INTEREXAMINER RELIABILITY AND CRANIAL OSTEOPATHY

Steve E. Hartman and James M. Norton

We assess the mechanism purported to underlie the health treatment regime labeled "cranial osteopathy" or "craniosacral therapy." We then summarize all published reports on interexaminer reliability associated with this modality, reanalyze some previously published data, and critique Upledger's[1] often-cited study. Our own and previously published findings suggest that the proposed mechanism for cranial osteopathy is invalid and that interexaminer (and, therefore, diagnostic) reliability is approximately zero. Since no properly randomized, blinded, and placebo-controlled outcome studies have been published, we conclude that cranial osteopathy should be removed from curricula of colleges of osteopathic medicine and from osteopathic licensing examinations.

. . . for both patient and practitioner to be blind to the clinical realities is an unacceptable version of the "double-blind."[2]

THE PRIMARY RESPIRATORY MECHANISM

Cranial osteopathy and craniosacral therapy[3] are variants of a health treatment regime that originated with Sutherland.[4] Physicians (primarily osteopaths), physical therapists, occupational therapists, chiropractors, dentists, and others currently use forms of this method. Although some practitioners make distinctions among various types of cranial therapy, we believe observations recorded here pertain to all. Because Sutherland was an osteopath and because all variants known to us have their roots in his ideas, here we refer to them all as "cranial osteopathy."

The biological model usually called upon to explain the various diagnostic and therapeutic ministrations performed by practitioners of cranial osteopathy has been given the name "craniosacral mechanism" or "primary respiratory mechanism" (PRM)[4(p24)] and includes the following elements:[3(pp10–11),4(pp51–53),5,6(pp165–166),7,8(p23)]

(1) inherent rhythmic motility of the brain and spinal cord;
(2) rhythmic fluctuation of cerebrospinal fluid (CSF);
(3) articular mobility of cranial bones;
(4) mobility of intracranial and intraspinal dural membranes; and
(5) mobility of the sacrum between the ilia.

According to the model, intrinsic rhythmic movements of the brain (independent of respiratory and cardiovascular rhythms) cause rhythmic fluctuations of CSF and specific relational changes among dural membranes, cranial bones, and the sacrum. Because several of these elements provide the biological/mechanistic underpinnings for the cranial rhythm—the focus of this review—we examine the primary respiratory mechanism in light of published, peer-reviewed research.

Element 1: Inherent Rhythmic Motility of Brain and Spinal Cord

Pressure pulses caused by respiratory and cardiac rhythms (transmitted to the cranium by the venous and arterial systems, respectively) produce minor movements of the brain. It is also true that some glial cells possess small

Steve E. Hartman, PhD, is in the Department of Anatomy and James M. Norton, PhD, is in the Department of Physiology, College of Osteopathic Medicine, University of New England. Correspondence concerning this article should be addressed to Steve E. Hartman, PhD, Department of Anatomy, College of Osteopathic Medicine, University of New England, Biddeford, ME 04005; e-mail: shartman@une.edu.

amounts of actin and myosin and are somewhat motile (as are many cells). However, in spite of assertions to the contrary by proponents of the PRM,[4(pp51–52),5,6(p165),7,8(pp23–24,34–35,42,86),9(pp4–5),10] the brain and spinal cord cannot be capable of intrinsically derived movement as organs (see also Becker)[11] because neurons and glial cells lack the requisite microstructure (in particular, dense arrays of actin and myosin filaments). Claims of "a subtle, slow, wormlike movement,"[7] "coiling and uncoiling of the cerebral hemispheres,"[6(p165)] "rhythmic expansion and contraction of the brain and spinal cord,"[5] and "dilation and contraction of the [cerebral] ventricles"[4(p52)] are scientifically groundless.

Element 2: Rhythmic Fluctuation of CSF

Again, although respiratory and cardiovascular rhythms produce minor movements of CSF, it is not these fluctuations in which cranial practitioners take special interest. Their focus is on a purportedly independent rhythm palpable throughout the body, produced by brain movement (element 1),[4(p51–52),6(p165),7,8(p35)] by rhythmic variations in CSF production,[3(pp11–12)] by extracranial muscles,[11,12] as an amalgam of other physiological rhythms,[13,14,15] or by other factors. Practitioners believe that these pulses within the cranium are translated, through movements of dural membranes (element 4) and bones of the cranium (element 3), to the surface of the head, where they can be palpated by properly trained individuals. This palpable rhythm goes by several names, including "cranial rhythmic impulse"[16] and "craniosacral rhythm." We are aware of 6 published attempts to assess quantitatively the reliability with which practitioners could measure, through palpation, parameters relevant to cranial osteopathic diagnosis. Five of these reports focused on frequency of the cranial rhythm and published interexaminer reliabilities of approximately zero; the sixth[1] is badly flawed. Here we summarize all of these reports, reanalyze some of their data, and give Upledger[1] its own section.

Element 3: Articular Mobility of Cranial Bones

Movement between the bases of the sphenoid and occipital bones "is an essential part of Sutherland's functioning model [the PRM]"[3(p10)] and many practitioners claim that movement is possible here throughout life.[3(pp10,16),4(pp24,31),6(p165),7,9(p23),10,17] However, research on large samples of fresh tissue (unembalmed) or living tissue (CT scans) has shown that these two bones undergo complete fusion at their bases between the ages of 12 and 19 (see Table 1; see Becker[11] for another treatment of this issue). Similarly, some practitioners claim that, even *after* ossification, "palpable deformation of the sphenobasilar junction can be appreciated even in the elderly"[10] and that adult humans show "intra-osseous bone flexibility throughout the cranial base."[21] This suggests the palpable deformability of the heavily mineralized matrix of solid bone. These claims are so completely lacking in scientific support that they border on ridiculous.

