Commentary

Clinical Trials for Acupuncture and Other Traditional Chinese Veterinary Medicine Treatments

Erin Mayo DVM

ABSTRACT
Recent emphasis on the importance of practicing evidence-based medicine (EBM) can be seen as an opportunity for traditional Chinese veterinary medicine (TCVM) practitioners to participate in the process of validation of acupuncture (AP) and other traditional Chinese veterinary medicine (TCVM) treatments. To do so requires an understanding of how to properly design, conduct and report clinical trials. Of particular importance are a succinct answerable research question, an adequate number of animals, clear inclusion and exclusion criteria, randomization of subjects, blindedness of evaluators, sham treatments, control groups, objective measurements of outcomes as much as possible and statistical analysis of data. Current clinical trial design methods may be well suited to investigate Chinese herbal medicines, but blinding and adequate control groups can be challenging to incorporate into clinical AP trials. Through the use of modifications, such as blinded evaluators and sham AP techniques, these difficulties can be mitigated. Including specific treatments for different TCVM pattern diagnoses, within a biomedical diagnoses , can ensure clinically reliable treatments, transparency and reproducibility of results by other researchers. Reviewing "The Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA)" and adaptations for veterinary TCVM studies, before creating the clinical study design, can ensure that all components are included. Clinical researchers need to be cognizant of the importance of a rigorous study design to ensure high quality results that are clinically relevant, thus improving overall patient care and contributing to the knowledge base of EBM.

Key words: Research, acupuncture, traditional Chinese veterinary medicine, TCVM, Chinese herbal medicine, clinical trials, sham acupuncture, evidence-based medicine, STRICTA

ABBREVIATIONS

AP	Acupuncture
CAM	Complementary and alternative medicine
EBM	Evidence-based medicine
RCT	Randomized controlled trials
TCVM	Traditional Chinese veterinary medicine

Recent emphasis on the importance of practicing evidence-based medicine (EBM) has thrust the use of acupuncture (AP), Chinese herbal medicines and other complementary and alternative medicine (CAM) treatments into a negative light by some, with suggestions that there is a paucity of scientific evidence of efficacy.[1] The scientific biomedical community defines "evidence of efficacy" as significant positive effects confirmed by established scientific research methodology. The current scientific methods were originally developed to investigate pharmacological agents and may be poorly suited to evaluate AP and other TCVM and CAM treatments, thus requiring adaptations.[2]

The legal and moral imperative to provide safe and effective treatment options to patients is always important for veterinarians to consider. The practice of veterinary medicine may still be as much an art, as it is a science. What constitutes standard of care varies among individuals, but all practitioners are accountable to certain professional standards. Some members of the conventional biomedical community may seem to disparage TCVM and other CAM practices, but others simply want all types of practices held to the same standard. How veterinarians choose to treat a particular patient or diagnosis must always be done with the ability to justify the treatment choice.

In a recent letter to the Journal of the American Veterinary Medical Association, Dr. Richard Palmquist highlights some of the difficulties that are encountered with CAM research, beyond those of conventional research methodology.[3] Financial support of CAM research is limited and often focuses on topics that are important or of interest to specific researchers in academic institutions. Without other clinicians interested in research of AP and other TCVM treatments, little

From: Ewing, NJ

attention and funding may be directed towards the validation of these procedures. Dr. Palmquist further suggests that "lack of evidence cannot be taken…as evidence of a lack of effectiveness", and the process of validation "begins by looking".[3]

Interest in the addition of new treatment methods often begins with veterinarians in clinical practice, as evidenced by increasing enrollment numbers of private practitioners in TCVM training programs. TCVM practitioners can play a vital role in the validation process of AP and other TCVM treatments. Most TCVM practitioners have experiential evidence of success with AP and other TCVM treatments for a variety of conditions, but this does not satisfy the requirements for EBM. The current pressure from the conventional biomedical community can be viewed as an opportunity to scientifically prove the effectiveness of TCVM treatments. The ability to better understand and improve a chosen treatment method can be achieved through clinical research. As TCVM practitioners provide the medical community with thought-provoking results from pilot clinical studies, increased interest in TCVM treatments will stimulate academic and private institutions to perform larger studies. The end result may not only be evidence for the effectiveness of AP and other TCVM treatments, but integration of TCVM practices into conventional medicine and improved patient outcomes.

The hierarchy of the strength of evidence in research is outlined in Figure 1.[4] The opinion of experts based on clinical observations and experience is the first stage of evidence development. Animal research may be more directly applicable to veterinary medicine and higher on the hierarchy, than for humans as shown in Figure 1. Case reports or case series are the next tier from the bottom in the hierarchy and are published to describe clinical phenomena. TCVM practitioners most commonly write case reports and case series. Case series are usually considered stronger evidence than a case report, but neither qualifies as EBM.

