



ELSEVIER

Acupuncture for low back pain: traditional diagnosis and treatment of 148 patients in a clinical trial

Hugh MacPherson^{a,*}, Lucy Thorpe^b, Kate Thomas^b, Mike Campbell^c

^a Foundation for Traditional Chinese Medicine, York YO10 5DD, UK

^b Medical Care Research Unit, School of Health and Related Research, University of Sheffield, Sheffield, UK

^c Institute of Primary Care and General Practice, School of Health and Related Research, University of Sheffield, Sheffield, UK

Summary *Aims:* To assess patterns of diagnosis, including concordance, and treatment within a clinical trial of traditional acupuncture for low back pain. *Setting:* In a pragmatic randomised controlled clinical trial, 148 patients with low back pain, of between 4 weeks and 12 months duration, were randomised to the offer of individualised acupuncture and received up to 10 treatments. *Methods:* Standardised diagnosis and treatment records were completed by practitioners for 148 patients. The diagnosis was based on three pre-defined low back pain syndromes. For a subgroup of patients, one of the six practitioners then independently re-examined the patients, blind to the original diagnosis. The diagnostic inter-rater reliability was assessed in terms of percentage congruent classifications and Cohen's Kappa. Structured interviews of practitioners established further details about practice styles. *Results:* The most commonly diagnosed syndrome associated with low back pain was Qi and Blood Stagnation (88% of patients), followed by Kidney Deficiency (53%) and Bi Syndrome (28%), with more than one syndrome being identified for 65% of patients. For the subgroup examined twice, practitioner concordance was reasonable: between 47 and 80% of classifications were congruent, while Kappa values lay between 0 ('the same as chance') and 0.67 ('good'). Practitioners provided 1269 treatments in total, using 177 different acupuncture points. Most commonly used channels were Bladder and Gall Bladder, and the commonest points were BL-23 and the two lowest Huatuoji-aji points. Auxiliary treatments were utilised by all practitioners to varying degrees. *Conclusions:* Diagnostic concordance among practitioners was reasonable, and clear themes emerged for treatment. Further research is required to develop a flexible trial protocol with scope for individualised treatment.

© 2003 Elsevier Ltd. All rights reserved.

Introduction

Concerns about the adequacy of treatment in clinical trials of acupuncture are widely discussed in the literature.^{1,2} Treatment protocols based on stan-

* Corresponding author. Tel.: +44-1904-321394; fax: +44-1904-321388.

E-mail address: hm18@york.ac.uk (H. MacPherson).

standardised formula of points have been criticised on the basis that such treatments do not reflect current everyday practice.³ However what constitutes current everyday practice cannot easily be defined for several reasons. First there are many different styles of acupuncture. Second most of the styles practised in the West have, as integral, the use of individualised treatments, where the prescription of points varies between patients, and for each patient, over time. Third, the boundary of what constitutes acupuncture treatment also varies between styles and between practitioners: for some, acupuncture only involves the insertion of needles whereas for others, acupuncture is part of a package of care that includes a range of auxiliary techniques and self-help advice as well as a focus on the wider aspects of patients' lives that relate to the self-healing process.

In a previous literature review of textbook information on the treatment of low back pain with acupuncture, 16 traditional Chinese medicine texts were examined.⁴ Despite variability between texts, the authors identified three common traditional acupuncture syndromes associated with low back pain, along with a group of commonly prescribed points for each. They also discussed the limitations of their literature survey, including selectivity in their choice of texts and limitations of the texts themselves.

To establish a more standardised treatment for use in a clinical trial, clinical interviews and practitioner surveys have been proposed.^{5,6} However, the goal of establishing a standardised treatment may not always be feasible or desirable, given the potential danger that creating something new, and not seen in normal practice, will be of limited use when generalising trial results to a wider population. When practised within the framework of traditional Chinese medicine, acupuncture involves a diverse repertoire of diagnostic frameworks and treatment interventions. While attempts by Westerners to systematise acupuncture have assisted in its transmission to the West, such efforts have tended to reduce the richness and diversity of the clinical encounter.⁵

The enormous diversity in the practice of acupuncture in China⁷ and the West⁸ shows that this medical tradition is not a static or tightly bounded system of medicine. There are no clear boundaries as to what constitutes the practice of acupuncture as it continues to adapt and evolve over time. It is therefore the nature of acupuncture practice itself, both in China and in the West, which makes it difficult to standardise without impinging on the integrity of the tradition. For this reason it is necessary, when considering what constitutes an

adequate acupuncture treatment in a clinical trial, to clearly identify the practitioners, the patients and the setting. It is only with this well-defined context that judgements can be made about the appropriateness and adequacy of treatment as well as the generalisability of the results.