Likewise, although palpable "articular mobility [at vault sutures] . . . is . . . the basis for important diagnostic and therapeutic procedures"[8(p32)] (see also Sutherland),[4(p23)] movement here also is impossible in most adults. Examination of many hundreds of specimens has revealed that, by age 30 or so, most vault sutures also have begun to ossify.[23,24,25] It is difficult to imagine the basis for claims such as: "sutural obliteration does not appear to occur normally during the aging process"[6(p165)] or "motion . . . persists throughout life."[8(p32)]

Elements 4 and 5: Mobility of Dural Membranes and the Sacrum

Whereas elements 1 through 3 of the primary respiratory mechanism easily are invalidated, the relationship of elements 4 and 5 to biological reality or to parameters of biomedical importance is more difficult to assess. Although elements 4 and 5 also may be of dubious merit,[12,26,27] they require no further attention here because they cannot be playing the biomedical role ascribed to them by practitioners of cranial osteopathy if elements 1 through 3 are invalid. Furthermore, others[28,29] have reported results inconsistent with one of the correlates of elements 4 and 5—the contemporaneous mechanical linking of cranial and sacral movements.

Although cranial osteopathy is widely practiced, there is little more scientific support now for its presumptive mechanism of action than when it was first presented more than 6 decades ago.[11,12,26,30] In fact, elements of the PRM are contingent on anatomical and physiological phenomena that, if proven real, would require reconsideration of some of what is now considered impossible in those disciplines (e.g., motility of neural organs and palpable deformability of solid bone).

Ordinarily, manual diagnoses and treatments will be more meaningful and effective (respectively) if based on a working understanding of relevant anatomical relationships and physiological principles. Even then, only if relevant parameters are measured reliably can there be

Table 1. Status of Union between Sphenoid and Occipital Bones

According To:

Peer–Reviewed Literature	Practitioners of Cranial Osteopathy
Complete ossification by age 16–19 (n = 100, age range 0–20 years)[18]	"[T]his cartilaginous articulation has a slight amount of mobility throughout life."[17]
Complete fusion in 95% of females by age 16 years and 95% of males by age 18 years (n = 189, age range 0–18 years)[19]	"[F]lexible synchondrosis accommodates the flexion and extension activity of the cranial base which continues throughout life."[3(p16)]
No sphenooccipital synchondrosis persisted in any patient past the age of 13 years (n = 253, age range 1–77 years)[20]	"[T]he sphenobasilar juncture is a synchondrosis and is capable of motion throughout life."[21]
Complete fusion in all females by age 17 years and all males by age 19 years (n = 157, age range 10–20 years)[22]	"[A]n intervertebral disc at the sphenobasilar junction . . . up to twenty–five or thirty years, and thereafter a mere movable articulation"[4]

Note: "n" is size of age-graded sample examined

any assurance that diagnosis will lead to an appropriate treatment. Setting aside the likelihood that the PRM itself is invalid, relatively little attention has been given to interexaminer reliability associated with cranial osteopathy, and it is to this that we now turn.

THE CRANIAL RHYTHM

Except for Upledger,[1] all published reports of interexaminer reliability associated with cranial osteopathy have focused on frequency of the cranial rhythm. Therefore, we concentrate now on elements 2 and 3 of the primary respiratory mechanism: the rhythmic movement of cerebrospinal fluid and cranial bones. For practitioners of cranial osteopathy, the cranial rhythm (CR) plays a central role in diagnosis and treatment.[3(p243),5,6,7,8(pp86,107),13, 17,31,32(p95)] For example: "utilization and management of this fluctuation becomes of prime importance"[8(p107)] and "all of craniosacral therapy . . . is predicated on the therapist being able to monitor the patient's craniosacral rhythm. Attempting to use these techniques without the sensitivity to monitor the craniosacral rhythm guarantees that the treatment will fail."[32(p195)] It is presumably because the cranial rhythm is at the mechanistic center of cranial osteopathy and plays a central role in diagnosis and treatment that it has so frequently been chosen by practitioners for tests of interexaminer reliability. Again, only if reliably measurable could such a pa-

rameter be related predictably to meaningful features of a patient's physiology and possibly useful as a biomedical tool.

Estimates of Interexaminer Reliability for the Cranial Rhythm

The CR is believed to exhibit several features measurable through palpation, including frequency, amplitude, regularity, symmetry, and quality. Of these, almost all quantitative estimates of interexaminer reliability have been for frequency. We are aware of five blinded attempts to measure practitioners' ability to assess CR rate through palpation.[28,29,33,34,35] For each of these reports (see Table 2), subject samples were small (ranging from 9 to 40). Examiners were variously trained and their experience at palpating the CR ranged from periods of 11 months to at least 20 years.

Several coefficients have been used to assess interexaminer reliability for determination of CR frequency (see Table 2, next-to-last column): intraclass correlation coefficients ICC(1,1) and ICC(2,1),[36] Pearson product-moment correlation coefficients, and coefficient alpha.[37] Instead of the published product-moment correlations (r), we report coefficients of determination (r^2—the percentage of variance shared by 2 variables) because they are better measures of reliability[38] and will support more useful comparisons. Of coefficients published so far, we believe that ICC(2,1) is most appropriate for comparative purposes[36] and have

Table 2. Interexaminer Reliabilities (Final Column) and Other Relevant Information Associated with Published Measurements of CR Frequency

Authors	N	Ages of Subjects (Years)	Examiners' Degree(s)	Examiners' Experience	Reliability Measure Used	Value of Reliability Measure	
Drengler & King[33]	10 E 10 S	adults[a]	DO[a]	5 to 20 years[a]	ICC(2,1)[h] alpha		−.0009 −.04
Hanten et al.[34]	2 E 40 S	22–54	PT students[b]	11 months	ICC(2,1) ICC(1,1)		?[c] .22
Norton[28]	6 E[d] 9 S	22–28	DO	"extensive"	ICC(2,1)[h]	head[e] sacrum[e] head-sacrum[f]	.14 .04 .05[j] .08[k]
					r^2 [i]	head sacrum head-sacrum[f]	−.08 −.10 −.09
Rogers et al.[29]	2 E 28 S	18–49	PT & RN	5 & 17 years	ICC(2,1)	head feet	.08 .19
					r^2 [i]	head feet head-feet[f]	.01 .05 −.005
Wirth-Pattullo & Hayes[35]	3 E 12 S	10–62	PT	4–6 years	ICC(2,1)		−.02
Upledger[1]	4 E[g] 25 S	3–5	DO	?	ICC(2,1)[l] r^2 [i]		.59[m] .41[m]

Note: "N" is sample sizes for examiners ("E") and subjects ("S"); "DO" is doctor of osteopathy; "PT" is physical therapist; "RN" is registered nurse; ICC(1,1) and ICC(2,1) are intraclass correlation coefficients

