Target Trial Emulation: Does surgical versus non-surgical management of cranial cruciate ligament rupture in dogs cause different outcomes?

Camilla Pegram [a,*], Karla Diaz-Ordaz [b], Dave C. Brodbelt [a], Yu-Mei Chang [c], Anna Frykfors von Hekkel [d], Chieh-Hsi Wu [e], David B. Church [d], Dan G. O'Neill [a]

[a] Pathobiology and Population Sciences, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Herts AL9 7TA, UK
[b] University College London, Department of Statistical Science, Gower Street, London WC1E 6BT, UK
[c] Research Support Office, The Royal Veterinary College, Hatfield, Herts AL9 7TA, UK
[d] Clinical Science and Services, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Herts AL9 7TA, UK
[e] Statistical Sciences Research Institute, University of Southampton, University Road, Highfield, Southampton SO17 1BJ, UK

ARTICLE INFO

ABSTRACT

Target trial emulation applies design principles from randomised controlled trials to the analysis of observational data for causal inference and is increasingly used within human epidemiology. Using anonymised veterinary clinical data from the VetCompass Programme, this study applied the target trial emulation framework to determine whether surgical (compared to non-surgical) management for cranial cruciate ligament (CCL) rupture in dogs causes improved short- and long-term lameness and analgesia outcomes.

The emulated target trial included dogs diagnosed with CCL rupture between January 1, 2019 and December 31, 2019 within the VetCompass database. Inclusion in the emulated trial required dogs aged $\geq$ 1.5 and < 12 years, first diagnosed with unilateral CCL rupture during 2019 and with no prior history of contralateral ligament rupture or stifle surgery. Dogs were retrospectively observed to have surgical or non-surgical management. Informed from a directed acyclic graph derived from expert opinion, data on the following variables were collected: age, breed, bodyweight, neuter status, insurance status, non-orthopaedic comorbidities, orthopaedic comorbidities and veterinary group. Inverse probability of treatment weighting (IPTW) was used to adjust for confounding, with weights calculated based on a binary logistic regression exposure model. Censored dogs were accounted for in the IPTW analysis using inverse probability of censoring weighting (IPCW). The IPCWs were combined with IPTWs and used to weight each dog's contribution to binary logistic regression outcome models. Standardized mean differences (SMD) examined the balance of covariate distribution between treatment groups.

The emulated trial included 615 surgical CCL rupture cases and 200 non-surgical cases. The risk difference for short-term lameness in surgically managed cases (compared with non-surgically managed cases) was −25.7% (95% confidence interval (CI) −36.7% to −15.9%) and the risk difference for long-term lameness −31.7% (95% CI −37.9% to −18.1%).

The study demonstrated the application of the target trial framework to veterinary observational data. The findings show that surgical management causes a reduction in short- and long-term lameness compared with non-surgical management in dogs.

1. Introduction

Veterinarians are increasingly encouraged to apply 'evidence based principles' in their clinical decision-making (Holmes and Cockcroft, 2004) but the paucity of relevant and reliable published evidence on veterinary interventions has limited the potential clinical welfare gains

for dogs (Dean, 2017). Randomised controlled trials (RCTs), along with their synthesis in the forms of systematic reviews and meta-analyses, are considered as the gold standard method for assessing the effectiveness of treatment interventions and are a valuable source of information on which to base clinical decisions (Balshem et al., 2011; Wareham et al., 2017). The primary advantage of RCTs is that randomisation of the exposure aims to ensure that the treatment effect is unconfounded because the groups (treatment and controls) should be balanced in both observed and unobserved confounders (Pfeiffer, 2010). However, randomisation only addresses confounding at baseline, whilst post-randomisation confounding can still occur. Additionally, RCTs often require large groups, can be costly and ethically challenging, the required duration can be long if time from intervention to outcome is prolonged and the eligibility criteria may result in trial participants not representing the wider population of interest (Pfeiffer, 2010). Therefore, observational data are increasingly recognised to represent a valuable alternative resource for information to estimate real-world causal effects, especially in the absence of available randomised experiments (or in complement), and might even reflect the "usual conditions" under which a treatment would be taken more accurately (Maringe et al., 2020).

Causal inference refers to an intellectual discipline that considers the assumptions, study designs, and estimation strategies that allow researchers to draw causal conclusions based on data (Hill and Stuart, 2015). In more simple terms, causal inference describes the process of drawing a conclusion that a specific treatment or exposure (i.e., intervention) was the "cause" of the effect (or outcome) that was observed (Frey, 2018). When analysing observational data to answer causal questions, an observational study can be conceptualised as a conditionally randomised experiment, given the measured covariates (Hernán and Robins, 2020). Causal inference from *large* observational databases can be viewed as an attempt to emulate the randomised experiment - the target experiment or target trial - that would answer the question of interest (Hernán and Robins, 2016).

Hernán and Robins (2016) have described a framework for research into comparative effectiveness using large observational databases that aims to make the target trial explicit i.e. the "target trial emulation framework" (Hernán and Robins, 2016). In broad terms, the framework explicitly defines the "target trial" as the trial you would like to conduct if it were feasible and then describes how to emulate this target trial using observational data. Observational data are subject to various confounding, selection and information biases that can result in underestimation or overestimation of the effect of interest (Dohoo et al., 2009; Hammerton and Munafò, 2021). The target trial emulation, in conjunction with other statistical techniques, such as inverse probability of treatment weighting (IPTW) later defined in the methods, aim to address these possible biases.

The VetCompass database holds large volumes of anonymised veterinary clinical data with the potential for use to answer causal questions of interest under the no unobserved confounding assumption i.e., no unmeasured variables affecting both exposure and outcome. Cranial cruciate ligament (CCL) rupture is one of the most frequent specific cause of lameness in dogs (Johnson et al., 1994) and is clinically managed either surgically or non-surgically (Kirkness, 2020). Surgical management is considered the gold standard treatment of choice for CCL rupture (Fauron and Perry, 2017), with multiple surgical techniques available that aim to stabilise the affected stifle and return the dog to full function where possible (Kirkness, 2020). The range of surgical techniques can be broadly categorised into two groups based on their approach: osteotomy-based techniques and suture-based techniques (ACVS., 2021). Of the osteotomy-based techniques, tibial plateau levelling osteotomy (TPLO) and tibial tuberosity advancement (TTA) are the most commonly performed, whilst extra-capsular suture stabilisation (also known as "lateral suture") is the most common of the suture-based techniques (ACVS., 2021).

It is reported, based on limited evidence, that the clinical outcomes

for dogs managed surgically are superior to dogs managed non-surgically (Comerford et al., 2013; Wucherer et al., 2013), although reasonable functional outcomes for small breed dogs managed non-surgically have been reported (Pond and Campbell, 1972; Vasseur, 1984). However, outcomes in dogs across the range of sizes managed surgically versus non-surgically have not previously been directly compared.

Two previous retrospective studies evaluated non-surgical management alone for CCL rupture in dogs. A US-based study on non-surgically managed dogs (n = 85) attending a referral veterinary medical teaching hospital reported that 85.7% of dogs weighing ≤ 15 kg compared with 19.3% of dogs weighing > 15 kg were considered clinically normal or improved after an average follow-up of 3 years (Vasseur, 1984). An older UK-based study of non-surgically managed dogs (n = 107) reported that 90% of small breed dogs compared with 78% of large breed dogs had no detectable lameness reported by the owner; although the time to assessment was not clearly specified (Pond and Campbell, 1972). Inference from both studies is limited by relatively small sample size and neither study compared non-surgical to surgical management. Additionally, it is possible the dogs in these older studies are no longer fully representative of the current UK dog population. A more recent study evaluated short-term and long-term outcomes for overweight dogs with bodyweight > 20 kg treated surgically (n = 21) compared with non-surgically (n = 19) for CCL rupture. A successful outcome was defined as ground reaction force > 85% of the value for a clinically normal dog and owner questionnaire responses indicating an improvement of ≥ 10% in lameness and quality of life scores between pre- and post-intervention questionnaires. A successful outcome was reported in 68%, 93%, and 75% dogs managed surgically and 47%, 33%, and 64% of dogs managed non-surgically at 12, 24, and 52 weeks after enrolment in the study, respectively (Wucherer et al., 2013); thus suggesting improved outcomes in dogs managed surgically, albeit less marked at 52 weeks. However, the study included only 40 dogs and focused on overweight, larger breed dogs limiting inference to the overall canine population.