Case control studies are observational studies, often used in epidemiology, that have a research question and hypothesis about a clinical phenomenon. Observing and comparing animals with and without a conventional disease or TCVM pattern to identify risk factors that might contribute to the development of the disease is typical for case control studies. Cohort studies are observational longitudinal studies, also often used in epidemiology, to follow 1 or more groups over time to observe some specific difference (e.g. Does a group of animals exposed to a substance have a higher incidence of liver disease compared to a group of animals not exposed to a substance).

Clinical randomized controlled trials (RCT) are used to test the efficacy of medical interventions such as TCVM treatments and may be small pilot studies involving 10-20 animals/group or multicenter studies involving a larger number of animals to ensure significance of results. Large clinical RCT are time consuming and expensive to perform. They can be useful to develop clinical guidelines for EBM, but repeatable results by different researchers are the strongest

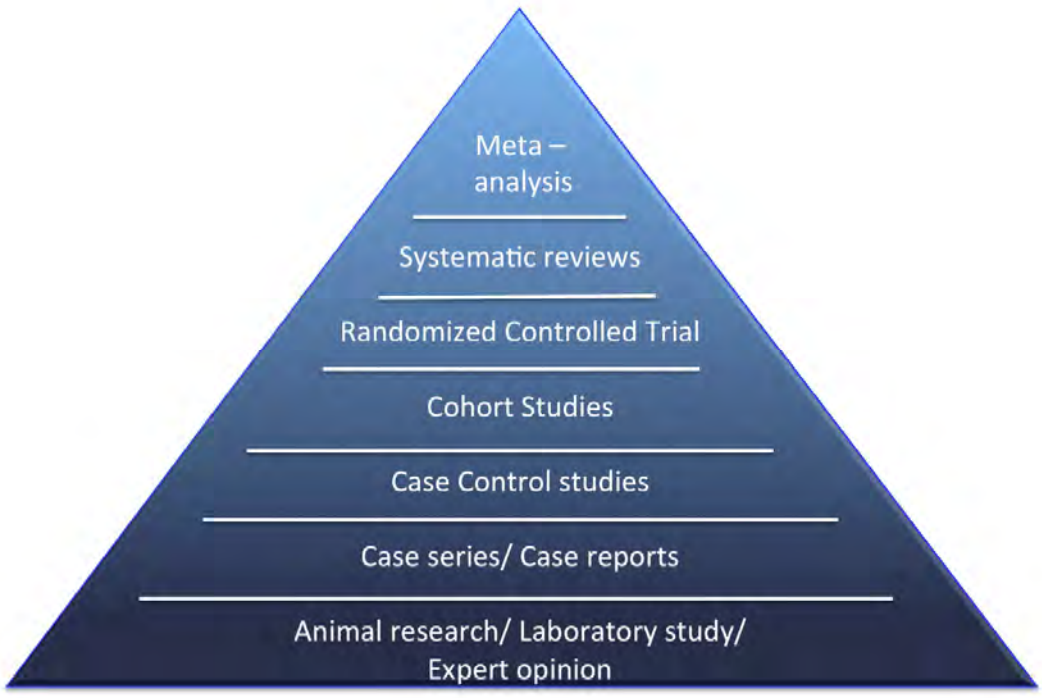

Figure 1: Hierarchy of research evidence

evidence. The results of systematic reviews and meta-analysis of RCT on a subject form the strongest EBM. Much of what is considered standard of care in conventional veterinary medicine and TCVM is not EBM and more clinical research is needed.

The World Health organization (WHO) "Guidelines for Clinical Research on Acupuncture" states that there are 3 criteria for quality acupuncture studies: validity, reliability and statistical significance.[5] To achieve this, the researcher must be knowledgeable about currently accepted study designs and their challenges and shortcomings, when used to investigate AP and other TCVM treatments. Accepted research design modifications that address the unique aspects of AP are available, but may vary with the specific research question. The essential components of a high quality clinical study must include: (1) a clear, concise research question, hypothesis and objective, (2) an adequate number of study subjects, (3) clear inclusion and exclusion criteria, (4) experimental and control groups, (5) randomization of group allocations, (6) objective outcome criteria and measurements, (7) blindedness and (8) adequate statistical analysis.

The Research Question, Hypothesis and Objective(s) of a Clinical Study

The first step in planning a clinical study is to determine the primary research question. Research questions arise from knowledge about known and unknown aspects of a subject.[6] Initially several research questions may become apparent, but a primary research question should be identified and will determine the basis of the hypothesis and objectives of a clinical study. The breadth of the topic should be considered when formulating the research question. Questions that are too simple may yield inconclusive results and questions that are too broad may be unable to answer with a single study. Breaking a problem down into smaller questions is usually best.

