A PILOT STUDY TO DEVELOP AND VALIDATE A TRADITIONAL CHINESE MEDICINE (TCM) QUESTIONNAIRE

A Health Status Instrument for TCM Assessment in Patients with Osteoarthritis (OA) of the Hip or Knee

Ping Wang

A thesis submitted to Auckland University of Technology in partial fulfilment of the degree of Master of Health Science

December 2004

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CERTIFICATE OF AUTHORSHIP

"I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the qualification of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgment is made in the acknowledgments."

Ping Wang

06 December 2004

ACKNOWLEDGEMENTS

I would like to formally acknowledge my supervisors, Peter Larmer and Joan Campbell, for their expert guidance and support. Their guidance and encouragement has led me to achieve the success of this study, and in so doing I had a wonderful and very enjoyable learning experience.

Many thanks to Neil Binnie for his professional advice on the statistical analysis, David Parker, Marcus Henning, and Shem Foo for their comments on the first draft, the staff from the Library of the Auckland University of Technology, and the staff from the Faculty of Health for their kind assistance.

I would also like to thank the subjects for their participation in this research, the practitioners and receptionists for their time and willingness to help. In particular I would like to thank Dr Alexander Chan, Dr Hong Kong, and Mrs Sue Hinds for their generous help with subject recruitment.

I would like to acknowledge the ethic approval from the Auckland University of Technology Ethics Committee (AUTEC) for this research. Approval Number: 03/63, Date of Approval: 15th of July 2003.

I would also like to acknowledge the Medical Outcomes Trust and QualityMetric Incorporated, who granted a royalty free license to reproduce the SF-36® Health Survey, perform data collection, perform data entry, and/or use scoring algorithms and normative data in this research. License Number: F1-020803-12953, Date of License: 13th of March 2003.

And I would like to acknowledge the SPSS 12.0.1 for Windows computer package from the Computer Lab of the Auckland University of Technology for statistical analysis.

And most of all, I would like to thank my wife Bessie for her endless support and encouragement, and my children James and John, for their understanding and patience throughout the study.

ABSTRACT

Background

Research suggests acupuncture is potentially an effective treatment for osteoarthritis (OA) of the hip or knee. Essential for the evaluation of Chinese acupuncture treatment is the availability of a reliable and valid measurement. However, currently there is no appropriate measurement instrument validated within traditional Chinese medicine (TCM) concepts and frameworks.

Objectives

To develop and validate a TCM questionnaire as a health status instrument for TCM assessment in patients with OA of the hip and knee.

Methods

The TCM questionnaire was developed from TCM theory and clinical experience. The questionnaire was examined by experts for content and face validity and pre-tested on a volunteer sample of three subjects. The developed questionnaire was validated on a convenience sample of ten subjects from six different clinical settings in Auckland region. The practitioner or receptionist from the selected clinical sites handed out the questionnaire package to their patients who fulfilled the study criteria. Each patient (subject) completed the questionnaire on their arrival and the re-testing questionnaire at a two-week interval.

The reliability of the questionnaire was estimated by examining the internal consistency reliability (Cronbach's alpha statistic) and test-retest reliability (Intra-class correlation coefficients). The content validity of the questionnaire was examined by literature review, interviews with patients, and experts' judgement. The construct validity was estimated by the methods of known groups, correlations between scales, and

correlations with the SF-36 health survey. The success of the grouping or scaling of the questionnaire was estimated by examining the item (i.e. question) internal consistency and item (i.e. question) discriminant validity.

Results

The TCM questionnaire scales corresponded to the "eight principal syndromes", "ten questions", and "eight patterns of OA" within TCM concepts and frameworks. Internal consistency reliability (Cronbach's alpha) was above .70 for all scales on both occasions of the first test and the second test. Test-retest reliability (intra-class correlation coefficient) for each scale was also above .70 for all scales, except the exterior (EXT) scale, which was .44. Moderate associations were found between the age of subjects and the scores of the interior (INT) scale and summary (SUM) scale. There was a significant difference between the groups of use and non-use of on-going medication in the EXT scale scores on the first test, $p = .012$. However, this significant difference was not found on the second test. As expected, strong or moderate associations were found between the TCM questionnaire and SF-36 comparable scales.

Conclusions

The TCM questionnaire was developed within TCM concepts and frameworks. The questionnaire contains 23 items with two main scales (the EXT scale and the INT scale) and one additional scale (the SUM scale). It takes approximately five minutes to complete and is entirely self-administered. Results from this pilot study indicate that this TCM questionnaire might have adequate reliability and validity. Therefore, the questionnaire has potential usage as an outcome measurement instrument for the assessment of TCM in the patients with OA of the hip or knee. For this application to be possible, the questionnaire needs further development and validation with a larger sample of patients who have a variety of OA conditions.

CHAPTER 1 INTRODUCTION

1.1 Statement of the problem

Osteoarthritis (OA) is a leading cause of chronic disability amongst the elderly, in large part due to knee and hip involvement (Felson, 1988; Guccione, 1997; Sharma, 2001). Due to limitations of current Western Medicine (WM) in the treatment of OA, interest has arisen in acupuncture for pain relief (Yamauchi, 1976). Because of the different health concepts and understanding of OA in WM and Traditional Chinese Medicine (TCM), any research to evaluate the effectiveness of Chinese acupuncture treatment must be conducted within TCM frameworks. However, a literature search failed to identify any TCM assessment or measurements instrument developed and validated within TCM concepts and frameworks, despite TCM and acupuncture having been practised for over 4000 years (Cheng, 1997).

1.2 Significance of the problem

Recent health care trends place increasing emphasis on evidence-based practice (EBP). Central to identifying the best evidence is the availability of research that establishes reliable, valid, and appropriate measurement (Jenkinson & McGee, 1998). Although there are a number of health status instruments already in existence, they do not apply to

the TCM theoretical model of health (Lewith, Jonas, & Walach, 2002). If no existing measurement instrument can adequately reflect the construct in question, then it is good practice not to use a surrogate measurement but to construct a new measurement instrument (Wittmann & Walach, 2002). Therefore, development of an appropriate measurement or assessment is an essential step in the evaluation of TCM acupuncture in the treatment of OA.

1.3 Purpose and structure of the pilot study

The purpose of this study is to develop and validate a TCM questionnaire as a health status instrument for TCM assessment in patients with OA of the hip or knee.

To develop and validate the TCM questionnaire, a comprehensive and critical review of literature is performed in Chapter Two. This review includes OA in different models of WM and TCM, requirements of a measure instrument, as well as relevant TCM concepts and frameworks.

Methodology applied in development and validation of the questionnaire is detailed in Chapter Three. The methods include how to develop and validate the questionnaire. The questionnaire development involves item (often a question) selection, questionnaire formats, and pre-testing. The questionnaire validation involves subject selection, data collection, and the methods of statistical analysis.

Results of this pilot study are presented in Chapter Four. The results include the data completeness of the questionnaire, distribution of responses to items, most important

results of reliability and validity tests, and tests of success of question grouping (i.e. scaling).

Discussions of the results in terms of the purpose of the study are detailed in Chapter Five. The discussion involves critical interpretation of the results, comment about the clinical significance, limitations of this study, and recommendations for future research.

CHAPTER 2 REVIEW OF LITERATURE

2.1 Introduction

The objective of this review is to outline the different understanding of osteoarthritis (OA) of the hip and knee in Western Medicine (WM) and Traditional Chinese Medicine (TCM). In light of these features, further objectives are to identify the different concepts of OA patient health status in TCM and WM, the limitations of current measurement instruments, and to consider how OA might be measured within TCM concepts and frameworks. This review also explores the limitations of the WM treatment and the potential benefit of TCM acupuncture therapy that might ameliorate OA pain and disability.

2.2 Osteoarthritis in Western Medicine (WM)

2.2.1 Prevalence and incidence

Osteoarthritis (OA), also often called degenerative joint disease or osteoarthrosis, is the most common form of arthritis (Hochberg, 2001; Kelsey & Hochberg, 1988; Lawrence, Hochberg, Kelsey, McDuffie, & Medsger, 1989). OA is also a leading cause of chronic disability amongst the elderly, in large part due to knee and hip involvement (Felson, 1988; Guccione, 1997; Sharma, 2001). It is estimated that at least 85% of persons 70 to

79 years of age have diagnosable OA (Cooper, 1994). The proportion in the population aged 65 years and older is expected to rise (Thompson, 1988). This change in population structure, together with the acknowledged association between OA and increasing age, means that in the future OA is likely to be recognised as a major public health problem and strain on healthcare resources (Badley, 1991, 1995; Finocchiaro, Abramson, & Ryan, 1996; Forbes, 1997; Gabriel, Crowson, & O'Fallon, 1995; Olshansky & Cassel, 1997; Verbrugge & Patrick, 1995).

The prevalence of OA of the hip vary from 3.1% (age 55-74 years) (Tepper & Hochberg, 1993) to 10% (age over 85 years) (Danielsson & Lindberg, 1997). OA of the hip is more frequent in males than females in the 45-64 year age group (Silman & Hochberg, 1993).

OA of the knee is more prevalent than OA of the hip. The prevalent rates vary from 3.8% (ages 25-74 years) (Lawrence, Hochberg, Kelsey, McDuffie, & Medsger, 1989) to 40% (ages 75-79 years) (Kellgren & Lawrence, 1958). OA of the knee is more frequent in females than males in the 55-64 year age group (Kellgren & Lawrence, 1963).

2.2.2 Clinical characteristics and classification

Clinically, the condition of OA of the hip or knee is characterised by the gradual development of the hip or knee joint pain, bony and soft tissue swelling, morning stiffness, stiffness related to inactivity, restricted range of movement and problems such as bursitis or tendonitis (Moskowitz, 1992). Severely affected hip or knee joints may also present with rest or night pain, instability, muscle wasting and joint deformities (Dieppe, 1995).

Anxiety, depression and feelings of social isolation may also present in patients with OA, and it is known that pain, disability and handicap may be determined or mediated by such psychological factors (Salaffi, Cavalieri, Nolli, & Ferraccioli, 1991; Summers, Haley, Reveille, & Alarcon, 1988).

Radiographic evidence of OA may not correspond with presenting symptoms. Many people with radiographic evidence of OA have few or only mild symptoms. In the over 55 year age-group 27% of males had radiological OA and 43% of these had symptoms; 45% females had radiological OA and 57% of these had symptoms (Lawrence, Bremner, & Bier, 1966). It is estimated that there is one case of symptomatic OA for every 18 cases of radiological OA (Badley, Thompson, & Wood, 1978; Van Saase, Van Romunde, Cats, Vandenbroucke, & Valkenburg, 1989). This evidence suggests it is wise to include both radiological and clinical criteria in any definition of OA.

The American College of Rheumatology (ACR) classification criteria for OA of the knee and hip (Altman, Alarcon, & Appelrouth, 1991; Altman, Asch, & Bloch, 1986) have been widely accepted. These classification criteria may not be used as diagnostic criteria because sensitivity and specificity are not at a level of 100%. Furthermore, many of the signs and symptoms come and go and are strongly influenced by other factors, such as general health. Therefore, it might not be appropriate or realistic that one set of clinical diagnostic criteria be developed for OA (Dieppe, 1995). However, these criteria are valuable in standardising the reporting of series of cases, in performing investigational studies, and in improving consistency in communication (Moskowitz & Holderbaum, 2001).

2.2.3 Clinical pathology and risk factors

OA is no longer considered only “degeneration” or “wear and tear” of the articular cartilage, but rather involves the entire joint, including the subchondral bone, ligaments, capsule, synovial membrane, and periarticular muscles (Dieppe, 1999; Liang & Fortin, 1991). These processes can alter joint shape beyond normal tolerance or alter remodelling at the bone-cartilage interface and result in OA (Doyle, 1986b).

OA joint pathological changes include the presence of focal or complete lesions of the articular cartilage, various degrees of bone remodelling, bone fragmentation, and exposure of underlying bone (Bullough, 1992; Radin, Schaffler, Gibson, & Tashman, 1995). The pathological change of OA may also include varying degrees of joint capsular and synovial membrane thickening, joint inflammation, ligament and tendon damage, and muscle pathology and atrophy (Doyle, 1986b).

The following are general risk factors that can cause OA (Hochberg, 2001; Moskowitz & Holderbaum, 2001):

- Age and gender
- Obesity and cartilage degeneration
- Occupation, sports, and joint injury
- Genetics and nutrition

- **Age and gender**

Age is the highest risk factor for OA (Cicuttini & Spector, 1997; Peyron & Altman, 1992). The Framingham study revealed that OA of the knee increased in both prevalence and incidence with age (Felson et al., 1987). OA of the hip seems to be more

prevalent in the older age group (age 55-74 years) (Tepper & Hochberg, 1993). However, there is some evidence to suggest that OA does not occur as a direct consequence of normal ageing and studies have shown that articular cartilage from patients with OA differs in a number of ways from cartilage of normal elderly individuals (Brandt & Fife, 1986; Sandell, 1995).

Gender is often associated with the pattern of distribution of OA. In a longitudinal study, OA of the knee was seen to occur more frequently in women than in men, after analyses were adjusted for age, body mass index, smoking, injury, chondrocalcinosis, and physical activity (Felson et al., 1997). Studies suggest that involvement of the knee joints constitutes the most usual pattern for women and, in contrast, men are more likely to have OA affecting the hips (Peyron & Altman, 1992; Tepper & Hochberg, 1993; Van Saase, Van Romunde, Cats, Vandenbroucke, & Valkenburg, 1989).

- **Obesity and cartilage degeneration**

Obesity has been strongly linked to OA of the knee (Silman & Hochberg, 1993). A study conducted over a 40-year period showed that weight loss significantly lowered the rate of knee OA (Felson, Zhang, Anthony, Naimark, & Anderson, 1992). In contrast to the knee, OA studies of the hip suggest that obesity has a less consistent association (Hochberg & Lethbridge-Cejku, 1997; Tepper & Hochberg, 1993; Van Saase, Vandenbroucke, Van Romunde, & Valkenburg, 1988).

Joints developing OA may show evidence of hyaline cartilage degeneration in early adulthood. Contrary to currently held views that these areas degenerate secondary to excessive loading, the most interesting theory is that these areas possibly degenerate because of under use rather than overuse (Bland, 1983).

- **Occupation, sports, and joint injury**

Occupations requiring repetitive use of particular joints over long periods may also subsequently develop hip and knee OA (Cooper, Campbell, & Byng, 1996; Cooper, McAlindon, & Coggon, 1994; Vingard, Alfredsson, Goldie, & Hogstedt, 1991). Occupational knee bending appears strongly linked to the development of OA (Andersen & Felson, 1988). Occupational lifting may also strongly link to a high risk of hip OA (Coggon, Kellingray, & Inskip, 1998).

Participation in sport has been associated with an increased risk of hip and knee OA (Hochberg & Lethbridge-Cejku, 1997; Peyron & Altman, 1992). A study revealed that increased sports activities of all kinds were associated with an increased risk of hip OA (Vingard, Alfredsson, Goldie, & Hogstedt, 1993). Both hip and knee OA have also been shown to be more prevalent among former soccer players (Roos, Lindberg, & Gardsell, 1994).

A strong association between a history of hip injury and hip OA has been found (Tepper & Hochberg, 1993). A study also found that a history of knee injury was significantly associated with knee OA (Davis, Ettinger, Neuhaus, Cho, & Houck, 1989).

- **Genetics and nutrition**

Genetics has a role in the development of OA. Some people may be born with defective cartilage or with slight defects in the way that joints fit together. A cross-sectional family study of patients undergoing total hip or total knee replacement suggested that siblings had an increased risk for both hip and knee replacement compared with spouses (Chitnavis et al., 1997).

Nutrients from the diet and other sources may prevent or delay the development of OA. Data from the Framingham Osteoarthritis Study showed a threefold increased risk of OA knee progression among people with the lowest vitamin C intake compared with those having a higher intake (Uitterlinden, Burger, & Huang, 1997). Inadequate levels of vitamin D also appears to be an important factor in OA progression (Keen, Hart, Lanchbury, & Spector, 1997; McAlindon, Felson, & Zhang, 1996; Uitterlinden, Burger, & Huang, 1997).

2.2.4 Medical treatment and its limitations

The goals in the management of OA are to reduce pain, maintain and/or improve joint mobility, and limit functional impairment (Batchlor & Paulus, 1992). In 1995, the American College of Rheumatology (ACR) published recommendations for the medical management of OA of the hip and knee (Hochberg et al., 1995a, 1995b). Those guidelines outlined the use of non-pharmacological modalities, including patient education, physical and occupational therapy, as well as the use of pharmacological agents and joint operation.

Pharmacological agents used in OA joint pain relief include anti-inflammatory agents, non-anti-inflammatory analgesic agents, anti-spasmodics, and depocorticosteroids (Lozada & Altman, 2001). Despite the potency and efficacy of these medications in the use of these medications in patients with OA, tolerance and toxicity remains a primary consideration (Batchlor & Paulus, 1992). This is especially so in the elderly and in those with concomitant illness, who may be particularly vulnerable to tolerance or toxicity (Schlegel & Paulus, 1986).

Analgesics and non-steroidal anti-inflammatory drugs (NSAIDS) are the most commonly used pharmacological agents in the treatment of OA (Dieppe, 1990; MacFarlane, 1992; Schnitzer, 1993). However, they can be associated with a variety of clinically important adverse effects (Griffin, Brandt, Liang, Pincus, & Ray, 1995; Griffin, Piper, Daugherty, Snowden, & Ray, 1991; Griffin, Ray, & Schaffner, 1988; Schlegel & Paulus, 1986). Furthermore, they frequently prove ineffective (Simon, 2000), and do not arrest the pathologic changes in cartilage or bone (Brooks & Day, 1991). Indeed, some may have deleterious effects on articular cartilage metabolism (Felson, 1993), and may contribute to degenerative changes in cartilage and bone (Palmoski & Brandt, 1980, 1982, 1983). Evidence from Cochrane systematic reviews of the recent studies of NSAIDs use for OA in the knee (Watson, Brookes, Kirwan, & Faulkner, 1998) and the hip (Towheed, Shea, Wells, & Hochberg, 1998) also concluded that there was insufficient evidence to distinguish any one NSAID as superior in action.

Patients with a painful flare of OA of the knee may benefit from intra-articular steroid therapy. However, a joint should not be injected more than three or four times in one year (Schnitzer, 1993), due to the possibility of cartilage damage from either repeated injections or overuse of steroids (Moskowitz, Davis, Sammarco, Mast, & Chase, 1970).

Patients with severe symptomatic hip and knee OA pain that has failed to respond to medical therapy, and who have significantly impaired function may be referred for surgery (Brandt, 1993). However, surgery is not always an option for some patients who have primacy of treatment for other co-existing conditions, have an extremely poor operative risk, or may require a lengthy waiting period.

Other treatments such as exercise, manipulation, heat and cold, laser therapy, transcutaneous nerve stimulation (TNS) diet and pain behaviour modification might also be effective in OA pain and functional limitations management (Clarke, 1991; Doyle, 1986a; Lozada & Altman, 2001). However, most of these non-pharmacological treatments have not undergone rigorous scientific scrutiny (Clarke, 1991).

2.3 Osteoarthritis in Traditional Chinese Medicine (TCM)

2.3.1 Overview of TCM

Traditional Chinese Medicine (TCM) is different from Western Medicine. TCM understands the body in a different way to the western biomedical model. TCM has been practised for more than 4000 years (Cheng, 1997). Influenced by the ancient Chinese philosophical concepts of yin yang and wu xing (see Appendix A), TCM views the human body as the human microcosm within the universal macrocosm, and understands disease within this philosophical paradigm. The body in TCM is viewed as a holistic system in which internal organs (zang fu) are connected by channels (jing luo) through which fundamental substances flow (qi, xue, jin yue, jing, and shen). See Appendix A "Fundamental Concepts and Frameworks in TCM" for definitions of these concepts.

TCM holds that these multi-layered channel (jing luo) systems are distributed over the whole body. They are linked with each other and connect the exterior, interior, upper and lower portions of the human body, thus making the body an organic whole. Along the channels lie about 365 acupuncture points (Yang et al., 1985) known in Chinese as "shu xue" which means both "transportation" and "hole". Acupuncture points are the

specific sites through which the qi and blood of the zang fu organs and channels is transported to the body surface (Cheng, 1997; Yang et al., 1985). Acupuncture points can be stimulated by needling, cupping, massage (tuina) and moxibustion (see Appendix A). This stimulation regulates the qi and blood of the zang fu organs and channels and allows disease within the body to be treated.

The aim of TCM treatment is to prevent disease before it occurs, and once disease is present to restore and maintain harmony within body systems.

2.3.2 Causes and mechanisms of OA in TCM

Harmony and balance of yin yang, wu xing, the internal organs (zang fu), fundamental substances (qi, blood, body fluids, jing and shen), and the channel systems creates balance and health within the body. Once disharmony occurs, external factors (pathogens) can invade the body and cause disease.

OA in TCM is described by the single inclusive phrase of "Bi syndromes" (Fang et al., 1985). Bi means obstruction. Bi syndromes are characterised by obstruction of qi and blood in the channels due to invasion of external pathogenic factors such as wind, cold, heat, and damp, which cause symptoms of pain, numbness, stiffness, heavy sensation of the joints and limitation of movement (Campbell, 2002; Cheng, 1997; Maciocia, 1997).

The invasion of external climatic factors is due to a pre-existing or temporary deficiency of the body's defensive qi and blood that allows the wind, cold, heat and damp to penetrate. This pre-existing or temporary deficiency of body's qi is relative rather than absolute. Thus, Bi syndromes are usually affliction of the channels alone, not the

internal organs (Ni, 1995). However, in chronic Bi syndromes or in the elderly, pre-existing internal factors such as a deficiency of qi and blood, and a deficiency in organs such as the kidney and the liver become important contributory factors to the development of the disease (Campbell, 2002). Six principal factors contribute to the development of Bi syndromes (e.g. OA). These include:

- Wind, cold and damp exogenous factors
- Seven emotions
- Improper diet
- Overstrain
- Retention of phlegm
- Blood stasis and trauma

- **Wind, cold and damp exogenous factors**

More than 2,000 years ago, the earliest of the extant medical classics in China called the "Huang Di Nei Jing" (Yellow Emperor's Classic of Internal Medicine) states in Chapter 43: "A combination of three pathogens - wind, cold and damp - invades the body, leading to obstruction and causing Bi" (Ni, 1995, pp. 160).