Using anonymised veterinary clinical data from the VetCompass Programme (VetCompass, 2019), the present study aimed to implement the target trial emulation framework to answer causal questions for veterinary first opinion observational data. More specifically, this study aimed to compare the effects of surgical management relative to non-surgical management on short- and long-term lameness outcome for CCL rupture in dogs. Duration of analgesia prescription was also assessed as a secondary outcome. Whilst an RCT might be possible to conduct for this specific research question, there would be ethical concerns in randomising dogs to either surgical or non-surgical management when surgery is still considered "gold standard" and a body size difference in outcome may exist (Comerford et al., 2013; Wucherer et al., 2013; Fauron and Perry, 2017). Likewise, other more general disadvantages of RCTs previously mentioned (cost, duration, representativeness etc) would apply in this setting (Pfeiffer, 2010).

Based on previous evidence (Pond and Campbell, 1972; Vasseur, 1984; Wucherer et al., 2013), it was hypothesized that surgical management causes a reduction in the presence of lameness at short- and long-term follow-up compared with non-surgical management. Since the study utilises observational data, target trial emulation and causal inference analyses were adopted to attenuate most biases. The degree to which explicit target trial emulation from human observational studies aligns with results from randomised controlled trials is not fully known, but the evidence is encouraging (Hernán et al., 2008; Cain et al., 2016; García-Albéniz et al., 2017; Labrecque and Swanson, 2017; Admon et al., 2019). It has been stated that the principles of target trial emulation should be adopted regardless, as they encourage known good practices (Labrecque and Swanson, 2017). To the authors' knowledge, this is one of the first studies to adopt the target trial emulation framework for veterinary observational data, whilst also the first to focus on estimating the treatment effect of surgical versus non-surgical

management of CCL rupture in dogs, and will provide veterinarians and owners with an evidence-base to guide their clinical decision making in relation to CCL rupture management in dogs.

2. Material and Methods

2.1. Data source and power calculation

VetCompass collates de-identified electronic patient record (EPR) data from primary-care veterinary practices in the UK for epidemiological research (VetCompass, 2019). The CCL study population included all available dogs under primary veterinary care at clinics participating in the VetCompass Programme during 2019. Dogs under veterinary care were defined as those with at least one EPR (free-text clinical note, treatment or bodyweight) recorded during 2019. Available data fields included a unique animal identifier along with species, breed, date of birth, sex, neuter status, insurance status and bodyweight, and also clinical information from free-form text clinical notes, summary diagnosis terms (The VeNom Coding Group., 2019) and treatment with relevant dates. VetCompass operates under an 'opt-out' client consent process for data-sharing, therefore data from dogs where owners have opted out are not included (VetCompass., 2023).

The study design applied target trial emulation (Hernán and Robins, 2016). Sample size calculations estimated that approximately 224 surgical cases and 54 non-surgical cases were required, at a minimum, assuming 30% of surgical cases and 50% non-surgical cases had long term lameness, 80% power, 95% confidence and a ratio of surgical to non-surgical cases of 4:1 (Vasseur, 1984; Christopher et al., 2013; Taylor-Brown et al., 2015), using ClinCalc Sample Size Calculator (www.clincalc.com). A 4:1 ratio was estimated, based on a previous primary-care study that reported 69 – 78% dogs aged < 12 years with CCL rupture were managed surgically (Taylor-Brown et al., 2015). Ethics approval was obtained from the RVC Social Science Ethical Review Board (reference number SR2018–1652).

2.2. Case definition, finding and covariates

Incident cases were included in the study, with these cases defined as dogs that were first diagnosed with CCL rupture in either stifle between January 1, 2019 and December 31, 2019. Candidate cases were identified by applying search terms relevant to the diagnosis and management of CCL rupture in the clinical notes (acl, ccl, cranial draw*, cruciate rupture, cruciate ligament, tta, tplo, lateral sut*, extracapsular sut*) (Taylor-Brown et al., 2015). The search findings were merged, and a subset of candidate cases, randomly presented through the online database using the RAND function in SQL Server (Microsoft Learn., 2022), had their clinical notes examined manually in detail to identify whether they met the case definition (Pegram et al., 2023). For dogs that met the case definition, demographic data were extracted automatically from the VetCompass database, with further data relating to clinical management extracted manually from the EPR (Table 1).

Based on existing evidence and expert opinion, a directed acyclic graph (DAG) was constructed using DAGitty version 3.0 (Textor et al., 2016), Fig. 1, that encapsulates beliefs about the causal relationships relevant to the question of interest. The DAG was used to identify which variables should be controlled for (Greenland et al., 1999), therefore data on the following variables were collected: age, breed, bodyweight, overweight status, neuter status, insurance status, non-orthopaedic comorbidities, orthopaedic comorbidities and veterinary group (Fig. 1). Long-term lameness was evaluated as a primary outcome, with short-term lameness and analgesia prescription secondary outcomes (with the same covariate set believed to apply, based on existing evidence and expert opinion, for analgesia prescription and short-term lameness as for long-term lameness).

Table 1

Definition and categorisation of demographic and clinical data extracted from the anonymised electronic patient records of dogs diagnosed with cranial cruciate ligament (CCL) rupture (n = 815) attending primary-care veterinary practices in the VetCompass™ Programme in the UK.

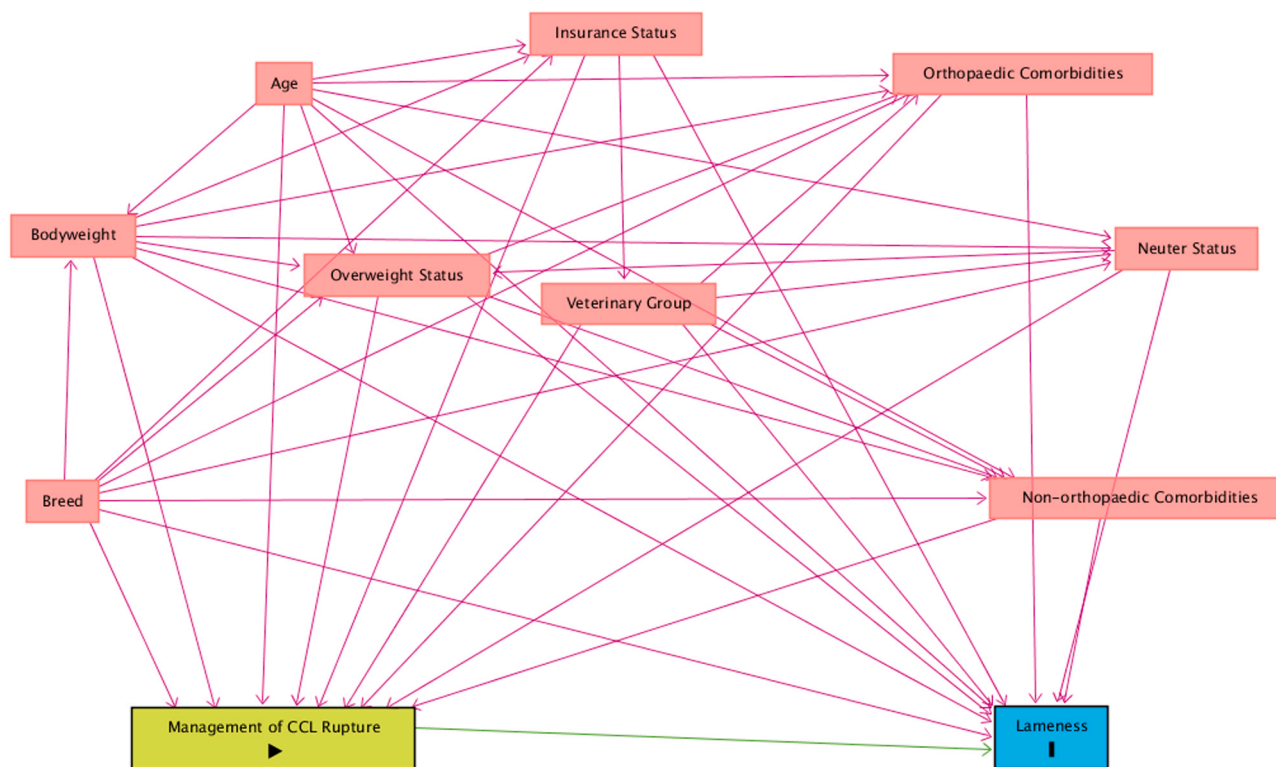| Demographic and clinical data extracted	Definition	Categorisation
Age	Age at diagnosis (years).	Included as continuous variable (based on the linearity assumption).
Breed	Breed information entered by the participating practices was cleaned and mapped to a VetCompass breed list derived and extended from the VeNom Coding breed list (The VeNom Coding Group., 2019).	Included individual breeds with at least 15 dogs managed surgically or 5 dogs managed non-surgically, to allow sufficient study power. Remaining dogs grouped as either "Purebred – Other" or "Crossbred".
Bodyweight	The median of all bodyweight (kg) values recorded for each dog after reaching 18 months old.	"< 15 kg", "≥ 15 kg" or "not recorded".
Overweight status	Overweight status required information recorded within the electronic patient record indicating that the dog was obese or overweight within the year prior to CCL rupture diagnosis (Rolph et al., 2014; Pegram et al., 2021).	"overweight/obese", "ideal bodyweight", "underweight" or "not recorded".
Neuter status	The final available electronic patient record neuter value.	"neutered" or "entire".
Insurance status	Status at the final available electronic patient record.	"insured" or "non-insured".
Veterinary group	The practice groups involved in the study.	Categorised as 1–5.
Orthopaedic comorbidities at diagnosis	Orthopaedic comorbidity as diagnosed by a veterinary surgeon or within one month prior to CCL rupture diagnosis.	Included as a distinct disorder if diagnosed in at least 10 dogs managed either surgically or non-surgically, otherwise recorded as "other comorbid disorders".
Non-orthopaedic comorbidities at diagnosis	Presence of non-orthopaedic comorbidities diagnosed at or within one month prior to CCL rupture diagnosis.	"Yes" or "No".
Lameness	Binary outcome (presence or absence). Short-term lameness was defined as evidence of lameness in the limb diagnosed with CCL rupture reported in the electronic patient record at 6 weeks – 4.5 months after diagnosis of CCL rupture. Long-term lameness was defined as evidence of lameness in the limb diagnosed with CCL rupture reported in the electronic patient record at 7.5 months – 16.5 months after diagnosis.	"lameness", "no lameness" or "lameness not recorded".
Analgesia	Final prescription date for analgesia dispensed specifically for CCL rupture was extracted and prescription at 3-, 6- and 12-months after diagnosis recorded as a binary outcome.	"prescription", "no prescription" or "prescription information not available".