The researcher must determine what clinical unknown can or should be investigated and the need for an investigation. A good research question should be: (1) feasible, (2) interesting, (3) novel, (4) ethical and (5) relevant (FINER criteria).[6] The feasibility of a research question includes consideration of potential patient recruitment problems, the time, effort and costs needed to determine a reliable answer and other practical considerations. If patient recruitment rate or sample size estimates are unknown, a pilot study may be needed to estimate the number of patients and time needed to complete the study. Pilot or feasibility studies are small-scale studies that can provide practical information used to justify a larger scale study that is more likely to be considered EBM. In clinical studies the effects of exclusion criteria on patient numbers and the numbers of clients (in veterinary medicine) that may decline or discontinue involvement and be lost to follow-up are first considered. Medical record reviews of current and past cases can be useful to provide an estimate of potential patient numbers. Another practical consideration is ensuring that the needed skills, equipment and personnel are already available to execute the study. Collaborators may be needed for some aspects of the study (e.g. a statistician to aid with data analysis). Pilot or feasibility studies also offer insights into the commitment and costs needed for a larger project. The results of previous pilot studies serve as justification for further exploration of the research question.

Interest in the research question is dependent on the individual researcher and potential funding agencies. In general, most clinical researchers select a research topic that is personally interesting as well as interesting to peers and the community.[6] Creating a research question that is interesting to specific funding agencies is essential to secure financial support for the study.

Novelty of the research question may be important to research foundations and scientific publications, because of a quest to produce and publish new information. Repeatability of study outcomes by different researchers is an essential part of the scientific process, so not all studies need be original, but should provide additional and novel information.[6] Previous studies may have had design flaws or been performed with too few subjects and these deficiencies can be rectified in a new study. The results of repeated clinical studies can support or call into question the data from previous studies.

Ethics in research is a complicated issue in both human and animal studies. While an in-depth discussion is beyond the scope of this paper, most TCVM practitioners should be able to determine whether a potential research question falls within currently accepted ethical boundaries. One of the most common ethical issues to be considered is the use of placebo treatments or withholding treatments. In cases where there is no accepted treatment, a placebo may be acceptable. In cases of severe illness with a known effective treatment, withholding the treatment is clearly unethical. While this may initially seem straightforward, TCVM practitioners are often faced with situations where it may be unclear whether withholding a conventional treatment would be acceptable. In these cases, consultation with peers, an ethics committee or institutional review board is a viable option to ensure the preservation of ethical standards. If funding is sought, the funding source will likely review the protocol for ethical criteria.

Relevance is essential.[6] A good clinical research question must be answerable and have clinical relevance. Even with a good clinical question, a poorly planned or executed study will yield results that are not relevant or reliable. Statistically significant differences found in a study may not be adequate enough to be clinically relevant or useful in clinical practice.

The PICOT format can be helpful to ensure all pertinent aspects of the research question, hypothesis and objectives have been included.[6] PICOT stands for: 1) population (patients), 2) intervention, 3) comparison

group, 4) outcomes of interest and 5) time. The population is the specific group of patients to be evaluated. The intervention is what will be performed on (or administered to) patients. The comparison group is to whom the experimental group will be compared (e.g. the control group). Outcomes of interest include what are intended to be affected, improved, measured and accomplished. The time is the time period(s) during the study when the outcomes will be assessed.

Once the primary clinical research question has been established, the hypothesis and objective(s) of the clinical study can then be determined. The hypothesis should explain the expected changes in the outcome(s) and is what will be statistically tested.[6] Traditionally the hypothesis is composed of 2 contrasting statements, the null and alternative hypotheses. The null hypothesis (N_0) states the outcome the researcher is predicting does not occur (e.g. no significant effect from the intervention). The researcher is actually hoping to reject the null hypothesis through statistical tests. The alternative hypothesis (H_1) states what outcome the researcher is predicting to occur (e.g. a significant effect from the intervention). The predicted outcome is thus what is "tested" with the clinical trial and forms the basis for the statistical analysis. The null hypothesis must be disproved in order for the alternative hypothesis to be accepted. If a question does not readily translate into a hypothesis, it may be too broad or complicated and may need to be more specific. Some questions may be descriptive in nature and will not translate into a hypothesis. Questions about prevalence of disease are an example. To answer these questions, one can use other study designs (e.g. observational studies) that provide data that may then generate a different research question and hypotheses for further investigation.

The primary objective of a study is a statement about how the study will answer the research question.[6] Objectives should include key features of the research question, such as the treatments to be used, the target study population and the expected outcome(s). The hypothesis can be similar to the objective, but with one key difference. The hypothesis may change during the design process, but the objectives usually remain established from the start.

An example of a TCVM research question, hypothesis and objective is as follows:

1. Question: Is the Chinese herbal medicine *Xiao Ying San* an effective alternative to L-thyroxine supplementation for the treatment of canine hypothyroidism?
2. Null Hypothesis: The treatment of hypothyroid dogs with *Xiao Ying San* for 8 weeks will result in no significant improvement of clinical signs, serum free thyroxine (FT_4) and serum thyroid-stimulating hormone (TSH) levels, compared to baseline levels and be ineffective compared to L-thyroxine supplementation.
3. Alternative Hypothesis: The treatment of hypothyroid dogs with *Xiao Ying San* for 8 weeks will result in significant improvements of clinical signs, serum free thyroxine (FT_4) and serum thyroid -stimulating hormone (TSH) levels, compared to baseline levels and is an effective alternative to L-thyroxine supplementation.
4. Objective: The primary objective of the study is to evaluate the improvement in clinical signs, increase of total serum free thyroxine (FT_4) levels and reduction of serum thyroid stimulating hormone (TSH) levels, after 8 weeks of *Xiao Ying San* administration, as compared to pre-treatment measures and 8 weeks of L-thyroxine supplementation in hypothyroid dogs.