Wind, cold, and damp exogenous factors are general terms for wind, cold, and damp climatic conditions which, in excess in the body, cause pathology. Pathogenic wind tends to move and change rapidly; pathogenic cold tends to coagulate qi and blood in the channels and impede their flow; pathogenic damp is heavy and lingering in nature, and tends to go downward (Campbell, 2002; Cheng, 1997; Kaptchuk, 2000). Pathogenic wind, cold, and damp can also develop into febrile (heat) Bi syndromes, due to prolonged stagnation of qi and blood. Febrile Bi syndromes may also develop directly from pathogenic damp heat (Fang et al., 1985).

- **Seven emotions**

The seven emotions refer to joy, anger, sorrow, pensiveness, grief, fear and fright. In TCM, intense and prolonged emotions directly affect the corresponding zang fu organs (Campbell, 2002; Maciocia, 1996; Ying et al., 1995). For example, prolonged anger results in stagnation of liver qi, which changes the flow of fundamental substances in channels resulting in Bi syndromes or worsening of any existing Bi syndromes. Therefore, emotional problems are also contributing factors in Bi syndromes.

- **Improper diet**

Improper diet leads to deficiency of qi and blood. The decline and deficiency of qi and blood will cripple the body's resistance to attack by external pathogens. According to zang fu theory, the transformation of qi and blood depends on the function of spleen. Spleen deficiency is the primary cause of large number of internal disorders (Campbell, 2002; Ying et al., 1995), including Bi syndrome due to qi and blood deficiency. Blood moistens and nourishes tissues; qi is responsible for functional responses within the body.

- **Overstrain**

Overstrain includes physical and mental components. Physical overstrain results in qi exhaustion and causes lassitude as well as injury to tendon, muscle and bone. Mental overstrain consumes and impairs the heart blood as well as damaging the spleen qi, causing palpitations, insomnia, loose bowel motions and so on. Therefore, excessive sport or work activities may predispose to the development of Bi syndromes. The constant repetition of a certain movement in one's work is also an obvious predisposing

factor, as this causes stagnation of qi and blood in an area which becomes more prone to invasion by exterior pathogenic factors (Maciocia, 1997).

- **Retention of phlegm**

Retention of phlegm or fluids often results from the exogenous factors, improper diet, or the seven emotions (Campbell, 2002; Cheng, 1997; Ying et al., 1995). These factors can cause an accumulation of fluid within the body, which coalesces to form phlegm (tan). Once the retention of phlegm is formed, it can follow the qi circulation, inward into the zang fu organs and outward to the skin, muscles, tendons and bone. Stagnation of phlegm in the channels and joints may lead to numbness of the joints, altered sensation, and restricted flexion and extension.

- **Blood stasis and trauma**

Blood stasis is a pathological state resulting from abnormal blood outside the vessels that fails to disperse, or from the impeded circulation of blood itself. For instance, deficiency of qi results in weak circulation of blood; stagnation of qi results in impeded circulation of blood; cold blood in the TCM concept results in coagulation of blood (Campbell, 2002; Cheng, 1997; Maciocia, 1996; Ying et al., 1995).

Blood stasis may also be caused by trauma. Stagnation of blood in local areas results in local swelling, pain and cyanosis. Therefore, trauma in a joint may lead to the development of Bi syndromes due to stagnation of qi and blood.

2.3.3 TCM diagnosis and the pattern identification (bian zheng) of OA

As noted previously, according to TCM theory, the human body is a holistic entity, with its interior and exterior connected by the channel systems. Therefore, pathological changes (disharmony or imbalance) in the human body are inevitably shown on the external body as abnormities in complexion, the tongue and, the pulses and can be felt by palpation of the body such as abdominal palpitation (Campbell, 2002; Deng, 1999; Kaptchuk, 2000; Maciocia, 1996). A TCM practitioner seeks to understand the internal disharmony by asking questions, observing, listening and palpating (the "four examinations") and then integrating the information into TCM theory in order to make a diagnosis. This process is called "pattern identification" (bian zheng).

According to the aetiology, pathophysiology and clinical symptoms, Bi syndromes may be classified into different patterns by means of the "eight principal syndromes" (see Appendix A). The patterns of OA in TCM practice may be complex due to the combination of several "basic patterns" as set out in Table 2.1.

Table 2.1 The patterns of OA in TCM

Pattern	Symptoms	Tongue	Pulse	Aetiology and Pathophysiology
Wind Bi or Wandering Bi	Pain moving from joint to joint, chills and fever.	Thin tongue coating	Superficial pulse	Due mainly to invasion by pathogenic wind, which is characterized by constant movement and change from joint to joint.
Cold Bi or Painful Bi	Severe pain in a joint, pain alleviated by warmth.	Thin white tongue coating	Tight pulse	Due to retarded circulation of qi and blood in the channels and collaterals caused by excessive cold. The severe pain of cold Bi is caused by contraction of the body.
Damp Bi or Fixed Bi	Soreness, heaviness, numbness and swelling of the joints, pain aggravated on humid or rainy days.	White and sticky tongue coating	Slow and soggy pulse	Damp is characterized by heaviness. Excess damp invades the limbs and joints, causes obstruction of circulation of qi and blood, and results in numbness and heaviness.
Febrile Bi	Severe pain and hot-red-swollen joints, fever and thirst.	Yellow tongue coating	Slippery and rapid pulse	The pathogenic factors in the channels and collaterals are transformed into heat.

(table continues)

Table 2.1 (continued)

Pattern	Symptoms	Tongue	Pulse	Aetiology and Pathophysiology
Phlegm and Blood Stagnation Bi	Severe, often stabbing pain, stiffness and deformity of bone in the joints.	Purple tongue with greasy coating	Slippery or rough pulse	The pathogenic factors lead to retention of body fluids, which turn into phlegm, and this further obstructs the joints and channels. Obstruction of qi, blood, and body fluids caused by phlegm leads to stasis of blood. The stasis of blood in the channels causes severe pain and pronounced stiffness due to stagnant blood failing to nourish and moisten sinews and ligaments.
QI and Blood Deficiency Bi	Dull pain and weakness of joints, poor appetite, loose stools.	Pale tongue	Thin pulse	Invasion of pathogenic factors occurs easily if the body's condition is weak, leading to malnourishment of the joints.
Yang Deficiency Bi	Pain and weakness of joints with cold feeling, frequent urination.	Pale tongue	Thin and weak pulse	Chronic Bi depletes the yang qi of body and causes deficiency of kidney and spleen yang. Pain and weakness of joints is due to the yang qi not warming and promoting the blood circulation.
Yin Deficiency Bi	Pain and weakness of joints with hot feeling, tinnitus, and dizziness.	Red tongue without coating	Thin and rapid pulse	Chronic Bi depletes the yin of the body and causes deficiency of kidney yin. The kidney yin nourishes bones and when it is deficient the bones are deprived of nourishment. Liver blood nourishes the sinews and when the Liver is deficient the sinews and tendons are not nourished.

(Campbell, 2002; Cheng, 1997; Fang et al., 1985; Maciocia, 1997)

2.3.4 TCM treatment for OA

TCM treatment is based on pattern identification. Although OA is a single diagnosis in western medicine, it is not a single diagnostic pattern in the TCM context. Therefore, different patients with OA will have different patterns, and the same client may also have different patterns at different stages in the evolution of their OA. Each of these patterns are treated in different ways, as it is said, “one disease, different treatments” (Ying et al., 1995, p.9).

The common TCM treatment approaches for OA include acupuncture, moxibustion (burning of the herb Artemisia Vulgaris), herbal medicine, and massage (tuina). Regardless of which treatment approach is applied, the treatment principles are consistent: eliminating the pathogenic factors, regulating the circulation of qi and blood, and restoring the balance of yin and yang. The treatment principles and some appropriate points for the different patterns of OA are given in Table 2.2.

Table 2.2 The different patterns of OA (Bi syndromes), the corresponding treatment principles and some appropriate acupuncture points

Pattern	Treatment Principles	Some Appropriate Acupuncture Points
Wind Bi	Eliminating wind and regulating circulation of qi and blood	Fengchi (GB20), Geshu (BL17), Xuehai (SP10), Zhongfeng (LR4)
Cold Bi	Expelling cold by warming the channels	Shenshu (BL23), Guanyuan (CV4), moxa applicable
Damp Bi	Resolving dampness and activating blood circulation in the channel	Yinlingquan (SP9), Zusanli (ST36), Zulingqi (GB41), Waiguan (TE5)
Febrile Bi	Clearing heat and regulating blood circulation	Dazhui (GV14), Quchi (LI11), Hegu (LI4), Xuehai (SP10)
Phlegm and Blood Stagnation Bi	Resolving phlegm and removing blood stasis	Fenglong (ST40), Yinlingquan (SP9), Xuehai (SP10), Geshu (BL17), Sanyinjiao (SP6), Neiguan (PC6), Shaohai (HT3)
Qi and Blood Deficiency Bi	Reinforcing qi and blood, regulating the channels	Zusanli (ST36), Guanyuan (CV4), Pishu (BL20), Shenshu (BL23), moxa applicable
Yang Deficiency Bi	Warming kidney and spleen yang	Shenshu (BL23), Mingmen (GV4), Pishu (BL20), Yanglingquan (GB34), moxa applicable
Yin Deficiency Bi	Nourish kidney essence and liver blood	Taixi (KI3), Ququan (LR8), Dashu (BL11), Shenshu (BL23)

(Campbell, 2002; Cheng, 1997; Fang et al., 1985; Maciocia, 1997)

2.4 Acupuncture for OA of the hip and knee: A systematic review

2.4.1 Introduction

Acupuncture has been practised in China for more than 4000 years (Cheng, 1997). It is now practised in over 100 countries around the world (World Health Organization, 1985). It is reported that on any one day, at least 5000 patients receive acupuncture in New Zealand (Campbell, 2004). Acupuncture techniques are based on traditional needle acupuncture, and may be combined with modern technology such as electro-acupuncture, laser acupuncture, microwave acupuncture and injection acupuncture. Acupuncture point selection also may include body acupuncture points, and/or micro-system acupuncture points such as scalp, hand, and ear acupuncture points. Acupuncture in China is referred to as "acupuncture and moxibustion" because both are based on the same channel system of point selection and traditionally integrated into practice.

Documentation of the TCM diagnosis and acupuncture treatment for symptoms of OA can be found in the classics of the Neijing written about 300 B.C (Ni, 1995). The details of TCM understanding of OA have been previously discussed in Chapter 2.3 "Osteoarthritis in TCM".

The usage of acupuncture in the treatment of pain is becoming more widely accepted in the west, and due to the limitations of current western medical therapies in the treatment of OA, interest has arisen in acupuncture for pain relief (Yamauchi, 1976). Furthermore

it has been reported that acupuncture has fewer side effects in comparison with conventional therapy (Norheim, 1996; White, Hayhoe, Hart, & Ernst, 2001).

An increasing number of clinical trials of acupuncture as a treatment for OA of the hip or knee have been reported (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001; Mclndoe, Young, & Bone, 1995; Takeda & Wessel, 1994). However, there is inconsistency in the results of these trials and this may confuse both the patient with OA and the practitioner when making their treatment decisions.

In order to condense the large amount of information from individual trials, evaluate the studies critically, and identify areas of future research, the author conducted a systematic review of randomised controlled trials (RCTs) of acupuncture in patients with OA of the hip or knee.

A systematic review is defined as a scientific tool which can be used to summarise, appraise, and communicate the result and implications of otherwise unmanageable quantities of research (NHS Centre for Reviews and Dissemination, 1996). Systematic reviews are overviews of primary studies and they should be rigorously designed, unbiased and reproducible. As a result the reviews are supposed to have greater power than individual studies and therefore allow more meaningful conclusions.

Since RCTs are widely accepted as the most reliable method of determining the effectiveness of specific therapies (Prescott, Counsell, & Gillespie, 1999), this review focused on the RCTs rather than other types of clinical trials.

This review also focused purely on acupuncture, and therefore excluded studies with modalities such as electro-acupuncture, laser acupuncture, or warming acupuncture with moxibustion.

To determine the effectiveness of acupuncture for OA of the hip or knee, it is necessary to assess the methodological quality and acupuncture quality of the review studies, and to identify areas for future study. The four objectives of the review were listed as follow:

- To assess the methodological quality of the RCTs of acupuncture in patients with OA of the hip or knee
- To assess the acupuncture quality of the RCTs of acupuncture in patients with OA of the hip or knee
- To determine the effectiveness of acupuncture for patients with OA of the hip or knee
- To identify areas for future research in acupuncture which is within TCM concepts and frameworks.

2.4.2 Review methods

2.4.2.1 Search strategy

Searches were performed in November 2003 for controlled trials of acupuncture for OA of the hip and knee, using 11 electronic databases (see Table B.1 in Appendix B) and 2 search terms (“acupuncture and osteoarthritis and knee”, and “acupuncture and osteoarthritis and hip”).

In addition, the following journals and relevant acupuncture trials reported in the past four years, from January 1999 to December 2002, were hand searched (Table B.2 in Appendix B):

- Four general and internal medicine journals: Annals of Internal Medicine, Archives of Internal Medicine, British Medical Journal, American Journal of Medicine
- Three rheumatologic journals: Arthritis and Rheumatism, Seminars in Arthritis and Rheumatism, Rheumatology
- Four acupuncture journals: Acupuncture in Medicine, American Journal of Acupuncture, Chinese Acupuncture and Moxibustion, Shanghai Journal of Acupuncture and Moxibustion
- Two alternative and complementary medicine journals: Complementary Therapies in Medicine, Alternative Therapies in Health and Medicine.

The first seven journals were chosen because: 1) they were more likely to have articles on OA; 2) they had high methodological quality for journal articles (Lee, Schotland, Bacchetti, & Bero, 2002). The other journals were chosen because they are related specifically to acupuncture.

Reference lists for any review articles retrieved were also searched.

2.4.2.2 Inclusion criteria

In order to avoid 'selection bias', all retrieved articles were reviewed according to the following two stages of inclusion and exclusion criteria:

- **Stage one**

Inclusion criteria

- Any language
- Clinical trials
- Any types of acupuncture treatment
- Study population was patients with OA only.

- **Stage two**

Inclusion criteria

- English or Chinese language
- Clinical RCTs
- Acupuncture is defined as a basic form of acupuncture that used traditional needle insertion into the acupuncture points
- Assessed acupuncture for OA of the hip or knee exclusively
- Published as a full-text article.

2.4.2.3 Exclusion criteria

- Case series
- Non-randomising uncontrolled studies
- Published as abstracts only
- Editorials, news, or correspondence sections
- Duplicate publication (i.e., the same trial with results from different lengths of follow-up published twice). In these cases, only the most recent article was selected for inclusion

- Trials in which one form of acupuncture was compared with another form
- Other forms of acupuncture (i.e. electro-acupuncture, TENS, laser acupuncture) or acupuncture combined with other therapies rather than basic form of acupuncture.

Full copies of all included studies were obtained. Due to constraints on time and resources, only one reviewer assessed each included article. The retrieved articles were coded and recorded on Endnote. The studies met the above inclusion criteria are presented in Table 2.3. The studies met the exclusion criteria are presented in Appendix B (Table B.3).

2.4.2.4 Data extraction and quality assessment

In order to extract relevant data from included studies accurately and without bias, data extraction forms were developed based on the factors of the quality assessment. Owing to time and resource constraints, the same reviewer performed the data extraction and no attempt was made to contact study authors for further information on missing or unclear data.

- **Methodological quality assessment**

The quality of the methods in the selected articles was assessed using the Jadad scale (see Table 2.3 for details) because of its validity (Jadad et al., 1996). Jadad scale assesses three aspects: randomisation, blinding, and reporting of dropouts and withdrawals. These three aspects are generally applicable to all reports of acupuncture research and directly related to the reduction of bias. This scale is simple, short, reliable and apparently valid (Jadad et al., 1996). It seems best to rely on one method that is

known to be valid, i.e. that of Jadad, until other methods that may be more sensitive to acupuncture methodology have been validated (White, Trinh, & Hammerschlag, 2002).

The code and points were marked in Jadad scale as follows: (A) study described as randomised, 1 point; (A1) additional point for appropriate method, 1 point; (A2) inappropriate randomisation method, reduce 1 point; (B1) subject blinded to intervention, 1 point; (B2) observer blinded to intervention, 1 point; (C) description of withdrawals and dropouts, 1 point.

The randomisation sequence generation was considered adequate if selection bias was prevented by use of, for example, a table of random numbers, random numbers generated by computer, coin tossing, or shuffling cards. Quasi-randomisation allocation procedures, such as allocation by hospital record number or birth date, or alternation, are considered as inappropriate randomisation methods. Subjects and acupuncturists are only considered to be "blind" if there is an expectation that they would have been unable to distinguish between the treatments applied to different groups. Observer blinding was scored if specified in the text. In trials in which key outcomes are self-reported (e.g., visual analogue scale, questionnaires), the observer is considered to be blind if the subject was blind.

In order to analyse the comprehensive methodological issues, an additional checklist of items was developed. The checklist of items includes: sample size, study design, allocation concealment, baseline comparison, reporting of a sample size justification, or reporting on an evaluation of the power of the study, and reporting of an intention-to-treat analysis.

Allocation concealment should be sufficient by which foreknowledge of treatment assignment is prevented and the allocation procedure is kept tamperproof (Mulrow & Oxman, 1997). Baseline comparison between groups should be tested to demonstrate their similarity in terms of important demographic characteristics and for key population characteristics likely to be related to prognosis (Bensoussan & Myers, 1996). Sample size justification should be based on a specified clinically meaningful difference, degree of variation in the primary outcome variable and permissible level of false positive and false negative error (Bensoussan & Myers, 1996). Intention-to-treat is preferred with all subjects remaining in their allocated groups for the final analysis (Bensoussan & Myers, 1996).

- **Acupuncture quality assessment**

The studies were further examined as to whether a TCM diagnosis was applied, whether the point selection was individualised and modified according to TCM diagnosis, or whether the point selection was a formula with a set of predefined points. In addition the needle sensation (de qi), acupuncture "dosage" and control procedures that could influence the result of the study were also considered.

- **Data synthesis, meta-analysis or evidence weighting**

Clinical homogeneity amongst the studies were assessed by comparing the retrieved studies with respect to the sites of OA, population of study, quality of methodology, quality of acupuncture treatment, control method, outcome measures and timing of follow-up. Meta-analysis or the statistical pooling of data can only be undertaken on data of high quality where sufficiently similar data items exist (Thomas & Fitter, 2002). The initial protocol for this review anticipated that results from several studies could be combined in a meta-analysis, but this was precluded by the heterogeneity of the studies.

Therefore, a best-evidence synthesis (Slavin, 1995) method was use. This method is used in some Cochrane systematic reviews (Van Tulder, Koes, & Bouter, 1997) and systematic review for OA of the knee (Ezzo et al., 2001). It is based on four levels of evidence and weighs the results according to the methodological quality:

- Strong evidence: multiple, relevant, high-quality RCTs with generally consistent outcomes
- Moderate evidence: one relevant, high-quality RCT, and one or more relevant, low-quality RCTs with generally consistent outcomes
- Limited evidence: one relevant, high-quality RCT, or multiple relevant, low-quality RCTs with generally consistent outcomes
- Inconclusive evidence: only one relevant, low-quality RCT, no relevant RCTs, or RCTs with inconsistent outcomes.

"Relevant" was defined as having at least one of the primary outcomes. "Generally consistent" was defined as two-thirds or more of the studies having the same result (positive or negative). "Multiple" was defined as more than one RCT. The maximum quality score was five in the Jadad scale. Therefore RCTs with a score less than three were considered low quality studies, while those with scores of three or more were classified as high quality.

2.4.3 Results of review

2.4.3.1 Literature search

The databases and hand-searched literatures are summarised in Tables B.1 and B.2 (see Appendix B). This systematic search resulted in the identification of 39 articles meeting the stage one criteria (Figure 2.1). After stage two criteria were applied a further 34 trials were excluded, and only five articles (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001; Mclndoe, Young, & Bone, 1995; Takeda & Wessel, 1994) representing 200 patients with OA remained. The excluded articles and the reasons for exclusion are presented in Table B.3 (see Appendix B).

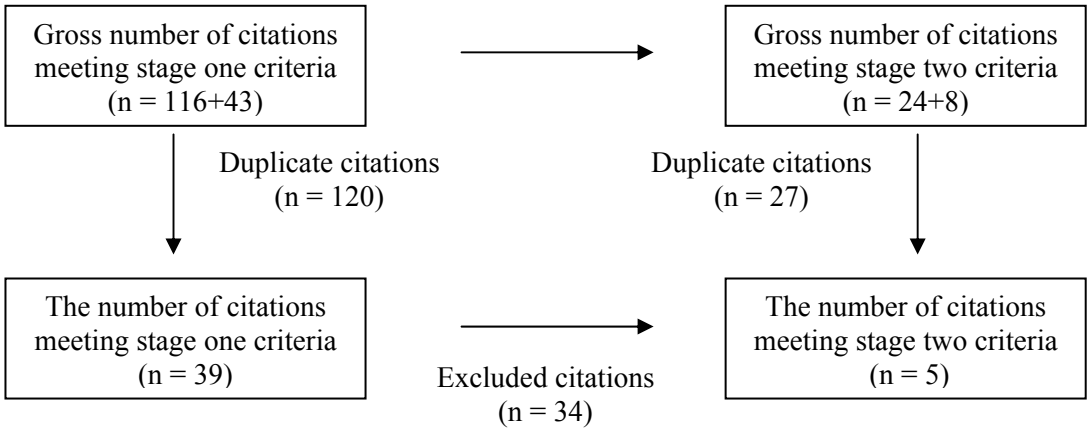

Figure 2.1 Search procedure for controlled trials of acupuncture for OA of the hip or knee

Twenty-seven of the 39 articles were found in the Cochrane Library database (Issue 4, 2003) and in the PubMed database (November 2003). The five included articles were found in the Cochrane Library database (Issue 4, 2003) and in the EBSCOhost database (November 2003). No RCTs of OA of the hip or knee were published in the journals of

Chinese Acupuncture and Moxibustion and Shanghai Journal of Acupuncture and Moxibustion from January 1999 to December 2002.

