Measuring Practitioner Opinion on Adverse Reactions to Acupuncture

Jillian M McDowell[1] Dip Phty, MPhty

Gillian Johnson[2]* Dip Phty, MSc, PhD

Leigh Hale[2] BSc (Physio), MSc (Physio), PhD

1 Private practitioner

2 Centre for Physiotherapy Research, School of Physiotherapy,
University of Otago, Dunedin, New Zealand

ABSTRACT

Background: The terminology associated with the categorisation and reporting of adverse reactions to acupuncture (ARA) does not necessarily fit well with other universal adverse reaction reporting systems. Evaluation of practitioners' interpretation of ARA concepts and associated terminology is needed to ensure high quality reporting. This study reports the survey of acupuncture practitioners' opinion regarding ARA terminology, using a custom- designed Adverse Reactions to Acupuncture Questionnaire (ARAQ). The questionnaire was administered on two separate occasions to examine the consistency of question response in the context of usual clinical practice over time. Method: Twelve female acupuncturists (11 physiotherapists and one general practitioner), mean age 46.83 (±8.3) years completed the initial and follow-up ARAQ administered eight months apart. Analysis: Intra-rater agreement analysis was performed using intra-class coefficients (ICCs) for the visual analogue scales (VASs) and linear weighted Cohen's κ coefficients for the ranked questions. A systematic decision rule process analysed the repeated responses for the word categorisation task. Results: A statistical level of intra-rater agreement ($P \leq 0.05$) was achieved in 77% of the VAS questions (0.76–0.88) on repeated administration of the ARAQ. In the word categorisation task, 41% of acupuncture-related symptoms attained entry to identical adverse event domains on repeated responses. Whilst overall hierarchical weighting of preference responses were predominantly unchanged for the ranking questions the majority of κ coefficients for individual ranking tasks were low. Conclusion: The levels of practitioner intra-rater agreement in the VAS questions and word categorisation task relating to ARA performed most consistently over time. It is suggested that the styles of questions be carefully considered in future questionnaire development of this nature. The variation in agreement may be as a result of the style of question however it is acknowledged that the responses may also be confounded by changing opinions of experts as they acquire new or different knowledge.

KEYWORDS   questionnaire, question styles, opinion, acupuncture, adverse events, nomenclature.

Introduction

Acupuncture is a treatment modality not without risk[1] and the assessment of such risk is reliant on the accurate reporting of negative outcomes by practitioners. In turn, for the adverse reaction to acupuncture (ARA) report to be valid, there must be a consensus of opinion on the terminology and the definitions used to describe them. Standardised terms holding the same meaning and weighting between individual practitioners are needed for this process. Such agreement plays an important role in improving the quality of information available for analysis in adverse reporting systems[2], and a lack of consensus on terminology would appear to be a major deficit in the area of adverse reactions to acupuncture, preventing meaningful assimilation of data.[3]

While the term adverse reaction to acupuncture (ARA) has been defined in retrospective[4-7] and prospective studies[8–18], there is little factual information available regarding the views and opinions of practising acupuncturists, including their conceptualisation of ARA nomenclature or their views on ARA reporting related issues. The frequency of ARA events has been extensively reported.[8,9,11,14–17,19,20] However, the wide variation in the terminology, particularly in the definition of an ARA, limits any ability to draw comparisons between such studies.[3] The problem is compounded further by the loose use of nomenclature whereby terms such as 'adverse reaction', 'adverse event', 'adverse effect', 'complication' and 'side effect' are used synonymously and interchangeably throughout the literature.[3] While the concept of an 'adverse event' and an 'adverse reaction' have accepted definitions within the drug literature[21] they appear to be less consistently utilised in the acupuncture field. The clinical weighting of symptoms is also problematic; minor sequelae following acupuncture such as nausea, faint and fatigue[10] have been similarly categorised as adverse reactions together with the more serious consequences of infection[22], cardiac tamponade[23] and pseudoaneurysm.[24] Severity and seriousness must also be considered in context; a migraine following acupuncture, for example, may be a severe response but not necessarily a serious response.