Creation of the hypothesis and objective may seem simple at first, but in TCVM studies, problems arise because different AP and Chinese herbal medicine treatments are required for different TCVM patterns associated with a conventional diagnosis, such as canine hypothyroidism. Canine hypothyroidism can be due to Liver *Qi* Stagnation, *Yang Qi* Deficiency and *Qi-Yin* Deficiency.[7] *Xiao Ying San* is contraindicated in animals with *Yin* Deficiency.[8] To accurately evaluate the effectiveness of *Xiao Ying San* and do no harm to the animal, the research question must be modified to state: "Is the Chinese herbal medicine *Xiao Ying San* an effective treatment for canine hypothyroidism associated with *Yang/Qi* Deficiency?" and the hypotheses and objectives must then be modified accordingly. The criteria for diagnosis of *Yang/Qi* Deficiency and evidence for clinical improvement on the TCVM examination and clinical laboratory tests must be clearly outlined in the outcome measures portion of the study design discussed below.

Adequate Animal Numbers

Adequate animal numbers is essential for meaningful statistical analysis and reliability of results.[4] Patient recruitment will partly depend on the prevalence of the problem to be studied. Conditions with low prevalence may require a long study period to enlist a sufficient number of study subjects. Besides the frequency of occurrence of the specific disease, another potential obstacle is client consent. The caretakers of animals to be included in a study must clearly understand that they will not have a choice of treatment (e.g. the animals will be randomly assigned to a group). Animals may be more difficult to recruit, when caretakers are not allowed the freedom to choose a treatment. Recruitment problems can be overcome with proper client communications and having all clients sign a release form stating their understanding and acceptance of the terms. A financial incentive may also be necessary for recruitment. In pilot studies the number of animals is determined by the researcher, based on their experience and personal judgment, and are usually small (e.g. 10-20 animals/group). The results of pilot studies are not

considered EBM, as larger clinical trials by more than 1 research center will then be needed for EBM. In larger clinical trials, the number of animals needed to achieve significance between groups can be determined by preliminary data from a pilot study or a statistical power analysis.

Inclusion and Exclusion Criteria

Another important aspect of the clinical trial design is the process of determining specific inclusion and exclusion criteria for subjects.[4] These criteria will identify the target study population (e.g. hypothyroid dogs). The inclusion criteria will specify which dogs are to be included in the study. The exclusion criteria will be what factors will exclude them from the study, such as co-morbidities that might affect the results.

Since it not possible to recruit every possible animal that fits the criteria, a sample population can be studied based on the inclusion and exclusion recruitment criteria. A sample population is a small number of animals selected from a larger population that is then used to make estimations or predictions about specific traits in the larger population.[4] The sample population must be appropriately representative of the study population for the estimations to be accurate. If the sampling process is flawed, bias will be introduced and the conclusions will not be generalizable to the study population. Recruiting a sufficiently large number of animals and using inclusion and exclusion criteria to narrowly define the target study population will increase the possibility that the sample population will be representative of the target population.[4] Using specific selection criteria will ensure less variation in the sample population, especially when patients are to be recruited from multiple sources. Reporting the selection criteria and process will also allow clinicians to apply the study findings to specific patients in the their own practice.

Inclusion and exclusion criteria define the target study population and can include age, weight, breed or diagnosis requirements. They may also include what diagnostic procedures will be used to establish the before and after clinical status of the patient. While not considered the standard approach, it is becoming more commonplace for researchers to include both conventional biomedical and traditional Chinese medicine (TCM) and TCVM diagnostic techniques, along with the previously discussed pattern diagnoses. Inclusion criteria for a study population of hypothyroid dogs for a TCVM study could include clinical signs of hypothyroidism and *Yang Qi* Deficiency, serum free (FT_4) levels below 0.8µg/dl and serum TSH levels above 40 µU/L. Including both the conventional and TCVM diagnoses will provide an integrated screening method of patient selection, when TCVM treatments are being studied.

It is important to ensure a clear and consistent approach to diagnosis, including standard questionnaires or other methods, and strive for consistency between practitioners in the process of diagnosis and treatment formulation.[2] Including both conventional and TCVM methods can offer opportunities to explore differences in patient responses, when they may have the same conventional diagnosis, but different TCVM pattern diagnosis or *vice versa*. There is an inherent variability between how TCVM practitioners evaluate patients, such as interpretation of tongue and pulse changes. Developing systematic and standardized approaches to the TCVM evaluation of patients will keep the diagnostic and treatment formulation process consistent and transparent. Additionally, conventional diagnostic tests will offer a set of standardized data.

