Raising the Bar for Animal Clinical Trials

In the context of clinical trials, scientific standards demand a degree of consistency, characterised by well-defined measurement using standardised instruments. In the words of William Thomson, Lord Kelvin of Glasgow in 1889, "When you cannot measure it, when you cannot express it in numbers… you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be" [1]

Recent advances in veterinary science mean that pharma companies can reap the benefits of health-related quality of life (HRQL) surveys and electronic data capture (EDC) for animal clinical trials, as well as those targeted at the human drug market[2]. As well as providing a more complete picture of the animals' true experience, an outcome increasingly sought by regulators, this has potentially huge implications for the outcomes, effectiveness and cost-benefit profile of trials.

Setting the Scene

Although a fraction of the size of the human drug development market, there is a growing demand for new, clinically validated therapies for chronic conditions in companion animals; e.g. in oncology, cardiology, endocrinology and osteoarthritis.

Moreover, as far back as 1929, Nobel Laureate August Krogh proposed the potential contribution of naturally occurring diseases in animals as models of human disease[3]. This vision was realised more than 30 years later when clinical trials were conducted in client-owned dogs with naturally occurring cancer. Currently, clinical trials in pet dogs and cats are at the forefront of translational efforts to discover and develop drugs for a variety of human diseases, in addition to cancer. These natural models of disease might ultimately decrease the failure rate in human clinical trials and accelerate the delivery of effective treatments to the human clinical market, thus saving money and time. Consequently, the requirement for companion animal trials within the human sector is predicted to increase year on year.

Faced with rising R & D costs, clinical investigators in both the human and animal therapeutic drug industry are constantly driven to make clinical trials more efficient, potentially saving development time and money. In the human field, strategies include the utilisation of study outcomes which measure, in a scientifically robust way, what is important to the patient (e.g. increase in wellbeing, physical and social functioning) as well as indicators of disease improvement or deterioration. This process further benefits from electronic data capture to achieve maximum efficiency.

The Benefits - and Challenges - of Measuring HRQL

HRQL surveys take the form of structured questionnaires which can be either generic or disease-specific. Disease-specific instruments may be more responsive to clinical change, but generic instruments measure the impact of anything that will affect quality of life (QOL), and can be valuable indicators of a range of impacts associated with disease and its treatment. They have been shown to provide an effective alternative when particular disease-specific instruments do not exist, and may be the only option when a patient is suffering from more than one condition. Many drug companies have successfully used HRQL surveys as patient-reported outcomes (PRO) to measure the impact and effectiveness of their drugs in human clinical trials. This is accepted practice, and guidelines in several therapeutic areas cite the measurement of HRQL as a desirable component of any clinical trial, with data used to support claims in approved medical product labelling, as either a primary or secondary outcome measure.

In support of the former, The Heart Failure Guideline Panel, sponsored by the Agency for Health Care Policy and Research (AHCPR), determined that changes in intermediate outcomes constituted insufficient evidence of the effectiveness of treatment and that the relevant outcomes were those experienced directly by patients. HRQL is about the value of treatment to the patient and its use is encouraged by the FDA. Similarly, and increasingly in veterinary drug trials, the use of valid, reliable and responsive HRQL measures is actively encouraged by the FDA to provide optimal information regarding treatment effect from the animal's perspective.

A PRO represents the effect of the disease on health and functioning from the patient's perspective, and is a report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. Accordingly, it is impossible to have a PRO for animals.

In non-verbal populations, the FDA encourages observer reports, but only where these include behaviours that can be observed. For example, an owner cannot validly report a dog's pain intensity, but can report dog behaviour thought to be caused by pain. Similarly, for this reason, asking an owner to rate their dog's HRQL on a 0 – 10 scale cannot be regarded as a valid measure, but a questionnaire instrument that draws such a conclusion from owner-reported observations of behaviour can. This is an important point because there are veterinary instruments currently in use that do utilise that invalid global estimate to report a HRQL outcome.

Building Appropriate Instrumentation

The quote from Lord Kelvin at the start of this paper hints at the rigorous thinking that the development of any instrument of measurement demands. Instrument development is an iterative process, in which instruments are refined and re-tested with new populations in new contexts and for new purposes. The development of instruments to measure HRQL is a costly and time-consuming undertaking, but the important contribution of the psychometric approach to such instrument development is widely recognised[4].