In this paper, we describe the diagnosis and treatment of patients who participated in the York Acupuncture for Back Pain trial. Given that the acupuncturists delivering the acupuncture were given a free hand to treat as they normally do, a feature of a fully pragmatic trial design, what we present here are descriptions of acupuncture in everyday practice. We were interested in how acupuncturists treat their patients with low back pain as well as the extent of variability between practitioners. We were also interested in the reliability of practitioners' traditional acupuncture diagnoses, an area ripe for investigation.

Methods

The York Acupuncture for Back Pain trial was a pragmatic randomised controlled trial,^{3,9} designed to evaluate the clinical and cost effectiveness of acupuncture for patients with low back pain, the protocol for which has already been published.¹⁰ This trial involved 241 patients with an episode of low back pain lasting more than 4 weeks and less than a year. All patients were referred by their GP to the researcher (L.T.) who, after obtaining informed consent and taking baseline data, randomised the patients, in the ratio of two to one, to the offer of acupuncture. Of the 241 patients, 160 were randomised to the offer of acupuncture. As the back pain of 10 patients resolved prior to treatment, 150 received acupuncture. Two of these patients were excluded from the analysis below as a practitioner who withdrew from the study treated them. Treatment of the 148 patients was delivered by six local acupuncturists who provided up to 10 treatments per patient usually on a weekly basis. The trial protocol allowed participating acupuncturists to tailor their treatments to individual patient needs. The six practitioners in the trial were all members of the British Acupuncture Council, and had been in practice for at least 3 years.

Prior to commencement of the trial, a series of meetings were held between the practitioners and two researchers (L.T. and K.T.). The purpose was to establish an acceptable shortlist of low back pain syndromes, agree the content of the standardised reporting booklets and to share personal styles and experiences in the treatment of low back pain. In developing the short-list of syndromes to be used

by practitioners at the first consultation, we also drew on a previous literature review⁴ and, through discussion, achieved consensus on a short-list of what we have called “key” syndromes: Qi and Blood Stagnation, Bi Syndrome and Kidney Deficiency, with the added option of an “Other”. For Bi Syndrome, we included sub-categories of Cold Bi and Damp Bi. For Kidney Deficiency, we listed the sub-categories of Kidney Yang Deficiency and Kidney Yin Deficiency. We also agreed a list of key symptoms to be included in the checklist for each category and sub-category. Practitioners agreed to complete this diagnostic process at the first consultation of each patient.

Some estimate of the reliability of the practitioners’ diagnoses was considered to be useful. Given the difficulties of getting practitioners together, the ideal of all practitioners diagnosing all the patients was clearly impractical. However one of the six practitioners, who was also one of the researchers (H.M.), independently diagnosed a subgroup of the other practitioners’ patients (chosen by the availability of patient and practitioner) while remaining blind to the original diagnosis. This independent diagnosis was only made on the key syndromes and undertaken as soon as possible after the first consultation.

A measure of agreement or concordance on key syndromes between the practitioner-researcher (H.M.) and each of the other five practitioners was made using two methods. First, in an approach recommended by Steinijans et al.,¹¹ we assessed the percentage of congruent classifications, i.e. exact agreements. Secondly we calculated Cohen’s Kappa, where Kappa is a measure of agreement between raters which is corrected for chance. A Kappa of zero would mean there was no agreement above chance while a Kappa of one would mean complete agreement. Other Kappa values between zero and one have no absolute definitions, however, Landis and Koch¹² suggested words

to describe this range which have subsequently been modified by Altman¹³: a Kappa of less than 0.2 is “poor”, 0.21-0.4 is “fair”, 0.41-0.6 is “moderate”, 0.61-0.8 is “good” and 0.81-1.0 is “very good”.

In standardised patient treatment booklets, practitioners reported on all acupuncture points specifying any specific needling techniques. Using a checklist format, auxiliary techniques were also reported, including moxibustion (moxa for short), massage, acupressure, cupping and Chinese herbs, as well as self-help treatments the practitioners had prescribed, such as specific exercises, and lifestyle advice (i.e. advice around diet, exercise and rest). Data were analysed using the statistical package SPSS.