Fig. 1. Directed acyclic graph (DAG) based on existing evidence and expert opinion to estimate our belief about the total effect of management of cranial cruciate ligament rupture (surgical versus non-surgical) on short-term and long-term lameness. The same causal structure was believed to apply with analgesia prescription as the outcome of interest.

2.3. Target Trial specification and emulation

A target trial of interest was specified and emulated using EPR data (Danaei et al., 2013; Hernán and Robins, 2016, 2020). The protocols of the target trial, and the trial emulation, are summarised in Table 2.

2.4. Descriptive analysis

Demographic data were described. Date of follow-up for descriptive analysis was date of final available EPR. Continuous variables were assessed graphically for their distribution and summarised using median, interquartile range (IQR) and range if non-normally distributed. Chi-square test was used to compare categorical variables and the Student's t-test or Mann–Whitney U test for univariable comparison of continuous variables between management groups as appropriate (Kirkwood., 2003).

2.5. Statistical analysis of the emulated trial

Lameness and prescription of analgesia at follow-up were compared between dogs managed surgically and dogs managed non-surgically for CCL rupture. For the analysis to be causal, the assumptions of consistency, no interference, no unobserved confounding and positivity should hold (Hernán and Robins, 2020). The consistency assumption implies that a dog's potential outcome under their observed exposure history is the outcome that will actually be observed for that dog (Rehkopf et al., 2016) i.e. the values of treatment under comparison correspond to well-defined interventions that correspond to the versions of treatment in the data (Hernán and Robins, 2020). No interference refers to the assumption that the potential outcomes of one dog are unaffected by the treatment assignment of other dogs (Hudgens and Halloran, 2008). Positivity refers to the assumption that the probability of receiving each treatment conditional on measured covariates is greater than zero

(Hernán and Robins, 2020). Positivity violations occur when certain subgroups (defined by a combination of covariates) in a sample rarely or never receive some treatments of interest. Positivity violations can be diagnosed through basic descriptive analyses, examination of the distribution of IP-weights and standardised mean differences (discussed in more detail below) (Petersen et al., 2012).

To emulate randomisation at the baseline point of diagnosis with CCL rupture, the following variables were considered sufficient to control for confounding: age, breed, bodyweight, insurance status, neuter status, orthopaedic comorbidities at diagnosis, non-orthopaedic comorbidities at diagnosis, overweight status and veterinary group attended (as defined and categorised in Table 1 and based on the DAG in Fig. 1). Therefore, confounder selection was DAG-driven rather than data-driven. Although many methods for covariate selection exist, DAGs explicitly consider the role of each variable in relation to the exposure and outcome and demonstrate knowledge, theories and assumptions about the causal relationship between variables in a simple and transparent way (Tennant et al., 2021).

Inverse probability of treatment weighting (IPTW) was used to adjust for confounding. IPTW is a propensity-score based method, with the propensity-score defined as the probability of treatment assignment conditional on observed baseline characteristics (Austin, 2011). For IPTW, a pseudo-population is created by weighting each individual in the population by the inverse of the conditional probability of receiving the treatment level they indeed received (Hernán and Robins, 2020). For example, if an individual dog managed surgically received an IPTW of 4 (reflecting that given their measured covariates, they were not very likely to receive surgery), their outcome would be "up-weighted" in the analysis. So the pseudo-population would include three copies (four appearances in total) of this individual dog in the surgical management group. The goal is to balance covariates between the two treatment groups (Barter, 2017).

To derive the weights using IPTW, a logistic regression model was

Table 2

Specification and emulation of target trial to estimate the effect of surgical or non-surgical management for cranial cruciate ligament (CCL) rupture on lameness and analgesia prescription as outcomes.

Protocol Component	Target trial description	Emulated trial using veterinary electronic patient records
Research question	To estimate the effect of surgical versus non-surgical management of cranial cruciate ligament (CCL) rupture on lameness, at approximately 12 months follow-up (long-term lameness), with a secondary outcome at 3 months (short-term lameness). Other secondary outcomes are the effect on analgesia prescription at 3, 6 and 12 months follow-up.	Same as target trial.
Eligibility criteria	Dogs aged ≥ 1.5 years and < 12 years first diagnosed with **unilateral** cruciate ligament rupture during 2019 and with no prior history of contralateral cruciate ligament rupture or stifle surgery.	Same as target trial. Candidate cases identified through relevant search terms and a random subset of candidate cases have clinical notes examined manually in detail to identify dogs that meet the case definition.
Treatment strategies	1. Surgery (using any method) within 4 weeks of diagnosis 2. No surgery in the 4 weeks after diagnosis	Receive surgery within a grace period (4 weeks between CCL rupture diagnosis and surgery). Non-surgical management defined as dogs not receiving surgery in the 4 weeks after diagnosis (with or without any additional strategies such as rest, analgesia and physiotherapy).
Assignment procedures	Eligible dogs randomly assigned to either strategy at diagnosis. Owners and veterinarians involved in the dog's care aware of the strategy to which they have been assigned.	Dogs non-randomly assigned to either strategy at diagnosis and treatment strategies retrospectively observed. All observed confounding factors adjusted for to aim for conditional exchangeability of the groups defined by initiation of the treatment strategies.
Follow-up period	Follow up starts at enrolment (which happens when dogs are first diagnosed), equivalent to treatment assignment and ends at 12 months.	Follow up starts at diagnosis, which does not correspond to treatment assignment and ends at approximately 12 months.
Censoring	Loss to follow up, death, or administrative censoring i.e., end of study observation period (12 months). Dogs that switch from non-surgical to surgical management (after the 4 week grace period and within the follow-up of 12 months) censored at date of surgery. Dogs with a subsequent contralateral CCL rupture censored at date of subsequent rupture.	Same as target trial.
Outcome	Primary outcome - presence of long-term lameness post-baseline. Secondary outcomes – presence of short-term lameness post-baseline and analgesia prescription for CCL rupture at 6- and 12-months post-baseline.	Same as target trial.