Control Groups

Animals are divided into experimental and control groups.[4,10] Having 1 or more adequate control groups is essential. The control group of animals can receive standard conventional treatments and comparison groups of animals can receive various combinations of AP and/ or other TCVM treatments alone or combined with conventional treatments. If the control group receives an accepted conventional treatment, the treatment protocol needs to be appropriate and adequate for the condition, as per current literature and clinical trials.[9] The control group may also receive a placebo or sham treatment in which patients receive no treatment or all aspects of the treatment except the "active ingredient".[10,11] Placebos or sham procedures are inert or ineffective treatments that simulate the actual treatment. They are useful to reduce the non-specific positive or negative effects of treatment administration that may affect the outcome. The placebo or sham procedure must only produce minor effects (not enough to be therapeutic) to allow comparison of outcomes between an investigational treatment and no treatment. They also allow investigators to demonstrate that the investigational treatment offers effects beyond that of the natural course of disease. As previously discussed, having a control group that receives a placebo, sham procedure or no treatment must be ethical. In TCVM clinical RCT, the control group often receives the conventional standard of care.

Randomization to Eliminate Selection Bias

Randomization of the allocation of patients to treatment and control groups is essential to eliminate selection bias and minimize the effect of known and unknown confounding variables.[4,10] Since it is almost impossible to anticipate and control every variable and difference in a test subject population, randomization can evenly distribute any potential confounding variables among all the groups, thus cancelling their effect on the final outcomes. Free research randomizer programs are available to randomly assign patients to treatment and control groups.

Outcome Measures and Criteria

Another important step is the determination of outcome measures, the parameters evaluated to demonstrate an effect.[4,10] Effect is the overall change

from a baseline value (e.g. a 50% reduction in pain) and is independent of other factors (e.g. the control results). The effects are changes that occur in any circumstance, not just in comparison to a control group. The effects to be measured will depend on the condition being investigated, though they generally are a combination of objective and subjective parameters. Developing objective outcome criteria is a challenge in clinical research. Clinicopathological test results are objective, but some treatments may be effective (e.g. result in improvement of clinical signs) and not change clinicopathological tests. For clinical examination findings, a scale of numbers with specific well-defined criteria (e.g. criteria for the comparative pain scale from 1-10) can help establish some objective guidelines and provide data that can be statistically analyzed.

When determining what outcome measures to monitor during a trial, it is important to determine how much effect is expected, otherwise known as the power of the study.[4,9,10] This can be estimated from preliminary data of a pilot study. In some situations the demonstration of a large effect can be accomplished with a small sample number. In AP and other TCVM studies, when the effect may be small, large sample numbers are needed. This is especially true when using sham AP techniques as the control or comparison group.[10,11] Sham AP techniques have been shown to have up to 50% of the effect of true treatment instead of the expected 30% associated with a pharmaceutical placebo.[9] This can lead to vast differences in sample size calculations. Acupuncture treatment is commonly expected to yield a 70% treatment effect, so if the difference in treatment effect between AP and sham AP is only 20% and not 40%, this will vastly change the number of samples needed to demonstrate the difference.

In some situations the ultimate effect of AP and other TCVM treatments may be difficult to measure in the short term. The final clinical outcome of TCVM treatments may not be short-term changes in biomarkers, but in the over-all course of the disease. It may be better to define the TCVM effect as the ability to produce subtle overall changes in the internal environment to allow the body's natural healing to take place, rather than producing dramatic changes in a handful of enzymes or specific tissues.[12] This may require a long follow-up period. In general, follow-up monitoring periods should be a minimum of 3 months and preferably a year.[9]

Blinding to Eliminate Bias

Blinding (masking) is needed to eliminate investigator, evaluator and client bias.[4,9,10] The currently accepted double-blind design in clinical RCT involves blinding the client (patient in human medicine) and either the practitioner or an evaluator or both. The TCVM practitioner can be blinded in studies of Chinese herbal medicines, as treatments and placebos can be prepared to appear the same by a 3rd party and unknown until after the evaluation process. Blinding presents problems in AP and *Tui-na* research, since an acupuncturist will know if a sham treatment has been administered. A 3rd party must be used for the outcomes evaluation in AP and *Tui-na* research.[9] If clients' opinions are included as an outcome measure, then client blindedness is important, as bias for or against AP or other TCVM treatments can affect outcomes. Designing objective methods of outcomes assessment from blinded evaluators is the best option. When the acupuncturist is not responsible for evaluating treatment outcomes, systematic bias associated with lack of blinding is minimized.

Statistical Analysis

A well stated null hypothesis (or alternative hypothesis) can help justify the statistical analysis used to compare outcomes between experimental groups.[10] For clinical RCT, statistical comparison of differences between groups is essential. Probability values (*p*-value) indicate statistical difference. The lower the *p*-value, the stronger the evidence; *p*-values below (<) 0.05 are usually considered significant, while *p*-values <0.01 are considered very significant. Statistical significance may not always equate to clinical relevance as previously discussed.

