COMMENTARY

FABLES OR FOIBLES: INHERENT PROBLEMS WITH RCTS

INTRODUCTION

For 50 years, the accepted standard by which the usefulness of a therapeutic treatment is judged has been the randomized controlled trial (RCT), building from Hippocrates' premise 2000 years ago that experience combined with reason was the therapy of choice for patients; that is, any treatment plan should both seem reasonable in theory and then be tested experimentally. Assuming that threats to both internal and external validity could be ruled out, the RCT became what is commonly regarded as the highest quality of clinical outcome study that could be mounted to allow inferences about cause and effect relationships to be drawn. The thinking was that the more rigorous and fastidious the design, the more credibility could be attached to the conclusions drawn from the outcomes of the study and the more likely the intervention was thought to have brought about those outcomes.[1] One of the strongest proponents of the RCT through the 1950s and 1960s was the British epidemiologist Archie Cochrane, who held that this type of experimental approach was essential for upgrading the quality of medical evidence.[2] In common hierarchical schemes of clinical experimental design, the RCT has been ranked the highest in rigor, as shown in Table 1.[3] Even greater rigor has been presumed to occur with the statistical combination and weighting of the results of multiple RCTs in a meta-analysis to generate a more conclusive estimate of effect size.[4,5]

From the point of view of clinical practice, however, especially in areas in which physical treatments are applied, the principles of fastidious treatments and blinding begin to wear thin and in a few recent examples regarding spinal manipulation, appear to have fallen apart completely. This difficulty is by no means confined to physical treatments, as the literature pertaining to the use of medications has also suggested that the inexperienced use and/or uncritical acceptance of the results of RCTs can lead to confusion. In this presentation, a few representative samples will be introduced as 7 case studies, which ironically would be ranked among the lowest in experimental rigor by the aforementioned hierarchy of clinical evidence.[3]

1. REDUCTION OF META-ANALYSES TO SUBJECTIVE VALUE SCALES

In their efforts to compare 2 different preparations of heparin for their respective abilities to prevent postoperative thrombosis, Juni et al[6] have demonstrated that diametrically opposing results can be obtained in different meta-analyses, depending on which of 25 scales is used to distinguish between high-quality and low-quality RCTs. The root of the problem is evident from the variability of weights given to 3 prominent features of RCTs (randomization, blinding, and withdrawals), as shown in Table 2 by the 25 studies which have compared the 2 therapeutic agents. In 1 study, a third of the total weighting of the quality of the trial is afforded to both randomization and blinding, whereas in another study cited in the article, none of the quality scoring is derived from these 2 features. Widely skewed intermediate values for the 3 aspects of RCTs under discussion are apparent from the 23 other scales presented. The astute reader will immediately suspect that sharply conflicting conclusions might be drawn from these different studies, and these are amply borne out by the statistical plots shown in Figure 1. Here, each of the meta-analyses listed resolve the 17 studies they have reviewed into high-quality and low-quality strata, based on each of their scoring systems. It can be seen that 10 of the studies selected show a statistically superior effect of 1 heparin preparation, low-molecular weight heparin (LMWH), over the other but only for the low-quality studies. Seven other studies reveal precisely the opposite effect, in which the high-quality but not the low-quality studies display a statistically significant superiority of LMWH.

Depending on which scale is used, therefore, it is possible to either demonstrate or refute the clinical superiority of one clinical treatment over another. In this manner, therefore, all the rigor and labor-intensive elements of the RCT and its interpretation by the meta-analysis are simply reduced to the subjective and undoubtedly capricious human element of value judgment through the arbitrary assignment of numbers in the weighting of experimental quality.[6]

2. OCCULT "SALAMI" PUBLICATIONS

At times, authors of studies have been known to present their data in more than 1 forum in the scientific literature, resulting in what has become referred to as mass-produced or "salami" publications. Because the exclusivity of such data is unknown, it will be oversampled by the unsuspecting author of a meta-analysis or systematic literature review and

Table 1. *Hierarchy of experimental designs*[3]

> 1. Control group outcomes study (including RCTs).
> 2. Single-subject experiment, replicated single-subject experiments.
> 3. Single-group outcome study.
> 4. Systematic case study.
> 5. Anecdotal case report.

Designs are presented in descending order of rigor.
RCT, Randomized control trial.

Table 2. *Scoring the quality of clinical trials: Variability of weights given to key methodological domains*[6]