Table 2 (*continued*)

Protocol Component	Target trial description	Emulated trial using veterinary electronic patient records
Causal contrasts of interest	Per protocol effect	*Observational analogue of a per protocol effect*: The intention to treat was not known from the data; In each arm of the emulated trial, patients who deviate from the protocol (after the grace period) are censored at their time of deviation.
Estimands	Marginal risk differences in presence of lameness and analgesia prescription at follow-up between arms.	Same as target trial.
Analysis plan	Per protocol analysis, including only dogs that complied with the trial protocol, with adjustment for post-randomisation confounding (to deal with protocol deviations).	Same as target trial with adjustment for baseline covariates using inverse probability of treatment weighting. Censoring due to loss to follow up or administrative censoring addressed using inverse probability of censoring weighting.
Adjustment variables	Other than variables needed to adjust for protocol deviations, there is no other adjustment for baseline variables as interest is on marginal risk difference (randomisation breaks the links between treatment and known or unknown confounders).	Variables used to adjust for confounding: age at first diagnosis of CCL rupture, breed, bodyweight, insurance status, orthopaedic comorbidities, non-orthopaedic comorbidities, overweight status and vet group.

first fitted, with treatment (surgical versus non-surgical) as the outcome regressed on the main term confounding variables described above (Table 1). Biologically plausible interaction terms were added to the model and their effect on the Akaike information criterion (AIC) and standardised mean differences (SMDs) were assessed for inclusion, with interaction terms included if they improved covariate balance (Austin and Stuart, 2015). SMD examines the balance of covariate distribution between treatment groups (Zhang et al., 2019). For each covariate, SMD between pre- and post-IPTW were calculated, with SMD < 0.1 indicating good covariate balance between the two treatment arms (Austin, 2009; Yang and Dalton, 2014). The linearity assumption was assessed by visually inspecting the scatter plot between the continuous predictor (age) and the logit values (Osborne and Waters, 2002). The model generated predicted probabilities of receiving either treatment for each dog, which were then used to calculate stabilised inverse probability (IP) weights (Xu et al., 2010). When a treated patient has an extremely low or high propensity score, an extreme weight is created. Extreme weights can increase the variability of the estimated treatment effect, leading to potentially biased results. The use of stabilised weights was achieved by replacing the numerator (which is 1 in the unstabilised weights) with the crude probability of exposure (i.e. given by the propensity score model without covariates) (Allan et al., 2020; Chesnaye et al., 2022).

Censored dogs were accounted for in the IPTW analysis using inverse probability of censoring weighting (IPCW). IPCW was first developed in the 1990 s by Robins et al. (Robins and Finkelstein, 2000) and aims to reduce bias introduced by informative censoring (Jiménez-Moro and Gómez, 2014), with the assumption of exchangeability and correct model specification (Howe et al., 2011). IPCW compensates for censored subjects by giving more weight to subjects with similar characteristics who are not censored (Dong et al., 2020). To perform IPCW, a binary logistic regression model at each follow-up time point (short- and long-term lameness and analgesia prescription at 3-, 6- and 12-months) was fitted, with censor as the outcome regressed on treatment and the confounding variables described. The model generated predicted

probabilities of being censored, which were used to calculate IP of censoring weights. Thus, separate censor models were created at each fixed time point. Different censoring models can be built according to the different reasons for censoring e.g., missing outcome data due to death/administrative censoring and censoring due to protocol deviation (Hernán and Robins, 2020). Due to the data structure in the current study, these two processes were combined, thus the resulting effect estimate is the probability of the outcome given that the dog remains adherent to treatment assigned during the grace period and remains alive or under veterinary care at the 3-, 6- and 12-month time points.

These IP of censoring weights were combined (by multiplication) with the stabilised IP weights generated from IPTW and used to weight each dog's contribution to binary logistic regression outcome models (for the presence or absence of lameness at short- and long-term follow up and the presence or absence of analgesia prescription at 3, 6 and 12 months) (Robins and Finkelstein, 2000; Robins et al., 2000; Hernán et al., 2001; Jiménez-Moro and Gómez, 2014; Hernán and Robins, 2020). Separate models were created for each outcome (short-term lameness, long-term lameness and analgesia prescription at 3- and 12-months), based on the weighted data and regressed on treatment. The robust (or sandwich) variance estimator was used to obtain valid standard errors (Zou, 2004). Effect modification was assessed by adding biologically plausible interaction terms to the outcome model (namely an interaction between management type and each of the following in turn: bodyweight, breed, age, overweight status, insurance status, orthopaedic comorbidities and non-orthopaedic comorbidities) and evaluating their effect on the confidence intervals and AIC.

Following assessment of the DAG, evaluation of the distribution of combined IP treatment and censoring weights, the effect of interaction terms in the propensity-score model and assessment of effect modification in the outcome model, the SMDs were used as a means of model evaluation. The SMDs compare the distribution of measured baseline covariates between treated and untreated subjects, assessing whether the propensity score model has been adequately specified (Austin, 2011; Zhang et al., 2019).

The variables for which corresponding SMDs were >0.1 were added to outcome models as independent variables. To ensure that the estimand was marginal, these outcome models were used to predict potential outcomes under both surgical management and non-surgical management, and the mean differences calculated (a weighted g-computation).

There may be some confounders in a study that are unknown or not measured and hence unobserved. Sensitivity analysis can examine the extent to which results are affected by values of unmeasured variables (Thabane et al., 2013). E-values were computed in the current study, with an E-value defined as the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to fully explain away a specific treatment–outcome association. This technique does not require specification of the prevalence of unmeasured confounders and does not make assumptions about their nature (VanderWeele and Ding, 2017). An online e-value calculator has been developed, in which the following are specified prior to calculation: outcome type, estimate type, number of diseased and non-diseased exposed dogs, number of diseased and non-diseased unexposed dogs, alpha level for confidence interval and the true causal effect to which to shift the estimate. The number of diseased and non-diseased (exposed and unexposed) dogs was calculated using the adjusted risks calculated after IPTW (VanderWeele and Ding, 2017; Mathur et al., 2018).

Missing data were handled using the missing-indicator method, which uses a dummy variable in the statistical model to indicate whether the value for that variable is missing (Burton and Altman, 2004; Miettinen, 2012). The missing-indicator method assumes that the confounder variable is only a confounder (simultaneously associated with treatment and outcome) when observed, and not when missing. Additionally, we assume that there is no interaction between the missing

indicator and the fully observed confounder in the true propensity score. The missing-indicator method is biased under standard missing at random assumptions, however in the propensity score context, the missing-indicator method can be used in a principled way (Blake et al., 2020). Data were checked for internal validity i.e., to check for consistency between variables available within the study and cleaned in Excel (Microsoft Office Excel 2013, Microsoft Corp.), with analyses conducted using R version 4.0.2 (R Core Team, Vienna, Austria). The "IPW" package was used to generate IP weights (and validated manually) (van der Wal and Geskus, 2011), with code for IPCW derived from Hernán and Robins (2020). The "survey" package was used for binary logistic regression outcome modelling (Oberski, 2014).

3. Results

The denominator population included 2250,741 dogs under primary veterinary care in the VetCompass database during 2019. CCL search terms yielded 32,372 candidate cases, of which 4122 (12.7%) were manually reviewed by the lead author. Of these, 815 (19.8%) were confirmed CCL cases eligible for the emulated trial i.e., fitting with the inclusion and exclusion criteria. Specifically, the emulated trial included 615 (75.5%) surgical cases and 200 (24.5%) non-surgical cases. Of the surgical cases, the most common surgical methods used were TPLO (266, 43.3%), lateral suture (139; 22.6%) and TTA (114; 18.5%). There were 19 (9.5%) non-surgical cases that went on to have surgery within the 12 month follow-up period and were censored at date of surgery. There were no dogs initially assigned to surgical management that died before having surgery. There were 3/815 (0.4%) dogs that died within the grace period, with evidence in the EPRs of non-surgical management at baseline, and so were included in the non-surgical management group. Likewise, 4/815 (0.5%) dogs were lost to follow-up within the grace period that had not had surgery (and with evidence that non-surgical management was recommended at baseline), and so were included in the non-surgical management group. The median follow-up time from first diagnosis (given a cut-off time of 12 months) was 12.0 months (IQR 7.1–9.3, range 0.0 – 12.0) and did not significantly differ in dogs managed non-surgically (median 12.0 months, IQR 8.2 – 12.0, range 0.3 – 12.0) compared to dogs managed surgically (median 12.0 months, IQR 6.6 – 12.0, range 0.0 – 12.0) (p = 0.560). The median time between diagnosis of CCL rupture and date of surgery was 7.0 days (IQR 0.0 – 15.0, range 0.0 – 30.0).

3.1. Demography

The median age of surgical cases (7.1 years, IQR 5.0 – 8.7, range 1.5 – 11.9) was younger than the median age of non-surgical cases (8.2 years, IQR 6.2 – 10.2, range 1.5 – 11.9) (p < 0.001). The median bodyweight of surgical cases (17.6 kg, IQR 10.0 – 28.9, range 2.6 – 62.2) was heavier than the median bodyweight of non-surgical cases (12.1 kg, IQR 8.3 – 24.0, range 2.6 – 65.1) (p < 0.001). The most common breeds amongst surgical cases were the Labrador Retriever (6.8% of surgical cases; 42), Jack Russell Terrier (6.3%; 39), West Highland White Terrier (3.9%; 24) and Staffordshire Bull Terrier (24; 3.9%) in addition to 175 (28.5%) crossbreds. The most common breeds amongst non-surgical cases were the Jack Russell Terrier (8.5%; 17), Cocker Spaniel (5.5%; 11), English Springer Spaniel (4.5%; 9) and Yorkshire Terrier (4.0%; 8) in addition to 58 (29.0%) crossbreds (Table 3).