2.4.3.2 Methodological quality

Table 2.3 summarises the quality ratings of all included studies. Four of the five studies were of high quality and scored three or more points (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001; Takeda & Wessel, 1994). Only one study was of poor quality (Mclndoe, Young, & Bone, 1995). One study gained the maximum score of five points (Fink, Kunsebeck, Wipperman, & Gehrke, 2001).

Table 2.3 Jadad score of the RCTs on acupuncture for OA of the hip or knee

Author, year	A	A1	A2	B1	B2	C	Total	Result
Christensen et al., 1992	1	1			1	1	4	+
Takeda & Wessel, 1994	1			1	1	1	4	=
McIndoe, Young, & Bone, 1995	1						1	=
Haslam, 2001	1	1				1	3	+
Fink, Kunsebeck, Wipperman, & Gehrke, 2001	1	1		1	1	1	5	=

Note. + = a positive difference between groups; = = refers to no difference between groups; A = study described as randomised, 1 point; A1 = additional point for appropriate method, 1 point; A2 = inappropriate randomisation method, reduce 1 point; B1 = subject blinded to intervention, 1 point; B2 = observer blinded to intervention, 1 point; C = description of withdrawals and dropouts, 1 point.

Sufficient detail in the randomisation procedure was reported in three studies (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001). Subject blinding was achieved in only two studies (Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Takeda & Wessel, 1994). Observer blinding was reported in three studies (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Takeda & Wessel, 1994).

The additional methodology checklist (Table 2.4) presents sample size of the studies which ranged from 29 (Christensen et al., 1992) to 67 (Fink, Kunsebeck, Wipperman, & Gehrke, 2001). Dropouts were under 20% in three of the four reported studies (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Takeda & Wessel, 1994). Only one study reported dropouts above 20% in the control group (Haslam, 2001). Two subject-and-observer-blinding, sham controlled studies had allocation concealment. All studies have baseline comparison between groups. No sample size justification was reported in the studies. Only two studies preformed intention-to-treat analysis (Haslam, 2001; Takeda & Wessel, 1994).

Table 2.4 Additional checklist of methodology of the RCTs on acupuncture for OA of the hip or knee

Author, year	Sample Size / Dropouts	Design	Allocation Concealment	Baseline Comparison between Groups	Sample Size Justification	Intention-to-treat Analysis
Christensen et al., 1992	29 / 3	RCT, observer-blind, cross-over, waiting list controlled	No	No difference	Not reported	Not reported
Takeda & Wessel, 1994	40 / 2 (1 from control group)	RCT, double-blind, sham-controlled	Yes	No difference	Not reported	Replaced by new recruits
McIndoe, Young, & Bone, 1995	32 / Not reported	RCT, intra-articular injections controlled	No	No difference	Small size in each group (n=16)	Not reported
Haslam, 2001	32 / 8 (7 from control group)	RCT, advice and exercise controlled	No	No difference	Small size in each group (n=16)	Replaced by new recruits
Fink, Kunsebeck, Wipperman, & Gehrke, 2001	67 / 5 (4 from control group)	RCT, subject- and observer-blind, sham controlled	Yes	No difference	Not reported	Not reported

2.4.3.3 Acupuncture quality

No studies included information on the TCM diagnosis (Table 2.5). However TCM acupuncture points were applied in four studies (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001; Takeda & Wessel, 1994). One study used a triple needling technique on the affected area (Mclndoe, Young, & Bone, 1995).

Formula point selection was applied in all studies. In three studies needle sensation (de qi) was mentioned in the study group but not in the control group (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Takeda & Wessel, 1994).

Acupuncture "dosage" was varied among the studies. As shown in Table 2.5, the patients in most of the studies attended 6-10 sessions during a treatment, with the exception of one study (Mclndoe, Young, & Bone, 1995) which has only three sessions. The frequency of sessions in all studies was 1-3 per week. The duration of a session in most studies was 20-30 minutes, with the exception of (Mclndoe, Young, & Bone, 1995) which did not specify the duration of the session.

Both sham acupuncture controlled studies had results of no difference between groups (Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Takeda & Wessel, 1994).

Table 2.5 Summary of acupuncture quality of the RCTs on acupuncture for OA of the hip or knee

Author, year	TCM Diagnosis	Point Selection	Needle Sensation (De Qi)	Acupuncture Manipulation and Dosage	Control Procedure
Christensen et al., 1992	None or not reported	Formula, ST34, 35, 36, SP10, Neixiyuan, LI4	De qi was pursued in acupuncture group	Manipulation in 20 minutes. Twice a week for 3 weeks	Waiting list
Takeda & Wessel, 1994	None or not reported	Formula, ST35, SP9, GB34, Extra31, 32	De qi or full depth of the needle (30mm) in acupuncture group and control group	Five minutes manipulation in 30 minutes. Three times a week for 3 weeks	Sham controlled with no active points and use of shallow needling
McIndoe, Young, & Bone, 1995	None or not reported	Formula, triple needling to reach the greater trochanter	Not reported	Two minutes manipulation for each needle. A total of 3 times at weekly intervals	Intra-articular steroid injection
Haslam, 2001	None or not reported	Formula, 4 ahshi points plus GB29, 30, 34, 43, ST44, LI4	Not reported	Ten seconds manipulation every 5 minutes in 25 minutes. Once a week for 6 weeks	Exercise
Fink, Kunsebeck, Wipperman, & Gehrke, 2001	None or not reported	Formula, 6 ahshi points plus GB30, 31, 34, BL37, 54, ST40	De qi was pursued in acupuncture group, but not reported in the control group	Two or 3 times manipulation in 20 minutes. Ten sessions for 3 weeks	Sham control with "placebo" points, and use of same depth needling but no manipulation

2.4.3.4 Outcomes: Treatment efficacy, safety and cost

Radiological evidence of OA of the knee or hip was required for sample selection in all five studies. Three studies focused on OA patients awaiting knee or hip replacement (Christensen et al., 1992; Haslam, 2001; Mclndoe, Young, & Bone, 1995).

In four of five studies, two or more outcome measures were applied (Table 2.6). The most frequently used outcome measures were pain intensity (VAS, MPQ, PRI) and function (WOMAC, walking, hip function index). Only one trial used overall assessment of patient satisfaction on the Carlsson's Comparative Scale and the Bullinger's Everyday Life measurement (Fink, Kunsebeck, Wipperman, & Gehrke, 2001).

In four studies, significant improvements ($p < .05$) for pain and function tests were reported (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001; Takeda & Wessel, 1994). One study showed no functional improvement in either study or control groups (Mclndoe, Young, & Bone, 1995). However when compared with control groups, only two studies reported positive results (Christensen et al., 1992; Haslam, 2001).

Two high-quality studies in patients with OA of the knee describe differing results for pain and function (Christensen et al., 1992; Takeda & Wessel, 1994). Both studies had different control methods (Table 2.5).

Two studies in patients with OA of the hip, compared acupuncture with other different types of active therapies (Haslam, 2001; Mclndoe, Young, & Bone, 1995). One high-quality study showed positive results when comparing acupuncture with advice and exercises (Haslam, 2001). One high-quality study in OA of the hip showed that TCM acupuncture was not more effective than placebo acupuncture using different needle placement (Fink, Kunsebeck, Wipperman, & Gehrke, 2001).

The positive effects of acupuncture were maintained at approximately two months follow-up in two studies (Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001). However, one study described an attenuation of the benefit of acupuncture at six-month follow-up (Fink, Kunsebeck, Wipperman, & Gehrke, 2001). In one long-term follow-up study, the improvement was maintained by administering monthly maintenance treatments of acupuncture (Christensen et al., 1992).

The reduction of analgesics consumption was reported in three studies (Christensen et al., 1992; Haslam, 2001; Mclndoe, Young, & Bone, 1995). Only one of five studies reported acupuncture had mild side effects, such as worsening of the pain, nausea or dizziness (Christensen et al., 1992). The study also reported the reduction of treatment cost compared to a knee arthroplasty operation.

Table 2.6 Summary of study population and outcome measures in the RCTs on acupuncture for OA of the hip or knee

Author, year, (Jadad Scores)	Inclusion Criteria	Pain Outcome	Other Outcomes	Follow-up	Analgesics Consumption	Safety	Cost
Christensen et al., 1992, (4)	OA patients awaiting knee replacement	Reduced only in study group, VAS (+)	HSS (+); walking 50m (+); climbing 20 steps (+)	Forty-nine weeks (+)	NSAID reduced (+)	Mild side effects	Reduced
Takeda & Wessel, 1994, (4)	Radiologic evidence of OA of knee	PRI of MPQ (=)	Pain stiffness, and function index of WOMAC (=)	None	Not reported	Not reported	Not reported
McIndoe, Young, & Bone, 1995, (1)	OA patients awaiting hip replacement	MPQ (=)	Mobility not improved in both groups (=)	None	Analgesics consumption reduced in both groups but (=)	Not reported	Not reported
Haslam, 2001, (3)	OA patients awaiting hip replacement		WOMAC (+)	Eight weeks follow-up (+)	Analgesics consumption reduced in both groups	Not reported	Not reported
Fink, Kunsebeck, Wipperman, & Gehrke, 2001	Radiological evidence of OA of hip	VAS (=)	Overall assessment (=); Quality of life (=); Hip function index (=)	Six weeks (-). Six months follow-up (-)	Not reported	No adverse effects	Not reported

Note. + = positive difference between groups; = = no difference between groups; VAS = visual analogue scale; PRI = pain-rating index; MPQ = the McGill pain questionnaire; WOMAC = the Western Ontario and McMaster Universities Osteoarthritis Index; HSS = the Hospital for Special Surgery (New York) Score.

2.4.4 Discussion of review

2.4.4.1 Methodological quality

- **Quality scores**

Four of five of the RCTs were rated three or more points (high-quality) on the Jadad scale (Table 2.3). In contrast the systematic review of acupuncture treatment for OA of the knee (Ezzo et al., 2001) reported that more than half of the RCTs were rated less than three points (low-quality). One explanation could be that the trials in this review were from more recent publications (published after January 1990). These trials may have been better quality.

- **Randomisation**

Some known and unknown prognostic factors may affect the study result when randomisation was inadequate. Inadequate randomisation allocation for unknown variables in one study (Takeda & Wessel, 1994) may possibly have led to the result of no significant differences between acupuncture and sham acupuncture groups, due to both group receiving de qi and this may have affected the result. In this study the authors Takeda & Wessel (1994) commented that when analyses used de qi (rather than treatment) as the group factor, then significant differences between groups were found for the WOMAC pain index and the pressure threshold scores.

- **Blinding**

Blinding, or masking the group allocation, is primarily used to control expectation of effects. Therefore, it is necessary to blind any individual who participates at any stage of a trial. Six roles in a study can theoretically be blinded, which includes subject, acupuncturist, observer, analyst, other participants, and the researcher testing the success of blinding (White, Filshie, & Cummings, 2001). However, the acupuncturist cannot be completely blinded to the true or sham treatment, except where the practitioner has little experience as an acupuncturist (Vincent & Furnham, 2001), or where special types of acupuncture point stimulation, such as with ultrasound are used (Lewith, Walach, & Jonas, 2002).

The blinding of the acupuncturists was performed in two studies where acupuncturists were blinded to one of two diagnoses (Godfrey & Morgan, 1978) or one of two treatment protocols (Allen, Schnyer, & Hitt, 1998). However, these methods of therapist blinding may confound the result because the acupuncturist and patient relationship could be an important factor in the treatment (Hopwood & Lewith, 2002). From the point of view of TCM and any other complementary medicine, the job of the practitioner is to promote the patient's self-healing capacity, and the practitioner and patient are in a partnership relationship during the treatment process (Cassidy, 2001). During treatment the acupuncturist may need immediate feedback from the patient as a guide for needling manipulation (Xi, 1985) or for further treatment (White, 2002). Furthermore, the acupuncturist would expect to change the prescription at almost every treatment. It has been suggested that videotaping treatment sessions could provide a way of detecting differences in acupuncturist behaviour so that behaviour and communication may be kept the same way between the patient and practitioner, and bias

may be minimised (Vicent & Richardson, 1986). Further studies are needed to evaluate ways of blinding.

- **Control method**

It is also difficult to design an appropriate control group in an acupuncture trial. The waiting list controlled study, showed that acupuncture can reduce the pain and analgesic consumption in OA patients waiting for arthroplasty surgery (Christensen et al., 1992). However, as there was no placebo or sham controlled group in this study, the specific effects of acupuncture remain unknown.

In order to test the specific effects of acupuncture, two sham control studies were conducted (Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Takeda & Wessel, 1994) (Table 2.5). Both of these studies reported that the acupuncture study group and control group had significantly reduced pain and physical disability in the OA joint, but that there were no significant differences between groups.

One study controlled by using an inactive point and shallow needling (Takeda & Wessel, 1994). The validity of this type of control is questionable. Firstly, it is difficult to judge what constitutes an inactive point. Ah Shi points (tender points) have no name and no fixed location, and extra points (special points) have a name and fixed location but do not relate to the twelve channels or the eight extraordinary channels. These types of points are regularly used in OA patients (Cheng, 1997). Secondly, the shallow needling technique is one of the TCM acupuncture techniques (qianzhenshu) and commonly used in Japanese acupuncture. As the authors Takeda & Wessel (1994) discussed, even superficial insertion of needles in the control group subjects may have been adequate to produce a physiological response in individuals with less fat around the knee.

Another shame control trial on acupuncture treatment for OA of the hip controlled by a different needle placement and no manipulation applied to the needle (Fink, Kunsebeck, Wipperman, & Gehrke, 2001). The needle placement in the control group was in the "area of the hip joint", but at least 5 cm away from the classical acupuncture points and their interconnecting line (meridians) and also clear of painful pressure points (Ah Shi or trigger points). The number of needles, the time frame (20 min), and the depth of needling were the same in both sham control and acupuncture groups. This study gained the maximum score of five points on the Jadad scale. However the validity of this study is also questionable. Firstly, in this study, the authors had not reported whether de qi sensation was elicited in the control group. It is suggested that the de qi sensation is an important factor for the effectiveness of acupuncture, and needle retention is an important technique in TCM acupuncture to promote de qi sensation (Xi, 1985). Although in Fink et al (2001)'s study, there was no manipulation or twisting applied to the needle in the control group, the de qi sensation may also be elicited when the needle is retained for the duration of 20 minutes. Secondly, in Fink et al (2001)'s study, the needle placement in the control group in the affected area may also have treatment effects. In TCM acupuncture, needle placement in the affected area is called sanci or yangci, and is one kind of needling techniques (Xi, 1985). Therefore, the authors' conclusion of non-specific effects of traditional Chinese acupuncture in OA of the hip may require further scrutiny.

The other two controlled studies selected active controls using other therapies (Haslam, 2001; Mclndoe, Young, & Bone, 1995). Both of the studies did not perform subject or therapist blinding, so could not control the placebo effects. However, both studies have

practical value, since efficacy, analgesic consumption, adverse effects, and cost effectiveness could be compared.

- **Sample size**

Since individual patients will vary in their response to the treatments, any trial should have as large a sample size as possible. A large sample size has a greater chance of detecting, as statistically significant, whether a worthwhile effect exists if it is found in the trial (Altman, 1991).

A sample size ("power") calculation or justification is based on a specified clinically meaningful difference, degree of variation in the primary outcome variable and permissible level of false positive and false negative error (Bensoussan & Myers, 1996). A false positive error (type I error) means that one of the treatments is believed to be better than the other but in reality the two treatments are equivalent. In contrast to false positive error, a false negative error (type II error) occurs where the two treatments are believed to be equivalent but in reality one is better.

There was no sample size justification performed in the five RCTs. In order to see how much chance the trials had of detecting a true difference between acupuncture and control groups in the pain reduction and functional improvement for the patients with OA, the power of each study was calculated retrospectively.

Power is equal to one minus the type II error probability and can be calculated using the G*power software package (Buchner, Erdfelder, & Faul, 1997). If a type I error (α) is equal to 0.05 then the power should be greater than 0.8. If the power is low, the study may fail to detect a treatment difference. This could mean there was no treatment

difference or because the study was too small to detect the difference. Even if the finding was positive there may be a problem with interpretation because of publication bias (Crichton, 1993).

Only two studies included sufficient information to allow the power calculation (Haslam, 2001; Takeda & Wessel, 1994). The power of these two studies is calculated as below:

Haslam (2001)'s study had 16 patients receiving acupuncture treatment and 16 patients receiving advice and exercises treatment. Based on Haslam's data, the mean of WOMAC scores of the advice and exercises control group was 831 points with a standard deviation of 235 points, the response of the acupuncture group was 696 points. Using the G*power software package (Buchner, Erdfelder, & Faul, 1997), and setting two-sided tests with $\alpha = .05$, then the power in Haslam's study is only .35. So in Haslam's study, the sample size has a probability of only .35 to detect a difference of improvement effect between groups.

Takeda & Wessel (1994)'s study had 20 patients receiving acupuncture treatment and 20 patients receiving sham acupuncture treatment. Based on Takeda & Wessel's data, the mean of pain rating index of McGill Pain Questionnaire (PRI) scores of the sham acupuncture control group at three weeks was 14.30 points with a standard deviation of 12.15 points, the response of the acupuncture group was 6.5 points. Using the G*power software package (Buchner, Erdfelder, & Faul, 1997), and setting two-sided tests with $\alpha = .05$, then the power in Takeda & Wessel's study is only .51. So in Takeda & Wessel's study, the sample size has a probability of only .51 to detect a difference of improvement effect between groups.

For the remaining three studies (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Mclndoe, Young, & Bone, 1995), there is inadequate information for a retrospective power calculation.

2.4.4.2 Acupuncture quality

- **TCM diagnosis**

As with any kind of therapy, the best results are based on correct diagnosis. Traditional Chinese acupuncture also needs the correct TCM diagnosis to obtain good results from treatment. It is unclear which types of acupuncture had been applied in the five reviewed studies. Although four of the trials used classical TCM acupuncture points and Ah Shi points, no information on the TCM diagnosis could be found (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001; Takeda & Wessel, 1994).

Two possible explanations may account for this. One is that musculo-skeletal pain problems are frequently treated in a formulaic manner. Although this is not "good" acupuncture practice in terms of TCM diagnosis and individualisation of treatment, this repeatable formulaic manner can be more easily applied in acupuncture research that uses RCTs as the method. Since the relatively simple local musculo-skeletal approach may be effective in pain management, the TCM diagnosis is ignored. Another possible explanation is that TCM diagnosis needs basic TCM theory, syndrome differentiation (diagnosis), and experienced TCM practitioners. However the concepts and knowledge of TCM are not easily understood and accepted within the western bio-medical culture.

- **De qi**

De qi refers to the needle sensation. It is often described as a feeling of heaviness and numbness, occasionally combined with a cold or warm feeling at the acupuncture site and may travel to other parts of the body. De qi may also refer to the feeling of tightness and heaviness under the tips of fingers of the acupuncturist (Xi, 1985). De qi is considered to be an important amplifying factor for the effectiveness of acupuncture (Cheng, 1997; Huang, 1999; Xi, 1985). One study reported that the subjects experiencing de qi regularly during treatment responded better to acupuncture than those who did not experience this sensation (Takeda & Wessel, 1994). These authors suggested that further studies in acupuncture needs to be undertaken with better control of de qi.

- **Point selection and acupuncture "dosage"**

The acupuncture treatment used in the study should have correct point selection and adequate "dosage". Acupuncture point selection is based on the diagnosis and the patient's condition. Acupuncture dosage may include the number of points; depth of needing; duration of needling; strength, duration and repletion of stimulation; number and frequency of sessions.

The point selection was formulated in each study but varied among the studies. The dosage of treatment was also variable within the studies. According to the common practice of experienced Chinese acupuncturists (Qiu, Kong, & Tu, 1988), a treatment must consist of at least 10-15 sessions at an incidence of 3-7 per week, each session lasting for at least 10-30 minutes or longer. One study used only three treatment sessions to treat OA patients awaiting hip replacement (Mclndoe, Young, & Bone, 1995). This study does not reflect clinical practice that would be acceptable to many

TCM acupuncturists, which seriously limits the conclusion that can be drawn from it. In the remaining four studies, the number and frequency of sessions were considered slightly inadequate, with 6-10 sessions at an incidence of 1-3 per week.

Two studies with high quality scores on the Jadad scale had positive results with only six treatment sessions at a frequency of 1-2 per week (Christensen et al., 1992; Haslam, 2001). This suggests that six treatment sessions, at a frequency of 1-2 per week, appears to be a good cut-off point for minimum adequate treatment. However, there were insufficient numbers in this review to reach any conclusion about the best application. A dose-response relationship has not emerged in terms of treatment duration and frequency of acupuncture.

Three ways of establishing correct point selection and dosage of acupuncture currently have been suggested. These include a literature review (Birch, 1997), surveying and consulting experts to establish a consensus (Sherman, Hogeboom, & Cherkin, 2001), or conducting controlled comparative clinical trials looking at every relevant variable of treatment (White, Filshie, & Cummings, 2001). However, most TCM acupuncturists would agree that it is vital to use treatment that is individualised to each patient at each session. Therefore, the established point selection and dosage of treatment, if any, should be used as an optimal treatment or guideline only. Individualised acupuncture treatment involving point selection and dosage of treatment still require further investigation.

2.4.4.3 Outcome measures of acupuncture treatment

Pain and function of the OA joint are considered as the primary outcome measures in most studies of OA. Since TCM acupuncture is claimed as a holistic therapy in terms of TCM theory, it is reasonable to measure the overall health status outcomes following acupuncture treatment. Only one trial assessed overall patient satisfaction assessment and quality of life, but the measures of overall assessment and quality of life may not be adequate measures within the TCM concepts (Fink, Kunsebeck, Wipperman, & Gehrke, 2001). Because of the different understanding of OA and different health concepts in WM and TCM, any research to evaluate the effectiveness of Chinese acupuncture treatment must be conducted within TCM frameworks. Since there are no existing measures of the health status within the TCM frameworks and concepts, a new measurement needs to be developed and validated.

2.4.4.4 Treatment efficacy, safety and cost

- **Treatment efficacy**

The existing evidence suggests that acupuncture may play a role in the treatment of OA of the knee and hip. However, there is insufficient evidence to confirm the effectiveness of acupuncture alone for the treatment of OA of the hip or knee.