An integral aim to the questionnaire design was to examine the quality of measured opinion, by assessing the level intra-rater agreement of a variety of question designs on two separate occasions. Visual analogue scales (VASs), ranking tasks and word categorisation tasks were used in the affective (factors that were likely to impact on the reporting of an ARA and level of agreement with statements about when an ARA should be reported) and reporting domains of the questionnaire. A key research concept explored in this questionnaire was whether or not acupuncturists conceptualise the same paradigm when considering issues relating to ARA and whether or not their opinion and viewpoints are consistently held over time.

In this paper, the development of a new ARAQ is described and the responses of health professionals who were considered expert in the field of acupuncture to three question styles contained in the questionnaire are assessed for intra-rater agreement longitudinally over a period of eight months on the repeated administration.

Methods

DEVELOPMENT OF THE ADVERSE REACTION TO ACUPUNCTURE QUESTIONNAIRE (ARAQ)

A literature search using the databases of Medline, AMED, CINAHL, PubMed, PEDro, DARE and the Cochrane Database of Systematic Reviews, from their inception to March 2005 and limited to English language and human, was carried out to gain an appreciation of the opinions and themes regarding ARA. Search keywords used were: acupuncture, adverse effect, adverse reaction, complication, questionnaire, terminology, survey and opinion. Of the 455 papers reviewed no independently validated general health instruments, questionnaires or disease activity indexes applicable to screening for adverse reaction to acupuncture or to the interpretation of the associated terminology could be located. Eight definitions[9,10,15,17,19,20] and 52 keywords[3,5-7,17,20,23,25-34] of relevance in the context of ARA were identified from these papers and these definitions formed the basis of the item generation in the development of the questionnaire.

QUESTIONNAIRE STRUCTURE

The ARAQ incorporated a range of tools to measure opinion and maximise response rates. The affective and reporting domains of the ARA were designed to capture the range of opinion, levels of preference and interpretation and incorporated three different question styles comprising two VAS questions, a word categorisation task, and three ranking questions.

The two VAS questions used semantic differentials with opposing phenomenon at the extremes of the scale[35] in which respondents had to indicate their response. Horizontal lines (100mm)[35] were selected and the anchor phrases exactly the same/completely different, strongly agree/strongly disagree or not at all/absolutely were used because their meanings were distinct, with maximum distance between meanings to avoid overlap.[36] The first VAS question comprised five scales and examined the synonymy between interchangeable ARA terms (adverse event, side effect, adverse effect, complication and medical error) and that of adverse reaction. The second VAS question examined respondent's views pertaining to their reporting practice of an ARA over six different scales.

A purpose-designed word categorisation task based on the methodology and criteria outlined by Fernandez[37] was utilised to examine the ability of the respondents to determine membership of descriptors to the adverse reaction domains (malpractice, complication, adverse reaction or known side effect). In this task, respondents were required to assign 39 acupuncture-related signs, symptoms and sequelae (S, S & S) to one of four ARA domains or, to the alternative options of don't know or other–please describe using pre-coded responses.

The three ranking questions explored participant opinion utilising levels of preference. In the first question respondents were required to rank their first, second and third preferences for an operational definition of an ARA out of eight possible options. The second question explored the respondent's perceived levels of seriousness related to six word domains (malpractice, complication, adverse reaction & adverse effect, adverse event, side effect) according to their perceived level of seriousness. The third question required respondents to prioritise six clinically relevant factors influencing the decision to report an ARA.

The language used throughout the questionnaire was targeted at a professional group with above average verbal skills and the layout of the questions was spaced to ensure none were missed due to visual clutter. Instructions for question responses were included within the questionnaire and the keywords in both the instructions and the questions were placed in bold typeface. A research assistant separated the consent forms from the questionnaires to ensure anonymity from the researcher.