Scale	No. items	Randomi-zation	Blinding	With-drawals
Andrew, 1984	11	9.1	9.1	9.1
Beekerman et al, 1992	24	4.0	12.0	16.0
Brown, 1991	6	14.3	4.8	0
Chalmers et al, 1990	3	33.3	33.3	33.3
Chalmers et al, 1981	30	13.0	26.0	7.0
Cho and Bero, 1994	24	14.3	8.2	8.2
Colditz et al, 1989	7	28.6	0	14.3
Deltsky et al, 1992	14	20.0	6.7	0
Evans and Pollock, 1985	33	3.0	4.0	11.0
Goodman et al, 1994	34	2.9	2.9	5.9
Gotzsche, 1989	16	6.3	12.5	12.5
Imperiale and McCullough, 1990	5	0	0	0
Jadad et al, 1996	3	40.0	40.0	20.0
Jonas et al, 1993	18	11.1	11.1	5.6
Kjeijnen et al, 1991	7	20.0	20.0	0
Koes et al, 1991	17	4.0	20.0	12.0
Levine, 1991	29	2.5	2.5	3.1
Linde et al, 1997	7	28.6	28.6	28.6
Nurmohamed et al, 1992	8	12.5	12.5	12.5
Onghena and Van Houdenhove, 1992	10	5.0	10.0	5.0
Poynard, 1988	14	7.7	23.1	15.4
Reisch et al, 1989	34	5.9	5.9	2.9
Smith et al, 1992	8	0	25.0	12.5
Spitzer et al, 1990	32	3.1	3.1	9.4
ter Riet, 1990	18	12.0	15.0	5.0

thus will be given more weight than it merits. One such instance has been reported in the evaluation of nonsteroidal anti-inflammatory drugs (NSAIDs) in treating rheumatoid arthritis, in which 44 publications of 31 clinical trials were found to result in an oversampling of at least 18%. Twenty of these studies were published in 2 different sources, 10 studies were published in 3 different sources, and 1 study was published in 5 different sources. The fact that these data were published elsewhere was not noted in 32 of the 44 articles. Even more unsettling is the finding that in about half of the articles, the first author and total number of authors were different, and there appeared to be important discrepancies between versions of the same trial.[7]

Further evidence is shown in studies of risperidone, an antipsychotic agent. In this instance, 20 articles plus unpublished reports actually represented only 9 trials.[8] Finally, a report from Tramer et al[9] has described how 84 trials involving 11,980 patients using ondansetron for postoperative emesis resulted from only 70 trials employing 8645 patients. It was believed that the duplicate data led to a 23% overestimation of the efficacy of ondansetron.

Here, it is clear that the "one man, one vote" principle of systematic data review has been violated, such that clinical observations derived from the RCTs of certain authors have been given excessive credibility. Care must be taken to ensure that the data incorporated into an analysis of the effect of a particular treatment in an RCT are scored only once, a highly formidable if not impossible task.

3. MANIPULATION OF EXPERIMENTAL RESULTS

One of the more startling analyses of RCTs has been presented by Johansen and Gotzsche,[10] who reviewed a meta-analysis comparing fluconazole and amphotericin B, 2 antifungal agents. To begin, in 3 large trials comprising 43% of the patients identified for meta-analysis, the results from amphotericin B were combined with the results for nystatin, known to be an ineffective drug for fungal infections. Worse, 79% of the patients in these trials were randomized to receive amphotericin orally, which is perplexing and disturbing, since amphotericin B is known to be poorly absorbed and is normally administered intravenously.

When questioned more closely about the sources of their data, 12 of the 15 authors were found to be less than fully compliant, with 1 suggesting that the trial was "old" and that

the primary data resided with the drug manufacturer, another claiming that sufficient time was lacking to respond, and a third professing the lack of access to the database because of a change of affiliation. The final surprise, which appeared to belie the validity of this entire undertaking, was the fact that Pfizer, the manufacturer of the superior drug, provided employment to 12 of the 15 authors in studies involving 92% of the total number of patients evaluated. It would appear that the intention all along was to manipulate the trials to favor the successful pharmaceutical product.