[a] personal communication; King, 2000
[b] personal communication; Hanten, 2000
[c] original data no longer available; personal communication, Hanten, 2000
[d] not all subjects evaluated by all examiners; see Norton[28] for details
[e] computed by us from largest subset of data[28] for which ICC(2,1) is strictly appropriate (all subjects measured by all examiners; examiners 1–6, subjects 6–9; his Table 1)
[f] measured simultaneously at 2 locations
[g] 3 examiners paired with Upledger,[1] each for a subset of subjects
[h] computed by us using data supplied by authors
[i] although r was reported, we have converted these to r^2; in this table, a minus sign appearing before an r^2 signifies that the parent correlation coefficient was negative
[j] computed by us from largest subset of Norton's data for which ICC(2,1) is strictly appropriate; examiner 1 at cranium, 2 at sacrum (simultaneously), for 8 subjects (see his Table 1)
[k] examiner 2 at cranium, 1 at sacrum (simultaneously; see his Table 1)
[l] computed by us from data in Upledger's Table 7, over all subjects and examiners
[m] $p < .001$

calculated this coefficient from raw data supplied by several authors who did not report it themselves (see Table 2). We used SYSTAT 9[39] to calculate product-moment correlations and mean squares required for computation of intraclass correlations. The ICC(2,1) coefficients, themselves, were computed using the appropriate equation[36(p423)] and a computer program written by us for that purpose (Microsoft FORTRAN, Version 5.0).

Based on these 5 reports, our overall impression is clear and appears to hold for subject samples of from 9 to 40; for examiners trained as osteopaths, physical therapists, and registered nurses; for examiners with experience ranging from periods of 11 months to 20 years; and for 4 different measures of reliability: frequency of the cranial rhythm could not be measured reliably. Reliabilities were so small that choice of reliability coefficient, statistical significance, and statistical power all are irrelevant. In fact, except for one report,[34] all showed one or more coefficients that were *negative*.

Of these 5 reports, we further analyzed raw rate data from Drengler and King[33] and Norton.[28] In both cases, there is little correspondence among rate assessments of different examiners and reported interexaminer reliabilities were negative and essentially zero (see Table 2). However, there is striking consistency *within each examiner* (see, e.g., Table 3); that is, some examiners tended to perceive rates that were relatively high (e.g., examiner 6 in Table 3 and Fig. 1; examiners 1 and 7 in Fig. 2) and others, *examining the same subjects*, tended to perceive rates that were relatively low (e.g., examiner 3 in Table 3 and Fig. 1; examiners 6 and 9 in Fig. 2). For the largest

of these data sets,[33] standard deviations of the 10 rate measurements *within each subject* were high (averaging 3.0 over all 10 subjects), reflecting the widely divergent assessments made for each subject by the 10 different examiners. On the other hand, standard deviations of the 10 rates determined *by each examiner* were less than half as large (averaging 1.4 over all 10 examiners), reflecting the degree to which each examiner's rate assessments tended to cluster around a particular value. Exactly the same pattern was found using Norton's[28] data (see Table 3). That is, for both of these samples, different examiners tended to measure very *different* rates *for the same subject* but a single examiner tended to measure very *similar* rates *for different subjects*. It is apparent that the data are patterned—but not as predicted by practitioners of cranial osteopathy. As distributed, measurement variance provides no evidence to support a central tenet of cranial osteopathy—that the cranial rhythm is a characteristic of the patient and is measured by the practitioner. Rather, whatever the cranial rhythm represents biologically, based on these data, it appears that perceived rates are characteristics of practitioners, not patients.

We can conceive of only 2 explanations for this finding. Perceived cranial rhythms may be primarily an amalgam of each examiner's own cardiovascular, respiratory, or other rhythms.[15] To date, there is little scientific support for this idea.[1,15,34,35] The only alternative we can conceive of is that the rhythm is the perceptual product of psychological phenomena manifested in the examiner. It is not unusual for individuals to perceive an imaginary sensation only because they have been told to perceive

Table 3. CR Rates (Pulses/Minute) Determined by 6 Examiners at the Crania and Sacra of 4 Subjects[28(Table 1)]

Subjects	Examiners	E1	E2	E3	E4	E5	E6
S6	C	4.68	2.96	2.60	4.06	4.10	5.23
	S	4.97	2.91	2.15	4.08	3.93	5.38
S7	C	6.32	3.79	2.47	6.13	4.93	7.35
	S	7.26	3.48	2.20	4.26	4.72	6.82
S8	C	5.06	3.57	2.06	5.53	4.09	4.78
	S	5.22	2.91	2.14	4.72	3.50	4.34
S9	C	4.32	3.35	2.20	4.61	3.63	5.71
	S	4.35	3.44	2.15	4.09	2.94	7.06
Mean	C	5.10	3.42	2.33	5.08	4.19	5.77
	S	5.45	3.19	2.16	4.29	3.77	5.90