STUDY SAMPLE

A purposive sample of 12 health professionals deemed to be expert in acupuncture on the basis of their academic qualifications, clinical and teaching experience in the area of acupuncture were approached to take part in this study. Initial consent by the respondents to be involved was given by electronic mail followed by written informed consent. The group was surveyed on two occasions eight months apart to (1) reduce any practice effect of undertaking the questionnaire and (2) to pragmatically provide a reasonable time gap in which to request busy clinicians to dedicate time to repeat the questionnaire. A secondary rationale was to longitudinally validate the different questionnaire approaches for future use with a larger national group. Participants were blinded to their responses in the initial ARAQ and those of their peers when undertaking the follow-up ARAQ. The content of the initial and follow-up questionnaire was identical. Ethics approval for this study was granted by the local Human Disability and Ethics Committee.

STATISTICAL ANALYSIS

Demographic data from the initial ARAQ were presented as means/standard deviations or percentage values. The VAS data from the initial and follow-up ARAQs were presented as median and inter-quartile ranges (IQR). The initial and follow-up responses to the word categorisation task were evaluated by two decision rule processes.[37] The first process singled out unclassifiable descriptors by identifying those S, S & S terms in which the relative frequency of the sum of responses for the options of don't know, multiple response and no responses exceeded 60%. In the second step, the remaining descriptors were assigned to one of four ARA domains using a three step decision rule based on criteria of absolute frequency, relative frequency and uni-modality characteristics.[37] In step one, the absolute frequency of each S, S & S had to exceed one response under the domain or be removed before the second step. Step two required the relative frequency of the S, S & S categorisations to exceed 50% under a domain and finally, step three excluded any S, S & S which achieved greater than 33% assignation under any other domain (indicating bi or multi-modality). The percentage of common descriptors assigned to the same word domain in the initial and follow-up questionnaires was determined.

In the three ranking tasks the respondents' hierarchical 'order of choice' was obtained by allocating weighted points for each item. For example, in the first ranking question (Table 4) (seeking the most appropriate operational definition of an ARA), the first choice was weighted by five points, the second by three points and the third by a single point.

Intra-rater reliability assessing levels of agreement between the initial and follow-up responses for the VAS questions was carried out using single measures intra-class correlation coefficients (ICC) using a two-way random model, and for the ranked questions, linear weighted Cohen's κ coefficients, along with their associated 95% confidence intervals were used. The ICC values were interpreted according to the criteria of Fleiss (ICC < 0.40 poor; 0.40–0.75 fair to good; 0.75 and above excellent).[38] For the κ values; the criteria set out by Landis & Koch[39] (0.00 to 0.20 slight agreement; 0.21 to 0.40 fair agreement; 0.41 to 0.60 strong agreement) were used. Data analyses were performed using MedCalc version 11.4.1.0 (Medcalc Software, Mariakerke, Belgium). Alpha levels of significance were set at $p \leq 0.05$.

Results

DEMOGRAPHICS

Twelve female acupuncturists (11 physiotherapists and one general practitioner) with a mean age 46.83 (±8.3) years completed the initial and follow-up questionnaires. All respondents were practising within the private sector and had a mean of 14.5 (±4.5) years of clinical experience and of these, nine (75%) held professional qualifications in acupuncture, while three others (25%) held university or equivalent

qualifications. The respondents were involved in the teaching of acupuncture in programmes which had been formally endorsed by the Physiotherapy Acupuncture Association of New Zealand or the Medical Acupuncture Association of New Zealand. All respondents had observed an ARA in their own clinical practice and 25% had observed adverse reactions in their colleagues' patients. Ten (83%) respondents held an ARA reporting policy in their practice and 58% had reported an adverse reaction during the course of their career.

DATA COMPLETENESS

From the initial questionnaire (including demographic data) data for a total of 76 possible responses from 15 questions per questionnaire were tabulated. Missing data for responses were low (range 0 to 11% per question). A total 20 of a possible 912 responses were missing (2%). Fifty percent of respondents endorsed all 76 responses (100% complete data), 25% missed one response and 17% missed two responses.

From the follow-up questionnaire (excluding demographic data), data for a total of 65 possible responses from six questions per questionnaire were tabulated. Missing data for responses were low (range 0 to 15% per question). A total 50 of a possible 780 responses were missing (6%). Sixty-seven percent of respondents endorsed all 65 responses (100% complete data), 8% missed one response, and 8% missed out three responses. One respondent (8%) missed 34 responses.