4. FLAWED RCT NO. 1: MISREPRESENTATION OF THERAPIES AND OVERGENERALIZATION OF RESULTS

A widely publicized study by Cherkin et al[11], which appeared in *The New England Journal of Medicine*, represents an inaccurate depiction of the 3 treatments which are presumably compared (chiropractic care, physical therapy, and medical intervention). These are reduced, respectively, to a single side-posture manipulation, the McKenzie method, and an education booklet. While these applications are certainly indicated in a fastidious design, there is no justification for the authors, who found little difference in outcomes between the 3 interventions with greater costs
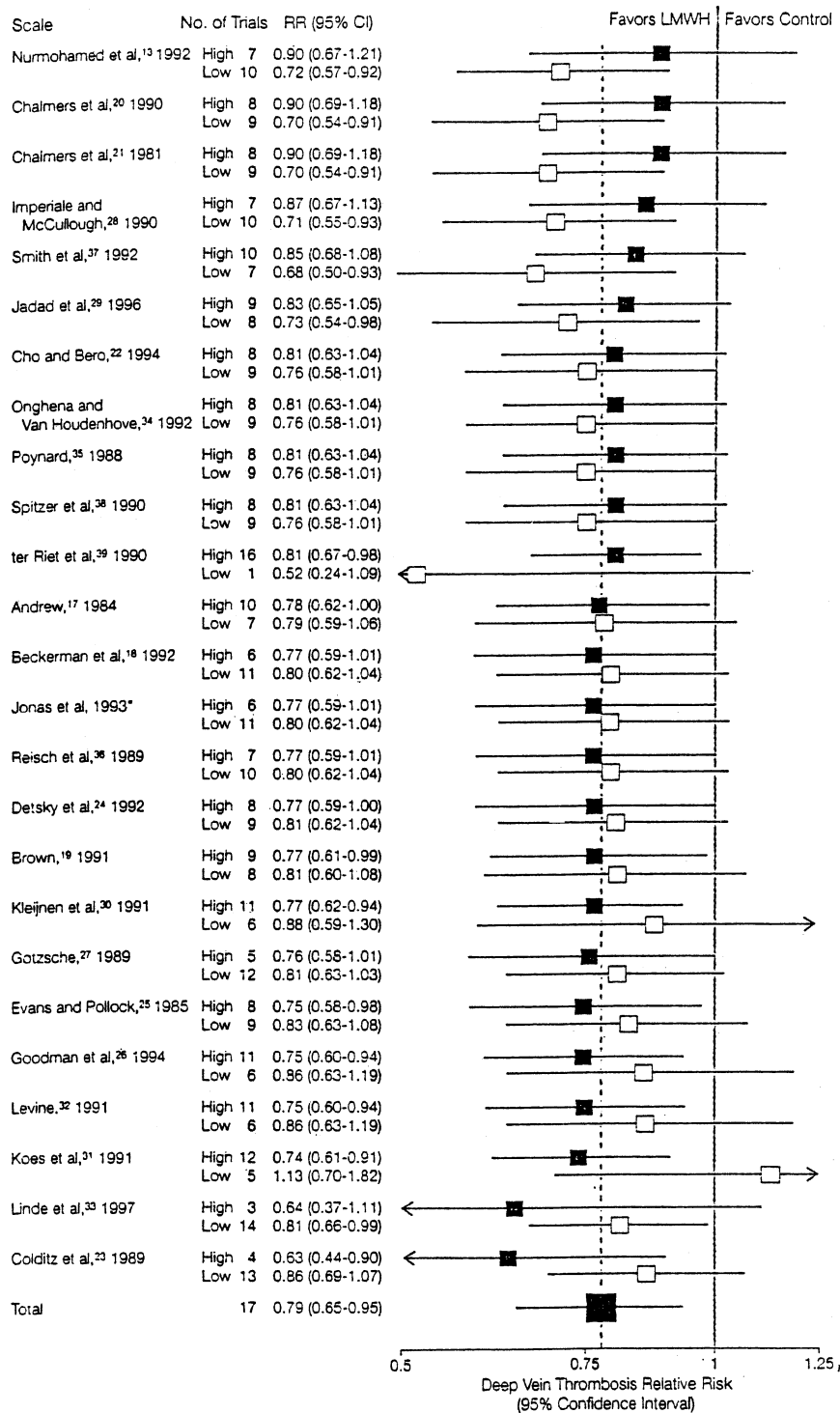
Scale		No. of Trials	RR (95% CI)
Nurmohamed et al,[13] 1992	High	7	0.90 (0.67-1.21)
	Low	10	0.72 (0.57-0.92)
Chalmers et al,[20] 1990	High	8	0.90 (0.69-1.18)
	Low	9	0.70 (0.54-0.91)
Chalmers et al,[21] 1981	High	8	0.90 (0.69-1.18)
	Low	9	0.70 (0.54-0.91)
Imperiale and McCullough,[28] 1990	High	7	0.87 (0.67-1.13)
	Low	10	0.71 (0.55-0.93)
Smith et al,[37] 1992	High	10	0.85 (0.68-1.08)
	Low	7	0.68 (0.50-0.93)
Jadad et al,[29] 1996	High	9	0.83 (0.65-1.05)
	Low	8	0.73 (0.54-0.98)
Cho and Bero,[22] 1994	High	8	0.81 (0.63-1.04)
	Low	9	0.76 (0.58-1.01)
Onghena and Van Houdenhove,[34] 1992	High	8	0.81 (0.63-1.04)
	Low	9	0.76 (0.58-1.01)
Poynard,[35] 1988	High	8	0.81 (0.63-1.04)
	Low	9	0.76 (0.58-1.01)
Spitzer et al,[38] 1990	High	8	0.81 (0.63-1.04)
	Low	9	0.76 (0.58-1.01)
ter Riet et al,[39] 1990	High	16	0.81 (0.67-0.98)
	Low	1	0.52 (0.24-1.09)
Andrew,[17] 1984	High	10	0.78 (0.62-1.00)
	Low	7	0.79 (0.59-1.06)
Beckerman et al,[18] 1992	High	6	0.77 (0.59-1.01)
	Low	11	0.80 (0.62-1.04)
Jonas et al, 1993*	High	6	0.77 (0.59-1.01)
	Low	11	0.80 (0.62-1.04)
Reisch et al,[36] 1989	High	7	0.77 (0.59-1.01)
	Low	10	0.80 (0.62-1.04)
Detsky et al,[24] 1992	High	8	0.77 (0.59-1.00)
	Low	9	0.81 (0.62-1.04)
Brown,[19] 1991	High	9	0.77 (0.61-0.99)
	Low	8	0.81 (0.60-1.08)
Kleijnen et al,[30] 1991	High	11	0.77 (0.62-0.94)
	Low	6	0.88 (0.59-1.30)
Gotzsche,[27] 1989	High	5	0.76 (0.58-1.01)
	Low	12	0.81 (0.63-1.03)
Evans and Pollock,[25] 1985	High	8	0.75 (0.58-0.98)
	Low	9	0.83 (0.63-1.08)
Goodman et al,[26] 1994	High	11	0.75 (0.60-0.94)
	Low	6	0.96 (0.63-1.19)
Levine,[32] 1991	High	11	0.75 (0.60-0.94)
	Low	6	0.86 (0.63-1.19)
Koes et al,[31] 1991	High	12	0.74 (0.61-0.91)
	Low	5	1.13 (0.70-1.82)
Linde et al,[33] 1997	High	3	0.64 (0.37-1.11)
	Low	14	0.81 (0.66-0.99)
Colditz et al,[23] 1989	High	4	0.63 (0.44-0.90)
	Low	13	0.86 (0.69-1.07)
Total		17	0.79 (0.65-0.95)