Note: E1–E6 are examiners; S6–S9 are subjects; "C" is cranium; and "S" is sacrum
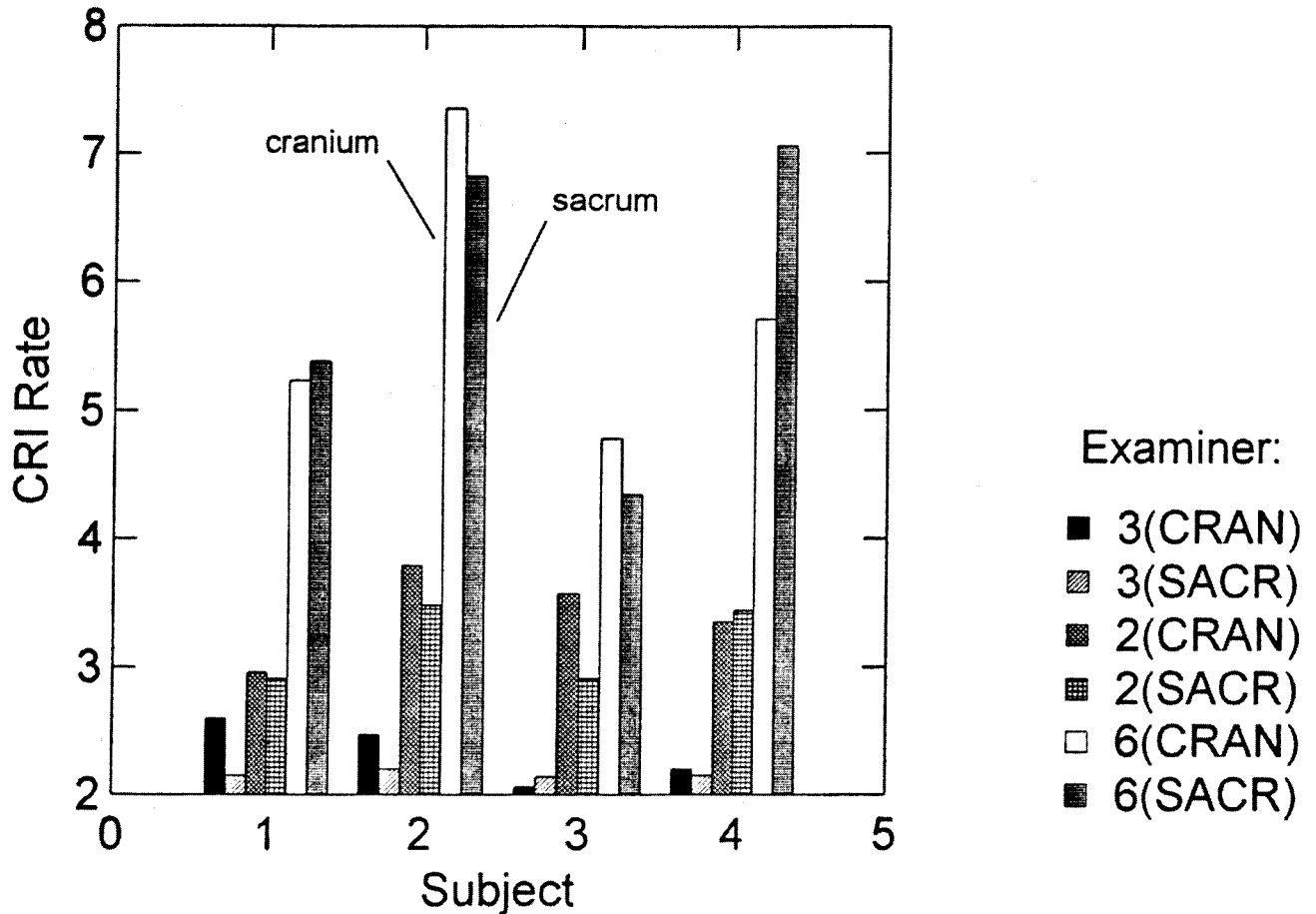
Figure 1. CR rate assessments (pulses/minute) of three examiners for same 4 subjects;[28] "CRAN" is cranium; "SACR" is sacrum.

it,[40] and "this is especially true when the stimulus is vague or ambiguous or when clear observation is difficult."[41(p37)] In particular, humans often perceive motion, that they themselves are the source of, as being externally produced ("automatisms")[42]—see especially Spitz[43] and numerous references therein—with "the *anticipation* of a given result being the stimulus which directly and involuntarily prompts the muscular movements that produce it."[44] Perhaps a cranial practitioner's expectations result in minute, subconscious contractions in hand/arm muscles of the practitioner, leading to tactile sensations that *appear* to confirm PRM-related expectations of cranial movements and rhythms in a patient.

In either case, data so far collected support the hypothesis that the cranial rhythm is not a reliably palpable, biological phenomenon occurring solely in the subject or patient and provide further evidence that this component of the primary respiratory mechanism is invalid. Whether perceived rates are a function of examiners' own physiological rhythms or examiners' psychologies, a "cranial rhythm" cannot be clinically useful because it apparently has nothing to do with the patient.

UPLEDGER (1977)

Upledger[1] evaluated interexaminer reliabilities and agreements for 19 palpatory diagnostic parameters used in his version of craniosacral therapy and compared his subjects' CR rates with the subjects' and examiners' cardiovascular and respiratory rates. For each of his 25 subjects:

a. the subject's and first examiner's cardiac and respiratory rates were measured,[1(Table 7)] then
b. the first examiner (1 of 4 osteopaths) evaluated the subject's CR rate[1(Table 7)] and the other 19 parameters (on a 5-unit scale),[1(app. B)] then
c. the subject's cardiac and respiratory rates were measured again, as were those of a second examiner, then
d. the second examiner reevaluated the subject's CR rate and the other19 parameters.

Upledger and one of three other osteopaths evaluated each subject. This yielded 25 pairs of CR rates and 25 sets of paired measurements for each of the 19 other pa-
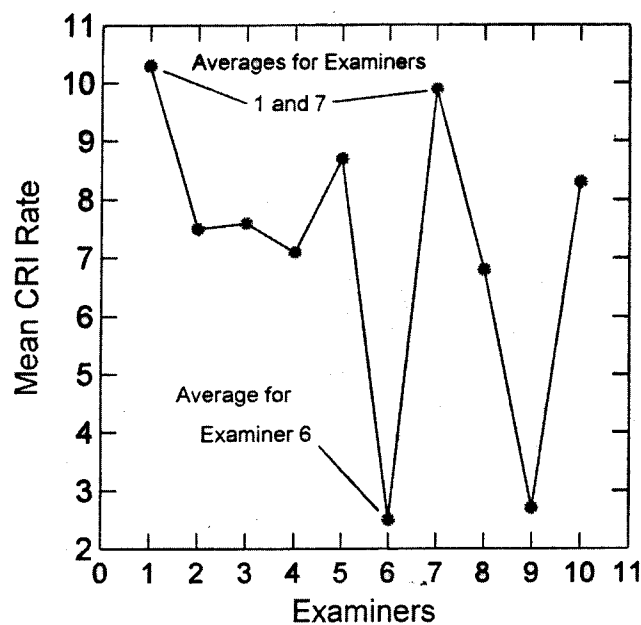
Figure 2. Mean CR rates (pulses/minute) for 10 examiners, all of whom measured same 10 subjects.[33]

Table 4. Product-Moment Correlations among First and Second Measures of Subjects' CR, Cardiovascular, and Respiratory Frequencies[1(Table 7)]

	CR 1	CR 2	CARD 1	CARD 2	RESP 1	RESP 2
CR 1	1.0					
CR 2	.64	1.0				
CARD 1	.06	.01	1.0			
CARD 2	.08	.12	.09	1.0		
RESP 1	.24	.15	.16	.53	1.0	
RESP 2	− .10	.01	.55	.23	.31	1.0

Note: "CARD" is cardiac rate; "RESP" is respiratory rate

rameters. These served as the basis for numerous estimates of agreement and reliability.[1(Tables 2–5)] This work has been published at least 3 times[1,3(pp345–356),45] and is frequently cited. This is the only published, detailed consideration of interexaminer reliability for diagnostic features (other than rate of the cranial rhythm) used in cranial osteopathy (or craniosacral therapy).