VISUAL ANALOGUE SCALES

The median values (IRQs) for the five VASs regarding synonymy with the term adverse reaction and the six VASs reporting practice of adverse reactions to acupuncture in the initial and follow-up questionnaires are presented in Tables 1 and 2 respectively. The ICC values for the responses examining the synonymy of terminology ranged from 0.17–0.80 (Table 1) and for statements on reporting practice from -0.58–0.88 (Table 2) with 77% of the scales in the two VAS questions reaching statistical levels of agreement ($p \leq 0.05$).

WORD CATEGORISATION TASK

In the word categorisation task, 16 out of a possible 39 (41%) descriptors achieved membership to identical domains of known side effect, malpractice and adverse reaction in the initial and follow-up ARAQ (Table 3). Two descriptors (headache & new symptoms) met the entry criteria for the known side effect domain, one descriptor met the entry criteria for adverse reaction (psychiatric disturbance) and five others (endocarditis, haematoma, hepatitis, pleural empyema & osteomyelitis) that of the malpractice domain in one but not on both occasions. No descriptors in either data set met the entry criteria for the complication domain in the initial or follow-up questionnaire.

RANKING TASKS

Consistent endorsement based on hierarchically weighted points was achieved for two out of three responses examining choice of an operational definition for an adverse reactions to acupuncture (Table 4) and five out of six responses for the perception of severity of terms (Table 5).

Consistent endorsement for all responses regarding factors influencing the decision to report an adverse reaction to acupuncture were also achieved (Table 6). The weighted Kappa values for the ranking preference for the operational definition of an ARA were fair to strong (0.23–0.31, Table 4) and slight to strong (-0.09–0.47) for the ranking of terminology in terms of severity (Table 5). The Kappa values for the factors influencing the decision to report an ARA also ranged from slight to strong over the six options (-0.08–0.55, Table 6).

Discussion

Our analysis showed that the responses from the VAS questions were more consistent over time compared to the performance of the ranking tasks as measured by intra-rater agreement levels between the initial and follow-up ARAQ. The application of stringent criteria to the word categorisation task also yielded clinically meaningful sets of descriptors for adverse event classification domains of known side effect, malpractice and adverse reaction (Table 3) with 41% of descriptors achieving entry to the end-point ARA. This latter result compares favourably with the original methodology described in the pain questionnaire on which it was based.[37]

The wide range of options offered in the ranking tasks and small sample size may explain the failure of the Kappa values to achieve stronger agreement levels. As a consequence, a recommendation for future questionnaires would be to reduce the number of options for the ranking questions in the case where more stringent analysis of the information is required. Alternatively more complete descriptors could have been supplied for the individual terms so as to remove any possibility of ambiguity. However an assumption was made that a perceived global lack of knowledge and poor clinical understanding of ARA issues would be minimised by confining the study cohort to expert acupuncturists. This assumption was also based on the premise that the more experienced a respondent is in thinking about a given topic the better they are able to think anew about that topic and to answer a relevant question.[40] It was also assumed that the expert opinion would be more likely to be fixed and stable over time compared with novice practitioners.

With an interval of eight months between the two questionnaires it is acknowledged that the relatively wide range of responses in the VAS questions (Tables 1 & 2) may represent a combination

TABLE 1   Synonymy of adverse reaction terms-median (inter-quartile range) for five visual analogue scale items in the initial and follow-up Adverse Reactions to Acupuncture Questionnaire (ARAQ) and intra-class correlation coefficients (ICC)

Synonymy with the term *adverse reaction*	Median (inter-quartile range)		ICC (95% confidence interval)
	Initial ARAQ	Follow-up ARAQ	
Adverse effect	4.50 (-0.40,4.85)	3.65 (1.88,4.48)	0.67 (-0.15,0.90)*
Averse event	2.00 (-1.45,4.60)	0.75 (-1.83,4.58)	0.33 (-1.70,0.81)
Complication	-1.5 (-2.95,2.90)	-0.55 (-2.63,0.88)	0.63 (-3.5,0.89)
Side effect	-2.00 (-3.5,-1.00)	-2.25 (-4.00,-0.08)	0.80 (0.29,0.94)**
Medical error	-4.70 (-5.00,-3.7)	-4.50 (-4.70,0.55)	0.17 (-1.37,0.74)