Favors LMWH | Favors Control

0.5   0.75   1   1.25

Deep Vein Thrombosis Relative Risk
(95% Confidence Interval)

Fig 1. *Scoring the quality of clinical trials: Dividing trials into high-quality and low-quality strata. From Juni P, Witsch A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282(11):1054-1060. Copyrighted ©1999, American Medical Association. Used with permission.*

associated with the side-posture or McKenzie treatments, to then state as a conclusion: "Given the limited benefits and high costs, it seems unwise to refer all patients with low back pain for chiropractic or McKenzie therapy."

First, one must be aware that there are several chiropractic techniques applicable to the management of low back pain; among them are low-force (the Logan Basic or Sacro-Occipital) techniques, flexion-distraction, use of a drop ta-
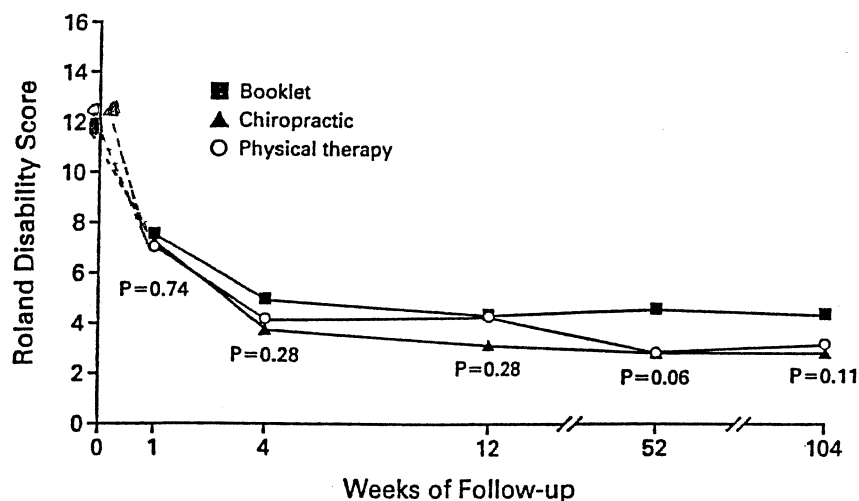
Fig 2. *Deficiencies of Cherkin study: Disability scores. From Cherkin DC, Deyo RA, Battle M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. New England Journal of Medicine 1998;339(15):1021-1029. Copyright © 1998, Massachusetts Medical Society. All rights reserved. Used with permission.*

ble, and traction. In this trial, only 1 high-velocity technique (side-posture) was applied, and it might not be equally effective for all patients. Furthermore, important ancillary procedures which are intrinsic to the chiropractic visit appear to have been denied to patients. In particular, extension exercises were forbidden, and patients were most likely not given any literature, even though these 2 options are considered to be parts of a customary chiropractic regimen for office visits. It appears that these 2 elements were permitted only in the other 2 arms of the trial. In short, the chiropractic treatment administered in this particular investigation appears to have been only a pale shadow of the actual therapy administered to patients in the real world. This would only add further irony to the inappropriate conclusion quoted from the authors above.

Additional problems with this trial surface with the examination of baseline characteristics regarding severity among the 3 groups tested, creating a bias in the outcomes. First, the percentage of patients who had prior chiropractic care for low back pain appears to be substantially lower for the chiropractic cohort (24%) than for the McKenzie and medical booklet groups (35% and 40%). This problem is only magnified by the authors' citation of another prominent investigation, noting that "the British study found the benefits of chiropractic to be most evident among patients who had previously been treated by chiropractors." Second, the chiropractic cohort indicates the highest percentage of patients who, because of low back pain and prior to their therapy, encountered more than 1 day of best rest (35% vs 24% and 22% for the McKenzie and medical booklet cohorts, respectively), more than 1 day of work lost (39% vs 41% and 30% for the McKenzie and medical booklet co-

horts, respectively], and more than a single day of restricted activity (72% vs 65% and 52% for the McKenzie and medical booklet cohorts, respectively).