Baseline comparisons between study groups are also important. The purpose of comparing groups before treatment initiation is to demonstrate that there is no group differences (beyond random chance) and selection bias has not been introduced. This is especially important if the study sample numbers are low and attrition is high. If the initial comparison shows no significant differences in the groups, there should be no significant bias introduced if several participants discontinue participation.

The appropriate statistical test will depend on the study design. Statistical methods to analyze the data are decided during the planning phase. Though it may be tempting, doing *post-hoc* analysis is generally discouraged.[4] Statistical analysis software programs are available, but novice researchers are encouraged to consult with a statistician to ensure the appropriate statistical tests are selected and the calculations are correctly performed.

Special Considerations for Acupuncture Research

A placebo should confer no specific treatment effect.[4] Sham AP procedures do not qualify as placebos, because they can confer significant physiologic effects, much greater than pharmacologic placebos.[9] There are many varieties of sham AP that include using: 1) acupoints different from the prescribed acupoints for the diagnosis, 2) non-acupoints, 3) non-penetrating devices at acupoints and 4) other pseudo-interventions such as inactivated lasers at acupoints. One AP sham method in humans involves using non-acupoints with needle penetration less than 4 mm.[10] Another sham AP technique is to tap toothpicks or other blunt devices on the skin at acupoints, but not penetrate the skin.[10]

Sham procedures in human clinical RCT are designed to eliminate bias by preventing the patient from knowing whether they received an active treatment or not, especially when the patient's opinion of the treatment outcome is requested. In both humans and animals, sham procedures may be used to reduce the chance of unknown factors affecting the outcome. When doing research on AP treatments, it may be difficult to determine if the difference seen in the groups is due to the AP needles in acupoints or other non-specific and effects associated with the AP treatment.[10]

Sham AP is done to control for any possible "non-specific" effects that may occur during treatment. These are the non-therapeutic effects that may occur from interaction of patient and practitioner, environmental effects or another unknown effect of AP administration.[10] The "specific" effect that is being investigated is the one that results in a change in the patient's clinical condition. Unfortunately, sham AP techniques appear to have more non-specific effects than pharmacological or other physical placebos, making it difficult to demonstrate significant differences between sham and true AP.[10-12] Studies that compare the effect of AP to standard conventional treatments often show greater differences, than those comparing AP to sham AP.[11] In one study, the effects from sham AP (using a non-penetrating device) showed greater physiologic effects (less pain and severity of symptoms) than an inert pill.[12] Therefore, it is not appropriate to label AP sham procedures as "placebo", since all AP sham procedures have been shown to produce significant non-placebo and non-specific effects.[9]

What constitutes "non-specific" and "specific" effects depends on the research question being asked. If the hypothesis is that a quiet environment results in a more effective treatment, the research question will focus on the specific effect of the environment and not on the non-specific effects of the acupoint prescription.[13] Some practitioners and researchers suggest that the "non -specific" effects are not really "non-specific". These researchers argue that the attempt to break down AP treatment (and all the associated ritual and interaction that accompanies it) into constituent parts is impossible, due to the complex nature of the interplay between the patient, practitioner, needle and environment.[13] Some would also say that it is the interplay that produces the results and removal or alteration of one of these aspects would generate altered results.[13] This is similar to the idea of the sum being greater than the parts.

Another source of controversy in AP clinical trial design is whether to use an individualized or pre-determined acupoint prescription protocol.[13] Some authors contend that "there is little evidence that individualized treatment strategies are superior to more standardized approaches….equally there is no evidence to suggest that there is any superiority in terms of effect between a Westernized and a traditional Chinese medicine format."[13] Standardized point selections help to increase internal validity (the ability to reproduce results in a given set of circumstances). The goal is to control all other variables except the independent variable (e.g. treatment). This makes it easier to establish a stronger relationship between cause (treatment) and effect (outcome). If the acupoint selection is tightly controlled, the results can be more easily compared.

The use of standard protocols (e.g. 1 treatment for every case with a specific conventional diagnosis) does not reflect the way TCVM is practiced by most veterinarians. The individualization of acupoint prescription to each patient's specific presentation and underlying TCVM patterns is a hallmark of many AP treatments. This may produce better outcomes, but individualized treatments cannot be easily compared, especially if other treatments like Chinese herbal medicine, *Tui-na* and Food therapy are also included. Negative results with standardized protocols could be due to: 1) lack of specificity of the treatment for the patient (e.g. treating for a conventional diagnosis and not the TCVM pattern), 2) over-simplification of the treatment or 3) patient unresponsiveness to AP.[2]

When deciding on acupoint selection, one does not have to simply choose either standardization or individualization, as there may be a spectrum of options. One option is a set of standard points with optional points that can be chosen by the practitioner for the specific case. Manualization is a technique in which a set of diagnostic guidelines and treatment designs are predetermined.[14] This allows freedom to individualize within a predetermined framework. Further allowance for individualization is to have no specified acupoint protocol except to exclude other treatments or have no acupoint protocol or exclusions of other treatment modalities. In general, when determining a protocol, simplification may be useful for "simple" problems (e.g. PC-6 for nausea or LI-4 for dental pain). When investigating more complicated clinical issues, however more complicated treatments protocols are usually necessary.