*P ≤ 0.05, **P ≤ 0.01   5 agree synonymous, -5 disagree synonymous

TABLE 2   Reporting practice of adverse reactions to acupuncture-median (inter-quartile range) for six visual analogue scales from initial and follow-up Adverse Reactions to Acupuncture Questionnaire (ARAQ) and their intra-class correlation coefficients (ICC)

Statement	Median (inter-quartile range)		ICC (95% confidence interval)
	Initial ARAQ	Follow-up ARAQ	
1 All Adverse reactions to acupuncture which the patient perceives as adverse should be reported	-1.90 (6.15)	-2.40 (8.50)	0.67 (-0.15,0.90)*
2 All adverse reactions to acupuncture which require medical attention should be reported	4.50 (1.90)	4.70 (1.15)	-0.58 (-7.34,0.59)
3 All adverse reactions to acupuncture that alter a patient's function for more than 24 hours should be reported	2.40 (3.65)	3.80 (5.38)	0.71 (0.10,0.92)*
4 All adverse reactions to acupuncture that are unexpected should be reported	-3.30 (6.20)	-2.00 (5.00)	0.63 (-0.13,0.89)*
5 All physiological reactions to acupuncture that are amplified should be reported	-2.50 (4.35)	-1.80 (4.50)	0.88 (0.58,0.97)***
6 All reactions that occur due to practitioner error should be reported as an adverse reaction to acupuncture	0.50 (5.85)	-0.55 (3.51)	0.74 (0.18,0.92)**

*P≤0.05, **P≤0.01, ***P≤0.001   5 agree with statement, -5 disagree with statement

TABLE 3   Word classification task assigning adverse reactions to acupuncture into domains

Known side effect	Adverse reaction	Malpractice	Unclassifiable signs, symptoms & sequelae
Aggravation of symptoms	Convulsion	Cardiac tamponade	Angina pectoris
Bruising	Psychiatric disturbance*	Endocarditis*	Galactorrhoea
Feeling cold	Seizure	Forgotten needle	Granuloma*
Headache*		Hepatitis*	
Insomnia		Osteomyelitis*	
Nausea		Peritonitis	
New symptoms*		Pleural empyema*	
Pain at the needle site		Pneumothorax	
Point bleeding			
Sleepiness/fatigue			

* denotes classification in one occasion but not two, whilst the remainder were classified into the same domain in the initial and follow-up questionnaire.

TABLE 4   Ranking order (hierarchical weighting points) for an operational definition of an adverse reaction to acupuncture from initial and follow-up Adverse Reactions to Acupuncture Questionnaire (ARAQ) and weighted Cohen's κ coefficients

Ranked definition preference	Initial ARAQ (weighting points maximum possible 60)	Follow-up ARAQ (weighting points maximum possible 60)	Weighted κ values (95% confidence intervals)
First preference: Any adverse effect possibly related to acupuncture making treatment necessary or severely interfering with the patient's well being.	1st (25)	1st (42)	0.23 (-0.11,0.58)
Second preference: A non-intended effect of acupuncture that may threaten the patient's life.	2nd (24)	2nd (39)	0.14 (-0.18,0.45)
Third preference: Any ill effect no matter how small that is unintended and non therapeutic, even if not unexpected.	3rd (15)		0.31 (0.02,0.60)
Those reactions observable in standard practice, of a systemic or local nature, distinct from therapist negligence.		3rd (12)	0.31 (0.02,0.60)

TABLE 5   Ranking opinion (hierarchical weighting points) regarding severity of terms from the initial and follow-up Adverse Reactions to Acupuncture Questionnaire (ARAQ) and weighted Cohen's κ Coefficients