One high-quality study of OA of the knee, showed positive results in comparing acupuncture with non-acupuncture (waiting list group) (Christensen et al., 1992). However, another high-quality study of OA of the knee reported negative results in comparing acupuncture with sham acupuncture (Takeda & Wessel, 1994). Therefore, there is limited evidence that acupuncture is more effective than waiting list (treatment

as usual), but not more effective than sham acupuncture in the treatment of OA of the knee. The first part of the findings is consistent with a previous review (Ezzo et al., 2001), but the second part of the findings differs from the Ezzo review, which suggests that there was strong evidence that real acupuncture was more effective than sham acupuncture. The difference may be due to two studies (Milligan, Glennie-Smith, Dowson, & Harris, 1981; Petrou, Winkler, Genti, & Balint, 1988) that were excluded from this review (Table B.3, see Appendix B).

One high-quality study of OA of the hip (Haslam, 2001) showed positive results in comparing acupuncture with advice and exercises. Therefore, there is limited evidence that acupuncture is more effective than advice and exercises in the treatment of OA of the hip.

The positive effects of acupuncture were maintained at approximately two-month follow-up in two studies (Fink, Kunsebeck, Wipperman, & Gehrke, 2001; Haslam, 2001). However the long-term effects (more than six months following treatment) need to be researched, and whether maintenance treatments of acupuncture are beneficial also needs to be assessed in future studies.

- **Safety and cost**

Safety and cost of a treatment are also important when making a treatment decision. Acupuncture is considered to be a safe and cost-effective treatment for patients with OA (Christensen et al., 1992; Fink, Kunsebeck, Wipperman, & Gehrke, 2001). However adverse effects, the reduction of analgesics consumption and cost need to be reported, so that a true comparison between treatment options can be made.

2.4.4.5 Previous works

Two systematic reviews (Ernst, 1997; Ezzo et al., 2001) relating to OA were found after the databases and the reference lists from papers obtained were searched (Tables B.1 and B.2, see Appendix B) . One review (Ernst, 1997) included other sites of OA as well as the hip and knee, while the other review (Ezzo et al., 2001) focused only on OA of the knee.

Both existing reviews have important limitations:

- Only studies prior to 1999 were included
- The classification of the types of acupuncture is broadly focused
- The definition of TCM acupuncture is not clear
- No studies are Chinese in origin (from China).

For patients wishing to make an evidence-based treatment decision it remains unclear whether acupuncture offers the most benefits for them. Uncertainty also remains about whether acupuncture alone is effective in the treatment for OA of the hip or knee.

2.4.4.6 Limitations of this review

This review has limitations due to time constraints and the financial requirements needs to fulfil this thesis.

This review was also limited due to the quality assessment relying on published information in reports only, not the trials themselves. However, some methodological deficiencies may lie in the reporting of trials rather than in their performance (Hill, LaValley, & Felson, 2002).

Only one reviewer conducted this review, therefore, blind assessment of studies and applying two or more independent quality assessors were not possible in the context of this review.

There are obvious difficulties with meta-analysis, not least being that pooling data from trials is difficult when using different inclusion and exclusion criteria, different control methods, and different qualities of acupuncture for the studies. Therefore the author considered that a meta-analysis was not justified.

The review included studies published in full text articles, and in English and Chinese only. Many studies were discarded because they only published the abstract with little information contributing to the review (n = 3) or published in other languages (n = 6). These studies may have lead to a different result in this review.

This review assessed acupuncture alone and many studies were discarded because they included electro-acupuncture (n = 5), moxibustion (n = 4) or other types of combination (n = 2). The review has not attempted to evaluate whether these combinations of treatment are effective or ineffective.

2.4.4.7 Conclusion of review

In summary, this review indicates that the RCTs of acupuncture in patients with OA of the hip or knee had problems with acupuncturist blinding, sham control, and lack of TCM diagnosis, assessment and measurement, resulting in limited evidence as to the usefulness of acupuncture in the treatment of OA hip or knee. The review does not

enable a definitive conclusion to be drawn about the effectiveness of acupuncture treatment in patients with OA of the hip or knee. However, there is limited evidence that acupuncture is more effective than waiting list (treatment as usual), advice and exercises, but not more effective than sham acupuncture in the treatment of OA of the knee. Due to the problems existing in the sham control studies, the specific effects of acupuncture alone remain uncertain.

2.4.4.8 Recommendations for future research

Recommendations for the conduct of future studies are made throughout this review. The specific points that need to be emphasised are:

- In relation to methodological quality, three areas need to be addressed:
    - Concealed allocation
    - Sample size justification
    - Analysis with intention-to-treat
- The use of TCM diagnosis, assessment, and outcome measurement needs to be addressed in future TCM acupuncture trial designs
- Further research is required to determine the specific treatment effects of acupuncture in treating OA of the hip or knee
- Future research is also needed into the relationship between the sensation of de qi and the result of the treatment
- There is a need for well-designed RCTs on acupuncture that measure its efficacy, safety, cost effectiveness and follow up for at least six months.

An appropriate measurement or assessment instrument is essential to evaluate TCM acupuncture in the treatment of OA of the hip or knee. However, the literature search

failed to identify any assessment or measurement instrument using TCM concepts and frameworks. To fill this gap in the literature, a study designed to develop and validate a TCM measurement instrument is required.

2.5 TCM concepts and frameworks

2.5.1 Fundamental concepts and frameworks in TCM

Fundamental concepts and frameworks in TCM are detailed in Appendix A. Among the most important concepts and frameworks related to a TCM assessment or measurement instrument are “four examinations” and “eight principal syndromes”. The “ten questions” is part of the “four examinations”. Because the “ten questions” is an important concept in development of a TCM questionnaire, this concept is also reviewed in detail.

2.5.2 TCM assessment, the pattern identification and the eight principal syndromes

TCM assessment refers to “four examinations”. The four examinations include inspection, listening and smelling, inquiry and palpation, which can be summarised as follows (Kaptchuk, 2000, pp. 171-213):

- Inspection of the client’s overall physical condition including complexion, tongue, secretions, excrement and emotional state
- Listening to the client’s voice and respiration, and smelling of any odours associated with the body or breath

- Inquiry for information on present and past complaints including: sensations of cold or hot; perspiration; headaches and dizziness; quality, nature and location of pain; thirst; appetite and tastes; bowel movement; bladder; patterns of sleep; family health history; work; lifestyle; childhood; and emotional life
- Palpation of the body to determine pain or sensitivity, including the taking of the pulses.

After the four examinations, the data is gathered and integrated into TCM theory in order to understand the aetiology, location and nature of pathophysiological changes. This process is called “pattern identification”. The “eight principal syndromes” (ba gang) is a general method by which this process happens (Campbell, 2002; Cheng, 1997; Kaptchuk, 2000). The “eight principal syndromes” refer to four pairs of syndromes: yin and yang; cold and heat; exterior and interior; and deficiency and excess.

2.5.3 TCM inquiry and the ten questions

Inquiry of the patient is of prime importance as this is how the majority of signs and symptoms of the specific complaint and underlying systemic imbalance are discovered and classified. In TCM, inquiry is often referred to as the “ten questions”.

The first list of “ten questions” seems to have been drawn up by Zhang Jing Yue in his Complete Book (Zhang, 1986). It has been modified and used up to the present day as an outline for inquiring about patient symptoms. This is a modified content of the “ten questions” (Deng, 1999, pp. 64-65):

"First ask hot and cold, second ask sweat;

Third ask head and body, fourth ask stools and urine;

Fifth ask food and drink, sixth ask chest;

Seventh ask hearing, eighth ask cause;

Ninth ask old disease, tenth ask cause;

In taking medicines, what changes appear;

Women especially ask the time of menses, slow, fast, block or flood.

For children add experience with measles and chicken pox."

The "ten questions" help the TCM practitioner form a more complete picture of the patient's condition and provide important insights into the pattern of disharmony existing within the patient.

However, it is still possible to ask an almost endless number of additional questions concerning these ten aspects of a person's health and illness. Furthermore, the ten questions cannot provide numerical data that is amenable to statistical analysis and to quantification. Therefore, the ten questions cannot be used as reliable and valid TCM measures until they have been refined and developed.

2.6 Different concepts of health in TCM and WM

TCM views health as a state of balancing of yin and yang (Ni, 1995) which is a state of harmony of the body system and the whole person in relation to the natural world. The functioning of "how it works" and the relationship among each part in this system is more important than the structure of the part itself or "what it is". Therefore, TCM views illness and disease as the "pattern of disharmony" rather than a single symptom

or disease itself. Thus within TCM it is possible to provide meaningful and understandable explanations of illness (Sobell, 1979), which includes no clearly diagnosable condition in the terms of WM (Vincent & Furnham, 2001) or symptoms that have a poor relationship with the degree of structural changes (Gascoigne, 2001).

In TCM, the body system includes the physical body, as well as the spirit, mind, and emotions. Within TCM the importance of the spirit, mind, and emotions may have a far greater bearing in the context of health of the patient than they realise (Vincent & Furnham, 2001).

By contrast, WM historically saw health as a state of absence of disease and abnormalities (Bowling, 2001b). However the World Health Organisation (WHO) in 1946 created a broader definition of health as total social, psychological and physical well-being, and thereby removed the emphasis of health as not only being merely absence of disease or infirmity (World Health Organisation, 1958).

The recent definition of health status in WM includes both functional status (e.g., activities of daily living) plus other aspects of health such as emotions and mood, symptoms (e.g., pain, sleep disturbance, fatigue), cognitive abilities, and social activities and roles (Froberg & Kane, 1989). However, WM places much emphasis on the physical body, remains uncomfortable with the concept of the psychosocial body and largely denies the existence of the energetic and spiritual bodies (Cassidy, 2001).

2.7 Requirements of a measurement instrument

The most important characteristics of a measurement instrument are reliability and validity (Bowling, 2001a; Osoba, 1998). Reliability refers to the accuracy and consistency of the measurement (Portney & Watkins, 2000). Validity of an assessment method refers to its capacity to measure what it is intended to measure (Portney & Watkins, 2000). A measurement may be reliable on the score, but it maybe consistently measuring the wrong outcome rather than what it is supposed to measure. The concepts of reliability and validity are described briefly here.

Reliability can be conceptually defined as an estimate of the extent to which a test score is free from random error (Dimitrov, Rumrill, Fitzgerald, & Hennessey, 2001), that is, to what extent observed scores vary from true scores (Portney & Watkins, 2000). True score for an individual is assumed to be invariant on repeated measurements (Hays, Anderson, & Revicki, 1998). The maximum reliability or zero error is expressed as a reliability coefficient of 1.00 (Portney & Watkins, 2000).

There are four basic categories of reliability estimation: inter-rater or intra-rater, equivalent forms, test-retest, and internal consistency reliability (Hays, Anderson, & Revicki, 1998; Portney & Watkins, 2000). Inter-rater reliability refers to a comparison of scores assigned to the same target (patient) by two or more raters. Intra-rater reliability refers to the stability of data recorded by one individual across two or more trials (Portney & Watkins, 2000). Equivalent forms reliability refers to the agreement between an individual's score on two or more measures designed to measure the same attribute. Test-retest reliability is a method to assess whether a measure is reliable over time, keeping all testing conditions as constant as possible (Portney & Watkins, 2000).

Unfortunately, many clinical variables may change over time. Therefore, a low test-retest reliability coefficient could also indicate a result of real change in health status or familiarity and greater understanding of the measurement. Internal consistency reliability for a scale is a function of the number of items and their co-variation, and is measured using the Cronbach's alpha statistic (Cronbach, 1951).

Validity, in contrast, places an emphasis on the nature of what are being measured (Streiner & Norman, 1995), and the ability to make inferences from test scores or measurements (Portney & Watkins, 2000).

There are also many types of validity. Essentially, there are four aspects to validity: face validity, content validity, criterion validity, and construct validity (Jenkinson & McGee, 1998). Face validity refers to whether items on a measurement instrument appear appropriate, make sense, and are easily understood. Content validity refers to a representative range of the content under study. Criterion validity refers to the ability of a measurement to correspond with an existing, valued measurement or "gold standard". Construct validity refers to the ability to determine if a measurement actually measure an abstract or construct. Construct validity is commonly validated through methods that include the known groups method, convergence (closely co-relate with related variables) and discrimination (does not co-relate with unrelated variables), factor analysis, hypothesis testing, and criterion validation (Portney & Watkins, 2000).

The determination of which forms of reliability and validity for a measurement can be made in a variety of contexts, depends on the measurement's intended purpose and a specific population (Portney & Watkins, 2000). Therefore, the different forms of

reliability and validity must be evaluated within the different contexts of the measurement's purpose and the specific population.

Another important desirable characteristic of a measurement is its responsiveness to change, particularly clinically important change (Bowling, 2001a; Streiner & Norman, 1995). There is an unresolved debate about whether responsiveness is an aspect of validity (Hays & Hadoron, 1992).

When a measurement contains multi-item scales, and the scales are scored using Likert's (1932) method of summated ratings, the scaling assumptions need to be tested (Ware, Kosinski, & Gandek, 2002). The scaling assumptions include whether the distribution of responses to items within the same scale and item variances are roughly equal, whether each item has a substantial linear relationship with the score for its scale (referred to as item internal consistency), and that whether each item clearly measures one concept more than other concepts (referred to as item discriminant validity) (Ware, Kosinski, & Gandek, 2002). See section 3.3.8.4 for further explanations of item internal consistency and item discriminant validity.

2.8 Limitation of current measurement instruments

There are different types of measurement instruments, which may include disease-specific (e.g. OA disease), domain-specific (e.g. social support, coping, self-esteem and other domains) and generic (e.g. health-related quality of life or health status of physical, mental and social heath) measurement (Bowling, 2001a). The generic instruments used in OA studies may include the Medical Outcome Study 36-item Short Form (SF-36) health survey (Ware & Sherbourne, 1992), and the Nottingham Health Profile (NHP)

(Hunt et al., 1981). The disease-specific instruments used in OA studies may include the Western Ontario MacMaster Universities Osteoarthritis Index (WOMAC) (Bellamy, Buchannan, Goldsmith, Campbell, & Stitt, 1988a, 1988b), the Arthritis Impact Measurement Scales (AIMS) and its revision (Meenan, Gertman, & Mason, 1980; Meenan, Mason, Anderson, Guccione, & Kazis, 1992), and the Health Assessment Questionnaire (HAQ) and its modifications (Fries, Spitz, & Kraines, 1980; Fries, Spitz, & Young, 1982). These instruments were developed well within WM model and place much emphasis on the pain and functional disability (Wolfe, 1995).

Taking the SF-36 health survey for an example, among the numerous existing measurement instruments available, the SF-36 health survey is the most commonly used health status measure in contemporary medicine (Kaplan, 1998). The SF-36 health survey includes eight health concepts: physical functioning (limitation in physical activities due to health problems), role-physical (limitation in work or other daily actives due to physical health), bodily pain (bodily pain and its interfere with normal work), general health (general health and its perceptions), vitality (energy level and fatigue), social functioning (limitation in social activities due to physical or emotional problems), role-emotional (problems with work or other daily activities due to emotional problems), and mental health (psychological distress and well-being) (Ware, Kosinski, & Gandek, 2002). The reliability and validity of the SF-36 health survey is well documented (Haley, McHorney, & Ware, 1994; McHorney & Ware, 1995; McHorney, Ware, & Raczek, 1993; Stewart, Hays, & Ware, 1988; Ware & Sherbourne, 1992).

In chronic diseases, such as OA, the measurement of pain and physical limitations, as well as quality of life are important. Use of SF-36 health survey as a measure in studies

of OA treatment would be useful. As Ware (1993) concluded, because the SF-36 health survey is short, well tested, and population norms exist, it may constitute a good generic core for use along with disease-specific outcome measures. However, SF-36 health survey may not be appropriate for TCM studies due to the difference of constructs in TCM health. For example, the vitality scale of the SF-36 health survey measures something quite different from "vital energy" (qi) as measured by TCM (Wittmann & Walach, 2002).

The literature search within the field of OA and acupuncture failed to identify any measurement or assessment instrument developed and validated within TCM concepts and frameworks. The lack of reflecting to the construct of TCM concepts and frameworks is a major limitation of current measurement instruments used to evaluate TCM acupuncture treatment in the treatment of OA. Due to the different understanding of OA and different health concepts in WM and TCM, the validation of the effectiveness of Chinese acupuncture treatment, research must be conducted within the TCM frameworks.

2.9 Summary of the review of literature

A literature review identifies the different understanding of OA and different health concepts in WM and TCM. The review suggests acupuncture is potentially an effective treatment for OA of the hip or knee. Essential to evaluate Chinese acupuncture treatment is the availability of a reliable and valid measurement. However, the literature search within the field of OA and acupuncture failed to identify any measurement or assessment instrument developed and validated within TCM concepts and frameworks.

2.10 Specific aims of the pilot study

To fill the gap in the literature of the lack of appropriate measuring instruments, a study was designed to develop and validate a TCM questionnaire as a health status instrument for TCM assessment in patients with OA of the hip or knee.

The following specific aims needed to be tested in this pilot study:

- The questionnaire is developed within TCM concepts and frameworks: The draft of the questionnaire is accepted by patients with OA of the hip or knee and acupuncture experts
- The questionnaire has acceptable reliability: The internal consistency reliability and test-retest reliability are above .70 (Cronbach, 1951; Lu, 1996; Nunnally, 1978; Streiner, 1993; Streiner & Norman, 1995)
- The questionnaire has acceptable validity: 1) the content validity is representative of or corresponds to the "eight principal syndromes" and "ten questions" in TCM concepts and frameworks. 2) The construct validity is supported by: a) strong or moderate associations between the subject's age and each of the related scale scores, b) there is significant difference between the groups of use and non-use of ongoing medication in the related scale scores, c) the correlations between scales are less than their reliability coefficients (Ware & Gandek, 1998), d) high or medium correlations between the comparable TCM questionnaire scale and SF-36 health survey scales. A strong association is defined as a Pearson correlation coefficient (r-value) between the questionnaire and above comparable groups of more than .70, while a moderate association is

defined as an r-value between .30 and .70 (Ware, Kosinski, & Gandek, 2002)

- The questionnaire has success of grouping (i.e. scaling). The success of grouping is supported if the individual question (i.e. item) is substantially linearly related to the total scale score (test of item internal consistency) and has stronger measures of its hypothesised constructs than of other constructs (test of item discriminant validity). Item internal consistency is satisfactory if an item correlates .40 or more with its hypothesised scale (after correlation for item-scale overlap) (Ware & Gandek, 1998). Item discriminant validity is supported if the correlation between and item and its hypothesised scale is significantly higher than the correlations between that item and all other scales. The default significance level for comparing two correlations is two standard errors (Ware & Gandek, 1998). The standard error of a correlation coefficient is approximately equal to one divided by the square root of the sample size.

CHAPTER 3 METHODOLOGY

3.1 Introduction

The methods used in this pilot study include: how a TCM questionnaire was developed, and how the questionnaire was validated. Because there are no existing criteria for the construction and validation of a TCM questionnaire, the generally accepted TCM concept and framework such as the "eight principal syndromes" and "ten questions" were used to develop the questionnaire. The questionnaire format was also designed in a critical but comprehensive manner. Next, the methods and procedures used in this validation study were adapted from another well-known WM questionnaire validation study, such as the methods of validation of SF-36 health survey (Ware & Gandek, 1998).

3.2 TCM questionnaire development

3.2.1 The content of questionnaire

To measure health status in TCM, the term must be operationally defined. Generally, TCM concepts of health status can be defined as a state of balancing of yin and yang (Ni, 1995). Since the concepts of yin and yang can be further divided into endless components (see Appendix A for detailed explanation), a major problem in the development of a TCM measure is the absence of criteria for the construction and

validation of health scales. It was not this thesis's purpose to create a set of scales that would totally measure the TCM concept of health status, but rather to address certain key components to assess the major dimensions of the concept. Therefore, the corresponding "eight principal syndromes", and "ten questions" were used in selecting items for each TCM questionnaire scale.

The "eight principal syndromes" are divided into eight components including exterior, interior, cold, heat, deficiency, excess, yin and yang. These components provide the basic foundation upon which a more precise evaluation of the patient's condition can be made (Campbell, 2002). The "eight principal syndromes" have been traditionally employed in the process of making a TCM diagnosis (Campbell, 2002; Deng, 1999; Kaptchuk, 2000).

Use of the "ten questions" in patients with OA of hip or knee, may be summarised into two groups of questions:

- The questions about the "local" conditions (e.g., joint of hip or knee)
- The questions about the "whole" conditions (e.g., physical and mental body).

Because pain is one of the major concerns in the OA patients, a subgroup of questions (Question A) about the pain of the hip or knee joint was generated from the first group of questions. Therefore, two main groups (Question A and B, and Question C) of questions have been developed (see Appendix C).

Two groups of questions were further subdivided into specific subcomponents according to the “eight principal syndromes”. Each subcomponent reflects important areas within the OA disorders.

3.2.2 Selection of items

3.2.2.1 The source of items

There is currently no standard question or standard response formats that have been used successfully in past TCM studies. Therefore, the items (i.e. questions) of the TCM questionnaire have been generated mainly from TCM theory and clinical experience.

According to TCM theory, “Bi syndromes” are due to an underlying deficiency of the body’s defensive qi and blood and deficiency of organs such as kidney and liver. These pre-existing factors allow the wind, cold, heat and damp to penetrate and cause obstruction of qi and blood in the channels (see Chapter 2.3 “Osteoarthritis in TCM” for more detailed explanation). Therefore, the items related to the underlying problems and the characters of invasive factors are more likely to be represented.

The choice of these items was also predicated on the clinical experience. Usually, the main symptoms include pain and other sensation in the OA joint, and the general conditions, are considered in making TCM assessment. Since clinical encounters are likely to be brief, the questions need to be reasonably short and phrased with respect to the daily clinical setting.

3.2.2.2 The wording of the questionnaire

The wording of the questionnaire was checked for any misunderstanding, ambiguity, and to ensure that it was jargon-free. For example, “hard bowel motion” instead of “constipation”, “ringing in the ears” instead of “tinnitus”, “feeling down related to OA” instead of “depression”. Because negatively worded items tend to have a low validity coefficient (Schriesheim & Hill, 1981), items which use words such as “not”, “never”, and which have words with negative prefixes, such as “in-“, “un-“, have been avoided. For example, “poor appetite” instead of “not good appetite”, “difficulty falling asleep” instead of “not easily falling asleep”.