Figure 2 depicts the actual outcomes of the 3 compared applications in the study through 12 weeks of follow-up. Curiously, the outcomes in the figure between weeks 0 and 1 were not shown in the original article but indeed represent the bulk of improvement in the 3 patient cohorts (the change from the baseline scores to those observed at 1 week of follow-up is depicted by the dotted line). In this chart, there does appear to be a tendency for the "chiropractic" group to show greater improvement at most of the weeks of follow-up evaluated, although statistically this is not borne out. Even with these abbreviated interventions, larger group sizes in this trial might have overcome what could have been a type II error and delivered statistically robust differences in both outcomes and baseline characteristics shown above. These are but a few of the deficiencies of this particular study, which have been outlined extensively elsewhere.[12-14] In summary, this study is a poor representation of therapies which have been successfully applied to live patients in physicians' offices worldwide. The deficiencies in its design undercut its validity to the point of compromising the reliability of the study as a whole. Indeed, the Royal College of General Practitioners, in a recent systematic review of the literature designed to update guidelines issued by the government of the United Kingdom for the management of low back pain (which themselves conflict with the Cherkin et al[11] study by citing spinal manipulation as a treatment of choice for low back pain[15]), has concluded that this RCT under discussion neither adds to nor detracts

from the evidence base regarding appropriate interventions for low back pain.[16]

5. FLAWED RCT No. 2: IMPROPER SHAM PROCEDURE

An equally widely publicized study appearing in *The New England Journal of Medicine* purported to add further negative evidence to the efficacy of spinal manipulation, stating that "the addition of chiropractic spinal manipulation to usual medical care for four months had no effect on the control of childhood asthma." This statement was based on the failure of active and sham-manipulated patient groups aged 7 to 16 years in a clinical trial to be differentiated in terms of their outcomes in both quality of life and airway function. What is indisputable is that there were major improvements from baseline to follow-up observed in each of the groups.[17]

The problem arises when one considers what was actually done in the sham procedures. Prolonged applications to no less than 3 distinct anatomical areas (gluteal, scapular, and cranial) to the patient are described. Admittedly, these are not high-velocity contact procedures, but this evades the issue. Two pieces of evidence strongly suggest that simple contact with patients through sham procedures will produce a significant effect. The first indicates that with respect to the reflexive inhibition of the alpha-motoneuron pool in human subjects, sham and active manipulative procedures display little difference. This is to suggest that cutaneous receptors, muscle spindles, and joint mechanoreceptors individually or in concert are significantly affected by so-called sham procedures.[18] The second demonstrates that 2 groups of children, aged 4 to 8 and 9 to 16, display profound changes in pulmonary functions, attitude and behavior scores, and cortisol levels following massage, as compared to a noncontact control group.[19] Thus, it would appear that physical contact with the patient is sufficient to trigger a cascade of physiological changes, which seem to have been erroneously dismissed in the asthma study. What appears to have been underemphasized by both the authors and most readers of the asthma study is that chiropractic encompasses a broad range of both high-velocity and low-force techniques together with ancillary procedures, many of which have obviously been embedded in the sham procedures described. In its attempt to craft a fastidious design, this trial gives the impression of missing the forest for the trees by attempting to portray the essence of chiropractic care as the lack of differentiation between the sham and manipulated experimental groups.

6. FLAWED RCT No. 3: INCONSISTENCIES BETWEEN PILOT AND FULL-SCALE TRIAL AND SHAM PROCEDURES

Another recently published RCT would have appeared to replicate the problems with the asthma trial by invoking a

Table 3. *Pilot[20] and full-scale[21] Dysmenorrhea studies: Inconsistencies*

	SMT		LFM	
	Full*	Pilot	Full*	Pilot
1. Pain: VAS (cm)				
Pre-	4.28	5.87	3.80	6.00
Post-	3.27	3.78	3.00	5.19
2. KDPGF$_{2a}$ (pg/mL)				
Pre-	128.6	133.9	130.2	142.8
Post-	124.3	116.1	119.2	126.3

SMT, spinal manipulative therapy; *LFM*, low-force manipulation; *VAS,* Visual Analogue Scale.
  *Indicates outcome at end of menstrual cycle No. 2, the closest approximation to the single menstrual cycle evaluated in the pilot study.

contact sham procedure and then failing to find a significant difference in outcomes between sham and actively manipulated patient groups—this time in women complaining of primary dysmenorrhea.[20]

What is curious in this instance, however, is that the same authors did find significant differences between the 2 experimental groups in their own pilot study published previously.[21] This is plainly apparent in Table 3, in which both pain and prostaglandin (KDPGF2a) levels are seen to decrease significantly in the active spinal manipulative therapy as opposed to the sham low-force manipulation group in the pilot study, whereas no such pattern can be detected in the full-scale investigation. However, a closer examination of the data explains at least what appears to have happened regarding the scales. Pain baseline levels in the full-scale study can be observed to be virtually 1.5 to 2 units less than the corresponding values in the pilot study. Since the baseline values in the full-scale study are close to the expected final outcome levels, their accurate measurement is a moot point. The reason is that the qualifying criteria for patients in the full-scale trial as opposed to the pilot were changed: instead of having to immediately report to the clinic with menstrual pain, patients were now allowed up to 48 hours to register for the trial, resulting in having many patients recording no pain at all during baseline measurements. Decreased prostaglandin levels at baseline also seem to be apparent for the patients in the full-scale trial, again raising the probability that finding a downward trend during the course of any treatment during the investigation would be less likely to occur.