By adopting a rigorous methodological approach to constructing HRQL instruments which assures their validity, reliability and responsiveness, developers and sponsors can be more confident that the product meets the standards required of an effective outcome measure for clinical trials. Psychometric methodologies for the development and testing of instruments are very well established[5,6] and can be summarised as follows:

- specify measurement goals (and hence the ideal measurement scale) – what is to be measured, in which population and for what purpose,
- develop a pool of potential items for the instrument,
- select suitable items from the pool and subject to expert validation,
- incorporate selected items into an instrument, with consideration given to layout, response options, instructions to respondent and administration,
- pre-test the instrument to ensure that the target respondent can use the instrument correctly,
- field-test the instrument, in order to evaluate its psychometric properties.

Instrument Properties

Validity (criterion, content and construct) is the most fundamental attribute of any instrument because it provides evidence that the instrument is able to measure that which it was designed to measure; its importance cannot be overemphasised. However, an instrument cannot be said to 'be valid': it can only be shown to have validity for particular purposes, with defined populations and in specified contexts[5]. Therefore, a key question to ask during instrument selection is 'what evidence is there that this instrument can measure what I want to measure in the population that I want to measure and for my particular purpose?'

- **Criterion validity** is the agreement of a new instrument with some existing `gold standard', however in the case of animal pain and HRQL no such gold standard exists, and so other forms of validity must be sought,
- **Content validity** focuses on the appropriateness and completeness of the items within the instrument and is deemed to be present when those items cover all the relevant aspects being measured without including any extraneous features. This is established during construction of the instrument and usually involves 'expert' input. The psychometric approach takes care of this,
- **Construct validity** is demonstrated when hypotheses regarding the attribute(s) in question are upheld by use of the instrument; e.g. we might hypothesise that the HRQL profile of healthy dogs will differ from that of ill dogs, and if an instrument supports that hypothesis then that is evidence for the construct validity of that instrument.

Reliability is a measure of whether an instrument can measure accurately and repeatedly what it is intended to measure, so that measurements of individuals on different occasions, or by different observers, or by similar or parallel tests, produce the same or similar results.

Responsiveness – Although validity and reliability have been used to evaluate instruments, some would argue that responsiveness is equally or even more important in the choice of an HRQL evaluative instrument such as is likely to be used in a clinical trial[7]. Responsiveness in a clinical instrument is that property which ensures that the instrument is sensitive enough to detect differences in health status that are important to the clinician or to the patient. Because the ability of an instrument to detect change influences the sample size for evaluating the effectiveness of treatment, the choice of a responsive instrument is an important economic consideration.

Utility – A useful clinical instrument must not only be valid, reliable and responsive but also 'practical and easy to administer, score and interpret'[8]. If it is not, people will not use it despite its validity, reliability and responsiveness. Systems that use EDC score highly in that regard.

Scoring and Interpretation of Data

HRQL measures can be single-index scores which will tell you the animal is better or worse, or they can be represented by a multi-dimensional profile with scores in different areas (domains) of QOL. Since HRQL is a complex multi-domain concept, one could argue that this should be better represented by a multi-dimensional profile to help understand the process of change and offer more flexibility with analysis. Another option is to combine both within a single instrument, as is the case in the human Sickness Impact Profile which yields sub-scale scores for its constituent domains, but can nevertheless be represented by a single aggregate total score across all domains. Currently, veterinary HRQL measures utilise one or the other, but not both.

Appropriate interpretation of data is fundamental to the success of a trial, but regardless of whether the primary endpoint for the clinical trial is based on individual responses to treatment, or on a group response, the FDA suggest that it is useful to display individual responses, often using an *a priori responder definition* (i.e., the individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit). The responder definition is determined empirically and will vary from trial to trial, but can be derived based on the owner's impression of change or appropriate weighting of domain scores in the HRQL instrument.

Important Design Features

Type of Measurement Scale

The measurement from an instrument can have nominal, ordinal or interval scale properties depending on the design of the instrument. Ordinal measurement, which provides information about how individuals relate to one another in relative terms, and interval measurement, with which individuals can be placed on a scale of equal units (like a ruler), are practicable and desirable for the assessment of

HRQL. Interval level measurement is more demanding to create[9] but provides more precise and meaningful measurement, and is likely to have increased sensitivity and responsiveness.

Response Options

Each item in a questionnaire is accompanied by an answer option or options. These may be dichotomous (yes/no), categorical (e.g. mild/moderate/severe), ordinal (e.g. numerical Likert-type scale) or even more complex. If responses are likely to lie on a continuum, it is important that respondents have the opportunity to answer in this way to ensure minimum loss of information since "the finer the distinction that can be made between subjects' responses, the greater the precision of the measure"[10]. For example, for a question about anxiety in a dog, more and better information will be captured by asking the owner to indicate an amount of anxiety than by simply asking if the dog is or is not anxious.