Data completeness were: breed 100.0%, age 100.0%, neuter status 100.0%, insurance status 100.0%, bodyweight 88.6%, vet group 100.0%, overweight status 39.9%, orthopaedic comorbidity at diagnosis 100.0% and non-orthopaedic comorbidities at diagnosis 100.0% (hence missing data for bodyweight and overweight status were handled using the missing indicator method). Data completeness for outcomes were: short-term lameness 70.6%, long-term lameness 54.6%, analgesia prescription at 3 months 87.0%, analgesia prescription at 6 months 76.8% and analgesia prescription at 12 months 56.2%.

Table 3

Count (%) of surgical cases (n = 615) and non-surgical cases (n = 200) for categorical variables recorded in dogs diagnosed with cranial cruciate ligament (CCL) rupture attending primary-care veterinary practices in the VetCompass Programme in the UK.

Variable	Category	Surgical no. (%)	Non-surgical no. (%)
Breed	Crossbred	175 (28.5)	58 (29.0)
	Purebred - other	191 (31.1)	65 (32.5)
	Labrador Retriever	42 (6.8)	5 (2.5)
	Jack Russell Terrier	39 (6.3)	17 (8.5)
	West Highland White Terrier	24 (3.9)	7 (3.5)
	Staffordshire Bull Terrier	24 (3.9)	7 (3.5)
	Golden Retriever	22 (3.6)	3 (1.5)
	English Springer Spaniel	22 (3.6)	9 (4.5)
	Cocker Spaniel	21 (3.4)	11 (5.5)
	Yorkshire Terrier	20 (3.3)	8 (4.0)
	Bichon Frise	20 (3.3)	6 (3.0)
	Rottweiler	15 (2.4)	4 (2.0)
Bodyweight (kg)	< 15	242 (39.3)	103 (51.5)
	≥ 15	307 (49.9)	70 (35.0)
	Not recorded	66 (10.7)	27 (13.5)
Neuter status	Entire	167 (27.2)	73 (36.5)
	Neutered	448 (72.8)	127 (63.5)
Insurance	Non – insured	171 (27.8)	110 (55.0)
	Insured	444 (72.2)	90 (45.0)
Overweight status	Overweight/obese	179 (29.1)	81 (40.5)
	Ideal bodyweight	46 (7.5)	19 (9.5)
	Not recorded	390 (63.4)	100 (50.0)
Orthopaedic comorbidity at CCL diagnosis	No recorded comorbidity	413 (67.2)	149 (74.5)
	Osteoarthritis	146 (23.7)	37 (18.5)
	Patellar luxation	26 (4.2)	10 (5.0)
	Other	30 (4.9)	4 (2.0)
Non-orthopaedic comorbidities at CCL diagnosis	No	592 (96.3)	180 (90.0)
	Yes	23 (3.7)	20 (10.0)
Veterinary Group	1	171 (27.8)	53 (26.5)
	2	204 (33.2)	60 (30.0)
	3	16 (2.6)	7 (3.5)
	4	115 (18.7)	29 (14.5)
	5	109 (17.7)	51 (25.5)

3.2. Descriptive analysis of clinical outcome

The following descriptive analysis of clinical outcomes reports confounded estimates i.e. prior to methods used to adjust for confounding and excludes dogs with missing outcome data (up to and including Table 4). A lesser proportion of dogs managed surgically had short-term lameness (32.2% versus 58.7%, risk difference (RD) −26.5%) and long-term lameness (16.7% versus 44.5%, RD −27.8%) than dogs managed non-surgically. Likewise, a lesser proportion of dogs managed surgically had an analgesic prescription at ≥ 3 months (27.4% versus 62.5%, RD −35.1%), ≥ 6 months (21.4% versus 52.2%, RD −30.8%) and at ≥ 12 months (17.4% versus 43.9%, −26.5%) than dogs managed non-surgically (Table 4).

3.3. Model evaluation

The standardised mean differences between groups for each of the covariates pre and post IPTW are shown in Table 5. Other than breed, the standardised mean differences in the weighted sample were all below 0.1 for each covariate, indicating well-balanced groups post weighting. Therefore, breed was added to weighted outcome models, to further control for any residual confounding (with the results in Table 6 including the adjustment for breed). These outcome models were then used to predict the potential outcomes under surgical management or

Table 4

Count (%) and unadjusted risk difference (RD) of lameness and analgesia outcomes in surgical cases (n = 533) and non-surgical cases (n = 176) of dogs diagnosed with cranial cruciate ligament (CCL) rupture attending primary-care veterinary practices in the VetCompass Programme in the UK. The RD estimates represent the risk in dogs managed surgically compared with dogs managed non-surgically. Dogs with missing outcome data (n = 231/815; 28.3% for short-term lameness, n = 370; 45.4% for long-term lameness, n = 106; 13.0% for analgesia prescription at 3 months, n = 189; 23.2% for analgesia prescription at 6 months and n = 357; 43.8% for analgesia prescription at 12 months) were excluded.

Outcome	Category	Surgical no. (%)	Non-surgical no. (%)	Unadjusted RD (%)
Short-term lameness	Short-term lameness	153 (32.2)	64 (58.7)	-26.5
	No short-term lameness	322 (67.8)	45 (41.3)	
Long-term lameness	Long-term lameness	56 (16.7)	49 (44.5)	-27.8
	No long-term lameness	279 (83.3)	61 (55.5)	
Analgesia prescription at 3 months	Prescription at ≥ 3 months	146 (27.4)	110 (62.5)	-35.1
	No prescription ≥ 3 months	387 (72.6)	66 (37.5)	
Analgesia prescription at 6 months	Prescription at ≥ 6 months	100 (21.4)	83 (52.2)	-30.8
	No prescription ≥ 6 months	367 (78.6)	76 (47.8)	
Analgesia prescription at 12 months	Prescription at ≥ 12 months	60 (17.4)	50 (43.9)	-26.5
	No prescription ≥ 12 months	284 (82.6)	64 (56.1)	

Table 5

Standardised mean differences (SMD) before and after applying inverse probability weighting. This table shows the SMD for each of the prespecified covariates pre and post weighting.

Variable	SMD before weighting	SMD after weighting
Breed	0.285	0.254
Age (years)	0.416	0.022
Bodyweight (kg)	0.306	0.036
Neuter status	0.202	0.040
Insurance	0.574	0.014
Overweight status	0.274	0.044
Orthopaedic comorbidities at diagnosis	0.216	0.030
Non-orthopaedic comorbidities at diagnosis	0.249	0.019
Vet group	0.213	0.048

non-surgical management to obtain marginal risk differences.

The median propensity score in dogs treated non-surgically was 0.66 (range 0.10–0.96), whilst the median propensity score in dogs treated surgically was 0.83 (range 0.21–0.98). The range of stabilised IP weights (when treatment weights were combined with censoring weights for each outcome) was 0.23–15.0. The linear assumption for age as a continuous covariate was met. Compared to the model including main terms only, inclusion of biologically plausible interaction terms did not improve model fit. Based on the SMDs, only breed did not sufficiently balance between treatment arms. Inclusion of interaction terms did not improve balance of breed (i.e., SMD was consistently > 0.1). Likewise, the decrease in the absolute value of SMDs for some covariates, when interaction terms were included, did not outweigh the increase in SMDs for other covariates. Therefore, the final models included the following covariates to generate propensity scores: age, bodyweight, overweight

status, neuter status, insurance status, non-orthopaedic comorbidities, orthopaedic comorbidities and veterinary group. Management and breed were included as separate terms in the outcome models. Significant effect modification was not evident.

3.4. Emulated trial results

After balancing covariates between the surgical and non-surgical dogs using IPTW, and accounting for censoring using IPCW, surgical management reduced risk of short- and long-term lameness compared with non-surgical management. Specifically, the risk difference for short-term lameness in dogs treated surgically versus non-surgically was −25.7% (95% CI −36.7 to −15.9), whilst the risk difference for long-term lameness was −31.7% (-37.9 to −18.1). Additionally, surgical management reduced risk of analgesia prescription, with a risk difference for analgesia prescription in dogs treated surgically versus non-surgically of −38.9% (-44.0.7 to −28.1) at 3 months, −34.1% (-40.4 to −24.0) at 6 months and −32.7% (-38.9 to −18.3) at 12 months following diagnosis (Table 6).