Although Upledger[1] did not publish interexaminer reliability for his measurements of CR rates, data in his Table 7 (column 4) have permitted us to do so. The intraclass correlation coefficient ICC(2,1) is not strictly appropriate for data from all 25 of his subjects because all were not measured by the same examiners.[36] However, we provide it for comparative purposes and, using Upledger's data, we believe the estimates of CR interexaminer reliability likely to be biased least by small sample size are those computed over all subjects. Over all 25 subjects and all 4 examiners, we calculated a coefficient of determination of .41 (probability that r = 0 < .001) and an ICC(2,1) of .59 (p < .001; see Table 2, final column). These are perhaps "too low to support [Upledger's] conclusion that the examination can be conducted with an acceptable degree of reliability,"[35] but they are statistically significantly greater than zero. Together with his published global estimate of 86% interexaminer agreement (compared to a chance expectation of 52%) for the other 19 parameters (his Table 6), these estimates reflect substantially greater reproducibility of CR rate measurements than anyone else has achieved (see Table 2).

Perhaps Upledger[1] achieved relatively high CR rate reliability because his subjects were so young (ages 3–5 years). In that case, if the cranial rhythm is real, there may have remained sufficient mobility among bones of his subjects' crania to permit reliable measurement of it. Alternatively, Upledger and his colleagues may have been more skillful in perception of the CR than other examiners reported on here. However, all 10 of Drengler and King's[33] examiners and all 6 of Norton's[28] examiners also were trained as osteopaths. Likewise, most practitioners represented here had considerable experience in monitoring cranial rhythms (see Table 2, column 5). Finally, one of the highest reliabilities published since Upledger's[1] report was achieved by two students in physical therapy,[34] each with less than a year's experience in palpating the CR (see Table 2).

Such factors may help explain the wide disparity between the CR rate interexaminer reliability manifested by Upledger[1] and all reported since, but only if the PRM is a real biological phenomenon. Given the paucity of data supporting this notion, we believe a more likely explanation for Upledger's apparent success is suggested by the fact that his study shows extraordinary evidence of carelessness and poor design.

(1) Although all other workers have reported tiny (or negative) CR rate interexaminer reliabilities, Upledger's were relatively high (e.g., r^2 = .41). By contrast (see Table 4), the first measures of his subjects' cardiac and respiratory rates were poorly correlated with respective second measures (r = .09 and .31). One possible explanation for this finding is that *actual* measurement reliability was similar for all rates, but CR rate reliabilities *appeared* greater because CR rates were simply more stable over the interval be-

tween first and second measures. Although some claim that the CR rate exhibits greater short-term temporal stability than cardiac and respiratory rates,[3](p6) we are aware of no published data that support this claim. Another possibility is that all rates were equivalently stable over the interval between first and second measurements but CR rate interexaminer reliability was much higher. However, it is difficult to imagine how a rhythm many cannot perceive and many others cannot detect without training (the CR) could be measured more reliably than cardiac and respiratory rhythms, which can be measured objectively and accurately by anyone who can count.

(2) Some cardiac and respiratory rate pairings presented in Upledger's Table 7 are difficult to explain. For example, when first measured, subject 19 had a cardiac rate of 120/minute and a respiratory rate of 24/minute. When measured again many minutes later (after evaluation of CR rate and the other 19 parameters), this subject showed a much *lower* cardiac rate (92/minute) and a respiratory rate that had almost *doubled* (40/minute). Similarly, subject 5's cardiac rate jumped from 84 to 120/minute while respiratory rate actually dropped slightly from 30 to 28/minute. These findings are, physiologically, very difficult to explain and suggest measurement error, transcription error, or some other form of carelessness.

(3) Upledger[1](p891) indicated that each child's CR rate was recorded "as counted for one minute." Elsewhere,[46] he reported that: "We did count cranial rhythmic impulses as well as heart and respiratory rate, but we only counted each for 15 seconds and multiplied by 4 to get the rates per minute." Indeed, examination of rates reported in his Table 7[1] shows that many more are divisible by 4 than would be expected by chance, suggesting that many rates were extrapolated from 15-second counts. However, many are *not* divisible by 4, suggesting that rates were recorded over different intervals on different occasions. Not only were perceived respiratory and CR frequencies too low for either to support accurate rate determinations using only 15-second counts, but this inconsistency could easily have affected results.

(4) Upledger[1] showed little correspondence between his subjects' CR rates and cardiac and respiratory rates of his subjects and examiners and said this would "help establish the CRI as an independent physiologic rhythm." However, as described earlier, these different rhythms were recorded over an unspecified time interval and none of the others was recorded simultaneously with CR rate. Considering this and the fact that first and second measures of cardiac and respiratory rates for subjects do not even resemble *one another* in a physiologically sensible way (see items 1 and 2), it is hardly surprising that neither resembles CR rates measured twice over a similar interval.

(5) We used raw data published in Upledger's appendix B to verify measurement reliabilities (product-moment correlations; personal communication, JE Upledger, 1998) for Upledger's 19 diagnostic parameters (his Tables 2–5). For Upledger paired with each of three other examiners (his Tables 2–4, columns 2) and for all four examiners and all 25 subjects combined (his Table 5, column 2), 26 of 76 (34%) of these reliabilities were misreported. Six of the differences between his and our correlations may have resulted from rounding errors but others were off by a great deal more. For example, for Dr Upledger with Dr Mitchell and parameters 13, 17, and 18, reported reliabilities were all zero (his Table 4, column 2). However, we calculated correlations of .92, –.90, and –.32, respectively.