All practitioners were interviewed at the end of the treatment period of the trial. This qualitative data on diagnosis and treatment provided context for the quantitative results.

Results

The patients’ mean age was 42 (range 20-64) and 38% were male and 62% female. On average their back pain had lasted 17.2 weeks (range 4-48 weeks). Background data on the six acupuncturists, who treated at least 20 patients each, are presented in Table 1.

Traditional acupuncture diagnosis

At the first consultation of the 148 patients, 86 (58%) were diagnosed as having both a primary and a secondary syndrome, 47 (32%) as having a primary syndrome only, 2 (1%) a secondary only, 2 (1%) as having two primary and one secondary, 6 (4%) with one primary and two secondary, 2 (1%) with two secondary syndromes only, and 4 (3%) with no syndromes diagnosed, an average of 1.7 syndromes per patient.

Table 1 Practitioner and treatment data in a clinical trial of acupuncture for low back pain.

	Practitioner						Mean
	A	B	C	D	E	F	
Duration of initial training	3	3	5	3	2	3	3.2
Location of initial training	UK	UK	China	UK	UK	UK	-
Years in practice	5	18	18	9	18	9	12.8
Number of patients treated	26	21	20	27	29	25	24.7
Number of treatments provided per practitioner	236	172	167	221	242	231	
Mean number of treatments per patient	9.1	8.2	8.4	8.2	8.3	9.2	8.6
Number of points used per practitioner	1450	1380	913	1530	2063	739	
Mean number of needles per treatment	9.1	12.0	9.8	10.4	10.4	6.0	9.6

Table 2 Diagnosis of key primary and secondary syndromes.

	Practitioner						Total
	A	B	C	D	E	F	
Primary syndromes							
Qi and Blood Stagnation	13	18	11	23	19	19	103 (70%)
Bi Syndrome	1	1	6	0	2	4	14 (9%)
Kidney Deficiency	8	3	2	3	8	2	26 (18%)
Secondary syndromes							
Qi and Blood Stagnation	7	2	3	4	7	4	27 (18%)
Bi Syndrome	5	2	4	3	2	11	27 (18%)
Kidney Deficiency	9	16	4	8	8	7	52 (35%)

Table 2 shows the key primary and secondary syndromes by practitioner. The predominant primary syndrome identified was Qi and Blood Stagnation, followed by Kidney Deficiency and then Bi Syndrome. The predominant secondary syndrome was Kidney Deficiency. A total of 88% of patients received a diagnosis of Qi and Blood Stagnation either as a primary or a secondary syndrome, 28% of Bi Syndrome and 53% of Kidney Deficiency.

A further diagnostic category of "Other" was also included as an option: a total of 20 patients had other syndromes diagnosed, most commonly Liver Qi Stagnation.

Diagnostic concordance

For the subgroup of patients given a blinded second diagnosis, the levels of agreement on syndromes, both congruent classifications and Cohen's Kappa, between the five practitioner pairs are presented in Table 3. A total of 87 (59%) comparisons were made

for the primary syndrome and 48 (32%) patients for the secondary syndrome.

Treatment data

The protocol allowed for up to 10 treatments per patient, the precise number being agreed between patient and practitioner. A total of 1269 treatments were provided, an average of 8.6 treatments per patient (range 1–10) and 9.6 needles per treatment (range 6–12). See Table 1 for variations between practitioners. The needles used were usually 25 or 40 mm long and between 0.20 and 0.30 mm in diameter.

From the practitioner interviews we know that needles were usually retained for around 20 min. However, all practitioners customised the treatment time for individual patients, usually within a range from 10 to 30 min, and exceptionally as low as 5 min and as high as 40 min. We also know that all practitioners attempted to attain *de qi* most of the

Table 3 Patient subgroup re-examined by single practitioner: inter-rater reliability: congruent classifications and Cohen's Kappa.

Practitioner pairs	Number of patients	Congruent classifications (%)	Cohen's Kappa	95% confidence intervals for Kappa	
Comparison of primary syndromes identified					
1	12	75	0.00	−0.98	0.98
2	17	71	0.59	0.28	0.89
3	20	60	0.41	0.10	0.73
4	15	47	0.02	−0.45	0.48
5	23	74	0.55	0.24	0.86
Comparison of secondary syndromes identified					
1	8	80	0.67	0.08	1.00
2	5	80	0.67	0.08	1.00
3	9	56	0.25	−0.30	0.80
4	11	64	0.44	0.01	0.88
5	15	60	0.38	−0.01	0.76

Table 4 Commonly used acupuncture points (used on at least 2% of all occasions).