There was no significant interaction between bodyweight (< 15 kg versus ≥ 15 kg), management type and lameness outcomes. The interaction was assessed by fitting to the outcome model on weighed data and the result was not significant and did not improve model fit (AIC). Specifically, the risk difference for short-term lameness in dogs < 15 kg treated surgically versus non-surgically was 8.9% (95% CI −29.3–47.1), whilst the risk difference for dogs ≥ 15 kg was −3.2% (-40.5–34.1). The risk difference for long-term lameness in dogs < 15 kg treated surgically versus non-surgically was 23.1% (95% CI −12.9–59.0), whilst the risk difference for dogs ≥ 15 kg was −1.4% (-39.5–36.7).

3.5. Sensitivity analysis

E-values were calculated to determine the risk ratio that an unmeasured confounder would need to have (conditional on the measured covariates) to fully explain away the treatment–outcome associations. Surgical treatment was used as the baseline group; therefore the E-values represent the risk ratio in dogs treated non-surgically compared to surgically. The E-values ranged from 2.92 to 5.35 (with lower confidence intervals 2.37–4.14) (Table 7).

4. Discussion

To the authors' knowledge, this is one of the first studies to adopt the

target trial emulation framework for veterinary observational data, whilst also the first to focus on estimating the treatment effect of surgical versus non-surgical management of CCL rupture in dogs. IPTW in conjunction with IPCW identified that surgical management reduces the risk of short- and long-term lameness compared with non-surgical management. Additionally, surgical management resulted in a reduced risk of analgesic prescription at 3-, 6- and 12-months compared with non-surgical management.

Historically, regression adjustment has been used more frequently than IPTW to account for differences in measured baseline characteristics between treated and untreated subjects (Austin, 2011). However, IPTW has become increasingly popular to adjust for confounding in observational studies, with a number of theoretical advantages proposed (Austin, 2011; Elze et al., 2017; Ali et al., 2019; Austin et al., 2021). First, IPTW allows for estimation of the marginal treatment effect, i.e. the average effect of treatment on the population (Armitage and Colton, 1998; Austin, 2011), whilst conventional covariate adjustment estimates the conditional treatment effect, i.e. the average effect of treatment on the individual. Therefore, if the objective of an observational study is to answer the same question as an RCT, the marginal effect is often of greater interest (Austin, 2011). Second, the standardised mean differences compare the distribution of measured baseline covariates between treated and untreated subjects, assessing whether the propensity score model has been adequately specified (Austin, 2011; Zhang et al., 2019). Conversely, goodness-of-fit tests used for conventional covariate adjustment do not determine the degree to which fitted regression models have successfully eliminated systematic differences between treated and untreated subjects (Austin, 2011).

After balancing covariates between the surgical and non-surgical dogs using IPTW, accounting for censoring using IPCW and adjusting further for breed, the marginal risk for short-term lameness and long-term lameness in dogs treated surgically was reduced by 25.7% and 31.7% respectively compared to non-surgically managed dogs. Clinically, these results highlight that surgical management of CCL rupture causes improved short- and long-term lameness outcomes compared to non-surgical management, and to a similar extent at both time points. That said, the absolute risks for short- and long-term lameness in dogs treated surgically (33.9% and 16.3% respectively) are not negligible and provide a benchmark for veterinarian-owner decision making. These results reflect findings from a previous study in which subjective lameness outcomes at 3 and 12 months respectively were improved by 20.6% and 11.4% in dogs managed surgically compared with those managed non-surgically (Wucherer et al., 2013). The long-term lameness improvement was less marked at 12 months in the previous study, compared to the current study, however this previous study used descriptive analysis methods only and was restricted to just 40

Table 6

Adjusted risk and corresponding risk difference (RD) (and 95% confidence intervals) for lameness and analgesia outcomes in dogs under UK primary veterinary care managed surgically (n = 615) or non-surgically (n = 200) for cranial cruciate ligament (CCL) rupture. The estimates represent the risk in dogs managed surgically compared with dogs managed non-surgically using inverse probability of treatment and censoring weighting to adjust for confounding.

Outcome	Adjusted risk in surgical cases (%)	Adjusted risk in non-surgical cases (%)	RD (%)	Lower CI (%)	Upper CI (%)
Short-term lameness	33.9	59.5	-25.7	-36.7	-15.9
Long-term lameness	16.3	48.1	-31.7	-37.9	-18.1
Analgesia prescription ≥ 3 months	27.5	66.4	-38.9	-44.0	-28.1
Analgesia prescription ≥ 6 months	21.0	55.2	-34.1	-40.4	-24.0
Analgesia prescription ≥ 12 months	17.4	50.1	-32.7	-38.9	-18.3

Table 7

Sensitivity analysis: E-values* (on the risk ratio scale, with lower confidence interval (CI)) for lameness and analgesia outcomes in dogs with cranial cruciate ligament (CCL) rupture under UK primary veterinary care managed non-surgically (n = 200) compared to surgically (n = 615). The risk differences for each outcome have been converted to a risk ratio for dogs managed non-surgically compared to surgically (n = 815).

Outcome	Risk Ratio	Lower CI	Upper CI	E-Value	Lower CI
Short-term lameness	1.76	1.50	2.06	2.92	2.37
Long-term lameness	2.95	2.34	3.72	5.35	4.14
Analgesia prescription at 3 months	2.42	2.06	2.84	4.27	3.55
Analgesia prescription at 6 months	2.62	2.15	3.20	4.68	3.73
Analgesia prescription at 12 months	2.87	2.30	3.58	5.19	4.05

* The e-value represents the risk ratio that an unmeasured confounder would need to have (conditional on the measured covariates) to fully explain away the treatment–outcome associations.

overweight dogs weighing > 20 kg.

There is some prior evidence that dogs weighing 15 kg or less have reasonable short- and long-term lameness outcomes when managed non-surgically (Pond and Campbell, 1972; Vasseur, 1984), however this effect by bodyweight was not evident in the current study. This may highlight surgical management of CCL rupture as superior, regardless of bodyweight, bringing into question the pervading veterinary view that smaller dogs have less requirement for surgery (Comerford et al., 2013; Taylor-Brown et al., 2015). However, reduced study power due to missing outcome data in the current study may in part explain this finding, which further studies could help to clarify.

The findings for analgesia prescription in the current study reflect those of lameness outcome, with the risk difference for analgesia prescription in dogs treated surgically versus non-surgically −38.9% at 3 months, −34.1% at 6 months and −32.7% at 12 months following diagnosis. Likewise, the counterfactual risks of short-term lameness and analgesia prescription at 3 months in dogs treated surgically and non-surgically are relatively similar, as are the counterfactual risks of long-term lameness and analgesia prescription at 12 months. This finding suggests analgesia prescription could be used as a useful and reliable proxy measure for lameness. Whilst the risk difference for lameness slightly increased from short- to long-term follow-up, the risk difference for analgesia prescription reduced, i.e. was greatest at 3 months following diagnosis. It is generally recommended that dogs are prescribed analgesia for a period of two weeks following surgery for CCL rupture to account for post-operative pain, provided there are no post-operative complications affecting this decision (Gruen et al., 2014). Whilst there is no definitive guidance for analgesia prescription in dogs managed non-surgically for CCL rupture, there is evidence that veterinarians opt for an initial 6–12 week period of recommended analgesia use (Vasseur, 1984; Comerford et al., 2013). Therefore, this might explain the slightly increased risk difference in analgesia prescription at 3 months relative to 6 and 12 months, as uncomplicated surgical cases are likely to cease analgesic treatment by two weeks after surgery, with residual usage more likely to relate to CCL-related pain rather than post-operative.

The current study shows that on average, surgical management leads to reduced lameness and analgesic prescription outcomes compared with non-surgical management. This finding is in line with the beliefs of many veterinarians about a substantial clinical benefit to dogs from cruciate surgery. All surgical methods were considered in the current study, and whilst the evidence for choosing one technique over another is limited (Bergh et al., 2014), further studies could evaluate clinical outcomes between different surgery types. Based on the SMDs, there was evidence of balance between the surgical and non-surgical groups for all covariates, other than breed. If there is evidence of imbalance (i.e. SMD > 0.1), mis-specification of the propensity score model should be evaluated (Zhang et al., 2019). It was reasoned that both breed and bodyweight could affect treatment decision-making, therefore breed was retained in the propensity-score model, but additionally added to the outcome model to further improve the control for confounding.

