 (-1.9, 1.8)	9	9

Table 5 All outcomes that were eligible for extraction related to hip dysplasia included in this systematic review that examined the efficacy of acupuncture for musculoskeletal conditions in dogs

Trial	Outcome description	Effect measure (95% CI)	# Treat	# Control
Bolliger et al., 2002	Mean difference in the change of mean braking impulse from 0 to three months between acupuncture and control	MD: 0 (-0.1, 0.1)	9	9
Bolliger et al., 2002	Mean difference in the reduction of mean peak propulsion from 0 to three months between acupuncture and control	MD: 0.4 (-1.6, 2.4)	9	9
Bolliger et al., 2002	Mean difference in the reduction of mean propulsion impulse from 0 to three months between acupuncture and control	MD: 0.04 (-0.2, 0.3)	9	9
Jaeger et al., 2006	Mean difference in reduction of mean pain score assessed by owner from time 0- 3 months between treatment and control*	MD: -1.3 (-2.1, -0.5)	36	42
Jaeger et al., 2006	Mean difference in reduction of owner assessed dysfunction score from time 0-6 months between treatment and control	MD: -1.3 (-2.1, -0.5)	36	42
Jaeger et al., 2006	Mean difference in reduction of total hip pain score assessed by clinical investigator from time 0-6 months between treatment and control	MD: -0.5 (-1.1, 0.07)	36	42
Jaeger et al., 2006	RR of improvement in lameness after 6 months assessed by clinical investigator	RR: 1.3 (0.8, 2.2)	20	22
Jaeger et al., 2006	RR of improvement in the signs of hip dysplasia according to owner's general impression of their dog's behavior in its daily life after 14 days	RR: 0.7 (0.4, 1.2)	36	42
Jaeger et al., 2006	RR of improvement in the signs of hip dysplasia according to owner's general impression of their dog's behavior in its daily life after 3 months	RR: 1.2 (0.8, 1.6)	36	42

Table 5 All outcomes that were eligible for extraction related to hip dysplasia included in this systematic review that examined the efficacy of acupuncture for musculoskeletal conditions in dogs

Trial	Outcome description	Effect measure (95% CI)	# Treat	# Control
Jaeger et al., 2006	RR of improvement in the signs of hip dysplasia according to owner's general impression of their dog's behavior in its daily life after 6 months	RR: 1.6 (1.2, 2.2)	36	42
Jaeger et al., 2006	RR of improvement in owner's assessment in change of mood at 14 days	RR: 0.9 (0.8, 1.1)	36	42
Jaeger et al., 2006	RR of improvement in owner's assessment in change of mood at 3 months	RR: 1.0 (0.9, 1.1)	36	42
Jaeger et al., 2006	RR of improvement in owner's assessment in change of mood at 6 months	RR: 1.1 (1.0, 1.2)	36	42
Jaeger et al., 2006	RR of improvement in owner's assessment in change of quality of life at 14 days	RR: 1.0 (0.9, 1.05)	36	42
Jaeger et al., 2006	RR of improvement in owner's assessment in change of quality of life at 3 months	RR: 1.1 (1.0, 1.2)	36	42
Jaeger et al., 2006	RR of improvement in owner's assessment in change of quality of life at 6 months	RR: 1.0 (incidence the same in both groups)	36	42
Jaeger et al., 2006	RR of improvement in change in appetite after 14 days	RR: 0.6 (0.2, 2.2)	36	42
Jaeger et al., 2006	RR of improvement in change in appetite after 3 months	RR: 1.5 (0.4, 5.0)	36	42
Jaeger et al., 2006	RR of improvement in change in appetite after 6 months	RR: 1.5 (0.4, 5.0)	36	42
Jaeger et al., 2007	Mean difference in the reduction of mean hip pain score evaluated by clinical investigator from 0 to 24 months between acupuncture and control	MD: -1.2 (-2.4, -0.02)	66	7

Table 5 All outcomes that were eligible for extraction related to hip dysplasia included in this systematic review that examined the efficacy of acupuncture for musculoskeletal conditions in dogs