One of the difficulties associated with AP research is the ability to guarantee the adequacy of the treatment being tested.[9] In pharmaceutical research, a dose of a chemical can be rationalized on the basis of pharmacokinetic data. Validity of acupoint prescriptions is not so easily proven, thus making it even more difficult to apply current standards of clinical research methodology. One proposed option of ensuring an appropriate treatment protocol is the Birch Relevant and Irrelevant Treatment Selection method.[9] This involves literature review, practitioner survey and consultation with expert panels to develop a consensus on appropriate acupoints for a particular condition or presentation, thus providing justification for the tested treatment beyond the investigator's personal opinion.

Internal *versus* External Validity

How a trial is designed will ultimately dictate its internal and external validity.[4,10] Internal validity is the ability of the trial results to demonstrate the treatment

effect within a tightly defined population. The inclusion criteria of these trials are very strict and the procedures are standardized. This eliminates bias and confounding results associated with unknown factors. The results should also be reproducible in the same set of circumstances, by other researchers. External validity is the ability of the trial results to translate to the general population. In these trials, the sampling criteria are much less rigorous to reflect the patient pool of the average practitioner. The treatment procedures may also be less standardized to reflect daily practice. These trials are more representative of actual clinical practice. However the results are more difficult to interpret because of the greater potential for confounding results and lack of reproducibility.

An example of this concept is efficacy and effectiveness trials.[11] Efficacy is the comparison of the treatment effects to placebo effects, thus demonstrating the difference between an intervention and a control in an ideal set of circumstances. Efficacy trials demonstrate therapeutic effects in an ideal setting with a homogenous sample and standard protocol of treatment with a "sham" or placebo comparison group. As discussed, the results should show a therapeutic effect from the AP needle and not that of some other component. This has good internal validity, but does not represent the general population that may ultimately receive the treatment or how the treatment may be administered. Effectiveness trials demonstrate therapeutic effects in actual clinical settings, as a practitioner would normally administer treatment and therefore are a more realistic approach. There are few exclusion criteria and no sham control/comparison group. With high external validity, these trial results may be more easily translated to daily practice, but may be marred by unknown confounding variables and bias and thus not be considered EBM. Therefore if TCVM practitioners are performing clinical research that may be time consuming and costly, studies must be well designed so the results are considered EBM and the effectiveness of TCVM treatments critically evaluated. Consultation with others experienced in designing clinical trials may be necessary to ensure a research plan that will produce valid results. Groups supporting TCVM research may offer guidance and usually have an outline of information to include when preparing a grant.[15]

Reporting and Publication

Once the data have been analyzed and conclusions drawn, the final step is reporting the information. The Standards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA) were designed to provide researchers with guidelines to improve the reporting of clinical trial methods and results, but are also useful when designing a clinical trial.[16] These instructions are an extension of the Consolidated Standards of Reporting Trials (CONSORT), which are guidelines for reporting all types of clinical RCT. The STRICTA standards provide additional information specific to AP trials, including many of the CONSORT guidelines for non-pharmacological treatments and pragmatic trials. The purpose of these recommendations is to ensure researchers are reporting information fully to improve transparency, reduce ambiguity and allow for successful replication of the study.[16,17] The guidelines are applicable across a wide range of trial designs in recognition of the variety of study types and level of individualization of treatments.

Getting Started

One may begin planning a clinical trial by carefully reviewing STRICTA and TCVM Clinical Trial Guidelines to ensure that the design contains all elements necessary to result in a high quality publishable study that will contribute to the EBM knowledge base.[16,17] The steps involved in conceiving, designing and conducting clinical research are outlined in Figure 2. The initial steps are the most critical. If the study is designed poorly, the results may be non-publishable and a waste of time and money. However, anyone who endeavors to conduct rigorous research must also realize that no study design is perfect. Compromises may have to be made due to money or time restrictions. Bias may be introduced because of an inability to recruit an adequately diverse study sample. Researchers must strive to follow the above methods of randomization, blinding and controlling to minimize factors that can produce inaccurate or questionable results. When compromises must be made, an alteration in the study plan may be necessary. Describing the limitations of a study identifies the flaws and potential alterations that could affect the results. Even when the results may not