(6) Also in his Tables 2–5 (final columns), Upledger[1] published percentages of agreement for each of the 19 diagnostic parameters, allowing a difference of up to 1 unit on his 5-unit scale. Again based on data from his Appendix B, 10 of the 76 values presented (13%) were miscalculated, being off by an average of about 9% and about 5% too high overall. (Although 5 of the 8 averages presented in Upledger's Table 6—each over all 19 parameters—are in error, none is off by more than 2 percentage points.)

In his abstract, Upledger[1] said: "these data would seem to support the reliability . . . of the examination findings." In his conclusion he said: "it is possible to achieve an acceptable degree of interexaminer reliability and percentage of agreement between examiners utilizing craniosacral examination methods and techniques . . . [this] lends considerable evidence to the existence of a real and perceptible craniosacral motion system." However, evidence of poor experimental design and extraor-

dinary lack of rigor call this central claim into question. His measured reliabilities can be meaningful and his claims justified only if the experimental blind between paired examiners was maintained scrupulously during all assessments of the cranial rhythm; otherwise, findings of the first examiner in each pair may have influenced (consciously or unconsciously) findings of the second examiner. With so much evidence for lack of scientific rigor, we cannot be confident that paired examiners were properly shielded from each other's findings. We are *not* suggesting that Dr Upledger intentionally misled his readers. However, before osteopaths (or others) can claim meaningful diagnostic and treatment control over the biomedical system labeled "cranial osteopathy" (or "craniosacral therapy"), it must be shown that biomedical parameters of this system can be measured reliably. Practitioners should *at least* be able to count cycles of the CR. After all, for both mechanistic explications and diagnostic and therapeutic judgements, practitioners rely on properties of this rhythm (e.g., amplitude, symmetry, and quality) that are even more derivative and subjective than its frequency. To date, only Upledger[1] has submitted results at all suggestive of reliability and we here provide reason to doubt the value of even these findings.

POSSIBLE POINTS OF CONTENTION

Some cranial practitioners have asked how we justify such a critical review of cranial osteopathy, given the wide array of other poorly understood (including more conventional) treatment modalities. Although much remains to be learned about mechanisms and effectiveness of many conventional treatments, the proportion of such treatments with these deficiencies is much overstated.[47] Most importantly, however, we are critical because the extraordinary biomedical implausibility of cranial osteopathy places it in a very different category than any standard, science-based method of health care known to us.

Some practitioners may disregard findings of poor interexaminer reliability for CR frequency summarized here because they are based on small samples of subjects and examiners. However, reliabilities of essentially *zero* across all of these small studies (except Upledger)[1] are ample support for our claims that: sample size is not at issue, the CR probably is not a biomedical feature of patients, and additional study using larger samples is unwarranted. It now should be up to practitioners to find *any* important diagnostic or therapeutic parameter of cranial osteopathy that can be measured reliably.

Some practitioners have argued that low reliabilities of CR rate measurements taken one after another are not surprising, given that: ". . . it seldom happens that a therapist practicing CranioSacral Therapy can touch a patient for more than a minute or two without having some therapeutic effect on this very sensitive craniosacral system."[46,21] Not only is this ad hoc assertion without scientific support, but we consider it suspiciously convenient: if true, it would render reliability of all diagnostic claims related to cranial osteopathy completely untestable. Also, this would leave unexplained the marked consistency of rate assessments across all subjects *within individual examiners*. In general, whether interexaminer reliability of *zero* results from irregularity of the rhythm itself or an inability of consecutive examiners to measure a stable CR and reach similar diagnostic conclusions, the implication is the same: none of the features of the CR can be useful sources of diagnostic information. Furthermore, others[28,29] examined comparability of CR rates measured *simultaneously* at 2 locations (head and sacrum, head and feet, respectively), and reported between-location product-moment correlations *below zero*.

Some practitioners have suggested that published reliabilities of essentially zero for measurement of CR rate are relatively unimportant, or even irrelevant, because rate for the cranial rhythm is less important clinically than its other qualities. However, if phases of the phenomenon labeled "cranial rhythm" or "primary respiratory mechanism" cannot even be counted, then it is unlikely that its other, more derivative and complex features (e.g., amplitude, symmetry, and quality) can be evaluated reliably either. We consider this to be the case, whatever the cause of the rhythm. Also, given that this one presumed biomedical parameter of cranial osteopathy has been the nearly unanimous choice for reliability testing and has failed utterly, we are suspicious of practitioners who now claim that this parameter was a poor selection because of its minimal clinical value.

Perhaps in recognition of the weakness of the mechanistic framework of cranial osteopathy and reported interexaminer reliabilities of zero, some practitioners deemphasize the PRM in their defense of cranial osteopathy, instead focusing on perceived clinical efficacy. They assert that properly trained cranial practitioners *must* be diagnosing reliably because many decades of clinical experience has shown treatment to be effective. We agree that, if cranial osteopathy were demonstrably efficacious, invalidity of mechanistic explanations and apparent lack of interexaminer reliability, although hard to explain, would be clinically moot. Unfortunately, although many clinicians (and patients)

have become convinced of the efficacy of cranial osteopathy, there are still no data, based on properly controlled research, supporting any claim that apparent symptom improvement following "cranial" treatment has ever involved more than, at most, a form of placebo effect. Recent summaries offer practitioners little cause for optimism.[48,49] Without careful scientific controls, weaknesses of perception and interpretation can fool both practitioners and patients into believing that a treatment is effective when it is not.[41(pp195–232),50,51,52,53,54] We believe that these and other natural, human, psychosocial influences help to explain how cranial osteopathy has achieved the 21st century without scientific support of any kind.

CONCLUSION

There is little science in any aspect of cranial osteopathy:

(1) there is no scientific support for major elements of the PRM;

(2) the only publication purporting to show diagnostic reliability with sufficient detail to permit evaluation[1] is deeply flawed and stands alone against 5 other reports that show reliabilities of essentially zero; and

(3) there is no scientific evidence of treatment efficacy.