Key terms	Initial ARAQ (weighting points maximum possible $n = 132$)	Follow-up ARAQ (weighting points maximum possible $n = 132$)	Weighted κ Coefficients (95% confidence intervals)
Malpractice	1st (122)	1st (111)	-0.09 (-0.22,0.03)
Adverse event	2nd (78)	2nd (81)	-0.29 (-0.59,0.00)
Adverse reaction	3= (72)	3rd (67)	0.08 (-0.32,0.49)
Complication	3= (72)	4th (65)	0.26 (-0.22,0.74)
Adverse effect	5th (64)	5th (49)	-0.02 (-0.31,0.27)
Side effect	6th (24)	6th (23)	0.47 (-0.01,0.96)

TABLE 6   Ranking opinion (hierarchical weighting points) of factors influencing the decision to report an adverse reaction to acupuncture from initial and follow-up Adverse Reactions to Acupuncture Questionnaires (ARAQ) and weighted Cohen's κ coefficients.

Factor	Initial ARAQ (maximum weighted points =132)	Follow-up ARAQ (maximum weighted points =132)	Weighted κ coefficients (95% confidence intervals)
Medical intervention	1st (120)	1st (120)	0.55 (0.17,0.94)
Loss of function	2nd (88)	2nd (88)	0.04 (-0.29,0.37)
Permanence	3rd (76)	3rd (83)	0.32 (-0.05,0.68)
Severity	4th (63)	4th (67)	0.04 (-0.29,0.37)
Duration	5th (36)	5th (34)	0.18 (-0.25,0.60)
Perception of the patient	6th (22)	6th (18)	0.08 (-0.33,0.48)

of altering data completeness levels, test-retest properties of the question design and a possible change in the experts opinion.[41] The observed changes in opinion may have arisen from external events such as media exposure, topics covered in continuing professional development programmes or the event of an ARA within their patient cohort. Hence, the lack of agreement in the performance attributes of the question styles examined in this study cannot be attributed solely to the question styles alone.

One limitation of the study was the ranking questions sought practitioner interpretation of the terminology itself without a supporting contextual framework. For example 'malpractice' was considered to be a more serious than an 'adverse event' (Table 5). It could be argued that many adverse events are more life threatening than some forms of malpractice. However for a practitioner the reporting of negligence or error (malpractice) might seem to have more serious repercussions than an adverse event which has a recognised low degree of preventability.[42] In another ranking question a clearly ranked preference for two operational definitions of an ARA emerged out of eight possible options provided to the respondents from the literature (Table 4).

In the word categorisation task, respondents were required to make a clinical decision for each descriptor and in doing so, filtered out rare outliers into the unclassifiable domain and incorporated the better known descriptors such as nausea[7,16] and bruising[7] into appropriate ARA domains (Table 3). One exception was that of fainting, which despite being one of the most frequently cited adverse symptoms to acupuncture[7] showed multimodal distribution across all domains and hence failed to achieve membership to any of the domain options. The word categorisation task was unique in its ability to describe the conceptualised domains with respondents showing that they were much more readily able to classify S, S & S which were more extreme (less and more severe) in nature. The descriptors in each adverse reaction domain were not definitive but serve to provide a formalised framework in which to communicate findings in a simple, effective manner for further discussion and policy formation.

The questionnaire includes questions pertaining to types of adverse events but also events of differing seriousness. Whilst the impetus for reporting reviewed in the questionnaire includes the factors of medical intervention, loss of function, severity, duration and patient perception the concept of seriousness may infiltrate all of these factors. The homogeneity of the population sample is seen as a strength in the design thereby reducing the confounding factor of variation in practitioner expertise. The lack of patient input as to what constitutes an adverse reaction is noted however the scope of the study was directed towards practitioner opinion rather than seeking consumer viewpoint or overall consensus agreement on statements relating to ARA.

Conclusion

The levels of practitioner intra-rater agreement in VAS questions and word categorisation task relating to ARA performed most consistently. It is suggested that the styles of questions be carefully considered in future questionnaire development of this nature. In the face of usual clinical practice over time the variation in agreement may be as a result of the style of question however it is acknowledged that the responses may also be confounded by the changing opinions of experts as they acquire new or different knowledge.