3.2.2.3 The length of items

Each item asks only a single and easy to understand question. The length of items was kept reasonably short with about 10-20 words only. It was found that, on average, items containing 10-20 words had validity coefficients almost four times higher than items with 70-80 words (Holden, Fekken, & Jackson, 1985).

3.2.3 Format of the questionnaire

3.2.3.1 A self-administered questionnaire

It is known that doctors’ judgments and patients’ estimations of their state of health are only weakly correlated and sometimes not correlated at all (Koller, Kussmann, & Lorenz, 1996). It is important that the patient’s feelings about their health are correctly identified. Patient self-reported symptoms such as pain and sensations are important in

TCM assessment. A patient self-administered questionnaire that measures TCM health status would be useful.

3.2.3.2 The four-week recall

There are two major types of recall biases might occur in responses to questions: the respondent reports too much (over-reporting) or too little (under-reporting) relative to their actual experiences (Lu, 1996). The over-reporting and under-reporting often result from the following two reasons:

- Memory declines over time, and symptoms occurring more recently are easier to remember than more distant events.
- There is a phenomenon often referred to as 'telescoping', which is a general tendency to report events from outside the time period being asked about in the question.

It was suggested that respondents were significantly more comfortable reporting their health-related behaviours over a four-week versus a two-week period (Yates, Wagner, & Surprenant, in press).

Therefore, a four-week recall period was adapted in the TCM questionnaire questions. Because it was intended to define periods sufficiently long to generate details of enough symptoms for assessment purposes, but short enough for patients to remember all or at any rate a high proportion of what their feelings and symptoms. The four-week recall period was applied in SF-36 (4-week recall, standard form version) health status survey (Ware, Kosinski, & Gandek, 2002) and the author thought that the previous four weeks

would capture a more representative and reproducible sample of recent health, but not be unduly affected by daily or momentary fluctuations (Stewart & Ware, 1992).

3.2.3.3 The scaling response formats

In considering approaches to the development of response scales, there are broadly four levels of response scales: nominal scales, ordinal scales, interval scales, and ratio scales (Jenkinson & McGee, 1998; Lu, 1996; Streiner & Norman, 1995). Nominal scales distinguish classes of objects, such as, “have knee pain” or “have no knee pain”. Ordinal scales are scales on which classes or objects are ordered on a continuum, for example, from “no pain”, “mild pain”, “moderate pain”, and “sever pain”, to “extreme pain”. By contrast, interval scales are ordered and the distances between values on one part of the scale are equal in distance to the distances between values on another part of the scale, for example, temperature scales. Ratio scales are similar to interval scales but with a meaningful zero point, so that the ratio of two responses has some meaning, for example, time scales.

Ideally measures should be developed using interval or ratio scales (Jenkinson & McGee, 1998), because the measures such as means, standard deviations, and differences among means can then be interpreted. However, measures of health status can never truly be regarded as fulfilling the requirement of these forms of measurement (Bowling, 2001b). In reality, we can never be certain whether the values in the measures are truly accurate. This is discussed in the following section.

The visual analogue scale (VAS) has been used extensively in medicine to assess a variety of constructs such as pain and functional capacity. The scale is usually a line of

100 mm with anchors like “no pain” and “pain as bad as it could be” at the opposite ends. Respondents are required to place a mark on the line corresponding to their perceived state. Although one can measure a response to an apparent accuracy of one per cent (a length measured in millimetres), there is no guarantee that the response accurately represents the underlying attribute to the same degree of resolution (Streiner & Norman, 1995). There is research showing that patients had more difficulty in using a VAS, than using numerical or adjectival scales which were rated as “none”, “mild”, “moderate”, “severe”, to “extreme” (Bosi Ferraz et al., 1990; Huskisson, 1974).

There is really no guarantee that the true distance between “none” and “mild” is the same as the distance between “severe” and “extreme”. The rating scales, such as visual analogue scale (VAS), numerical and adjectival scales are on an ordinal level of measurement. Thus, attempts are made to convert ordinal responses into interval level scales using weight responses rather than simply using arbitrary codes (such as 0 = none, 1 = mild, and so forth). However, there is a considerable body of evidence that weighting schemes do not generally improve the accuracy and reliability of most health status measures (Jenkinson, 1991; Lei & Skinner, 1980; Streiner, Goldberg, & Miller, 1993). It has been suggested that unless the distribution of scores is severely skewed, one can analyse data from rating ordinal level of scales as if they were interval levels without introducing severe bias (Streiner & Norman, 1995).

Based on the above considerations, the ordinal scaling response format (e.g., none, mild, moderate, severe, extreme) was chosen in the development of TCM questionnaire. This format can also reflect varying degrees of symptoms or imbalance a respondent has in TCM concepts of health status.

3.2.3.4 The five-category scale

There is evidence that reliability of scales drops as fewer categories are used (Nishisato & Torii, 1970). It is also suggested that the minimum number of categories used by raters needs to be in a range of five to seven (Nishisato & Torii, 1970; Streiner & Norman, 1995). Therefore a five-category scale from "none" to "extreme" was chosen in the TCM questionnaire format. The five-category scale was also expected to reduce the loss of information and still be within the patient's ability to discriminate.

3.2.4 Ethics and ethical approval

The major issue in ethical considerations for this questionnaire's development is the autonomy of the subject. The participant needs to be seen as an autonomous individual, who has the right to privacy and to decide whether to participate in this study.

To protect the participant's privacy and confidentiality, the questionnaires were prepared with unique numeric coding. All names and contact details were deleted from data, leaving only unique number and demographic information. Therefore, forms were data-entered and matched via unique number to demographic information.

The participants were asked to read and understand the information provided about this research before they signed the consent form (see Appendix E).

The TCM questionnaire and its research project were approved by the Auckland University of Technology Ethics Committee.

3.2.5 Pre-testing of the questionnaire

Two patients with OA of the knee and one patient with OA of the hip from a acupuncture clinic were interviewed to discuss whether items in the TCM questionnaire are relevant, clear, and whether all the main themes had been covered. To test whether item responses represented correctly what a respondent thought, the patients also were asked to think aloud as they formulated their responses.

Finally, the preliminary draft of the TCM questionnaire was examined by two highly experienced TCM acupuncturists. Their expert opinion contributed to the content validity study and face validity of the questionnaire. Their suggestions also offered the opportunity for additional items to be added. As a result the wording of the questions was altered slightly, and an item of "feeling down" was added to measure the state of the emotions. The developed questionnaire was then subjected to the validation study.

3.2.6 Summary

The TCM questionnaire (see Appendix C) was developed within TCM concepts and frameworks. The questionnaire contains 23 items with two main groups of questions (three subgroups of A. B. and C.). The two main groups of questions formed two main scales (the exterior scale which contains question item 1 to 8 and interior scale which contains question item 9 to 23), and one additional scale (the summary scale which contains question item 1 to 23). The questionnaire takes approximately five minutes to complete and is entirely self-administered. The developed TCM questionnaire was subject to the next stage of validation study.

3.3 TCM questionnaire validation

3.3.1 Study design

A repeated measures (one-way) experimental design is applied in this pilot study. One group of subjects is tested under all conditions and each subject acts as their own control (Portney & Watkins, 2000). This study design allows estimation of the TCM questionnaire's test-retest and internal consistency reliability, as well as construct validity.

3.3.2 Subjects for pilot study

Patients with symptomatic OA of the hip or knee were invited to join this pilot study. To be eligible, patients had to have definite radiographic evidence of primary OA in the hip and/or knee, and fulfil defined inclusion and exclusion criteria.

3.3.2.1 Inclusion criteria

- Fulfil American College of Rheumatology (ACR) combined clinical and radiographic criteria for OA of the hip and/or knee
- Proficient in English
- Sign the consent form

ACR criteria for OA of the knee (Altman, Asch, & Bloch, 1986)

- Knee pain
- Bony growths (osteophytes) shown on the knee joint on X-ray

- And at least one of the following three features:
    - Age greater than 50 years
    - Stiffness of the knee joint within 30 minutes of activity
    - A noise from the knee joint rubbing together.

ACR criteria for OA of the hip (Altman, Alarcon, & Appelrouth, 1991)

- Hip pain
- And at least two of the following three features:
    - A normal Erythrocyte Sedimentation Rate (ESR) blood test, less than 20 mm/hour
    - X-ray evidence of bony growths (osteophytes) on the thigh bone or hip joint
    - X-ray evidence of hip joint narrowing.

3.3.2.2 Exclusion criteria

- No radiographic (e.g., x-ray) evidence to confirm the OA of the hip and/or knee
- Having none or limited English language skills
- The consent form was not signed.

Because the TCM questionnaire was developed in English, any translation may affect the reliability and validity of the questionnaire (Bowling, 2001a; Ren, Amick, Zhou, & Gandek, 1998). There remains concern that certain features of the language, such as idioms, are very difficult to translate and make little sense within a different cultural

context (Ren, Amick, Zhou, & Gandek, 1998). Therefore, proficiency in English was required.

3.3.2.3 Sampling frame and generalisation

Participants for this pilot study were chosen through convenience sampling from the primary care population within the Auckland area in New Zealand.

In an attempt to establish external validity of generalisation, participants were recruited from six different clinical settings including three medical centres, two physiotherapy clinics, and one acupuncture clinic.

3.3.2.4 Sample size

Twelve participants were recruited in this pilot study. The sample size for full study validation of the TCM questionnaire needs to be estimated on the basis of this pilot study and study assumptions.

3.3.3 TCM questionnaire

3.3.3.1 The TCM questionnaire

The developed TCM questionnaire (see Appendix C) was used in this pilot study.

3.3.3.2 Collecting additional questions

There were additional un-scaled questions in the first questionnaires (see Appendix C) asking respondents about their health: whether they have on-going medication for OA, whether they have more than one OA joint, whether they have the OA joint as a primary health problem, and demographic characteristics (e.g., age and gender). These questions were used partly as a form of construct validity. For example, use of on-going medication for OA in the hip or knee might be expected to be higher in patients with low scores on TCM questionnaire, as medications are perceived to help the pain relief in patients with OA of the hip or knee.

3.3.4 Choosing the SF-36 health survey as a comparison form

SF-36 health survey was chosen as a comparison form to estimate the construct validity of the TCM questionnaire. It was based on the following four main considerations:

- The validity and reliability of SF-36 health survey has been well documented (McHorney & Ware, 1995; Stewart, Hays, & Ware, 1988)
- It consists of eight comprehensive aspects of health, which may be considered as a "holistic" measure commonly used in contemporary medicine (Kaplan, 1998)
- SF-36 Standard Australia/New Zealand Version 1.0 developed by the International Quality of Life Assessment (IQOLA) Project is available (Sanson-Fisher & Perkins, 1998)
- A licence to use the SF-36 health survey in this study was granted.

3.3.5 Data collection procedures

Data collection was performed from 16th of July 2003 to 13th of March 2004.

3.3.5.1 Preparing questionnaires

The TCM questionnaire (see Appendix C) was incorporated into a stamped addressed envelope, together with the SF-36 questionnaires (see Appendix D), a participant information sheet (see Appendix E.2), a consent form (see Appendix E.3) and a cover letter (see Appendix E.4). The questionnaires and the consent form have the same unique code for each participant.

3.3.5.2 Delivery of questionnaires

The practitioner or receptionist from each of the selected clinical sites handed out the envelopes containing the questionnaire package to their patients on arrival who fulfilled the study criteria for OA of the hip or knee. Appropriate patients were also selected as volunteers after seeing an advertisement (see Appendix E.1) in the clinic's waiting room.

3.3.5.3 Collecting questionnaires on the first test

Prior to participation in this study, participants were invited to read the "Participant Information Sheet" (see Appendix E.2) and signed an informed consent form (see Appendix E.3). The participants' consent forms were left at receptions to be collected weekly by the researcher during the study period. Confidentiality was ensured by numeric coding of the questionnaire.

Participants completed the TCM questionnaire and the SF-36 health survey after their appointment at the above clinics, and returned them in the stamped addressed envelopes.

3.3.5.4 Collecting questionnaires on the second test

To examine the test retest reliability, a copy of the TCM questionnaire (see Appendix C) and SF-36 health survey (see Appendix D), together with an instruction letter (see Appendix E.5) and a stamped addressed envelope, were sent to the respondent at a two-week interval. The two-week interval was expected to minimise memory effects, but was sufficiently short to assume that the OA underlying process was unlikely to have changed between two occasions of assessment.

3.3.5.5 Reminding non-respondent

For those who had not responded to the re-testing questionnaires, a reminder letter (see Appendix E.6), replacement questionnaires, and a stamped addressed envelope were mailed approximately one week later. Those who failed to respond to the replacement questionnaires were contacted by phone to determine their reasons for withdrawal. See Appendix F for the questionnaires sent and received checklist.

3.3.6 Scoring the TCM questionnaire

The TCM questionnaire items and scales were scored. A higher score indicates better health. The TCM questionnaire items and scales were scored in four steps (Ware, Kosinski, & Gandek, 2002):

- Data entry (item recording)

- Computing scale scores by summing across items in the same scale (raw scale scores)
- Transforming raw scale scores to a 0 – 100 scale (transformed scale scores, and allowed direct comparison with SF-36 scales)
- Scoring checks.

3.3.6.1 Data entry

The TCM questionnaire item responses were entered as coded in the questionnaire (None = 5, Mild = 4, Moderate = 3, Severe = 2, Extreme = 1). The following rules for handling some of the common coding problems were adapted from the SF-36 Health Survey Manual & Interpretation Guide (Ware, Kosinski, & Gandek, 2002):

- If a respondent marks two responses which are adjacent to each other, randomly pick one and enter that number
- If a respondent marks two responses for an item and they are not adjacent to each other, code that item "missing"
- If a respondent marks three or more responses for an item, code that item "missing"
- If a respondent misses more than 50% items in A or B or C, the questionnaire is withdrawn from the study
- If a respondent misses an item, code the item mean score for the sample as the missing item value.

3.3.6.2 Computing raw scale scores

The raw scale score was the sum of responses for all items in the scale (see Table 3.1).

Table 3.1 Scoring the TCM questionnaire scales

Scale	No. of Items[a]	No. of Levels	Sum Item Values	Lowest and Highest Possible Raw Scores	Possible Raw Score Range
EXT	8	33	1+2+3+4+ 5+6+7+8	8, 40	32
INT	15	61	9+10+11+12+13+14+15+ 16+17+18+19+20+21+22+23	15, 75	60
SUM	23	93	1+2+3+4+5+6+7+8+9+10+11+12+ 13+14+15+16+17+18+19+20+21+22+23	23, 115	92

Note. EXT = Exterior; INT = Interior; SUM = Summary

[a]Item numbers correspond to the TCM questionnaire in Appendix C.

3.3.6.3 Transformation of scores

The next step involved transforming raw scale scores to 0 to 100 scale scores using the formula (Ware, Kosinski, & Gandek, 2002) shown below:

$$\text{Transformed score} = \frac{\text{Actual raw score} - \text{Lowest possible raw score}}{\text{Possible raw score range}} \times 100$$

For example, an exterior raw score of 24 would be transformed as follows:

(24 – 8)/32 x 100 = 50

Where lowest possible score = 8, and possible raw score range = 32.

3.3.6.4 Scoring checks

Scoring checks included the following two steps:

- Data entry checked for any missing or mistyping data, and checked for the methods of handling the coding problems whether they followed the Data Entry rules (see section 3.3.6.1)
- The TCM questionnaire scale scores were calculated by hand for 30% of respondents and the results compared to those produced by the computer.

3.3.7 Scoring the SF-36 health survey

Scoring the SF-36 health survey was performed based on its scoring instructions from the SF-36 Health Survey Manual & Interpretation Guide, under the section 6: scoring the SF-36 (Ware, Kosinski, & Gandek, 2002).

3.3.8 Statistical analysis

3.3.8.1 Outline of statistical analysis

To address the specific aims of this pilot study, the following methods of analysis were used:

- Cronbach's alpha and intra-class correlation coefficient (ICC) were used to estimate reliability for each of the following dependent variables: exterior (EXT), interior (INT), and summary (SUM) scale score. An acceptable alpha or coefficient level was set at .70 (Cronbach, 1951; Lu, 1996; Nunnally, 1978; Streiner, 1993; Streiner & Norman, 1995) for comparison

among groups, and .90 for comparison among individuals (Lu, 1996; Nunnally, 1978; Streiner & Norman, 1995) (as discussed under "3.3.8.2 Tests of reliability" section)

- Repeated measures analysis of variance (two-way ANOVA) was used to test for differences (or bias) between test and retest for each of the following dependent variables: EXT, INT, and SUM scale score. A significant level was defined as $p < .05$
- Pearson's product-moment correlation coefficient (r) was used to estimate construct validity of each of the TCM questionnaire scales. The correlation estimated associations between subject's age and each of the following dependent variables: INT and SUM scale score; between scales of the TCM questionnaires; between comparable scales of the TCM and SF-36 questionnaires. A degree of association (r-value) was set as above .7 for strong association, below .3 for weak association, between .30 and .70 for moderate to substantial association (Ware, Kosinski, & Gandek, 2002) (as discussed under "3.3.8.3 Tests of validity" section)
- Single factor analysis of variance (one-way ANOVA) was use to test for difference between groups of each factor for each of the following dependent variables: EXT, INT, and SUM scale score. The factors included: gender, use of drugs, more OA joints, and OA is a primary health problem. A significant level was defined as $p < .05$
- Pearson's product-moment correlation coefficient was also used to estimate item internal consistency and item discriminant validity (as discussed under "3.3.8.4 Tests of grouping success" section). The success of item internal consistency was defined as item-scale correlation coefficient (r-value) of above .40 (after correction for overlap). The success of item discriminant

validity was defined whether the correlation between item and its hypothesised scale is significantly higher (two stand errors or more) than the correlations between that item and another TCM scale (except the SUM scale).

All data were analysed using the SPSS 12.0.1 for Windows computer package.

The rationale use of the above statistical analysis and the procedure of analysis are explained in the following details.

3.3.8.2 Tests of reliability

Mathematically, reliability refers to how much of the variation in a score is real or true as opposed to chance or random errors (Selltiz, Wrightsman, & Cook, 1976). The reliability of the TCM questionnaire was estimated by examining:

- Internal consistency reliability (Cronbach's alpha statistic)
- Test-retest reliability (Intra-class correlation coefficients).

Reliability coefficient value in this study referred to either Cronbach's alpha or intra-class correlation coefficients (ICC) from the reliability test. Reliability values of .70 and higher are acceptable for group comparisons (Cronbach, 1951; Lu, 1996; Nunnally, 1978; Streiner, 1993; Streiner & Norman, 1995). Reliability values below .5 are regarded as low and suggest that items in a scale are not all tapping the same underlying health concept (Bowling, 2001a). Reliability values above .9 are taken to mean that a measure has very high internal reliability and can be used for comparisons among individuals (Lu, 1996; Nunnally, 1978; Streiner & Norman, 1995), however it may also

indicate that the same question is being asked more than once (Jenkinson & McGee, 1998).

- **Internal consistency reliability**

Internal consistency reliability indicates the measurement of the same concept by different scale items. Internal consistency reliability was evaluated by Cronbach's alpha statistic (Cronbach, 1951). Cronbach's alpha statistic is the average of all possible split-half reliabilities adjusted to the original number of items (Cornbach & Warrington, 1951). Spilt-half reliability refers to the split-half correlation. If a scale is developed to measure a single characteristic, the two halves of the test of this scale will provide fairly equivalent results. For example, the score from the sum of the TCM questionnaire item 1, 3, 5, and 7, provides a result equivalent to the sum of the item 2, 4, 6, and 8 in the EXT scale.

The computation of Cronbach's alpha was based on the number of items on the scale and the ratio of the average inter-item covariance to the average item variance. Therefore, Cronbach's alpha was applied for estimating the TCM questionnaire scale internal consistency reliability.

Different items within a scale can be regarded as different measures, Therefore, intra-class correlation coefficients (ICC) (McGraw & Wong, 1996; Shrout & Fleiss, 1979) may also be appropriate to estimate internal consistency reliability. A two-way random model with consistency (rather than absolute agreement) was applied in the ICC reliability test. The two-way random effects model assumes that different assessments are randomly selected. In this model, the main effect of multiple assessments is incorporated into the estimate of total variability. The random effects model provides an

estimate of agreement because it assesses whether measures (e.g. items) are interchangeable (Hays, Anderson, & Revicki, 1998).

- **Test-retest reliability**

Test-retest reliability means the same participant undertake the same task or measurement at two or more times when no change is expected. Test-retest reliability is also called intra-rater reliability, which is statistic agreement of a measure when administered by the same rater on different occasions.

Test-retest reliability has traditionally been analysed using the correlation coefficients. However, it is suggested that simple correlation analysis is less appropriate for studying concordance, because it makes no allowance for chance agreement (Bowling, 2001a), but only covariance between the test-retest. For example, if scores were consistently higher (or lower) on the retest, the correlation coefficient would be a high positive. Furthermore, because correlation testing cannot separate out variance components, due to error or true differences in a data set, it is suggested that the correlation coefficient is not a true reliability coefficient (Portney & Watkins, 2000; Streiner & Norman, 1995).

The ICC, therefore has become a preferred index (Portney & Watkins, 2000; Streiner & Norman, 1995), because it reflects both the degree of correlation and agreement among ratings. It can also compare the variance between subjects, the variance between raters, and the variance between times with error variance.

The test and retest can be regarded as two raters or the same rater in a different administration time. The ICC is recommended to be used to evaluate rater reliability (Portney & Watkins, 2000; Shrout & Fleiss, 1979; Streiner & Norman, 1995). The ICC

was applied using two-way random model with absolute agreement (rather than consistency, due to stricter and identical patterns of scores are necessary for successful rating). The two-way random model is based on a repeated measures analysis of variance, with rater as the independent variable (Shrout & Fleiss, 1979). The ANOVA separates the total variance into effects due to differences between subjects, differences between raters, and error variance (Portney & Watkins, 2000). The F-ratio associated with the rater effects reflects the difference among raters, or the extent of agreement or disagreement among them (Portney & Watkins, 2000).