	Practitioner						Total of all six practitioners
	A	B	C	D	E	F	
Total Bladder channel points	663	747	193	344	868	288	3103 (38.4%)
BL-23	106	152	4	117	6	64	449
BL-26	118	4	-	3	157	7	289
BL-32	43	50	8	12	57	39	209
BL-40	76	41	28	31	143	19	338
BL-53	43	31	3	8	191	52	328
BL-54	62	63	17	23	67	11	243
BL-60	47	50	17	33	70	11	228
Total Gall Bladder channel points	105	153	165	285	343	150	1201 (14.9%)
GB-30	34	46	2	98	128	21	329
GB-34	16	28	84	71	155	7	361
Total Kidney channel points	19	96	10	20	33	82	260 (3.2%)
KID-3	17	89	9	19	33	71	238
Total Huatuojiayi points	418	164	293	435	526	12	1848 (22.9%)
Huatuojiayi-L3	73	10	37	63	72	-	255
Huatuojiayi-L4	103	42	110	68	162	1	486
Huatuojiayi-L5	104	12	91	47	220	1	475
Ahshi points	22	15	15	91	90	2	235 (2.9%)
Shiqizhuxia (extra point)	52	40	1	46	71	-	210 (2.6%)
All other points	171	165	236	309	132	205	1218 (15.1%)

time. However, *de qi* was not always sought, and not always attained. In sensitive patients, needles were sometimes inserted without obtaining *qi*.

A total of 177 different acupuncture points were used throughout the trial. In Table 4 we present the most commonly used points, that is points used sufficiently often to constitute at least 2% of the total. Points were used both bilaterally and unilaterally. These data do not identify the commonly used point prescriptions, where a prescription would normally include local and distal points as well as points focused on underlying syndromes.

Auxiliary treatments were utilised by all practitioners, but to varying degrees. Moxa was used in 17.7% of treatments, mostly in the form of a "moxa box" (10.5%), but also "moxa on the needle" (5.3%) and "sparrow-pecking" with a moxa stick (1.9%). Moxa was used by all practitioners, who reported using it to target Cold Bi or Kidney Yang Deficiency. Massage (42.2%), acupressure (12.8%), cupping (4.5%) and Chinese herbs (3.4%) were also used. Practitioners reported using massage and acupressure often briefly and early on in the treatment session, to enable accurate localisation of the back pain, to refine point selection and to build a stronger therapeutic relationship with the patient. Self-help was also prescribed adjunctively by all practitioners, commonly diet (11.3%), yoga exercises (3.3%), other specific exercises (3.0%), and relaxation exercises (2.6%). Practitioners reported prescribing yoga or stretching exercises to move a

patient's low back *qi*, relaxation exercises to calm the *shen*, and dietary change to strengthen the Spleen, tonify the *qi* and help clear Damp Bi.

Association between diagnosis and treatment

Whether diagnosis determined treatment was assessed for one specific relationship: that between a diagnosis of Kidney Deficiency, whether as primary or secondary, and the use of the point KID-3, the only commonly used Kidney point, and one which is strongly indicated for Kidney Deficiency.⁴ Analysis using the Chi-square test showed a highly significant association ($\chi^2 = 13$, d.f. = 1, $P < 0.005$). For the other two syndromes, Qi and Blood Stagnation and Bi Syndrome, the relevant acupuncture points used would be expected to be local to the area of pain, and therefore not amenable to tests of association with the syndromes.

Discussion

We found that practitioners often diagnosed multiple syndromes, with two or more syndromes identified for 65% of patients, with an average of 1.7 syndromes per patient. The diagnosis of multiple syndromes was also a feature of research by Sherman et al.¹⁴ They described low back pain diagnosis and treatment of 88 patients in the acupuncture

treatment arm of a trial and 73 patients in the teaching clinic of an acupuncture school: 1.7 syndromes per patient were identified in the trial and 2.3 syndromes per patient in the teaching clinic. In our study, Qi and Blood Stagnation was diagnosed for 88% of all patients, Kidney Deficiency (53%) and Bi Syndrome (28%). Similarly Sherman et al.¹⁴ found that Qi and Blood Stagnation and Qi Stagnation (a variation of Qi and Blood Stagnation) were the two most commonly diagnosed syndromes, followed by Kidney Deficiency. A similar result was also found by Hogeboom et al.¹⁵ in which six acupuncturists identified Stagnation of any type as the most common, followed by Kidney Deficiency.