Until mechanistic claims associated with the PRM have been validated, until diagnostic reliability has been established, and until properly randomized and placebo-controlled outcome studies have demonstrated symptom improvement following manipulation of relevant parameters, we believe cranial osteopathy should be excluded from required curricula of colleges of osteopathic medicine and from osteopathic licensing examinations.

*   *   *

While this manuscript was in press, another examination of interexaminer reliability for measurement of the cranial rhythm was published.[55] Again, 4 of 8 published coefficients were *negative* and, again, data suggested these 2 osteopathic physicians may have been imagining the cranial rhythm.

REFERENCES

1. Upledger JE. The reproducibility of craniosacral examination findings: a statistical analysis. *J Am Osteopath Assoc.* 1977;76:890–899.

2. Dodes JE. The mysterious placebo [quotation from: Jarvis WT. *Dubious Dentistry: A Dental Continuing Education Course.* Loma Linda, Calif: Loma Linda University; 1990]. *Skeptical Inquirer.* 1997;21(1):44–45.

3. Upledger JE, Vredevoogd JD. *Craniosacral Therapy.* Chicago, Ill: Eastland Press; 1983.

4. Sutherland WG. *The Cranial Bowl.* New York, NY: Free Press; 1939.

5. Ettlinger H, Gintis B. Cranial concepts. In: DiGiovanna EL, Schiowitz, S, eds. *An Osteopathic Approach to Diagnosis and Treatment.* 2nd ed. Philadelphia, Pa: Lippincott-Raven; 1997.

6. Greenman PE. *Principles of Manual Medicine.* 2nd ed. Philadelphia, Pa: Williams & Wilkins; 1996.

7. Lay EM. Cranial field. In: Ward RC, ed. *Foundation for Osteopathic Medicine.* Baltimore, Md: Williams & Wilkins; 1997:901–913.

8. Magoun HI. *Osteopathy in the Cranial Field.* 3rd ed. Fort Worth, Tex: Sutherland Cranial Teaching Foundation; 1976.

9. Lippincott RC, Lippincott HA. *A Manual of Cranial Technique.* Ann Arbor, Mich: Edwards Brothers; 1948.

10. Mitchell FL Jr. Clinical significance of cranial suture mobility. In: Retzlaff EW, Mitchell FL Jr, eds. *The Cranium and Its Sutures: Anatomy, Physiology, Clinical Applications and Annotated Bibliography of Research in the Cranial Field.* New York, NY: Springer Verlag; 1987:13–26.

11. Becker RF. Cranial therapy revisited. *Osteopath Ann.* 1977;5:316–334.

12. Ferguson A. Cranial osteopathy: a new perspective. *Am Acad Osteopath J.* Winter 1991:12–16.

13. McPartland JM, Mein EA. Entrainment and the cranial rhythm. *Altern Ther.* 1997;3:40–45.

14. Nelson KE, Sergueef N, Lipinski CM, Chapman AR, Glonek T. Cranial rhythm related to the Traube-Hering-

Mayer oscillation: comparing laser-Doppler flowmetry and palpation. *J Am Osteopath Assoc.* 2001;101:163–173.

15. Norton JM. A tissue pressure model for palpatory perception of the cranial rhythm. *J Am Osteopath Assoc.* 1991;91:975–994.

16. Woods JM, Woods RH. A physical finding related to psychiatric disorders. *J Am Osteopath Assoc.* 1961;60:988–993.

17. Greenman PE, Mein EA, Andary M. Craniosacral manipulation. *Man Med.* 1996;7:877–895.

18. Melsen B. Time and mode of closure of the spheno-occipital synchondrosis determined on human autopsy material. *Acta Anat.* 1972;83:112–118.

19. Madeline LA, Elster AD. Suture closure in the human chondrocranium: CT assessment. *Radiol.* 1995;196:747–756.

20. Okamoto K, Ito J, Tokiguchi S, Furusawa T. High-resolution CT findings in the development of the sphenooccipital synchondrosis. *Am J Neuroradiol.* 1996;17:117–120.

21. Ferguson AJ, McPartland JM, Upledger JE, Collins M, Lever R. Cranial osteopathy and craniosacral therapy: current opinions. *J Bodywork Mov Ther.* 1998;2:28–37.

22. Sahni D, Jit I, Neelam, Suri S. Time of fusion of the basisphenoid with the basilar part of the occipital bone in northwest Indian subjects. *Forensic Sci Int.* 1998;98:41–45.

23. Cohen MM Jr. Sutural biology and the correlates of craniosynostosis. *Am J Med Genet.* 1993;47:581–616.

24. Perizonius WRK. Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (A.D. 1883–1909). *J Hum Evol.* 1984;13:201–216.

25. Verhulst J, Onghena P. Cranial suture closing in *Homo sapiens*: evidence for circaseptennian periodicity. *Ann Hum Biol.* 1997;24:141–156.

26. Dove CI. The origin and development of craniosacral osteopathy. *Holistic Med.* 1988;3:35–45.

27. Ferguson A. Osteopathic approaches. *Br Osteopath J.* November 1987:8–10.

28. Norton JM. A challenge to the concept of craniosacral interaction. *Am Acad Osteopath J.* 1996;6(4):15–21.

29. Rogers JS, Witt PL, Gross MT, Hacke JD, Genova PA. Simultaneous palpation of the craniosacral rate at the head and feet: intrarater and interrater reliability and rate comparisons. *Phys Ther.* 1998;78:1175–1185.

30. Ferré C, Barbin JY. The osteopathic cranial concept: fact or fiction? *Surg Radiol Anat J Clin Anat.* 1991;13:165–170.

31. Greenman PE, McPartland JM. Cranial findings and iatrogenesis from craniosacral manipulation in patients with traumatic brain syndrome. *J Am Osteopath Assoc.* 1995;95:182–192.

32. Manheim CJ, Lavett DK. *Craniosacral Therapy and Somato-emotional Release: The Self-healing Body.* Thorofare, NJ: Slack; 1989.

33. Drengler KE, King HH. Interexaminer reliability of palpatory diagnosis of the cranium. *J Am Osteopath Assoc.* 1998;98:387.

34. Hanten WP, Dawson DD, Iwata M, Seiden M,

Whitten FG, Zink T. Craniosacral rhythm: reliability and relationships with cardiac and respiratory rates. *J Orthop Sports Phys Ther.* 1998;27:213–218.