Ideally, the test-retest reliability refers the average measures value in the intra-class correlation coefficient of both tests. In reality, the administration of the questionnaire on a single occasion is considered as the only interest. Therefore, the single measure values of the intra-class correlation coefficient were adapted in the questionnaire reliability.

To further examine the distribution of differences, the method of Bland and Altman (Bland & Altman, 1986) was applied. According to the method of the Bland and Altman, the differences between test and retest against their mean for each subject were plotted. The mean and standard deviation of the differences and limits of agreement were calculated.

3.3.8.3 Tests of validity

- **Content validity**

Content validity was examined in two ways (Hays, Anderson, & Revicki, 1998; Portney & Watkins, 2000; Streiner & Norman, 1995):

- By a literature review, to check whether the scales of questionnaire were covered by at least one question, and whether there were no irrelevant items
- By the questionnaire pre-testing, which had been done in the questionnaire development.

- **Construct validity**

Construct validity was examined in three ways in this study (Hays, Anderson, & Revicki, 1998; Portney & Watkins, 2000; Streiner & Norman, 1995):

- Known groups method
- Convergence and discrimination (correlations between scales) (Ware & Gandek, 1998)
- Criterion validation (correlations with the SF-36 health survey)

1) Known groups method

a) Subject's age:

It is known that the older people may have poorer health in general conditions within the TCM concepts (see section 2.3.2). Therefore, strong correlations were expected in the following groups:

- Between the age and the INT scale scores
- Between the age and the SUM scale scores.

b) Use and non-use of on-going medication groups:

It was expected that the EXT scale scores would be different between the groups using or not-using ongoing medication for the OA joint, because severe conditions of the OA

joint often require medications for pain or other symptom relief. Therefore, a single factor analysis of variance (one-way ANOVA) was applied (Hays, Anderson, & Revicki, 1998). The application of one-way ANOVA in the known groups method was also found in the SF-36 health survey validation study (Brazier et al., 1992). The significant level was defined as $p < .05$.

c) Other groups:

It was not expected in the following groups that different participant's gender and health condition would lead to significant difference of scores on the TCM questionnaire:

- Female vs. male
- Using on-going medicine for the OA joint vs. not-using of on-going medicine for the OA joint on the other scales (except the EXT scale)
- Involvement of more than one OA joint (hip or knee) vs. only one OA joint
- OA of the hip or knee as a primary health problem vs. OA of the hip or knee as not a primary health problem.

A one-way ANOVA model was conducted for each of these factors (each factor has two comparing groups described as above). The significant level was defined as $p < .05$.

2) Convergence and discrimination (correlations between EXT and INT scales)

The convergence and discrimination validity were examined by the correlation coefficients between scales on the TCM questionnaire. Adapting the method of Ware and Gandek (1998), the correlations among all scales were computed and compared with reliability estimates. The correlation of the scales between EXT and INT was expected to be less than their reliability coefficients.

3) Criterion validation (correlations with the SF-36 health survey)

The construct validity was also examined by the method of criterion validation (Hays, Anderson, & Revicki, 1998; Portney & Watkins, 2000). The validation involved the correlations between comparable scales on TCM questionnaire and SF-36 health survey. High or medium correlations between the following comparable scales were expected:

- The TCM EXT scale (questions about pain and other sensation in the OA joint) vs. the SF-36 bodily pain scale (questions about bodily pain and its interfere with normal work)
- The TCM INT scale (questions about the general body) vs. the SF-36 general health scale (questions about the general health and its perceptions)
- The TCM SUM scale (questions about pain and other sensation in the OA joint, and questions about the general body) vs. the SF-36 general health scale (questions about the general health and its perceptions).

The correlation coefficients in this construct validity study are referred to as the Pearson's product-moment correlation coefficients. The degree of the association was defined as the following (Ware, Kosinski, & Gandek, 2002):

- The high correlation coefficient (strong association) is .70 or above
- The medium correlation coefficient (moderate to substantial association) is between .30 and .70
- The low correlation coefficient (weak association) is .30 or below.

3.3.8.4 Tests of grouping (scaling) success

The 23-item TCM questionnaire consisted of two groups of questions including:

- Questions about OA joints (Question Item 1 to 8)

- Questions about the general body (Question Item 9 to 23).

These two groups of questions formed two main scales: EXT scale (Question Item 1 to 8) and INT scale (Question Item 9 to 23). The combination of these two groups of questions formed an additional scale: SUM scale (Question Item 1 to 23).

The success of these grouping (or scaling) for the TCM questionnaire was estimated by examining:

- Item internal consistency
- Item discriminant validity

- **Item internal consistency**

Item internal consistency was evaluated by the correlation between each item and its hypothesised scale (or group). If a scale is internally consistent, then each item needs to correlate with all other items within its hypothesised scale (Streiner & Norman, 1995). For example, the EXT scale contains 8 items (questions 1 to 8), thus, a high correlation between each item score of these 8 items and the EXT scale score would be required.

To estimate the item internal consistency, it is necessary to correct the overlap (Howard & Forehand, 1962). Correction for overlap applies when the item score is also in the scale score. For example, the correlation between the Item 1 score and the EXT scale score corrected for overlap means that the correlation was based on the Item 1 score and the sum score of item 2 to item 8 (without Item 1).

In almost all instances, the best correlation coefficient to use is the Pearson product-moment correlation (Nunnally, 1978). The item-scale correlations above .40 was

recommended as a high standard of internal consistency (Ware, Kosinski, & Gandek, 2002). A scaling success was counted whenever the correlation between an item and its hypothesised scale (corrected for overlap) equalled or exceeded .40. It is recommended that items below .20 correlations be discarded (Bowling, 2001a; Streiner & Norman, 1995).

- **Item discriminant validity**

Item discriminant validity means that the item measures one health concept more accurately than other health concepts. The success of item discriminant validity was determined by whether the correlation coefficient between the item and its hypothesised scale was significantly higher than the correlation between this item and another scale. The significant level was recommended at least two standard errors of the correlation coefficient (Ware, Kosinski, & Gandek, 2002). One standard error of a correlation coefficient is approximately equal to one divided by the square root of the sample size (Ware & Gandek, 1998).

For example, when the sample size is 10, the correlation coefficient between Item 1 score and the EXT scale score (corrected for overlap) must be at least .63 higher than the correlation coefficient between Item 1 score and the INT scale score. If this happens, the Item 1 discriminant validity would be successful. This highlights the difficult of applying this criterion to small samples, since it will be difficult for correlations to differ by more than .63.

CHAPTER 4 RESULTS

4.1 Data quality

4.1.1 Subjects

The sample of twelve participants completed the TCM and SF-36 questionnaires on the first occasion (Test 1). Eleven participants completed repeat tests of both TCM and SF-36 questionnaires on the second occasion (Test 2). One participant could not be contacted. Of the eleven participants, one (aged 83) left 18 items blank on a 23-item TCM questionnaire, and was therefore withdrawn from the study, giving the valid sample of ten who completed both TCM and SF-36 questionnaires on the first and second occasions. Of the ten participants, five male and five female, their age varied from 40 to 70 (Mean 63.1 ± Std. Deviation11.84).

4.1.2 Data completeness of the TCM questionnaire

One or more items were missing in five (25%) of the test and retest TCM questionnaires (two on the first test, three on the second test).

Missing value rates per item (i.e. question) ranged from 0% to 10% on the first test TCM questionnaires, but ranged from 0% to 30% on the second test (Table G.1 and G.2, see Appendix G.1). The highest rate of missing data was found in Item 7 (on Test 2).

The overall missing data was slightly higher on the second test than the first test (Table G.1 and G.2, see Appendix G.1).

4.1.3 Distribution of responses to items of the TCM questionnaire

In the test and retest, the distributions of responses to items are shown in the Table G.1 and G.2 (see Appendix G.1). These tables illustrate that not all of the responses categories were used for every item. The items within the EXT scale tended to have high frequency of response in categories 3 and 4. However, the items within the INT scale tended to have high frequency of response in categories 4 and 5.

Mean value per item ranged form 3.10 (Item 3 on the second test) to 4.60 (Item 12 on the second test). Ranges between minimum and maximum for per item value were from one to four (Tables G.3 and G.4, see Appendix G.2). The lowest range of one point was found in Item 7 on the second test, where response categories were between 4 and 5. Descriptive statistics for item and scale values of the TCM questionnaire are detailed in Appendix G.2.

4.2 Tests of reliability of the TCM questionnaire

4.2.1 Internal consistency reliability

Cronbach's alpha was above .70 for all scales on both occasions of the first test and the second test (ranged from .75 to .95) (Table 4.1). The INT scale had high alpha values of .90 on the first test and .93 on the second test. The SUM scale also had high alpha value above .90 on the second test, and slightly below .90 on the first test.

Table 4.1 Internal consistency reliabilities (Cronbach's alpha statistics) of the TCM questionnaire scales (n=10)

Scale	No. of Items	Test 1 Cronbach's Alpha	Test 2 Cronbach's Alpha
EXT	8	.75	.89
INT	15	.90	.93
SUM	23	.87	.95

Note. EXT = Exterior; INT = Interior; SUM = Summary.

Intra-class correlation coefficient (average measure value) for each scale had the same value as Cronbach's alpha (Table 4.2). Because the TCM questionnaire scale scores were based on all items within the scale rather than a single item, the average measure value of the intra-class correlation coefficient was used as scale's internal consistency reliability value.

Table 4.2 Internal consistency reliability (intra-class correlation coefficients) of TCM questionnaire scales (n=10)

Scale	No. of Items	Intra-class Correlation Coefficients[a]	95% Confidence Interval		Sig. (Two-tailed)
			Lower Bound	Upper Bound	
EXT	8	.75 (.89)	.42 (.74)	.93 (.97)	.000 (.000)
INT	15	.90 (.93)	.77 (.85)	.97 (.98)	.000 (.000)
SUM	23	.87 (.95)	.88 (.88)	.96 (.98)	.000 (.000)

Note. EXT = Exterior; INT = Interior; SUM = Summary

Result from the second test is presented in brackets

[a]Two-way random effects model where both people effects and measures effects are random (see methodology section for detailed explanation); Using a consistency definition where the between-measure variance is excluded from the denominator variance (see methodology section for detailed explanation); Because the average measure (i.e. all items within a scale) was the only interest for the use of the questionnaire, the average measurement value of ICC was used for the questionnaire scale's reliability value.

4.2.2 Test-retest reliability

Intra-class correlation coefficient for each scale was above .70 for all scales, except the EXT scale, which was .44 (Table 4.3).

Table 4.3 Test-retest reliability (intra-class correlation coefficients method) of TCM questionnaire scales (n=10)

Scale	Intra-class Correlation Coefficients[a]	95% Confidence Interval		Sig. (2-tailed)
		Lower Bound	Upper Bound	
EXT	.44	-.20	.82	.089
INT	.86	.53	.96	.001
SUM	.75	.27	.93	.005

Note. EXT = Exterior; INT = Interior; SUM = Summary

[a]Two-way random effects model where both people effects and measures effects in a different administration time are random (see methodology section for detailed explanation); Using an absolute agreement definition where the between-measure variance is excluded from the denominator variance (see methodology section for detailed explanation); Because single measure (i.e. single test) was the only interest for the use of the questionnaire, the single measurement value of ICC was used for the questionnaire scale's reliability value.

The two-way ANOVA of the TCM questionnaire scales between test and retest indicates that there was no significant mean differences (or no bias, $p > .05$) for scale scores between test and retest (Table 4.4). Therefore, the scale's intra-class correlation coefficient can be regarded as in agreement between test and retest (Portney & Watkins, 2000).

Table 4.4 The TCM questionnaire scale mean differences in the test and retest (n=10)

Scale	Test	Mean	Std. Deviation	F Statistics	Sig.
EXT	Test1	69.22	11.07		
	Test2	64.28	17.26	1.032	.336
INT	Test1	76.67	18.86		
	Test2	75.83	19.18	.062	.808
SUM	Test1	74.08	13.47		
	Test2	71.82	17.26	.402	.542

Note. EXT = Exterior; INT = Interior; SUM = Summary.

The plots (Figure G.1, G.2 and G.3, see Appendix G.5) of mean differences between test and retest against their mean (Bland & Altman, 1986) further displayed that:

- There was no relation between the difference and the size of each scale score
- There was considerable agreement between test and retest.

The "limits of agreement" (Bland & Altman, 1986) of each scale and 95% confidence interval for the scales score differences for an individual person ranged from -35.07 to 19.84 (Table 4.5). For example, the 95% confidence interval around the INT scale score for an individual patient would be about ± 21.10 points (or more precisely, within the limits of agreement between –21.51 and 19.84) on a 100 point scale with a reliability of .86 (Table 4.3 and Table 4.5).

Table 4.5 Limits of agreements and confidence interval for individual patient TCM questionnaire scale scores

Scale	Mean Differences (d)	Standard Deviation of Mean Differences (s)	95% of Difference Within 2 s (95% Confidence Interval)	95% of Differences Within the Limits of Agreement (d±1.96s)
EXT	-4.94	15.37	30.74	(-35.07; 25.19)
INT	-.83	10.55	21.10	(-21.51; 19.84)
SUM	-2.26	11.28	22.55	(-24.36; 19.84)

Note. EXT = Exterior; INT = Interior; SUM = Summary.

4.3 Tests of validity of the TCM questionnaire

4.3.1 Content validity

Table H.1 (see Appendix H) suggests that the TCM questionnaire scales were corresponding to the "eight principal syndromes" (ba gang) within TCM concepts and frameworks. Each scale was represented by one or more questions (content coverage), and there were no irrelevant items.

As presented in Table H.2 (see Appendix H) the patterns of OA in TCM were well captured by the TCM questionnaire scales. Items in TCM questionnaire scales corresponding to the Ten Questions and TCM inquiry are shown in Table H.3 (see Appendix H). The lowest and highest scores of each scale have been defined in Table 4.6.

Table 4.6 Content-based descriptions of the lowest and highest scale scores

	Definitions of Lowest and Highest Scores	
Scale	Lowest Possible (Floor)	Highest Possible (Ceiling)
EXT	Extreme pain and/or other sensations in OA joint due to invasion of exterior pathogenic factors such as wind, cold, heat, and damp.	No pain and/or other sensations in OA joint due to invasion of exterior pathogenic factors such as wind, cold, heat, and damp.
INT	Extreme pre-existing symptoms in general body as results of interior zang fu and/or qi blood, jin ye, shen disorders.	No pre-existing symptoms in general body as results of interior zang fu and/or qi blood, jin ye, shen disorders.
SUM	Extreme pain and/or other sensations in OA joint, and/or extreme pre-existing symptoms in the general body based on the "eight principle syndromes" within the TCM concepts.	No pain and/or other sensations in OA joint, and/or no pre-existing symptoms in the general body based on the "eight principle syndromes" within the TCM concepts.

Note. EXT = Exterior; INT = Interior; SUM = Summary.

4.3.2 Construct validity

4.3.2.1 Known groups method

- **Age groups**

The correlation values between the subjects' age and each of the scales score of INT and SUM scale were .34 to .56 with the exception of the SUM scale on the second test, which was .29 (Table 4.7). This result indicates moderate associations between the subjects' age and the scores of the INT scale, and between the subjects' age and the scores of the SUM scale.

Table 4.7 Pearson correlations between the subjects' age and the TCM questionnaire interior and summary scales (n=10)

	INT	SUM
Age	.56 (.34)	.52 (.29)

Note. INT = Interior; SUM = Summary

Result from the second test is presented in brackets.

- **Using and not-using on-going medication groups**

Table 4.8 shows that there was significant difference between the groups using and not using on-going medication in their EXT scale scores on the first test, F (1, 7) = 11.26, p = .012. However, the significant difference was not found on the second test.

Table 4.8 One-way ANOVA for the exterior scale by the factor of use of drugs

Group	No. of Subjects	Mean	Standard Deviation	F Statistics	Sig.
1	5	77.19 (75.25)	6.40 (2.91)		
2	4	64.84 (53.05)	3.93 (21.94)	11.262 (5.19)	.012 (.06)

Note. Group 1 = use of ongoing medicine for the OA joint; Group 2 = non-use of ongoing medicine for the OA joint

Result from the second test is presented in brackets.

- **Other groups**

As expected, significant differences were not found in the other factors. These factors included the gender groups (male or female), whether more than one OA joints was involved, whether on-going medicine was taken for the OA joint (except the EXT scale), and whether OA was a primary problem (Table G.9, see Appendix G.4).

4.3.2.2 Convergence and discrimination (correlations between scales)

The correlations of the scales between EXT and INT were less than their reliability coefficients in both test and retest (.16 vs. .75 & .90 in the first test; .70 vs. .89 & .93 in the second test) (Table 4.9). This result indicates that the EXT and INT scales measure different concepts (Ware & Gandek, 1998).

Table 4.9 Reliability coefficients and inter-scale correlations between independent scales of exterior and interior (n=10)

	EXT	INT
EXT	**.75 (.89)** [a]	
INT	.16 (.70)	**.90 (.93)** [a]

Note. EXT = Exterior, INT = Interior

Result from the second test is presented in brackets. Because the summary scale was generated from the combination of the exterior and interior scale, it was not tested for convergence and discrimination validity

[a]Scale internal consistency reliability (Cronbach's alpha) is presented in bold.

4.3.2.3 Criterion validation (correlations with the SF-36 health survey)

As expected, correlation coefficients, r-values, were between .30 and .70, or above .70 between the comparable scales of the TCM questionnaire and SF-36 health survey (Table 4.10). This result indicates there was a strong or moderate association between these comparable scales:

- The TCM EXT scale vs. the SF-36 bodily pain scale (r = .46 in the first test, and r = .38 in the second test)
- The TCM INT scale vs. the SF-36 general health scale (r = .80 in the first test, and r = .90 in the second test)
- The TCM SUM scale vs. the SF-36 general health scale (r = .79 in the first test, and r = .83 in the second test).

Table 4.10 Pearson correlations between the TCM questionnaire scales and the SF-36 health survey (n=10)

	EXT	INT	SUM
BPH	**.46 (.38)** [a]	.72 (.70)	.79 (.64)
GHH	.19 (.49)	**.80 (.90)** [a]	**.79 (.83)** [a]

Note. The TCM questionnaires scales: EXT = Exterior, INT = Interior, SUM = Summary; the SF-36 health survey scales: BPH = Bodily pain, GHH = general Health

Result from the second test is presented in brackets

[a]Correlations for criterion validation are in bold.

4.4 Tests of scaling (grouping) success of the TCM questionnaire

Tables G.7 and G.8 (see Appendix G.3) show the item to scale correlations of TCM questionnaire on the first and the second test. Correlations of item and its hypothesised scale corrected for overlap are accompanied by an ([a]).

A standard error of the correlation coefficient was approximately equal to one divided by the square root of the sample size (Ware & Gandek, 1998). Thus, where the sample size was 10, the two standard errors would be .63.

Because the SUM scale was generated from the main scales of EXT and INT, it was not tested for item discriminant validity.

Table 4.11 shows the scaling success rates for item internal consistency were above 73% in both tests, with the exception of the SUM scale, where success rates were 48% in the first test, and 91% in the second test. The scaling success rates for the item

discriminant validity of the EXT and INT scales were lower than 33% (Table 4.12). However, Tables G.7 and G.8 (see Appendix G.3) show that most correlations between items and hypothesised scales were higher (but not more than two standard errors) than the correlations between items and another scale. The result suggests that items where scaling success did not met the criteria (Ware & Gandek, 1998; Ware, Kosinski, & Gandek, 2002), needs to be revised or discarded in future studies.

Table 4.11 Summary results of tests of item internal consistency (n=10)

Scale	No. of Items	Range of Correlations[a]	Internal Consistency Tests	
			No.[b] of Success / Total	Success Rate (%)
EXT	8	.10; .84 (.12; .98)	6/8 (6/8)	75 (75)
INT	15	.15; .94 (.29; .96)	11/15 (14/15)	73 (93)
SUM	23	-.05; .86 (-.21; .94)	11/23 (21/23)	48 (91)

Note. EXT = Exterior; INT = Interior; SUM = Summary

Result from the second test is presented in brackets

[a]Range of correlations between items and hypothesised scales corrected for overlap

[b]Number of correlations between items and hypothesised scales corrected for overlap ≥ .40.

Table 4.12 Summary results of tests of item discriminant validity (n=10)

Scale	No. of Items	Range of Correlations[a]	Discriminant Validity Tests	
			No.[b] of Success / Total	Success Rate (%)
EXT	8	-.19; .35 (-.34; .84)	1/8 (0/8)	13 (0)
INT	15	-.5; .43 (.30; .73)	5/15 (0/15)	33 (0)

Note. EXT = Exterior; INT = Interior

Result from the second test is presented in brackets. Because the summary scale was generated from the main scales of exterior and interior, it was not tested for the item discriminant validity

[a]Range of correlations between items and other scales (the additional scales of deficiency and summary were not included)

[b]Number of significantly (> 2 SE) higher correlations between items and hypothesised scales.

4.5 Summary of findings

- Data completeness of the TCM questionnaire was above 70% (maximum missing value per item was 10% on the first test and 30% on the second test)
- Not all of the response categories were used for every item
- The items within the EXT scale tended to have a higher frequency of responses in the categories 3 and 4. However, the items within the INT scale tended to have a higher frequency of response in the categories 4 and 5
- Internal consistency reliability (Cronbach's alpha) was above .70 for all scales on both occasions of the first test and the second test. The INT scale has high alpha values of .90 on the first test and .93 on the second test
- Test-retest reliability (intra-class correlation coefficient) for each scale was above .70 for all scales, except the EXT scale, which was .44. There were no significant differences (or no bias) between the first and the second tests
- The TCM questionnaire scales were corresponding to the "eight principal syndromes" within the traditional Chinese medicine concepts and frameworks. Each scale was represented by one or more questions (content coverage), and there were no irrelevant items. The patterns of OA in TCM were well captured by the TCM questionnaire scales. The items in TCM questionnaire scales correspond to the "ten questions" and TCM inquiry
- Moderate associations were found between subjects' age and each of the scores of INT and SUM scales (the correlations coefficients were between .30 and .70, except the SUM scale on the second test that was .29)
- There was significant difference between the groups using and not-using ongoing medication in the EXT scale scores on the first test, $F(1, 7) =$

11.26, p = .012. However, a significant difference was not found on the second test

- Significant differences were not found in the other factors evaluated. These factors included the gender groups (male or female), whether there was more than one OA joint, whether there was use of on-going medicine for the OA joint (except the EXT scale), and whether OA was a primary health problem
- The correlations of the scales between EXT and INT were less than their reliability coefficients in both test and retest.
- As expected, high or medium correlations were found (correlation coefficient was between .30 to .70, or above .70) between the comparable scales: the TCM EXT scale vs. the SF-36 bodily pain scale; the TCM INT scale vs. the SF-36 general health scale; the TCM SUM scale vs. the SF-36 general health scale
- The scaling (grouping) success rates for the items' internal consistency were above 73% in the both tests, with the exception of the SUM scale, where success rates were 48% in the first test and 91% in the second test
- The scaling success rate for the item discriminant validity of the EXT and INT scales was lower than 33%. However, most correlations between items and the hypothesised scales were higher (but not more than two standard errors) than the correlations between items and another scale.