Trial	Outcome description	Effect measure (95% CI)	# Treat	# Control
Jaeger et al., 2007	Mean difference in the reduction of mean lameness score evaluated by clinical investigator from 0 to 24 months between acupuncture and control	MD: -1.1 (-3.1, 0.9)	66	7
Jaeger et al., 2007	Mean difference in the reduction of mean hip pain signs score evaluated by owners from 0 to 24 months between acupuncture and control	MD: -1.3 (-2.7, 0.09)	66	7
Jaeger et al., 2007	Mean difference in the reduction of mean dysfunction score evaluated by owner from 0 to 24 months between acupuncture and control	MD: -2.2 (-3.695, -0.705)	66	7
Jaeger et al., 2007	RR of improvement of owner's impression of dog's behaviour	RR: 5.7 (0.9, 35.3)	66	7
Jaeger et al., 2007	RR of improvement in lameness assessed by clinical investigator	RR: 5.7 (0.4, 82.3)	34	5
Jaeger et al., 2007	RR of improvement of owner's impression in the change of dog's quality of life	RR: 1.0 (0.8, 1.2)	66	7

Negative MD in pain score, dysfunction score, and hip pain score denotes numerical improvement in acupuncture group compared to control group. *MD was calculated for the time points 0 and 3 month instead of 0 and 6 months for pain score assessed by owner because control group change in pain score was not reported for the 6 month time period. For RR of improvement of orthopaedic evaluation, RR of improvement in owner questionnaire at one month Category 1, RR of improvement in owner questionnaire at one month Category 2, RR of improvement in owner questionnaire at one month Category 3, RR of improvement in lameness, RR of improvement in the signs of hip dysplasia according to owner's general impression of their dog's behavior in its daily life, RR of improvement in owner's assessment in change of mood, RR of improvement in owner's assessment in change of quality of life, RR of improvement in change in appetite, RR of improvement of owner's impression of dog's behaviour, and RR of improvement in lameness a RR > 1 favors acupuncture treatment.

Table 7: Summary of findings table for acupuncture compared to control for hip dysplasia in dogs**Patient or population:** hip dysplasia in dogs**Setting:** dogs of any breed**Intervention:** acupuncture**Comparison:** control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control	Risk with acupuncture				
Relative risk of improvement in locomotion based on clinical investigator assessment assessed with: Likert scale follow up: mean 11 months	361 per 1,000	469 per 1,000 (289 to 794)	RR 1.3 (0.8 to 2.2)	99 (3 RCTs)	⊕○ ○ VERY LOW ^{a,b,c}	
Relative risk of improvement in locomotion based on owner assessment assessed with: Likert Scale follow up: mean 15 months	448 per 1,000	717 per 1,000 (538 to 986)	RR 1.6 (1.2 to 2.2)	169 (3 RCTs)	⊕⊕ ○ LOW ^{c,d}	

Table 7: Summary of findings table for acupuncture compared to control for hip dysplasia in dogs**Patient or population:** hip dysplasia in dogs**Setting:** dogs of any breed**Intervention:** acupuncture**Comparison:** control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control	Risk with acupuncture				
Relative risk of improvement in quality of life based on owner assessment assessed with: Likert Scale follow up: mean 15 months	No meta-analysis possible; Jaeger 2006 QOF Likert scale measure results indicate that both treatment and control groups had complete improvement.			151 (2 RCTs)	⊕○ ○ VERY LOW ^{c,d,e}	
Mean difference in the reduction of mean pain scores assessed by clinical investigator assessed with: Pain score Scale from: 1 to 4 follow up: mean 15 months		The mean mean difference in the reduction of mean pain scores assessed by clinical investigator in the intervention group was 0.5 points lower (0.9 lower to 0.1 lower)	-	151 (2 RCTs)	⊕⊕ ○ LOW ^{a,c}	

Table 7: Summary of findings table for acupuncture compared to control for hip dysplasia in dogs**Patient or population:** hip dysplasia in dogs**Setting:** dogs of any breed**Intervention:** acupuncture**Comparison:** control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control	Risk with acupuncture				
Mean difference in the reduction of mean pain scores assessed by owners assessed with: Likert scale Scale from: 0 to 10 follow up: mean 15 months		The mean mean difference in the reduction of mean pain scores assessed by owners in the intervention group was 0.7 points lower (1.1 lower to 0.3 lower)	-	151 (2 RCTs)	⊕⊕ ○○ LOW ^{c,d}	
Mean difference in the reduction of mean lameness score assessed by clinical investigator assessed with: Lameness score follow up: mean 24 months	Insufficient data for meta-analysis, only one study reported this outcome			73 (1 RCT)	⊕○ ○○ VERY LOW ^{a-c,e}	