As for the asthma trial discussed above, it would have been far preferable to have a control group of patients having experienced no physical contact if chiropractic procedures were to be more accurately evaluated. The fact that a much larger group of chiropractors applied the sham procedure in the full-scale trial as opposed to a single practitioner in the pilot raises questions regarding the uniformity of training and reproducibility of contact procedures, the lack of which would have created a significant

scattering of patient outcome measurements. Final discrepancies between the pilot and full-scale trial which are mystifying include the application of an effleurage in the full-scale trial prior to administering either the sham or high-velocity procedure, the pretreatment obscuring the therapeutic effects being followed, and the lack of a 24-hour period of abstention from exercise in the full-scale investigation, which had been included in the pilot study. All these differences may have been related to difficulties of recruiting a sufficient number of patients for the full-scale as opposed to the pilot trial, underscoring how the constraints of an experimental procedure may carry the investigation even farther afield from what is presumed to occur in the physician's office.

To their great credit, the authors state their conclusions far more precisely and conservatively than those seen in the previously discussed trials: "The [results of this trial] are strong evidence that either the low force mimic maneuver was an insufficient placebo treatment or, in fact, that manual therapy does not relieve the pain in women with primary dysmenorrhea." The concern is that both sections, rather than simply the latter portion of this statement, can be carried into any future citations in research publications, as well as into the public consciousness.

7. FLAWED RCT NO. 4: EFFECTS MAY BE OBSCURED BY SMALL SAMPLES SIZES IN A TYPE II ERROR

In comparing patient groups given either high-velocity cervical spinal manipulation or low-level laser treatments as a control, Nilsson[22] observed a tendency of the manipulated group to fare better in terms of pain experienced, headache hours per day, and use of analgesics to alleviate discomfort (Fig 3). The first trial involving 39 patients showed a trend toward improvement in all categories but failed to reach the usual level of statistical significance. Upon increasing the total patient number to 54 with resumed recruitment, however, the investigators arrived at statistically significant differences in all 3 parameters ($P = .04$ to .03).[23] Had the aforementioned asthma[17] or low back pain trials[11] been repeated with larger patient numbers, trends which appeared in much of the data might have become statistically significant differences, overcoming a type II error. Clearly, the potential exists to misinterpret the results of an RCT if they are not reviewed from a multiplicity of viewpoints rather than accepting statistical numbers at face value.

From the preceding, we can appreciate that the following principles need to be maintained as a checklist with which to avoid being mislead by a published RCT:

1. Outcomes of meta-analyses depend on the scoring systems used for inputs.
2. A potential exists for corruption in the comparison of pharmaceutical agents.
3. Oversampling of data may occur from duplicate ("salami") publications.
4. Fastidious interventions in RCTs must not be confused with actual clinical treatments.
5. RCTs which include physical methods of intervention must be checked for inappropriate sham procedures.
6. Trends in RCTs may be obscured by type II errors produced by small sample sizes.
7. The results of RCTs must be confined to the parameters expressed within the investigation and not indiscriminately generalized to clinical practice.

Further concerns about the integrity of RCTs have been stoked by a recent review of 136 research projects addressing a malignant blood disease. The authors of this particular study found a disparity of positive results, depending on the funding source of research, reporting that 74% of the trials reviewed favored a new treatment when they were funded by a for-profit source and that figure being reduced to 47% when funding was provided by nonprofit sources. Moreover, inferior controls were found in 60% of occasions when a particular trial was supported by a for-profit entity but only 21% of the time when a nonprofit source provided funding. The authors were forced to conclude that the uncertainty principle (known as clinical equipoise) appears to have been violated, generating a bias in research.[24]

Adding to the leveling of the hierarchical playing field of experimental design discussed above in Table 1 is the intriguing observation from Benson and Hartz,[25] which suggests that observational studies since 1984 have risen sufficiently in quality to match the findings of the more lofty RCTs. In a search of both the Abridged Index Medicus and the Cochrane databases to identify 2 or more treatments for the same condition, the authors located 136 reports addressing 19 diverse treatments. They found that in most cases, estimates of the treatment effects from observational studies and RCTs were similar; in only 2 out of the 19 analyses did the magnitude of observational studies lie outside the 95% confidence interval for the combined magnitude of RCTs. Thus, there was little evidence that estimates of combined treatment effects from observational studies reported after 1984 were either consistently larger or qualitatively different from those obtained in the more fastidiously constructed RCTs.