CHAPTER 5 DISCUSSION

5.1 Introduction

Results from this pilot study are interpreted and discussed in caution due to the major limitation of the small sample of subject. This discussion starts with statements to achieve or fail the specific aims, and is followed by an analysis of the data quality of the TCM questionnaire. The estimation of the TCM questionnaire's reliability and validity is a main part of this study, therefore it is emphasised in discussion. The scaling success of item internal consistency and item discriminant validity is an important part of estimating the questionnaire's construct validity, and as it provides important information for individual question development, it has been discussed separately.

Limitations of this study are also identified and explained. The potential benefits of the "success" of a TCM questionnaire development, and recommendations for future studies are listed. Conclusions of the study are expressed at the end of this chapter.

5.2 Statements to achieve or fail the specific aims

Of the four specific aims set in the chapter of review of literature, one was achieved:

- A questionnaire was developed within TCM concepts and frameworks: The draft of the questionnaire was accepted by patients with OA of the hip or knee and TCM acupuncture experts.

The other three specific aims were partly achieved:

- The questionnaire has acceptable reliability: The internal consistency reliability or test-retest reliability was above .70, except the EXT scale, where test-retest reliability was 0.44
- The questionnaire has acceptable validity: The content validity was representative of or corresponds to the "eight principal syndromes" and "ten questions" in TCM concepts and frameworks. The construct validity was supported by: 1) moderate associations between subject's age and each of the scores of INT and SUM scales, except the SUM scale on the second test (.29), 2) there was significant difference between the groups using and not-using ongoing medication in the EXT scale scores in the first test, but not in the retest, 3) the correlations of the scales between EXT and INT were less than their reliability coefficients, 4) high or medium correlations between the comparable TCM questionnaire scale and SF-36 health survey scale
- The questionnaire has success rates of above 73% for item (i.e. question) internal consistency, but low success rates of below 33% for item discriminant validity.

5.3 Data quality

5.3.1 Sampling frame

This pilot study involved convenience sampling of patients with OA of hip and knee. The patients were selected from six health care settings, in order to include subjects from both genders, a wide range of age, and a wide range of OA conditions. However, due to patients requiring definite radiographic evidence of primary OA in the hip and/or knee and being proficient in English, the sample size in this study was small, with only ten useable sets of data.

5.3.2 Data completeness

The overall percentage of missing data was higher on the second test (6.09% vs. .87%) than the first test. This may be related to small sample size or patients being bored with the same questionnaire on the second test.

The highest rate of missing data was found in the Item 7 (10% in the first test, 30% in the second test), suggesting that this item ("cold sensation in the OA joint") may be somewhat difficult to interpret for patients. In contrast to the question of "hot sensation in an OA joint" (usually indicating inflammatory conditions), asking about a cold sensation is not common in western medical assessment. Thus, these patients may ignore the question about a cold sensation in the joint. However, cold sensation in the OA joint is very meaningful in TCM. For example, a cold feeling in an OA knee joint provides evidence of kidney yang deficiency. This concept may be more specific for

TCM and therefore result in a low response rate in comparison with the other items for patients used to treatment in WM.

5.3.3 Distribution of responses to items

The tables of distributions of responses to items show that not all of the response categories were used for every item (question). The TCM questionnaire was designed to assess a wide range of health conditions for patients with OA of the hip or knee. With a sample size of only ten in this study, it is not surprising to find that not all response categories were used for every item.

The lowest range of one point between minimum and maximum value was found in Item 7 indicating that the responses to "cold sensation in the OA joint" were only between "mild" and "none" categories. A possible reason for this may be due to people ignoring, denying, or not understanding what the cold sensation in their joints was meant to mean.

5.4 Reliability of the TCM questionnaire

As noted in the literature review chapter, there are four basic categories of reliability estimation: internal consistency, test-retest, inter-rater, and equivalent forms (or alternate forms) reliability (Hays, Anderson, & Revicki, 1998; Portney & Watkins, 2000). Because equivalent forms reliability refers to the agreement between an individual's score on two or more measures designed to measure the same attribute (Hays, Anderson, & Revicki, 1998), it is not appropriate to use the SF-36 health survey

as an equivalent form to estimate the TCM questionnaire reliability. The questionnaire was designed for self-administration, therefore inter-rater reliability was not required in this study. The results of internal consistency and test-retest reliability estimation will now be discussed.

5.4.1 Internal consistency reliability

The Cronbach's alpha values were above the accepted cut-off point of .70 (Cronbach, 1951; Lu, 1996; Nunnally, 1978; Streiner, 1993; Streiner & Norman, 1995) for the TCM questionnaire scales. This indicates that these scales are acceptable for comparisons among groups. The INT scale had alpha values above the standard of .90 (Lu, 1996; Nunnally, 1978; Streiner & Norman, 1995) and therefore can be used for comparisons among individuals. However, there was no strong conclusion for these scales internal consistency reliability due to the small sample in this study.

5.4.2 Test-retest reliability

The intra-class correlation coefficient for each scale was satisfactory for all scales, with the exception of the EXT scale, which was below the accepted cut-off point of .70 (Cronbach, 1951; Lu, 1996; Nunnally, 1978; Streiner, 1993; Streiner & Norman, 1995). Therefore, the test-retest reliability would be acceptable for all scales with exception of the EXT scale, which requires further development.

A two-week interval was applied in this test-retest reliability study. The initial purpose of the two-week interval was to minimise memory effects, and was sufficiently short to assume that the underlying OA process was unlikely to have changed between the two

assessments. The mean difference in the EXT scale score of 4.94 (mean 69.22 in Test 1; 64.28 in Test 2) indicates that there might be some real changes in the OA joint over the two-week interval. For example, pain or some other sensations in the OA joint might change during the two weeks. Therefore, the low test-retest reliability for the EXT scale may suggest that the scale is sensitive to change. Consequently, a shorter interval between test and retest, such as one-week interval with no treatment in between (with exception of continuation of using on-going medications for the OA joint), might be more appropriate in further studies developing a TCM questionnaire.

The 95% confidence interval for the three TCM questionnaire scales test-retest score differences for an individual OA patient ranged from ± 21.10 to ± 30.74. In comparison with the data adapted from the section 7:14 in the SF-36 Health Survey Manual & Interpretation Guide, which ranged from ± 12.3 to ± 28.0 (Ware, Kosinski, & Gandek, 2002), the limits of agreement for the TCM questionnaire scores are similar to those of SF-36 health survey, so they can be regarded as satisfactory.

5.4.3 Comparisons of internal consistency reliability and test-retest reliability

The test-retest reliability coefficients tended to be slightly lower than estimates based on the internal consistency method. The internal consistency estimates for the three scales (the EXT, INT, and SUM scales) ranged from .75 to .95 with a median of .85, compared with a range of .44 to .86 and a median of .65 for the test-retest estimates in this study. A similar trend was found in another study validating the SF-36 health survey questionnaire (Brazier et al., 1992). One of the possible reasons for this trend was that the scales were sensitive to relatively short-term changes such as those that would be

observed in a two-week interval between administrations (Ware, Kosinski, & Gandek, 2002).

5.5 Validity of the TCM questionnaire

As discussed in the literature review chapter, there are four aspects of validity: face validity, content validity, criterion validity, and construct validity (Jenkinson & McGee, 1998; Portney & Watkins, 2000).

Some authors suggest that responsiveness to change is also an aspect of validity (Hays, Anderson, & Revicki, 1998; Hays & Hadoron, 1992; Portney & Watkins, 2000). Because no changes to the treatment were expected, the responsiveness of questionnaire was not tested. The retest was to estimate the reliability of the TCM questionnaire, not looking for changes in the OA condition.

The validation of criterion validity is based on the correlations with a "gold standard", or criterion measurement that is already established or validated (Bowling, 2001b; Portney & Watkins, 2000). However, as noted in the literature review chapter, there are no measurement instruments established or validated in TCM concepts and frameworks. It is therefore not possible to estimate the questionnaire's criterion validity at this stage. The SF-36 health survey, used in this study, was for the purpose of supporting the TCM questionnaires construct validity.

The following three aspects of validity of the questionnaire will now be discussed in detail: face validity, content validity, and construct validity.

5.5.1 Face validity

Face validity was examined in pre-testing study. The result of pre-testing indicates that the questionnaire was appropriate in the TCM manner and was easily understood. However, assessments of face validity are usually considered subjective and scientifically weak because there are no standards for judging (Portney & Watkins, 2000). Therefore, face validity cannot be considered sufficient on its own, but it can be used as a part of validation purpose.

5.5.2 Content validity

Content validity refers to a representative range of the content under study. Testing content validity is a challenge in TCM because of the wide range of TCM concepts and the lack of defining precise concepts or standards.

In TCM, OA may be regarded as having several different patterns. Each pattern reflects different Chinese medicine concepts, such as, yin yang, zang fu, qi xue, jing, shen, body fluid, and their deficiency, excessive, exterior, and interior, etc (see Chapter 2.3 "Osteoarthritis in TCM" and Appendix A for detailed explanations). Despite these concepts having been effectively adapted in clinical practice over thousands of years, there is a lack of clearly defined and precise standards. For example, qi in Chinese medicine, is a commonly used concept, but defies the western concept as a standard term. Qi may refer to all relationships and movements within the universe, or it may more simply refer to "air" and "breath" (Campbell, 2002; Cheng, 1997; Kaptchuk, 2000; Maciocia, 1996). More over, in clinical practice most patients present with a complex

mixture of signs and symptoms. Thus, it is no surprise that different TCM experts seeing the same patient might have different diagnoses for the same patient.

Despite these disadvantages in TCM, TCM has practically proven its validity and efficacy in dealing with the health of human beings over thousands of years (Cheng, 1997; Ying et al., 1995). Conversely, the modes of thinking, strategic principles, and underlying philosophical concepts that the TCM practitioner must use to construct this highly adaptive, highly effective system of medical intervention and health care (Zhu & Rose, 2002) may be potentially useful for western society.

In western modern society, in order to obtain respect, evidence-base research and practice insists the effectiveness of TCM be consistently measured. Hence, it is worthwhile to develop valid instruments within the TCM frameworks and concepts. The TCM questionnaire was developed with these considerations.

Due to the lack of validated TCM questionnaires and the lack of defined standards, the development of this TCM questionnaire was constructed using textbooks, experts' opinions, "general" accepted concepts, and personal clinical experiences.

The "eight principal syndromes" (i.e. ba gang) is the most common TCM diagnostic method used (Campbell, 2002; Cheng, 1997; Kaptchuk, 2000). The "eight principal syndromes" refer to four pairs of syndromes: yin and yang; cold and heat; exterior and interior; and deficiency and excess.

The principal syndromes of yin and yang generally classify the essential nature of the disorder and also summarise the other three pairs (Campbell, 2002; Deng, 1999).

Exterior, heat and excess are summarised as yang syndromes, while interior, cold and deficiency are summarised as yin syndromes. However, the eight principal syndromes or patterns always occur in combinations. For example, the combination of deficiency and heat is common in elderly patients with OA of kidney yin deficiency. Thus, it might be a problem to classify the patient's pattern into a strict yin or yang syndromes. For example, if a patient presented with a kidney yang deficiency, this patient's syndrome would be yang due to yang disorder. Obviously, this is not a correct statement, because yang deficiency indicates a deficiency of cold and cold is generally classified as a yin syndromes (Campbell, 2002; Deng, 1999). Many textbooks, e.g. Deng (1999), describe the yin pattern, yang pattern, yin deficiency and yang deficiency separately. However, since a yin pattern overlaps yang deficiency, it might not be practical to separate these two. This can be confusing and difficult for a western medicine type evaluation. In Chinese theory patterns can evolve and change into other patterns. So there is no difficulty with the concepts. Measuring the evolution of these changes would be challenging.

In clinical practice, asking a patient with OA questions often includes two parts: the local OA joint and the general body condition. The pre-existing condition of general body deficiency (such as kidney and liver deficiency, or qi and blood deficiency) is commonly regarded as an important background factor contributing to the development of OA in TCM (see Chapter 2.3 for more detail). Therefore, the OA joint condition is regarded as the "end", "result", or "secondly" (biao) in TCM, while the general body condition is the "root", "causation", or "firstly" (ben).

The OA joint condition and the general body condition may also be regarded as the exterior syndromes and interior syndromes within the concepts of the eight principal

syndromes, which refer to the “superficial” aspect of the body and the “interior” aspect of the body (Campbell, 2002). The “exterior” includes skin, hair, flesh, and channels, etc, while the “interior” includes zang fu, qi blood, and marrow, etc (Campbell, 2002; Cheng, 1997; Deng, 1999).

The joint condition and the general body condition can be clearly defined in the TCM questionnaire as the exterior (EXT) scale and interior (INT) scale, rather than the other three pairs of the eight principal syndromes. The other three pairs can be regarded as subscales under the EXT and INT scales according to the combinations of the “eight principal syndromes”.

Therefore, the EXT and INT scales were chosen as two main scales in the TCM questionnaire. The SUM scale (i.e. the combination of the EXT and INT scales) was developed as an additional scale, because it was useful to obtain a single score for summarising the patient’s condition.

5.5.3 Construct validity

Because the measurement scales of EXT and INT are abstract variables, and the scales are not so obviously related to any actual variable, it makes the validity of these scales hard to verify. The validity of the TCM questionnaire could therefore, only address the “degree of confidence” (Streiner & Norman, 1995). The evidence to support or refute the theoretical framework behind the construct was gathered by methods including the known groups method, convergence and discrimination, and criterion validation (Portney & Watkins, 2000).

5.5.3.1 Known groups method

The theoretical construct of the SUM scale of the TCM questionnaire was defined as (Table 4.6): pain and/or other sensations in OA joint, and/or pre-existing symptoms in the general body based on the "eight principal syndromes" within the TCM concepts. Because age is an important factor that is generally associated with general body and local joint conditions (imbalances of yin yang qi blood in TCM), the older age may indicate more imbalances and thus a potentially lower score on the TCM questionnaire SUM scale.

Similarly, the concept of interior in TCM was defined as (Table 4.6): pre-existing symptoms in the general body as a result of interior zang fu and/or qi blood, jin ye, shen disorders. Because the general body conditions are generally associated with aging, older age may often relate to poorer general body health, and thus a lower score on the TCM questionnaire INT scale.

The results show that there are moderate associations between age and the scores of the INT and SUM scales. Thus, it gives supporting evidence to the construct validity of the TCM questionnaire.

The theoretical construct of the EXT scale of the TCM questionnaire was defined in Table 4.6: pain and/or other sensations in OA joint due to invasion of exterior pathogenic factors such as wind, cold, heat, and damp. Because pain in an OA joint is a major consideration for the use of on-going pain relieving medication, the use or non-use of on-going medication may be an indicator for the TCM questionnaire EXT scale.

However, a significant difference in the factor of use of medications was found only in the first test, not the second test. This might be due to the EXT scale measuring pain and other sensations rather than pain itself.

As expected, significant differences were not found in the other factors. These factors included the gender groups (male or female), whether there was involvement of more than one OA joints, whether on-going medicine was used for the OA joint (except the EXT scale), and whether OA was a primary problem. These factors were not expected to have relationships with conditions of either the OA joint (except use of medication) or the whole body. Because the EXT scale was designed to assess the OA joint condition, and the INT scale was designed to assess the whole body condition, scores on the EXT or INT scale should not be affected by factors of gender, number of OA joints, and whether OA is the primary problem. However, due to the lack of TCM literature address particularly on the relationship between these factors and OA of the hip and knee, the result cannot be a strong support to the questionnaire's construct validity.

5.5.3.2 Convergence and discrimination

Convergent validity means that any of two scales believed to reflect the same underlying concept or construct would correlate highly. By contrast, discriminant validity means that low correlations are expected from two scales believed to reflect the different underlying concepts or constructs (Portney & Watkins, 2000). For example, the EXT scale and the INT scale were designed to assess different concepts of the local joint condition (e.g. pain and other sensations in the OA joint) and the general body condition (e.g. the other symptoms of the internal body), therefore a low correlation was expected between these two scales.

The correlations between scales can provide some evidence to support or refute the convergent and discriminant validity (Ware & Gandek, 1998). A reliability coefficient can be thought of as a correlation between a scale and itself. The correlations of the scales between EXT and INT were less than their reliability coefficients. There was evidence of unique reliable variance measured by each scale (Guilford, 1954). Thus, the EXT and INT scales measure distinct concepts.

The item internal consistency and discriminant validity could be regarded as part of convergent and discriminant validity and will be discussed at the end of this chapter.

5.5.3.3 Criterion validations

Construct validity could also be supported by the correlations of those relevant criterion validations (Hays, Anderson, & Revicki, 1998; Portney & Watkins, 2000; Ware, Kosinski, & Gandek, 2002). Although there was no other "standardised" instrument available, it was possible using the comparable scales of SF-36 health survey to test parts of the TCM questionnaire.

There were strong or moderate associations between the comparable scales of TCM questionnaire and the SF-36 health survey questionnaire. This provides supportive evidence of the construct validity of the TCM questionnaire.

Because the TCM questionnaire and the SF-36 health survey questionnaire were constructed within different concepts (TCM vs. WM), both questionnaires could not be used for testing the alternate form's reliability. The correlations between these two

questionnaires were applied to test the relationship only. Although both forms were administered at the same time, it was assumed that there was no carry-over effect (such as memorising some questions from a questionnaire may affect testing of the another questionnaire) due to different concepts of these questions.

5.6 Tests of scaling success

The scaling success of item internal consistency and item discriminant validity can serve a part of the questionnaire's validation (Ware & Gandek, 1998; Ware, Kosinski, & Gandek, 2002), and can provide important information for individual question development. The items where item-scale correlations have not been satisfactory for item internal consistency or item discriminant validity need to be revised or discarded in future studies.

5.6.1 Item internal consistency

The results show that the scaling success rates in item internal consistency validity were above 73%, with the exception of the SUM scale, where success rates were 48% in the first test, 91% in the second test. The reasons for the large different success rates for the SUM scale in the results between the two tests may be due to the small sample size and some changes in responses to the individual items between the test and retest. The large difference in success rates for the SUM scale may also suggest that using a single rating to summarize and interpret the TCM questionnaire needs further development.

The success rates of above 70% in the item internal consistency validity suggest that most items within its scale are generally tapping into the same health concept (such as exterior, interior condition). Items with item-scale correlations (corrected for overlap) below .40 require further revision. However, there were inconsistent results in the test and retest due to the small sample. It might be wise to reconsider all items whose item-scale correlations were below .40 in both tests, as well as including a larger number of respondents.

5.6.2 Item discriminant validity

The item discriminant validity was only tested in the scales of EXT and INT. The item discriminant validity was not tested for the SUM scale due to this scale was generated from the EXT and/or INT scale.

Although the results show that the most correlations between items and hypothesised scales were higher (but not more than two standard errors of the correlation coefficient) than the correlations between items and other scales, the scaling success rates for the item discriminant validity of the EXT and INT scales were lower than 33%. The small sample size indicates that a large standard error would occur (Ware & Gandek, 1998), and therefore, it was hard to achieve the success rate of the item discriminant validity.

5.7 Limitations of the pilot study

In interpreting the results of this pilot study, the following limitations must be considered: repeated measures design, small sample of subjects, small range of OA

conditions, the two-week interval, ordinal scaling response format, five-category response format, and lack of defining Chinese standards for validation.

- **Repeated measures study design**

A repeated measures experimental design (one-way) was applied in this pilot study. This design has the ability to control the potential influence of individual differences, but also has disadvantages of potential learning or carryover effects (Portney & Watkins, 2000). The two-week interval used in this repeated measures design might reduce the learning or carryover effects, but could not guarantee the patient would remain in a stable health condition over the two-week period.

- **Small sample of subjects**

Sample of subject was relatively small, due to the following reasons:

  - Subjects who have no X-ray to confirm an osteoarthritic knee or hip joint were excluded from this research.
  - Subjects who were not proficient in English were excluded from this research (translation may have influenced the reliability of the questionnaire).
  - The study was conducted over a relatively short period of time (an eight-month period for data collection, from the ethics approval date of 15th of July 2003 to when the licence for permission to use SF-36 Health Survey expired, 13th of March 2004).

- **Small range of OA conditions**

Although subjects were recruited from different clinical settings including three medical centres, two physiotherapy clinics, and one acupuncture clinic, given the small number of subjects, it is not surprising that patients with a limited range of OA conditions responded to the TCM questionnaire.

This small range of OA conditions might affect the estimates of the TCM questionnaire reliability. Reliability is based on the proportion of the total observed variance that is attributable to error (Portney & Watkins, 2000), and therefore, if the total variance is small, the reliability coefficient will probably be low. Hence results from this study can only be applied to similar patients in the future.

- **The two-week interval**

The TCM questionnaire was sent to the respondent after a two-week interval. The initial purpose of this two-week interval was to minimise any memory effects, but sufficiently short to assume that the underlying OA process was unlikely to have changed between the two occasions of assessments. Given the higher internal consistency reliability and considering the fluctuation of subjective symptoms of OA, the trend of slightly lower test-retest reliability indicates that there might be some real changes in the health condition of those patients with OA during the two-week period.