Acknowledgement

Disclosure Statement

No competing financial interests exist.

Clinical Commentary

This study examines the practitioner's view of a definition of an adverse reaction to acupuncture, the language used to describe an adverse reaction and factors influencing their reporting practice. A questionnaire to measure the practitioner's opinion in a user-friendly manner was developed and the responses of 12 health professionals were examined on its repeated administration, eight months apart. Visual analogue scale type questions, ranking questions and a word categorisation task were used to measure and gauge opinion. Ultimately, agreement between practitioners on what constitutes an adverse reaction to acupuncture will serve to strengthen the quality of reporting systems.

References

1. Witt C, Pach D, Brinkhaus B, Wruck K, Tag B, Mank S, et al. Safety of acupuncture: Results of a prospective observational study with 229,230 patients and introduction of a new medical information and consent form. Forschende Komplementarmedizin und Klassische Naturheilkunde. 2009 Apr;16(2):91–7.

2. Vincent C. Reasons for not reporting adverse events: an empirical study. J. Eval. Clin. Pract. 1999;5(1):13–21.

3. White A. A cumulative review of the range and incidence of significant adverse events associated with acupuncture. Acupuncture in Medicine. 2004;22(3):122–33.

4. Norheim AJ, Fonnebo V. A survey of acupuncture patients: results from a questionnaire among a random sample in the general population in Norway. Complement. Ther. Med. 2000;8(3):187–92.

5. Norheim AJ, Fonnebo V. Acupuncture adverse effects are more than occasional case reports: results from questionnaires among 1135 randomly selected doctors, and 197 acupuncturists. Complement. Ther. Med. 1996;4:8–13.

6. Grabowska C, Squire C, MacRae E, Robinson N. Provision of acupuncture in a university health centre - a clinical audit. Complement. Ther. Nurs. Midwifery. 2003;9(1):14–19.

7. Bensoussan A, Myers SP, Carlton AL. Risks associated with the practice of traditional Chinese medicine: an Australian study. Arch. Fam. Med. 2000;9(10):1071–8.

8. Yamashita H, Tsukayama H, Hori N, Kimura T, Tanno Y. Incidence of adverse reactions associated with acupuncture. J. Altern. Complement. Med. 2000;6(4):345–50.

9. Yamashita H, Tsukayama H, Tanno Y, Nishijo K. Adverse events in acupuncture and moxibustion treatment: a six-year survey at a national clinic in Japan. J. Altern. Complement. Med. 1999;5(3):229–36.

10. White A, Hayhoe S, Hart A, Ernst E. Adverse events following acupuncture: prospective survey of 32 000 consultations with doctors and physiotherapists. Br. Med. J. 2001;323(7311):485–6.

11. Ernst G, Strzyz H, Hagmeister H. Incidence of adverse effects during acupuncture therapy - a multicentre survey. Complement. Ther. Med. 2003;11(2):93–7.

12. MacPherson H, Thomas K. Short term reactions to acupuncture - a cross-sectional survey of patient reports. Acupuncture in Medicine. 2005 September 1, 2005;23(3):112–120.

13. Endres H, Molsberger A, Lungenhausen M, Trampisch HJ. An international standard for verifying the accuracy of serious adverse event reporting: the example of an acupuncture study of 190,924 patients. Eur. J. Med. Res. 2004;9:545–51.

14. MacPherson H, Scullion A, Thomas KJ, Walters S. Patient reports of adverse events associated with acupuncture treatment: a prospective national survey. Quality and Safety in Health Care. 2004;13:349–55.

15. Melchart D, Weidenhammer W, Streng A, Reitmayr S, Hoppe A, Ernst E, et al. Prospective investigation of adverse effects of acupuncture in 97 733 patients. Arch. Intern. Med. 2004;164:104–5.

16. MacPherson H, Thomas K, Walters S, Fitter M. A prospective survey of adverse events and treatment reactions following 34 000 consultations with professional acupuncturists. Acupuncture in Medicine. 2001;19(2):93–102.

17. Odsberg A, Schill U, Haker E. Acupuncture treatment: side effects and complications reported by Swedish physiotherapists. Complement. Ther. Med. 2001;9(1):17–20.