In the rush to worship RCTs and extoll their fastidious construction, it is easy to forget what gave rise to performing the RCT in the first place, the astute clinical observation. Indeed, the epidemiologist David Sackett[26] has attempted to reconcile this dilemma by indicating that both observations taken in the doctor's office and rigorous experimental design are needed to build the evidence required for clinical treatment: "External clinical evidence can inform, but can never replace, individual clinical expertise, and it is this expertise that decides whether the external evidence applies
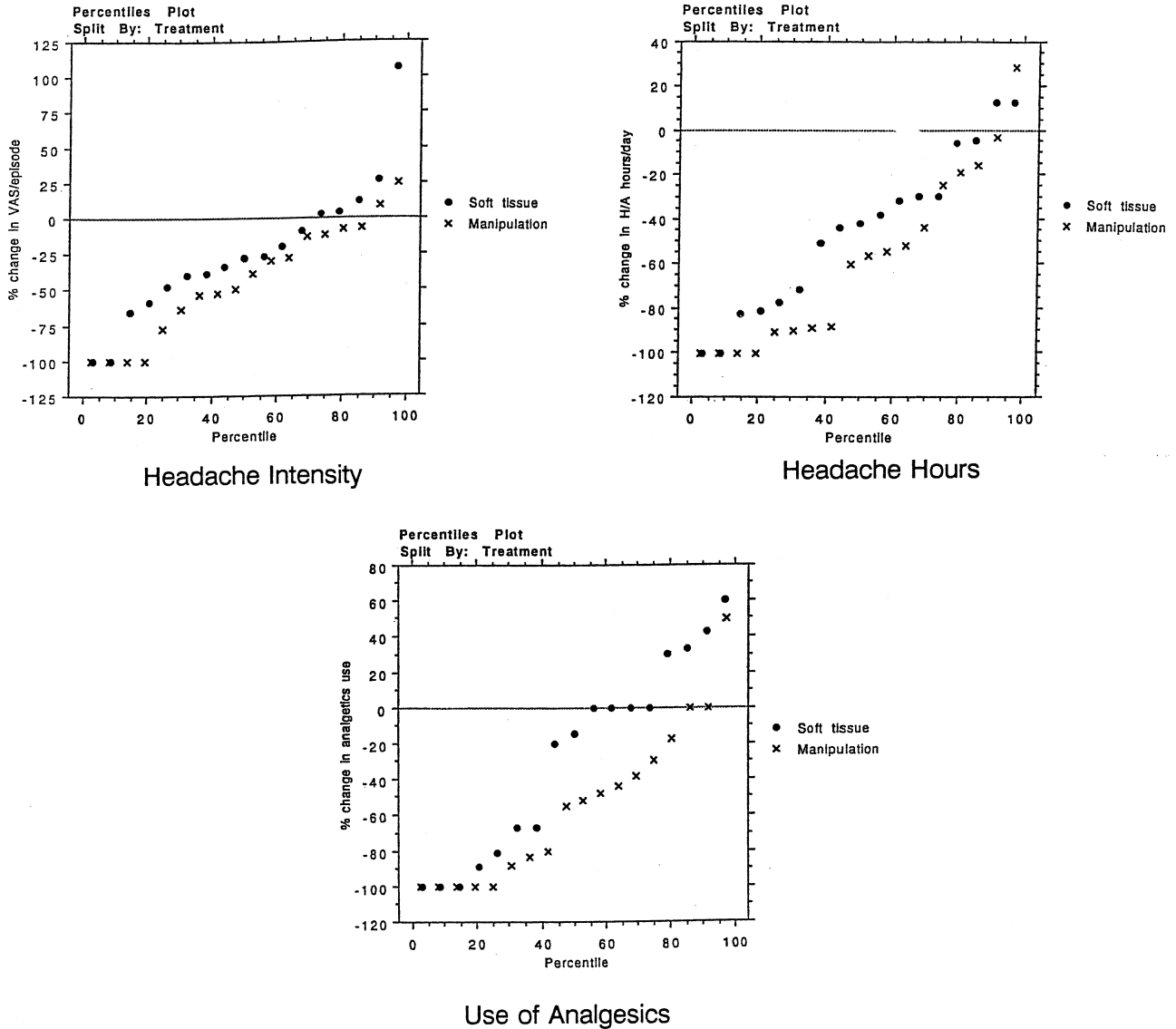
Headache Intensity

Headache Hours

Use of Analgesics

Fig 3. *Cervicogenic headache: Trend toward improvement with SMT. From Nilson N. A randomized controlled trial of the effect of spinal manipulation in the treatment of cervicogenic headache. JMPT 1995;18(7)435-440.*

to the individual patient at all and, if so, how it should be integrated into a clinical decision."

The problems of uncritically accepting evidence from randomized controlled trials and meta-analyses in clinical decision-making have been extensively reviewed elsewhere.[27-31] To build the proper documentation for evidence-based medicine, therefore, one needs to be able to evaluate RCTs realistically in the proper context. Some of the irregularities discussed in this report might tempt the clinical researcher to cast a jaundiced eye on RCTs per se; rather, he or she should simply be prepared to synthesize the proper design and interpretation of RCTs with sound observations gleaned from the individual patient.

CONCLUSION

The 7 case studies reviewed in this report combined with an emerging concept in the medical literature both suggest that reviews of clinical research should accommodate our increased recognition of the values of cohort studies and case series. The alternative would have been to assume categorically that observational studies rather than RCTs provide inferior guidance to clinical decision-making. From this discussion, it is apparent that a well-crafted cohort study or case series may be of greater informative value than a flawed or corrupted RCT. To assume that the entire range of clinical treatment for any modality has been successfully captured by the precision

of analytical methods in the scientific literature, indicates Horwitz,[32] would be tantamount to claiming that a medical librarian who has access to systematic reviews, meta-analyses, Medline, and practice guidelines provides the same quality of health care as an experienced physician.