Respondent Bias

In research involving any questionnaire instrument, anything that influences the responses of the participants away from an accurate or truthful response can have a large impact on the validity of a study. This could be as simple as the way questions are phrased or, in the case of HRQL, could be related to the complexity of the human-animal bond. Some owners may not wish to portray their dog as having a poor QOL, or indeed another might be looking for some justification for euthanasia.

Although respondent bias can be minimised through the use of very large numbers of trial subjects (incurring a significant cost implication to the trial), it makes more sense to limit such bias through careful construction of the questionnaire itself, and by hiding from the respondent the extent to which each response option represents the 'right' or 'wrong' answer. A range of established methods can be used for this purpose. Electronic data capture also helps by not allowing respondents to go back and check their previous responses.

Delivery

Capturing data electronically, as opposed to using pen and paper, has many advantages and is an increasingly well-accepted methodology within human clinical trials. EDC is generally accepted to improve the quality of data (no missing data, no out of range data, better respondent compliance) and should comply with 21 CFR 11.

Due to recent advances in veterinary medicine, EDC is now available for use in companion animal clinical trials, with all the attendant advantages that this brings. As well as providing a platform for HRQL assessment with automatic and instantaneous data capture, computation of scores and output delivery to the sponsor, there is the capacity to ask other questions of the owner that may be pertinent to the study, such as details regarding side-effects.

Removing Guesswork and Reaping Benefits

Being able to measure in a scientifically robust fashion how an animal "feels" hugely strengthens holistic understanding of the effects of drug therapies and provides empirical data on which to support better research outcomes, where fewer subjects and shorter trials add economic advantage.

In short, having a clinically significant "Doctor Dolittle" process to engage directly with animals to measure physical and emotional responses is a win/win scenario in commercial terms too.

References

1. Thomson W (1889) Popular Lectures and Addresses vol. 1, 'Electrical Units of measurement', delivered 3 May 1883
2. Reid J, Wiseman-Orr ML, Scott EM, Nolan AM (2013) Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. JSAP 54: 227-233
3. Krogh A (1929) The Progress of Physiology, American Journal of Physiology 90: 243 – 251
4. Cook KF, Monahan PO and McHorney CA (2003) Delicate balance between theory and practice: health status and item response theory (Editorial). Medical Care 41: 571-574
5. Streiner DL and Norman GR (2008) Health Measurement Scales: a practical guide to their development and use, 4th Edition. Oxford: Oxford University Press
6. Abell N, Springer DW and Kamata A (2009) Developing and Validating Rapid Assessment Instruments. Oxford: Oxford University
7. Guyatt G, Walter S and Norman G (1987) Measuring change over time: assessing the usefulness of evaluative instruments. Journal of Chronic Diseases 40: 17 1 -178
8. Landgraf JM and Abetz LN (1996) Measuring health outcomes in pediatric populations: issues in psychometrics and application. In: Spilker B (ed) Quality of Life and Pharmacoeconomics in Clinical Trials (2nd Edition) pp 793-802. Lippincott-Raven: Philadelphia, USA
9. Morton C, Reid J, Scott E, Holton L and Nolan A (2005) Application of a scaling model to establish and validate an interval level pain scale for assessment of acute pain in dogs. American Journal of Veterinary Research 66: 2154-2166
10. Bowling A (1991) Measuring Health: A Review of Quality of Life Measurement. Open University Press: Buckingham, UK

Jacky Reid BVMS, PhD, DVA, DipECVAA MRCA, MRCVS, Professor of Veterinary Anaesthesia, recognised RCVS specialist in anaesthesia and European veterinary specialist in anaesthesia and analgesia, is an honorary senior research fellow in the University of Glasgow. She is a founding director and CEO of NewMetrica, a research company set up to build on and continue research carried out by the University of Glasgow Pain and Welfare Group over two decades, to measure pain and health-related quality of life (HRQL) in non-verbal species. NewMetrica currently supplies scientifically robust questionnaire instruments to measure acute pain in dogs and cats and HRQL in dogs, the latter being web-based and suitable for EDC in clinical trials. A similarly web-based generic instrument to measure cat HRQL with a 'bolt-on' osteoarthritis disease-specific module is under construction and should be available early 2017. Please contact Jacky for further information by email Jacky.reid@newmetrica.com or phone +44 7876 683262