In terms of the reliability of the acupuncturists' diagnoses, the results of this study show reasonable agreement in terms of the percentage of congruent classifications: 47–75% for primary syndromes and 56–80% for secondary syndromes. Agreement is less good when based on the measure of reliability using Cohen's Kappa which lay in the respective ranges of 0.0 (no better than chance) to 0.59 ("moderate" agreement) and 0.25 ("poor") to 0.67 ("good" agreement). The apparent contradiction between these two levels of agreement can be explained in part by a weakness of the Kappa measure: when scores are distributed across the available response categories in a very unequal manner (in this case a high proportion of Stagnation of Qi and Blood as a primary syndrome), a high value of Kappa is almost impossible to attain, and a low value may not necessarily imply poor agreement.

With our design, congruence depended crucially on the diagnostic skills of the one practitioner-researcher who diagnosed the other practitioners' patients. Another factor that contributed to this level of agreement can be explained by what we know from the practitioners' interview reports: that the three key syndromes pre-selected at the outset were sometimes experienced as too constraining. Despite having agreed to this framework, practitioners also reported on the utility of channel based, *zang-fu* and location-based (e.g. trigger point) approaches. This limitation is reinforced by evidence from the literature that traditional acupuncture is not, and never has been, a "system of medicine" despite repeated attempts at standardisation,⁷ and is better understood as a fluid and evolving set of medical practices.¹⁶

In their recent research, Hogeboom et al.¹⁵ attempted to assess inter-rater reliability using a Latin square design in which six practitioners all diagnosed the same six patients, using diagnostic categories that they would have used in their normal practices. However, they were unable to do formal statistical tests on the reliability of the

categories Qi and Blood Stagnation and Kidney Deficiency because of the number of additional syndromes that were concurrently diagnosed for each patient. In their analysis of the reliability of some sub-groupings, they found generally poor levels of agreement.

Establishing robust data on concordance in this study was hampered by practical difficulties associated with arranging for all practitioners to undertake multiple diagnoses of each other's patient. It was feasible for only one of the six practitioners to act as an independent diagnostician for the majority of the other five's patients. Given the statistical needs for inter-rater reliability tests, and that three syndromes will yield nine response categories, the number of patients was relatively small. More practitioners involved as independent raters, as well as more patients receiving pairs of diagnoses, would have improved the statistical power of the results. Despite these limitations, the reliability in this trial is of a similar order to that found in those areas of biomedicine where multiple factors contribute to judgements on diagnosis, see for example, psychiatrists diagnosing DSM-III affective disorders¹⁷ and neurologists diagnosing cervicogenic headaches.¹⁸

From the data on treatment, we have found that points from the Bladder and the Gall-Bladder channels are much used, 38.4 and 14.9%, respectively, as well as Huatuoji points (22.9%). The most commonly used points were BL-23 and the two lowest Huatuoji points. From the recorded data and the practitioner interviews we know that points selected were often a combination of local points (such as BL-23, BL-26, BL-53, BL-54 and GB-30 as well as lower lumbar Huatuoji points) and distal points (such as BL-40, BL-60, GB-34 and GB-40). From the interviews, we know practitioners aimed to balance these points with others focused on underlying diagnostic categories, such as Bi Syndrome and Kidney Deficiency, key points for the latter being BL-23 and KID-3. There is considerable comparability between the acupuncture points used in this study and those documented in other research. Two of the five most commonly used points in this trial, BL-23 and BL-40, were the most commonly used points in the clinical trial that Sherman et al.¹⁴ undertook.

This study reports on six British Acupuncture Council practitioners all working in York in the UK. They all practice acupuncture primarily within the framework of traditional Chinese medicine. However there were important differences in their background and training: one practitioner trained for 5 years full-time in China, four attended one of two different 3-year part-time college courses in the UK, and one trained over 2 years with short

courses and an apprentice style clinical training in the UK as well as 1-month internship in China. In interpreting this data therefore, caution is needed when assessing how representative the treatments are.