Anthony Rosner, PhD
Director of Research and Education
Foundation for Chiropractic Research and Education
1330 Beacon Street, Suite 315
Brookline, MA 02446

REFERENCES

1. Bull JP. The historical development of clinical therapeutics. J Chronic Dis 1959;10:218-48.
2. Mechanic D. Bringing science to medicine: the origins of evidence-based practice. Health Aff 1998;17:250-1.
3. Blanchard EB. Biofeedback and the modification of cardiovascular dysfunctions. In: Gatchel RJ, Price KP, editors. Clinical application of biofeedback: appraisal and status. New York: Pergamon Press; 1979.
4. Beecher HK. The powerful placebo. JAMA 1955;159:1602-6.
5. Glass GB. Primary, secondary, and meta-analysis of research. J Educ Res 1976;7:177-88.
6. Juni P, Witsch A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999; 282:1054-60.
7. Gotzsche PC. Multiple publication of reports of drug trials. Eur J Clin Pharmacol 1989;36:429-32.
8. Huston P, Moher D. Redudancy, disaggregation, and the integrity of medical research. Lancet 1996;347:1024-6.
9. Tramer MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. BMJ 1997;315:635-40.
10. Johansen HK, Gotzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. JAMA 1999;282:1752-9.
11. Cherkin DC, Deyo RA, Battie M, Street J, Barlow W. A comparison of physical therapy, chiropractic manipulation, and provision of an educational booklet for the treatment of patients with low back pain. N Engl J Med 1998;339:1021-9.
12. Rosner AL. Evidence-based clinical guidelines for the management of acute low back pain: response to the guidelines prepared for the Australian Medical Health and Research Council. J Manipulative Physiol Ther 2001;24:214-20.
13. Freeman MD, Rossignol AM. A critical evaluation of the methodology of a low-back pain clinical trial. J Manipulative Physiol Ther 2000;23:363-4.
14. Chapman-Smith D. Back pain, science, politics and money. The Chiropractic Report 1998;12:1-4, 6-8.
15. Rosen M. Back pain. Report of a Clinical Standards Advisory Group committee on back pain. London: Her Majesty's Stationery Office; 1994. p. 46, 58, 60.
16. Royal College of General Practitioners. Unpublished update of CSAG guidelines [reference 15]. 1999.
17. Balon J, Aker PD, Crowther ER, Danielson C, Cox PG, O'Shaugnessy D, et al. A comparison of active and simulated chiropractic manipulation as adjunctive treatment for childhood asthma. N Engl J Med 1998;339:1013-20.
18. Dishman JD, Bulbulian R. Spinal reflex attenuation associated with spinal manipulation. Spine 2000;25:2519-25.
19. Field T, Henteleff T, Hernandez M, Martinez E, Mavunda K, Kuhn C, et al. Children with asthma have improved pulmonary functions after massage therapy. J Pediatr 1998;32:854-8.
20. Kokjohn K, Schmid DM, Triano JJ, Brennan PC. The effect of spinal manipulation on pain and prostaglandin levels in women with primary dysmenorrhea. J Manipulative Physiol Ther 1992;15:279-85.
21. Hondras MA, Long CR, Brennan PC. Spinal manipulative therapy vs. a low force mimic maneuver for women with primary dysmenorrhea: a randomized, observer-blinded, clinical trial. Pain 1999;81:105-14.
22. Nilsson N. A randomized controlled trial of the effect of spinal manipulation in the treatment of cervicgogenic headache. J Manipulative Physiol Ther 1995;18:435-40.
23. Nilsson N, Christensen HW, Hartvigsen J. The effect of spinal manipulation in the treatment of cervicogenic headaches. J Manipulative Physiol Ther 1997;20:326-30.
24. Djulbegovic B, Lacevic M, Cantor A, Fields K, Bennett CL, Adams JR, et al. The uncertainty principle and industry-sponsored research. Lancet 2000;356:635-8.
25. Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. N Engl J Med 2000;342:1878-86.
26. Sackett DL. Editorial: evidence-based medicine. Spine 1998;23:1085-6.
27. Feinstein AR, Horwitz RI. Problems in the "evidence" of "evidence-based medicine." Am J Med 1997;103:529-35.
28. Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. J Clin Epidemiol 1995;48:71-9.
29. Kaptchuk T. The double-blind, randomized, placebo-controlled trial: gold standard or golden calf? J Clin Epidemiol 2001;54:541-9.
30. Jonas W. The evidence house: how to build an inclusive base for complementary medicine. West J Med 2001;175:79-80.
31. Radford MJ, Foody JM. How do observational studies expand the evidence base for therapy? JAMA 2001;286:1228-30.
32. Horwitz RI. The dark side of evidence-based medicine. Cleve Clin J Med 1996;63:320-23.