35. Wirth-Pattullo V, Hayes KW. Interrater reliability of craniosacral rate measurements and their relationship with subjects' and examiners' heart and respiratory rate measurements. *Phys Ther.* 1994;74:909–920.

36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.

37. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.

38. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000:560.

39. Wilkinson L. *SYSTAT 9.* Chicago, Ill: SPSS; 1999.

40. Ellson DG. Hallucinations produced by sensory conditioning. *J Exp Psychol.* 1941;28:1–20.

41. Schick T Jr, Vaughn L. *How to Think about Weird Things: Critical Thinking for a New Age.* 2nd ed. London, England: Mayfield Publishing; 1999.

42. Hyman, R The mischief-making of ideomotor action. In: Sampson W and Vaughn L, eds. *Science Meets Alternative Medicine: What the Evidence Says about Unconventional Treatments.* Amherst, NY: Prometheus Books; 2000:95–116.

43. Spitz HH. *Nonconscious Movements: From Mystical Messages to Facilitated Communication.* Mahwah, NJ: Erlbaum; 1997.

44. Carpenter WB. On the influence of suggestion in modifying and directing muscular movement, independently of volition. *Notices of the Meetings of the Royal Institution of Great Britain.* 1852;1:147–153.

45. Upledger JE. The reproducibility of craniosacral examination findings: a statistical analysis. In: *1977 Yearbook of the American Academy of Osteopathy,* Colorado Springs, Colo: American Academy of Osteopathy; 1977:95–104.

46. Upledger JE. Research and observations that support the existence of a craniosacral system. Available at: http://iahe.com/press/P-MON.htm. Accessed April 23, 2000.

47. Imrie RH, Ramey DE. The evidence for evidence-based medicine. *Sci Rev Alt Med.* 2001;5:104–108.

48. Green C, Martin CW, Bassett K, Kazanjian A. A systematic review of craniosacral therapy: biological plausibility, assessment reliability and clinical effectiveness. *Complement Ther Med.* 1999;7:201–207.

49. Oppel L, Beyerstein BL, Mathias R, et al. *Craniosacral Therapy: A Review of the Scientific Evidence.* Prepared by the Alternative Therapy Evaluation Committee, for the Insurance Corporation of British Columbia; 1997.

50. Beyerstein BL. Why bogus therapies seem to work. *Skeptical Inquirer.* 1997;21(5):29–34.

51. Beyerstein BL. Social and judgmental biases that make inert treatments seem to work. *Sci Rev Alt Med.* 1999;3(2):20–33.

52. Beyerstein BL. Alternative medicine and common errors of reasoning. *Acad Med.* 2001;76:230–237.

53. Buckman R, Sabbagh K. *Magic or Medicine? An In-*

vestigation of Healing and Healers. Amherst, NY: Prometheus Books; 1995:119–152.

54. Gilovich T. *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life*. New York, NY: Free Press; 1991:125–145.

55. Moran RW, Gibbons P. Intraexaminer and interexaminer reliability for palpation of the cranial rhythmic impulse at the head and sacrum. *J Manipulative Physiol Ther*. 2001;24: 183–190.

IN BRIEF

FTC Charges Sellers of Cell Phone Radiatio Protection Patches with Making False Claims

The Federal Trade Commission has charged two companies that sold devices that purportedly protect users from electromagnetic radiation emitted by cellular telephones with making false and unsubstantiated claims. In separate court actions announced on February 20, the FTC alleges that Stock Value 1, Inc. and Comstar Communications, Inc. (Comstar) falsely represented that their products block up to 97%–99% of radiation and other electromagnetic energy emitted by cellular telephones, thereby reducing consumers' exposure to this radiation. According to the FTC, the defendants lacked a reasonable basis to substantiate their claims. The commission is seeking permanent injunctions, consumer redress, and other equitable relief.

"These companies are using a shield of misrepresentation to block consumers from the facts," said J. Howard Beales III, director of the FTC's Bureau of Consumer Protection. "There is no scientific evidence that their products work as they claim."

Stock Value 1, Inc., based in Boca Raton, Florida, and also known as SV1, and its president, Deborah Jenkins, marketed and sold two products—SafeT-Shield™ and NoDanger—that purportedly block electromagnetic energy emitted from cellular and cordless telephones to consumers throughout the United States. These products consist of metallic fiber patches that are placed over the earpieces of cellular and cordless telephones. The defendants advertised their products through TV, radio, and print ads, and on the Internet.

Comstar, based in West Sacramento, California, and its president, Randall Carasco, marketed and sold their products under the names WaveShield, WaveShield 1000, and WaveShield 2000. They advertised their products to consumers nationwide through TV, radio, and print ads, and on the Internet.

The complaints allege that the defendants, in both cases, failed to disclose in their ads that the vast majority of electromagnetic energy emitted by cellular and cordless phones comes from the antenna and parts of the phone other than the earpiece. The defendants allegedly also failed to disclose that the WaveShield, NoDanger, and SafeTShield™ products have no effect on this other electromagnetic energy. These facts, the FTC said, would be material to consumers' decision to buy or use their products.

Both complaints further allege that the defendants made false statements that their products had been scientifically "proven" and "tested," when in fact that was not the case.

According to a May 2001 report by the General Accounting Office, "Scientific research to date does not demonstrate that the radio frequency energy emitted from mobile phones has adverse health effects, but the findings of some studies have raised questions indicating the need for further investigation."

The FTC has issued a new Consumer Alert, "Radiation Shields: Do They 'Cell' Consumers Short?" which offers suggestions for cell phone users who want to limit their exposure to electromagnetic emissions from their phones. According to the FTC, there is no scientific proof that these so-called shields significantly reduce exposure from electromagnetic emissions. Consumers who want to limit their exposure can take steps such as:

- limiting cell phone use to short conversations;
- increasing the distance between the antenna and the head by using a hands-free set or a car phone with the antenna outside the car; and
- avoiding using cell phones where the signal is poor.

These cases were referred to the commission by the Good Housekeeping Institute, the consumer product evaluation laboratory of Good Housekeeping Magazine. Independent tests conducted by the Good Housekeeping Institute on SafeTShield™, WaveShield, and similar products found that the products did not reduce radiation exposure from cellular telephones.