The limitations of this study mirror previous VetCompass studies, and are largely based on the nature of retrospective analysis of electronic patient record data, including issues related to missing and misclassified data and application of a case definition to the data available (O'Neill et al., 2014). However, we used appropriate statistical methods to account for missing data and loss to follow-up. Expert opinion was sought in construction of the DAG, however it is possible unmeasured confounders could influence the risk differences calculated. E-values were calculated to quantify the minimum strength of association, on the risk ratio scale, that an unmeasured confounder would need to have with both the treatment and outcome, conditional on the measured covariates, to fully explain away the treatment–outcome effects (VanderWeele and Ding, 2017). The E-values ranged from 2.92 to 5.35, with the lower estimate i.e., 2.92 for short-term lameness as an outcome. Therefore, the risk difference for short-term lameness of −25.7 (and

corresponding risk ratio of 1.76) can be explained away by an unmeasured confounder that was associated with both treatment (management of CCL rupture) and outcome (short-term lameness) by a risk ratio of 2.92, but weaker confounding could not do so (VanderWeele and Ding, 2017). E-values should be interpreted in context and with other strengths and weaknesses of the study and design (VanderWeele and Ding, 2017), but these results suggest that the causal interpretation is strongest for long-term lameness as an outcome (E-value 5.35), whilst more moderate for short-term lameness.

We used a one-month grace period in the current study to ensure the treatment strategies were realistic i.e., dogs will not often have surgery for CCL rupture on the day of diagnosis. However, a consequence of this is that data for an individual dog can be consistent with both treatment strategies during the grace period i.e., they can be non-surgically managed whilst still having the opportunity to be surgically managed within 4 weeks. If a dog dies or is lost to follow-up during this period, randomly assigning the dog to one strategy or the "clone-and-censor" approach have been proposed to avoid bias (Hernán and Robins, 2016). A small number of dogs died or were lost to follow-up within the grace period in the current study (n = 7), and were assigned to the non-surgical arm. Although the outcomes were at 3, 6 and 12 months (so by definition could not occur within the grace period), the missing outcome data for these dogs could bias the results in either direction. In the present study, censoring weights were calculated at fixed time points. Censoring weights can be modelled in different ways, namely using a binary logistic regression model, pooled logistic regression model (or in the case of time-to-event data), Kaplan-Meier Estimator or Cox proportional hazards model, depending on the study design. Additionally, different forms of censoring, e.g. loss to follow-up and protocol deviation, can be modelled separately if the data structure and proportion of censored patients allows (Matsouaka and Atem, 2020; Murray et al., 2021). It should be noted that IPCW based on a fixed binary logistic regression model, with different types of censoring combined, can be less informative than other approaches, albeit less likely to result in extreme weights (Matsouaka and Atem, 2020).

5. Conclusions

Overall, this study demonstrated the application of the target trial framework to veterinary observational data. CCL rupture was used as the condition of interest, with the findings showing that surgical management causes a reduction in short- and long-term lameness, and analgesic prescription, compared with non-surgical management in dogs of all sizes. These findings can inform discussions between veterinarians and owners when deciding on treatment for CCL rupture.

CRediT authorship contribution statement

Camilla Pegram: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Anna Frykfors von Hekkel:** Conceptualization, Methodology, Writing – review & editing. **Chieh-Hsi Wu:** Methodology, Writing – review & editing. **David Church:** Funding acquisition, Writing – review & editing. **Dan O'Neill:** Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Software, Supervision, Writing – review & editing. **Karla Diaz-Ordaz:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Dave Brodbelt:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Yu-Mei Chang:** Conceptualization, Funding acquisition, Methodology, Supervision,

Writing – review & editing.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Data/analysis code availability

Acknowledgements

Ethics approval

Ethics approval was granted by the RVC Social Science Ethical Review Board (reference number SR2018–1652). (reference number URN 2015 1369).

Consent for publication

Authorised for publication

References

ACVS. 2021. Cranial Cruciate Ligament Disease [Online]. Available: https://www.acvs.org/small-animal/cranial-cruciate-ligament-disease [Accessed 17/11/2021].

Admon, A.J., Donnelly, J.P., Casey, J.D., Janz, D.R., Russell, D.W., Joffe, A.M., Vonderhaar, D.J., Dischert, K.M., Stempek, S.B., Dargin, J.M., Rice, T.W., Iwashyna, T.J., Semler, M.W., 2019. Emulating a novel clinical trial using existing observational data. predicting results of the prevent study. Ann. Am. Thorac. Soc. 16, 998–1007.

Ali, M.S., Prieto-Alhambra, D., Lopes, L.C., Ramos, D., Bispo, N., Ichihara, M.Y., Pescarini, J.M., Williamson, E., Fiaccone, R.L., Barreto, M.L., 2019. Propensity score methods in health technology assessment: principles, extended applications, and recent advances. Front. Pharmacol. 973.

Allan, V., Ramagopalan, S.V., Mardekian, J., Jenkins, A., Li, X., Pan, X., Luo, X., 2020. Propensity score matching and inverse probability of treatment weighting to address confounding by indication in comparative effectiveness research of oral anticoagulants. J. Comp. Eff. Res. 9, 603–614.

Armitage, P., Colton, T., 1998. Encyclopedia of biostatistics. J. Wiley.

Austin, P.C., 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat. Med. 28, 3083–3107.

Austin, P.C., 2011. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar. Behav. Res 46, 399–424.

Austin, P.C., Stuart, E.A., 2015. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat. Med 34, 3661–3679.

Austin, P.C., Yu, A.Y.X., Vyas, M.V., Kapral, M.K., 2021. Applying propensity score methods in clinical research in neurology. Neurology 97, 856–863.

Balshem, H., Helfand, M., Schünemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., 2011. GRADE guidelines: 3. rating the quality of evidence. J. Clin. Epidemiol. 64, 401–406.

Barter, R. 2017. The intuition behind inverse probability weighting in causal inference. Available from: https://www.rebeccabarter.com/blog/2017-07-05-ip-weighting/ [Accessed 03/12/2021].

Bergh, M.S., Sullivan, C., Ferrell, C.L., Troy, J., Budsberg, S.C., 2014. Systematic review of surgical treatments for cranial cruciate ligament disease in dogs. J. Am. Anim. Hosp. Assoc. 50, 315–321.

Blake, H.A., Leyrat, C., Mansfield, K.E., Seaman, S., Tomlinson, L.A., Carpenter, J., Williamson, E.J., 2020. Propensity scores using missingness pattern information: a practical guide. Stat. Med. 39, 1641–1657.

Burton, A., Altman, D., 2004. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. Br. J. Cancer 91, 4–8.

Cain, L.E., Saag, M.S., Petersen, M., May, M.T., Ingle, S.M., Logan, R., Robins, J.M., Abgrall, S., Shepherd, B.E., Deeks, S.G., 2016. Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. Int. J. Epidemiol. 45, 2038–2049.

Chesnaye, N.C., Stel, V.S., Tripepi, G., Dekker, F.W., Fu, E.L., Zoccali, C., Jager, K.J., 2022. An introduction to inverse probability of treatment weighting in observational research. Clin. Kidney J. 15, 14–20.

Christopher, S.A., Beetem, J., Cook, J.L., 2013. Comparison of long-term outcomes associated with three surgical techniques for treatment of cranial cruciate ligament disease in dogs. Vet. Surg. 42, 329–334.

Comerford, E., Forster, K., Gorton, K., Maddox, T., 2013. Management of cranial cruciate ligament rupture in small dogs: a questionnaire study. Vet. Comp. Orthop. Traumatol. 26, 493–497.

Danaei, G., Rodríguez, L.A.G., Cantero, O.F., Logan, R., Hernán, M.A., 2013. Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. Stat. Methods Med. Res. 22, 70–96.

Dean, R.S., 2017. Veterinary clinical trials are on trial. Vet. Rec. 181, 193.

Dohoo, I., Martin, S., Stryhn, H., 2009. Veterinary epidemiologic research. VER. Inc., Charlottetown, PE, Canada.

Dong, G., Mao, L., Huang, B., Gamalo-Siebers, M., Wang, J., Yu, G., Hoaglin, D.C., 2020. The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of independent censoring. J. Biopharm. Stat. 30, 882–899.

Elze, M.C., Gregson, J., Baber, U., Williamson, E., Sartori, S., Mehran, R., Nichols, M., Stone, G.W., Pocock, S.J., 2017. Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies. J. Am. Coll. Cardiol. 69, 345–357.

Fauron, A., Perry, K., 2017. Canine CRCL disease Part 3: treatment and prognosis. Veterinary Times 47, 16–22.

Frey, B.B., 2018. THE SAGE Encyclopedia of Educational Research, Measurement, and Evaluation.

García-Albéniz, X., Hsu, J., Hernán, M.A., 2017. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. Eur. J. Epidemiol. 32, 495–500.

Greenland, S., Pearl, J., Robins, J.M., 1999. Causal diagrams for epidemiologic research. Epidemiology 37–48.

Gruen, M.E., Roe, S.C., Griffith, E., Hamilton, A., Sherman, B.L., 2014. Use of trazodone to facilitate postsurgical confinement in dogs. J. Am. Vet. Med. Assoc. 245, 296–301.