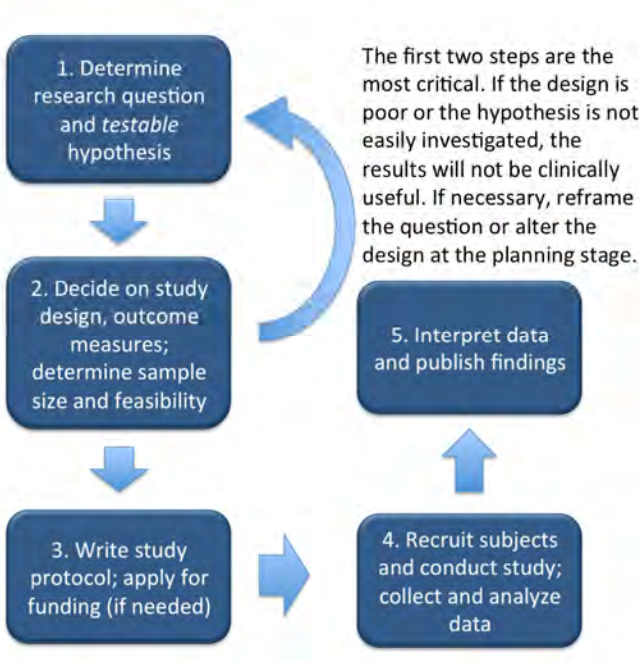

Figure 2: Process for designing clinical trials

be generalizable, they may provide an interesting starting point for another researcher that may have resources to conduct a more rigorously designed trial.

An important part of conducting effective clinical research is to establish a competent team to design and execute the study. This may involve consulting or corroborating with fellow clinicians and researchers to perform multi-center clinical trials or serve as blinded evaluators of outcomes. Continuing education meetings offer opportunities to network and find other TCVM veterinarians with similar clinical research interests. If there are particular areas in which one does not feel adequately skilled (e.g. study design, statistics or scientific writing), consulting experts in these areas may be needed. A common problem for inexperienced clinical researchers is to start with a project that is too large or time consuming to ever complete. A pilot study that can be performed with less time and expense might be the best way to start. The data from the pilot study may provide justification for further studies or new research questions. With a commitment to learn how to effectively design, execute and publish the results of high quality clinical studies, TCVM practitioners can elevate AP and other TCVM practices from testimonials of success to EBM status.

REFERENCES

1. Baker S, Baker G. Pragmatic versus philosophical evaluation of complementary and alternative medicine. J Amer Vet Med Assoc 2012; 241 (9):1146.
2. Schnyer R, Birch S, MacPherson H. Acupuncture practice as the foundation for clinical evaluation. Acupuncture Research, MacPherson H, Hammerschlag R, Lewith F, Schnyer R (ed). New York, NY: Elsevier 2008:153-180.
3. Palmquist R. Dialogue on homeopathy. J Amer Vet Med Assoc 2012; 242(3):307.
4. Holmes M, Cockcroft P. Handbook of Veterinary Clinical Research. Oxford, UK: Blackwell Publishing Ltd. 2008:10,21-22,99,107.
5. WHO, Guidelines for Clinical Research on Acupuncture, World Health Organization Western Pacific Region, Manila, Philippines, 1995. http://apps.who.int/medicinedocs/en/d/Jh2948e/
6. Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Research questions, hypotheses and objective. Canadian J of Surgery 2010; 53(4):278-281. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2912019/
7. Xie HS, Preast V. Xie's Veterinary Acupuncture. Ames, IA: Blackwell Publishing 2007:306-307.
8. Xie HS, Preast V. Xie's Chinese Veterinary Herbology. Ames, IA: Wiley-Blackwell Publishing 2010:365.
9. Birch S. Clinical research on acupuncture: Part 2. Controlled clinical trials, an overview of their methods. J Alt Comp Med 2004; 10(3):481-498.
10. Lin J, Chen C, Huang Y, Chen Y. How to design the control group in randomized controlled trials of acupuncture? Evid Based Comp Alt Med 2012; Article ID 875284:1-7. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3398668/
11. Witt C. Clinical research on acupuncture-concepts and guidance on efficacy and effectiveness research. Chin J Integrative Med 2011; 17(3):166-172.
12. Kaptchuk T, Stason W, Davis R et al. Sham device versus inert pill: randomized controlled trial of two placebo treatments. British Medical J 2006; 332(7538):391-97. http://www.bmj.com/content/332/7538/391
13. White P, Linde K, Schnyer R. Investigating the components of acupuncture treatment. Acupuncture Research. MacPherson H, Hammerschlag R, Lewith F, Schnyer R (ed). New York, NY: Elsevier 2008:133-152.
14. Schnyer R, Allen J. Bridging the gap in complementary and alternative medicine research: manualization as a means of promoting standardization and flexibility of treatment in clinical trials of acupuncture. J Alt Comp Med. 2002; 8(5):623-634.
15. TCVM Research Grant Applications. http://www.aatcvm.org/index.php/en/research-grants
16. MacPherson H, Altman D, Hammerschlag R et al. Revised standards for reporting interventions in clinical trials of acupuncture (STRICTA): Extending the CONSORT statement. PLoS Medicine. 2010; 7 (6):e1000261. http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000261
17. TCVM Clinical Trial Guidelines. Amer J of Trad Chinese Vet Med 2014; 9(1): http://www.ajtcvm.org/tcvm-clinical-trial-guidelines