18. White A. The safety of acupuncture - evidence from the UK. Acupuncture in Medicine. 2006 17 February 2008;24 (Suppl):S53–57.

19. Lewith GT, White P. Side effects associated with acupuncture and a sham treatment: perhaps we should take a closer look at what is really responsible? J. Altern. Complement. Med. 2003;9(1):16–19.

20. White A, Ernst E. Adverse events associated with acupuncture reported in 2000. Acupuncture in Medicine. 2001;19(2):136–7.

21. Medicines and Healthcare Products Regulatory Agency. Yellow Card Scheme. [online]; 2005 [updated 2005 Jan 17; cited 2005 Aug 7]; Available from: <http://www.yellowcard.gov.uk/>.

22. Shah N, Hing C, Tucker K, Crawford R. Infected compartment syndrome after acupuncture. Acupuncture in Medicine. 2002;20(2–3):105–6.

23. Cheng TO. Infective endocarditis, cardiac tamponade and AIDS as serious complications of acupuncture. Arch. Intern. Med. 2004;164:1464.

24. Kao CL, Chang JP. Pseudoaneurysm of the popliteal artery: a rare sequela of acupuncture. Tex. Heart Inst. J. 2002;29(2):126–9.

25. Chen CY, Liu GC, Sheu RS, Huang CL. Bacterial meningitis and lumbar epidural hematoma due to lumbar acupunctures: a case report. Kaohsiung Journal of Medical Sciences. 1997;13(5):328–31.

26. Ernst E, White AR. Acupuncture may be associated with serious adverse events. Br. Med. J. 2000;320(7233):513–4.

27. Saw A, Kwan MK, Sengupta S. Necrotising fasciitis: a life-threatening complication of acupuncture in a patient with diabetes mellitus. Singapore Med. J. 2004;45(4):180–2.

28. Ernst E, White AR. Prospective studies of the safety of acupuncture: a systematic review. Am. J. Med. 2001;10(6):481–5.

29. Yamashita H, Tsukayama H, White AR, Tanno Y, Sugishita C, Ernst E. Systematic review of adverse events following acupuncture: the Japanese literature. Complement. Ther. Med. 2001;9(2):98–104.

30. Alani RM, Busam K. Acupuncture granulomas. J. Am. Acad. Dermatol. 2001;45(6 Suppl):225–6.

31. Campbell A, Macglashan J. Acupuncture-induced galactorrhoea - a case report. Acupuncture in Medicine. 2005 September 1, 2005;23(3):146.

32. Cole M, Shen J, Hommer D. Convulsive syncope associated with acupuncture. Am. J. Med. Sci. 2002;324(5):288–9.

33. Drago F, Rongioletti F, Battifoglio ML, Rebora A. Localised lipoatrophy after acupuncture [comment]. Lancet. 1996;347(9013):1484.

34. Daivajna S, Jones A, O'Malley M, Mehdian H. Unilateral septic arthritis of a lumbar facet joint secondary to acupuncture treatment - a case report. Acupuncture in Medicine. 2004;22(3):152–5.

35. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. Res. Nurs. Health. 1990;13:227–36.

36. Newstead SE, Pollard P, Riezebos D. The effect of set size on the interpretation of quantifiers used in rating scales. Appl. Ergon. 1987;18(3):178–82.

37. Fernandez E, Boyle GJ. Affective and evaluative descriptors of pain in the McGill Pain Questionnaire: reduction and reorganization. Journal of Pain. 2001;2(6):318–25.

38. Fleiss JL. The design and analysis of clinical experiments. New York: Wiley; 1985.

39. Landis JR, Kock GG. An application of hierarchical Kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 1977;33(2):363–74.

40. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology. 1991;5:213–36.

41. Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. Health & Quality of Life Outcomes. 2004 Mar 23;2:16.

42. Davis P, Lay-Yee R, Briant R, Scott A, Schug S. Adverse events in New Zealand public hospitals II: preventability and clinical context. Journal [serial on the Internet]. 2003 Date; 116(1183).