Hammerton, G., Munafò, M.R., 2021. Causal inference with observational data: the need for triangulation of evidence. Psychol. Med. 51, 563–578.

Hernán, M.A., Robins, J.M., 2016. Using big data to emulate a target trial when a randomized trial is not available. Am. J. Epidemiol. 183, 758–764.

Hernán, M.A., Robins, J.M., 2020. Causal inference: what if. Boca Raton: Chapman & Hall/CRC.

Hernán, M.A., Brumback, B., Robins, J.M., 2001. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. J. Am. Stat. Assoc. 96, 440–448.

Hernán, M.A., Alonso, A., Logan, R., Grodstein, F., Michels, K.B., Stampfer, M.J., Willett, W.C., Manson, J.E., Robins, J.M., 2008. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiol. (Camb., Mass.) 19, 766.

Hill, J., Stuart, E.A., 2015. Causal inference: overview. In: Wright, J.D. (Ed.), International Encyclopedia of the Social & Behavioral Sciences, Second Edition. Elsevier, Oxford, pp. 255–260.

Holmes, M., Cockcroft, P., 2004. Evidence-based veterinary medicine 1. Why is it important and what skills are needed? Practice 26, 28–33.

Howe, C.J., Cole, S.R., Chmiel, J.S., Munoz, A., 2011. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. Am. J. Epidemiol. 173, 569–577.

Hudgens, M.G., Halloran, M.E., 2008. Toward causal inference with interference. J. Am. Stat. Assoc. 103, 832–842.

Jiménez-Moro, J.L., Gómez, J., 2014. Inverse probability of censoring weighting for selective crossover in oncology clinical trials. Paper SP02.

Johnson, J.A., Austin, C., Breur, G.J., 1994. Incidence of canine appendicular musculoskeletal disorders in 16 veterinary teaching hospitals from 1980 through 1989. Vet. Comp. Orthop. Traumatol. 7, 56–69.

Kirkness, H., 2020. Management of cranial cruciate ligament ruptures in dogs. Vet. Nurs. J. 35, 235–237.

Kirkwood, B., 2003. Essential medical statistics. In: Kirkwood, B.R., Sterne, J.A.C. (Eds.), Blackwell Science. Mass. Oxford, Malden.

Labrecque, J.A., Swanson, S.A., 2017. Target trial emulation: teaching epidemiology and beyond. Eur. J. Epidemiol. 32, 473–475.

Maringe, C., Benitez Majano, S., Exarchakou, A., Smith, M., Rachet, B., Belot, A., Leyrat, C., 2020. Reflection on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. Int. J. Epidemiol. 49, 1719–1729.

Mathur, M.B., Ding, P., Riddell, C.A., VanderWeele, T.J., 2018. Web Site and R Package for Computing E-values. Epidemiology 29, e45–e47.

Matsouaka, R.A., Atem, F.D., 2020. Regression with a right-censored predictor using inverse probability weighting methods. Stat. Med. 39, 4001–4015.

Microsoft Learn. 2022. RAND (Transact-SQL) [Online]. Available: https://learn.microsoft.com/en-us/sql/t-sql/functions/rand-transact-sql?view=sql-server-ver15 [Accessed 16/01/2023].

Miettinen, O.S., 2012. Theoretical epidemiology: principles of occurrence research in medicine. Theor. Epidemiol.: Princ. Occur. Res. Med. 359-359.

Murray, E.J., Caniglia, E.C., Petito, L.C., 2021. Causal survival analysis: a guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence. Res. Methods Med. Health Sci. 2, 39–49.

Oberski, D., 2014. lavaan. survey: An R package for complex survey analysis of structural equation models. J. Stat. Softw. 57, 1–27.

O'Neill, D.G., Church, D.B., McGreevy, P.D., Thomson, P.C., Brodbelt, D.C., 2014. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. PloS One 9, e90501.

Osborne, J.W., Waters, E., 2002. Four assumptions of multiple regression that researchers should always test. Practical assessment, research, and evaluation 8, 2.

Pegram, C., Raffan, E., White, E., Ashworth, A., Brodbelt, D., Church, D., O'Neill, D., 2021. Frequency, breed predisposition and demographic risk factors for overweight status in dogs in the UK. J. Small Anim. Pract.

Pegram, C., Brodbelt, D.C., Diaz-Ordaz, K., Chang, Y., von Hekkel, A.F., Church, D.B., O'Neill, D.G., 2023. Risk factors for unilateral cranial cruciate ligament rupture diagnosis and for clinical management in dogs under primary veterinary care in the UK. Vet. J. 292, 105952.

Petersen, M.L., Porter, K.E., Gruber, S., Wang, Y., Van Der Laan, M.J., 2012. Diagnosing and responding to violations in the positivity assumption. Stat. Methods Med. Res. 21, 31–54.

Pfeiffer, D.U., 2010. Veterinary Epidemiology An Introduction. John Wiley & Sons Ltd West Sussex, UK.

Pond, M., Campbell, J., 1972. The canine stifle joint I. Rupture of the anterior cruciate ligament: an assessment of conservative and surgical treatment. J. Small Anim. Pract. 13, 1–10.

Rehkopf, D.H., Glymour, M.M., Osypuk, T.L., 2016. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. Curr. Epidemiol. Rep. 3, 63–71.

Robins, J.M., Finkelstein, D.M., 2000. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 56, 779–788.

Robins, J.M., Hernan, M.A., Brumback, B., 2000. Marginal structural models and causal inference in epidemiology. Lww 550–560.

Rolph, N.C., Noble, P.-J.M., German, A.J., 2014. How often do primary care veterinarians record the overweight status of dogs? J. Nutr. Sci. 3.

Taylor-Brown, F.E., Meeson, R.L., Brodbelt, D.C., Church, D.B., McGreevy, P.D., Thomson, P.C., O'Neill, D.G., 2015. Epidemiology of cranial cruciate ligament disease diagnosis in dogs attending primary-care veterinary practices in England. Vet. Surg. 44, 777–783.

Tennant, P.W., Murray, E.J., Arnold, K.F., Berrie, L., Fox, M.P., Gadd, S.C., Harrison, W. J., Keeble, C., Ranker, L.R., Textor, J., 2021. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int. J. Epidemiol. 50, 620–632.

Textor, J., Van der Zander, B., Gilthorpe, M.S., Liśkiewicz, M., Ellison, G.T., 2016. Robust causal inference using directed acyclic graphs: the R package 'dagitty. Int. J. Epidemiol. 45, 1887–1894.

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Debono, V.B., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G., Goldsmith, C.H., 2013. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med. Res. Methodol. 13, 92.

The VeNom Coding Group. 2019. VeNom Veterinary Nomenclature [Online]. VeNom Coding Group. Available: http://venomcoding.org [Accessed January 28 2019].

van der Wal, W.M., Geskus, R.B., 2011. ipw: an R package for inverse probability weighting. J. Stat. Softw. 43, 1–23.

VanderWeele, T.J., Ding, P., 2017. Sensitivity analysis in observational research: introducing the E-Value. Ann. Intern Med. 167, 268–274.

Vasseur, P., 1984. Clinical results following nonoperative management for rupture of the cranial cruciate ligament in dogs. Vet. Surg. 13, 243–246.

VetCompass. 2019. VetCompass™ Programme [Online]. London: RVC Electronic Media Unit. Available: http://www.rvc.ac.uk/VetCOMPASS/ [Accessed March 5th 2019].

VetCompass. 2023. About [Online]. Available: https://www.rvc.ac.uk/vetcomp ass/about [Accessed 21/09/2023].

Wareham, K.J., Hyde, R.M., Grindlay, D., Brennan, M.L., Dean, R.S., 2017. Sponsorship bias and quality of randomised controlled trials in veterinary medicine. BMC Vet. Res. 13, 234.

Wucherer, K.L., Conzemius, M.G., Evans, R., Wilke, V.L., 2013. Short-term and long-term outcomes for overweight dogs with cranial cruciate ligament rupture treated surgically or nonsurgically. J. Am. Vet. Med. Assoc. 242, 1364–1372.

Xu, S., Ross, C., Raebel, M.A., Shetterly, S., Blanchette, C., Smith, D., 2010. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health 13, 273–277.

Yang, D., Dalton, J., 2014. A unified approach to measuring the effect size between two groups using SAS®. SAS Global Forum; 2012. Contract No.: Paper, 335-2012.

Zhang, Z., Kim, H.J., Lonjon, G., Zhu, Y., 2019. Balance diagnostics after propensity score matching. Ann. Transl. Med. 7.

Zou, G., 2004. A modified poisson regression approach to prospective studies with binary data. Am. J. Epidemiol. 159, 702–706.