Table 7: Summary of findings table for acupuncture compared to control for hip dysplasia in dogs**Patient or population:** hip dysplasia in dogs**Setting:** dogs of any breed**Intervention:** acupuncture**Comparison:** control

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control	Risk with acupuncture				
Mean difference in the reduction of mean lameness score assessed by owners assessed with: Lameness score Scale from: 0 to 10 follow up: mean 15 months		The mean mean difference in the reduction of mean lameness score assessed by owners in the intervention group was 0.8 points lower (1.2 lower to 0.4 lower)	-	151 (2 RCTs)	⊕⊕ ○○ LOW ^{c,d}	

*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI).

CI: Confidence interval; RR: Risk ratio; MD: Mean difference

GRADE Working Group grades of evidence

High quality: We are very confident that the true effect lies close to that of the estimate of the effect

Moderate quality: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low quality: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

Very low quality: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

- a. High risk of bias for blinding due to lack of blinding of personnel
- b. RR confidence interval crosses both no effect and appreciable benefit
- c. Inadequate sample size to detect a significant difference
- d. High risk of bias for blinding due to lack of blinding of owners
- e. Meta-analysis not possible, one study reported this outcome

References

Balaji B, Lakshmipathi GV, Sreenu M, Rao NV. 1998. Acupuncture therapy for experimental muscle and joint injuries in dogs. *The Indian Journal of Animal Sciences* 68.

Barnes PM, Powell-Griner E, McFann K, Nahin RL. 2004. Complementary and alternative medicine use among adults: United States, 2002. *Advance Data*:1–19.

Bolliger C, DeCamp CE, Stajich M, G.L. Flo, Martinez SA, Bennett RL, Bebachuk T. 2002. Gait analysis of dogs with hip dysplasia treated with gold bead implantation acupuncture. *Veterinary and Comparative Orthopaedics and Traumatology* 15:116–122.

Chan W-W, Chen K-Y, Liu H, Wu L-S, Lin J-H. 2001. Acupuncture for general veterinary practice. *Journal of Veterinary Medical Science* 63:1057–1062.

Epstein ME, Rodan I, Griffenhagen G, Kadrlik J, Petty MC, Robertson SA, et al. 2015 AAHA/AAFP Pain management guidelines for dogs and cats. *J Fel Med Sur.* 2015;17:251–72.

European Food Safety Authority (EFSA). 2010. Application of systematic review methodology to food and feed safety assessments to support decision making: systematic review methodology and food and feed safety risk assessment. *European Food Safety Authority Journal* 8:1637.

GRADEpro. GRADEpro GDT. Version June 1 2016. McMaster University, 2014.

Habacher G, Pittler MH, Ernst E. 2006. effectiveness of acupuncture in veterinary medicine: systematic review. *Journal of Veterinary Internal Medicine* 20:480–488.

Hayashi AM, Matera JM, Fonseca Pinto ACB de C. 2007. evaluation of electroacupuncture treatment for thoracolumbar intervertebral disk disease in dogs. *Journal of the American Medical Association* 231:913–918.

Hielm-Bjorkman A, Raekallio M, Kuusela E, Saarto E, Markkola A, Tulamo R-M. 2001. Double-blind evaluation of implants of gold wire at acupuncture points in the dog as a treatment for osteoarthritis induced by hip dysplasia. *Veterinary Record* 149:452–456.

Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JAC, et al. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal* 343:d5928.

Higgins JPT, Green S, editors. 2011. *Cochrane handbook for systematic reviews of interventions*. Version 5.1.0. The Cochrane Collaboration.

Jaeger GT, Larsen S, Sjøli N, Moe L. 2006. Double-blind, placebo-controlled trial of the pain-relieving effects of the implantation of gold beads into dogs with hip dysplasia. *Veterinary Record* 158:722–726.

Jaeger GT, Larsen S, Sjøli N, Moe L. 2007. Two years follow-up study of the pain-relieving effect of gold bead implantation in dogs with hip-joint arthritis. *Acta Veterinaria Scandinavica* 49:9.