The need for a comprehensive acupuncture trial protocol was discussed by the participating researchers and practitioners. A tension emerged between wanting to present a relatively simple and standardised framework, and wanting to develop a framework that incorporated the full complexity of acupuncture when individualised for patients and provided by experienced practitioners. The need for retaining sufficient replicability for the purposes of a clinical trial was accepted. In the interviews, practitioners reported that they selected the appropriate acupuncture points for treatment based on careful diagnosis and subtle palpation of the affected area. They emphasized that their needling of individual points required judgements about needling depth, nature of the *de qi* response, and specific needling techniques to move or strengthen *qi*. Such a trial protocol would require further research, utilising the results of this and other trials, with synthesis by clinical and trial experts, leading to a formalised consensus to ensure wider credibility among traditional acupuncturists.

Conclusion

Acupuncturists diagnosing patients with back pain commonly make multiple diagnoses, with Qi and Blood Stagnation being the most prevalent. Diagnostic concordance among practitioners was reasonable, given the pattern-based framework of traditional acupuncture. From the treatment data, some clear themes emerged. Most commonly used channels were the Bladder and Gall Bladder, while the most commonly used points were BL-23 and the lowest two lumbar Huatuoji points. Further research is needed to establish a flexible acupuncture trial protocol for patients presenting with low back pain.

Acknowledgements

Thanks are due to Liz Oswald for inputting the diagnostic and treatment data and Helen Wilkinson for administrative assistance. We also thank the participating patients and acupuncturists.

References

1. Birch S. Issues to consider in determining an adequate treatment in a clinical trial of acupuncture. *Complement Ther Med* 1997;5:8–12.
2. White AR, Ernst E. A trial method for assessing the adequacy of acupuncture treatments. *Altern Ther Health Med* 1998;4(6):66–71.
3. MacPherson H. Out of the laboratory and into the clinic: acupuncture research in the real world. *Clin Acupunct Oriental Med* 2000;1(2):97–100.
4. Birch S, Sherman K. Zhong yi acupuncture and low back pain: traditional Chinese medical differential diagnoses and treatments for chronic lumbar pain. *J Altern Complement Med* 1999;5(5):415–25.
5. Ergil KV. Editorial comment on Birch and Sherman. *J Altern Complement Med* 1999;5(5):427–8.
6. Sherman K, Lao L, Hopwood V, Kawahita K, MacPherson H. Matching acupuncture clinical study designs to research questions. *Clin Acupunct Oriental Med* 2002;3(1):12–5.
7. Hsu E. The transmission of Chinese medicine. Cambridge, UK: Cambridge University Press, 1999. p. 296.
8. MacPherson H, Kaptchuk T. Acupuncture in practice: case history insights from the West. Edinburgh: Churchill Livingstone, 1997. p. 482.
9. Thomas KJ, Fitter MJ. Evaluating complementary therapies for use in the National Health Service: "Horses for courses". Part 2. Alternative research strategies. *Complement Ther Med* 1997;5:94–8.
10. Thomas KJ, Fitter M, Brazier J, MacPherson H, Campbell M, Nicholl JP, Roman M. Longer term clinical and economic benefits of offering acupuncture to patients with chronic low back pain assessed as suitable for primary care management. *Complement Ther Med* 1999;7(2):91–100.
11. Steijnmans VW, Diletti E, Bomches B, Greis C, Solleder P. Interobserver agreement: Cohen's Kappa coefficient does not necessarily reflect the percentage of patients with congruent classifications. *Int J Pharmacol Ther* 1997;35(3):93–5.
12. Landis JR, Koch GG. The measure of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
13. Altman DG. Practical statistics for medical research. Chapman & Hall, 1991. pp. 404–408.
14. Sherman KJ, Cherkin DC, Hogeboom CJ. The diagnosis and treatment of patients with chronic low back pain by traditional Chinese medical acupuncturists. *J Altern Complement Med* 2001;7(6):641–50.
15. Hogeboom CJ, Sherman KJ, Cherkin DC. Variation in diagnosis and treatment of chronic low back pain by traditional Chinese medicine acupuncturists. *Complement Ther Med* 2000;9:154–66.
16. Scheid V. Chinese medicine in contemporary China: plurality and synthesis. Durham, USA: Duke University Press, 2002. p. 407.
17. Fuhrer R, Rouillon F, Lellouch J. Diagnostic reliability among French psychiatrists using DSM-III criteria. *Acta Psychol Scand* 1986;73:12–6.
18. Van Suijlekom HA, De Vet HC, Van Den Berg SG, Weber WE. Interobserver reliability in physical examination of the cervical spine in patients with headache. *Headache* 2000;40(7):581–6.