Joaquim JGF, Luna SPL, Brondani JT, Torelli SR, Rahal SC, de Paula Freitas F. 2010. Comparison of decompressive surgery, electroacupuncture, and decompressive surgery followed by electroacupuncture for the treatment of dogs with intervertebral disk disease with long-standing severe neurologic deficits. *Journal of the American Veterinary Medical Association* 236:1225–1229.

Kapatkin AS, Tomasic M, Beech J, Meadows C, Boston RC, Mayhew PD, Powers MY, Smith GK. 2006. Effects of electrostimulated acupuncture on ground reaction forces and pain scores in dogs with chronic elbow joint arthritis. *Journal of the American Veterinary Medical Association* 228:1350–1354.

Kaptchuk TJ. 2002. Acupuncture: theory, efficacy, and practice. *Annals of Internal Medicine* 136:374–383.

Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998;317:1185–90.

Mavridis D, Salanti G. 2014. How to assess publication bias: funnel plot, trim-and-fill method and selection models. *Evidence-Based Mental Health* 17:30.

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 4:1.

O'Connor AM, Sargeant JM, Gardner IA, Dickson JS, Torrence ME, Dewey CE, Dohoo IR, Evans RB, Gray JT, Greiner M, et al. 2010. The REFLECT statement: methods and processes of creating reporting guidelines for randomized controlled trials for livestock and food safety. *Journal of Food Protection* 73:132–139.

Pandis N. 2011. The evidence pyramid and introduction to randomized controlled trials. *American Journal of Orthodontics and Dentofacial Orthopedics* 140:446–447.

Quinn MM, Keuler NS, Lu Y, Faria MLE, Muir P, Markel MD. Evaluation of agreement between numerical rating scales, visual analogue scoring scales, and force plate gait analysis in dogs. *Vet Surg.* 2007;36:360–7.

Rose W, “A Scoping and systematic review of the efficacy of veterinary acupuncture” master’s thesis., University of Guelph, 2016, Guelph Atrium (201609).

Roudebush P, Allen TA, Dodd CE, Novotny BJ. 2004. Application of evidence-based medicine to veterinary clinical nutrition. *Journal of the American Veterinary Medical Association* 224:1765–1771.

Sargeant JM, Kelton DF, O’Connor AM. 2014. Study designs and systematic reviews of interventions: building evidence across study designs. *Zoonoses Public Health* 61:10–17.

Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Ann Int Med* 2010;152. Epub 24 March.

Schulz KF, Chalmers I, Hayes RJ, Altman DG. 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* 273:408–412.

Sharifi D, Sasani F, Bakhtiari J, Rezaei T. 2009. The effect of acupuncture therapy on the repair of the calcaneal tendon (tendo calcaneus communis) in dogs. *Turkish Journal of Veterinary and Animal Sciences* 33:181–184.

Thalheimer W, Cook S. 2009. How to calculate effect sizes from published research: a simplified methodology.

Tindle HA, Davis RB, Phillips RS, Eisenberg DM. 2005. Trends in use of complementary and alternative medicine by US adults: 1997-2002. *Alternative Therapies In Health And Medicine* 11:42–49.

Toews L. 2011. The information infrastructure that supports evidence-based veterinary medicine: a comparison with human medicine. *Journal of Veterinary Medical Education* 38:123–134.

Um S-W, Kim M-S, Lim J-H, Kim S-Y, Seo K-M, Nam T-C. 2005. Thermographic evaluation for the efficacy of acupuncture on induced chronic arthritis in the dog. *The Journal of Veterinary Medical Science* 67:1283–1284.

Vecino JA, Galvez A, Bonafonte JI, Labodia A, Sopena J, Sarrat R, Whyte J, Whyte A, Martinez-Sanudo MJ, Peiro JM. 2005. Osteoporosis and acupuncture. ISPRM:402.

Wojciechowska JI, Hewson CJ, Stryhn H, Guy NC, Patronek GJ, Timmons V. Evaluation of a questionnaire regarding nonphysical aspects of quality of life in sick and healthy dogs. *Am J Vet Res.* 2005;66:1461–7.

Yeates J, Main D. Assessment of companion animal quality of life in veterinary practice and research. *J Sm Anim Pract.* 2009;50:274–81.

Xie H, Preast V. 2007. Xie's Veterinary acupuncture. First. Iowa, USA: Blackwell Publishing.