- **Ordinal scaling response format**

The ordinal scaling response format (e.g., none, mild, moderate, severe, extreme) was chosen in the development of a TCM questionnaire. The ordinal scaling responses are not preferred for statistical analysis such as means, standard deviations, etc. (Jenkinson

& McGee, 1998). Although ideally measures should be developed using interval or ratio scales (Jenkinson & McGee, 1998), measures of health status can never truly be regarded as fulfilling the requirement of these forms of measurement (Bowling, 2001b). In reality, we can never be certain whether the values in the measures are truly accurate.

- **A five-category response format**

A five-category response format, from "none to extreme", was chosen in the development of the TCM questionnaire. The purpose of using this format was to reduce the loss of information and still be within the patient's ability to discriminate. However, it might also limit reporting the wider range and important characteristics of OA such as the nature, duration and frequency of the symptoms.

- **Lack of defining standards for validation**

There is lack of defining standards by which to validate a TCM questionnaire. Therefore, the textbooks, experts' opinions, "general" accepted concepts, and personal clinical experiences were used to construct this questionnaire. The construct validity was verified by the commonly used procedures (Portney & Watkins, 2000) including the known groups method, convergence and discrimination, and criterion validation. However, the construct validation is never fully realised, as these procedures only provide some degrees of evidence to support or refute the construct within the questionnaire.

5.8 Clinical significance

OA is a leading cause of chronic disability in older ages, because of knee and hip involvement (Felson, 1988; Guccione, 1997; Sharma, 2001). Due to the limitations of

current western medical therapies, interest has arisen in acupuncture for pain relief (Yamauchi, 1976). However, to validate the effectiveness of TCM acupuncture, research must be conducted within the TCM frameworks.

The potential benefits of the "success" of a TCM questionnaire development may also include:

- Provide researchers with appropriate new tools to accurately measure the effectiveness of TCM therapy in patients with OA of the hip or knee
- Provide practitioners with appropriate new tools to identify whether the TCM treatments are the most beneficial for their patients with OA of the hip or knee
- Provide patients with OA of the hip or knee with meaningful and "objective" information for their best treatment decision-making
- Provide third party payers and policy makers with "objective" evidence for the most effective TCM treatment delivered at the least possible cost in patients with OA of the hip or knee
- The use of conventional or western medical measurements can no longer be defended solely on the basis that they are more "objective" than TCM measurements, particularly when they are often conceptually less relevant.

5.9 Recommendations for future studies

Further research is required to further develop and validate the TCM questionnaire. The following recommendations and considerations need to be addressed in future studies: larger sample population, wider range of OA conditions, revise some questions,

consider additional questions, consider a one-week interval between assessments, factor analysis, responsiveness to changes, rash analysis, generalisability, and ongoing pursuit of validation.

- **Larger sample population and wider range of OA conditions**

Obviously, the next stage in the TCM questionnaire development and validation require a larger sample population and the wider range of OA conditions. The sample size for a reliability study can be estimated on the size of the reliability coefficient and the width of the confidence interval (Streiner & Norman, 1995). For example, if it is expected that the reliability coefficient will be .70, and the width of the confidence interval is ± .10 (with $\alpha = .05$), then the sample size will need to be 130 participants (Streiner & Norman, 1995).

- **Revise questions**

Those questions (items), which had the highest percentage of missing data, and/or which did not meet the success scaling on the either first test or second test, could be revised. These questions include:

- Question 3: “Pain in OA joint during the day”
- Question 5: “Stiffness in OA joint”
- Question 6: “Weakness in OA joint”
- Question 7: “Feeling cold in OA joint”
- Question 15: “Hard bowel motion”
- Question 19: “Ringing in the ears”.

- **Consider one-week interval**

Due to the limitation of the two-week interval in the evaluation of test-retest reliability, future studies may consider a one-week interval to minimise the real changes in OA conditions and to avoid learning and memory effects. The literature search failed to identify any studies that specifically address the appropriate time interval for evaluation of test-retest reliability.

- **Factor analysis**

Factor analysis is a common statistical technique for construct validation (Portney & Watkins, 2000). Factor analysis examines a correlation matrix and attempts to identify groups of questions such that there are strong correlations amongst all the questions within a group, but weak correlations between questions within the group and those outside the group (Fayers & Machin, 1998). Therefore, factor analysis could identify how many factors (groups of questions) lie in the TCM questionnaire. It is expected that two factors, such as “exterior” and “interior”, or “yin” and “yang” might be identified in the TCM questionnaire.

- **Responsiveness to changes**

Further studies are also required on the TCM questionnaire’s responsiveness to change. Responsiveness to change over time is an important aspect of validity (Hays & Hadoron, 1992). The results reported here indicate that the TCM questionnaire might be sensitive to variations in OA conditions, and would therefore be responsive to changes in patients with OA. More studies are needed to determine the responsiveness by applying the TCM questionnaire with treatments of known benefits.

- **Rash analysis**

To overcome the limitation of using ordinal scaling response format in the TCM questionnaire, the considerations of specialised statistical applications may be taken in further validation studies. For example, Rash analysis may be a way of transforming an ordinal scale into equal intervals (Andrich, 1988; Rash, 1980; Wright & Masters, 1982).

- **Generalisability**

In addition to the estimates of reliability, future studies of the TCM questionnaire require documentation of its generalisability. The concept of generalisability theory is a comprehensive estimate of reliability and was proposed by Cronbach and his associates (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). In the generalisability theory, factors contributing to measurement error are considered as separate components of variance, and distinguishable from random error (Portney & Watkins, 2000). Therefore, the generalisability theory makes it possible to statistically identify the most likely sources of error in the TCM questionnaire. By identifying these sources of error, errors can be further reduced and improvements made in the questionnaire measurement (Streiner & Norman, 1995).

- **Ongoing pursuit of validation**

Development of a TCM measurement is a consuming task, but a worthwhile effort. Because the TCM questionnaire was designed to measure health status within the TCM concepts and frameworks, the validation of the questionnaire needs to be emphasised. However there is no existing "standard" to validate a TCM questionnaire, as the validation is an ongoing process. In particular, future research needs to further identify

what types of questions are likely to best evaluate the construct of TCM concepts such as “exterior” and “interior”.

5.10 Conclusions

This is unique research. A literature review within the field of OA and acupuncture was unable to find any TCM measurement or assessment instruments currently in use, despite TCM and acupuncture having been practised for over 4000 years (Cheng, 1997). Currently, TCM and acupuncture is being increasingly integrated into the western management of patients. It is reported that on any one day, at least 5000 patients receive acupuncture in New Zealand (Campbell, 2004). Development and validation of a TCM measurement or assessment instrument is highly demanded, in particular for most common diseases, such as OA of the hip or knee.

The TCM questionnaire was designed as a new health status instrument for TCM assessment in patients with OA of the hip or knee. The questionnaire was developed and validated within TCM concepts and frameworks. The questionnaire contains 23 items with two main scales (the EXT scale and INT scale) and one additional scale (the SUM scale). It takes approximately five minutes to complete and is entirely self-administered.

Results from this pilot study indicate that this TCM questionnaire might have acceptable qualities of reliability and validity. This TCM questionnaire measures health status in a TCM manner that could be acceptable to patients. Given the small sample size, and responsiveness to changes not being documented, there is no strong conclusion of its qualities of reliability and validity. However, the results in this pilot study indicate that this TCM questionnaire has potential usage as an outcome measurement tool for the

assessment of TCM in patients with OA of the hip or knee. For this application to be possible, the questionnaire needs further development and validation with a larger sample of patients who have a variety of OA conditions.

APPENDIX A FUNDAMENTAL CONCEPTS AND FRAMEWORKS IN TCM

Introduction

The traditional concepts introduced in Section 2.3.1, in Chapter two, are defined here in Appendix A as they relate to the context of TCM. These concepts are historical, philosophical, theoretical and contextual. They enable a traditional acupuncturist to make linkages between philosophy, medicine theory, observation and clinical practice.

The fundamental concepts and frameworks include: yin and yang, wu xing, zang fu organs, fundamental substance, channel systems, "four examinations", "pattern identification", "eight principal syndromes", needling, moxibustion, tuina, and cupping.

- **Yin and yang**

Yin and yang was originally included in ancient Chinese philosophy. The most ancient expression of this concept seems to have been whether a place faces the sun or not (Kaptchuk, 2000; Maciocia, 1996; Ying et al., 1995). The place being exposed to the sun is yang, whereas the place not having sun exposure is yin. The meaning of yin and yang was then extended to represent the two opposite but interrelated aspects.

According to the theory of yin yang, yang is associated with qualities such as active, bright, external, functional, upward, hot, and hyperactive; yin is associated with qualities such as passive, dark, internal, substantial, downward, cold, and hypoactive, belong to yin (Ying et al., 1995). The concept of yin yang can be used to describe two closely related aspects that form a unit (Table A.1).

Table A.1 Some examples of yin and yang

Yin	Yang	Yin	Yang
Dark	Light	Zang organ	Fu organ
Cold	Hot	Blood	Qi
Night	Day	Substance	Function
Stillness	Movement	Lower part of body	Upper part of body
Interior	Exterior	Dull pain	Sharp pain
Winter	Summer	Weakness	Strength
Deficiency	Excess	Slow	Rapid
Falling	Rising	Heavy	Light
Hard	Soft	Dampness	Dryness

(Beinfield & Korngold, 1992; Campbell, 2002; Maciocia, 1996)

The relationship between yin and yang can be described briefly as follows: opposition, inter-dependence, mutual consumption and support, inter-transforming, and infinite divisibility (Cheng, 1997). Yin and yang are expressed as a symbol (Figure A.1), which shows the interrelationship between the two components.

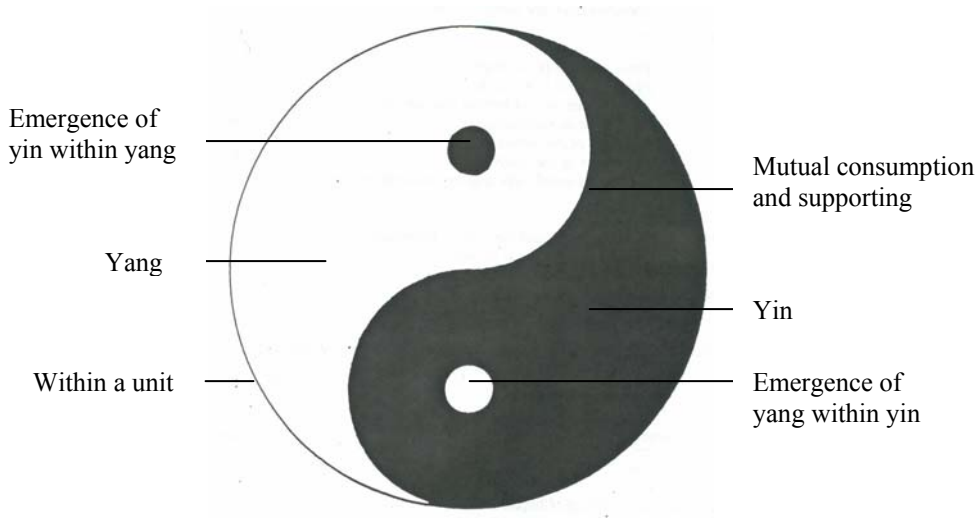

Figure A.1 Yin yang symbol (tai qi chuan)
(Campbell, 2002)

- **Wu xing**

Wu xing has different translations in English. The common translations, such as five movements, five phases, and five elements, may not be completely accurate. Wu xing in Chinese traditionally refers to the five categories of which wood, fire, earth, metal, and water (Table A.2) interact with each other to maintain a state of constant motion and change (Campbell, 2002; Cheng, 1997; Ying et al., 1995).

Table A.2 Classification and correspondences according to the wu xing

	Wood	Fire	Earth	Metal	Water
Season	Spring	Summer	Late Summer	Autumn	Winter
Colour	Green	Red	Yellow	White	Blue
Climatic Qi	Wind	Heat	Damp	Dryness	Cold
Yin Organ	Liver	Heart	Spleen	Lungs	Kidneys
Yang Organ	Gallbladder	Small Intestine	Stomach	Large Intestine	Bladder
Sense Organ	Eyes	Tongue	Mouth	Nose	Ears
Manifestation	Nails	Complexion	Lips	Skin	Hair
Emotion	Anger	Joy	Worry	Sadness	Fear
Human Sound	Shout	Laugh	Sing	Weep	Groan
Taste	Sour	Bitter	Sweet	Spicy	Salty
Body Tissue	Tendons	Blood Vessels	Muscles	Skin	Bone

(Campbell, 2002; Cheng, 1997; Kaptchuk, 2000; Ying et al., 1995)

The harmony of interaction within wu xing is based on the promoting (sheng cycle) and controlling relationships (ke cycle), which are showed in Figure A.2.

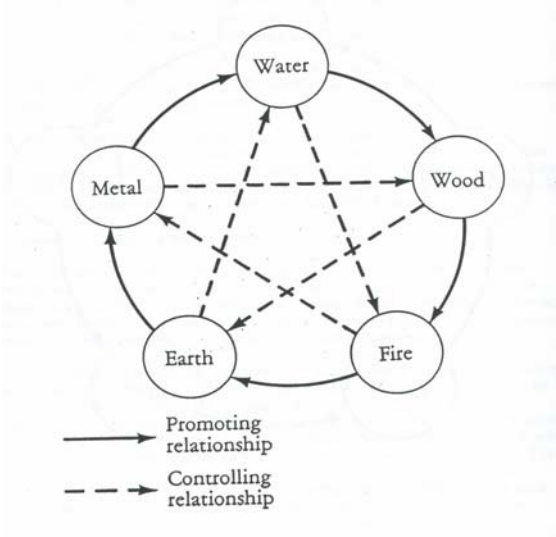

Figure A.2 The promoting and controlling relationships in wu xing (Campbell, 2002)

In TCM, wu xing is applied to explain and analyse the relationships within the internal organs, and the relationships between human beings and the natural world. In the context of the wu xing, illness is considered a process, not a static entity. For example, a patient with OA of the hip or knee may also present with loss of hair, reduced hearing and frequent urination. This patient has a Chinese "kidney imbalance". If the patient also complains of reddish eyes, anger, and difficulty falling asleep, the patient may also have "liver and heart problems", as it is said, "kidney water not generating liver wood" and "kidney water not controlling heart fire" in TCM.

- **Zang fu organs**

The concepts of the internal or zang fu organs in TCM have been developed mainly on the basis of observation of their outward physiological and pathologic manifestations. The zang organs are yin in character and have the functions of storage and

transformation of fundamental substances (qi, xue, jin yue, jing, and shen). By contrast, the fu organs are yang in character and have the functions of receiving and transporting food and substances. According to the theory of yin yang, each zang organ is associated with its paired fu organ (Table A.3).

Table A.3 The paired zang and fu organs in TCM

Zang (Solid Organ)	Fu (Hollow Organ)
Liver	Gallbladder
Heart	Small Intestine
Spleen	Stomach
Lung	Large Intestine
Kidney	Bladder
Pericardium (Xin Bao Luo)	San Jiao

Note. Pericardium is a zang organ in TCM, which has the function of protecting the heart. San jiao is pericardium's yang paired organ, which has the function of regulation of water and is often regarded as a water passage. In TCM, brain, marrow, bone, vessels, gallbladder and uterus are classified as extraordinary fu organs.

Although the organs are identified by their western anatomical names, TCM views their function on a far broader scope (Table A.4). For example, the "liver" in TCM is as closely related to its emotional function as its digestive function, while the "heart" not only promotes blood circulation but also has the function of "governing" mental and spiritual activity.

Table A.4 The main function and syndromes of zang fu organs

Organ	Functions	Syndromes
Liver	Stores the blood; Ensures the smooth flow of qi.	Migraine, symptoms worsen with anger, frustration, or pressure, premenstrual tension, irregular or painful menstruation, irregular bowel movement, easy to anger.
Heart	Controls blood circulation; Houses the shen or spirit.	Disturbed sleep, nervousness, intermittent pulse, mental disturbance, palpitations.
Spleen	Transporting, distributing and transforming nutrients; Keeps blood circulating within the vessels.	Weakness and fatigue, indigestion, constipation or loose stools, haemorrhages, prolapse of internal organs, heavy sensations in limbs.
Lung	Governs qi and respiration; Dominates dispersing and descending; Regulates the water pathways; Regulates vessels; Controls the rhythmic movements of the lung.	Shortness of breath, wheezing, coughing, dryness nose and throat, respiratory and skin allergies, fatigue, anaemia.
Kidney	Stores the Essence; Promotes growth, development, fertility of the human body; Governs water metabolism; Controls reception of qi.	Back pain, knee pain, low libido, diminished hearing, miscarriage, retarded growth, premature menopause, frequent urination, and weak bones.

(Beinfield & Korngold, 1992; Campbell, 2002; Cheng, 1997; Kaptchuk, 2000; Ying et al., 1995)

The concepts in TCM suggest that each internal organ is a system in its own right (Campbell, 2002). The internal organs can also be seen as a network of relationships (Beinfield & Korngold, 1992). Harmony among the zang fu organs is mainly based on the theories of yin yang, and wu xing. Equilibrium between the external and internal environments is maintained through the connection between the channels and the zang fu organs.

- **Fundamental substances: qi, blood (xue), body fluids (jin ye), essence (jing) and spirit (shen)**

Qi has been casually translated with the term “energy” or ‘life force”, but the concept of qi extends considerably further. Everything in the universe is composed of qi, including the fundamental substances and the phenomena of their changes or motions. In TCM, qi refers to a range of dynamic physiological process, in which the idea of movement is paramount (Campbell, 2002), such as activation, warming, defense, transformation, and containment.

Blood (xue) is mainly composed of nutritive qi and body fluids (Ying et al., 1995) and formed by the functional activities of zang fu organs such as the spleen, stomach, heart, lung, lung, liver and kidney (Cheng, 1997). This is in contrast to the Western medicine concept that blood is produced by the bone marrow. The main function of blood is to nourish the whole body. As well, In TCM the function of blood is also regarded as the material basis for mental and spiritual activities (shen).

Body fluids (jin ye) refer to all fluids normally secreted in the body that serve to moisten, nourish, and lubricate the body. Fluids can be separated into liquid (jin) and humour (ye). Liquid is thin and moistens the muscles, skin, nose, eyes, mouth and other orifices. Humour is thick and sticky and lubricates the joints, nourishing and moisturising the bone marrow, spinal cord and the brain.

Qi, blood and body fluids are closely related to each other. It is qi's activities that transform food and water into nutritive qi and body fluids and ultimately into blood. Blood and body fluids, are yin substances which are motionless in nature, and depend

on qi for their movement. Both blood and body fluids are derived from food and water, hence they are said to have "a common source", and can be transformed into each other.

Essence (jing) is an essential part of the body that determines growth, bone marrow, reproduction, development, and basic constitution. It has no equivalent in Western medicine. By contrast to qi, essence can be considered to be closely associated with slow developmental change of the individual through life (Kaptchuk, 2000; Maciocia, 1996).

Spirit (shen) in Chinese, encompasses human consciousness, healthy mental, and physical function (Ergil, 2001). Jing, qi, and shen are referred to in Chinese tradition as "the three treasures" (Campbell, 2002; Maciocia, 1996). In TCM, shen and the body are inseparable and part of the unity of the cosmos. Different aspects of shen, which are known as five spirits and seven emotions, are linked to specific organs in TCM. For example, "liver" is responsible for the ethereal soul (hun); "kidney" is responsible for the will (zhi); anger is related to the "liver"; joy is related to the heart; intellect belongs to the "spleen".

- **Channel systems (jing luo)**

The channel systems consist of the channels (jing mai) and collaterals (luo mai) and their connective parts. The channels include the twelve regular channels and their branches, the divergent, the musculotendinous, the cutaneous, and the eight extraordinary channels. The twelve regular channels include three yin and three yang channels of the hand and foot. The eight extraordinary channels intersect with the main channels at special connecting points known as master or confluent points and by this connection are able to influence flows within the main channel system. The collaterals

are the connecting branches between the channels including fifteen collaterals, minute collaterals and superficial collaterals.

The twelve regular channels are responsible for the main circulation of qi, blood, and body fluids. Qi and blood in the channels are regulated and controlled by the eight extraordinary channels. Through the connection of the collaterals qi and blood are transported to the whole body.

- **The "four examinations", "pattern identification" and "eight principal syndromes"**

The "**four examinations**" include inspection, listening and smelling, inquiry and palpation, which can be summarised as followed (Kaptchuk, 2000, pp. 171-213):

  - Inspection of the client's overall physical condition including complexion, tongue, secretions, excrement and emotional state
  - Listening to the client's voice and respiration, and smelling of any odours associated with the body or breath
  - Inquiry for information on present and past complaints includes: sensations of cold or hot; perspiration; headaches and dizziness; quality, nature and location of pain; thirst; appetite and tastes; bowel movement; bladder; patterns of sleep; family health history; work; lifestyle; childhood; and emotional life
  - Palpation of the body to determine pain or sensitivity, including the taking of the pulses.

After the four examinations, the data is gathered and integrated into TCM theory in order to understand the aetiology, location and nature of patho-physiological changes. This process is called "**pattern identification**". The "**eight principal syndromes**" **(ba gang)** is a general method by which this process happens (Campbell, 2002; Cheng, 1997; Kaptchuk, 2000). The "eight principal syndromes" refer to four pairs of syndromes: yin and yang; cold and heat; exterior and interior; and deficiency and excess.

- **Needling, Moxibustion, Massage (tuina) and Cupping**

The common methods of acupuncture point stimulation include needling, moxibustion, massage (tuina) and cupping. **Needling** method refers to insertion of hair-thin solid needles into points accompanied by manipulation to the needles. **Moxibustion** is a therapeutic approach by applying heat to points through burning an herb Artemisia Vulgaris (moxa) in the form of a cone or stick. **Tuina massage** applies pressure to the points or the channels and may be called acupressure. **Cupping** is attaching small jars to the skin surface to move stagnation qi and blood and to balance yin and yang.

APPENDIX B SUMMARY OF DATABASE AND HAND SEARCHES AND EXCLUDED ARTICLES

Table B.1 Summary of database searches

Database	Total Number of Citation Found	Citations Meeting Stage One Criteria	Citations Meeting Stage Two Criteria
The Cochrane Library (Issue 4, 2003)	58	27	5
CINAHL (Ovid, 1982-Nov 2003)	21	8	3
AMED (Ovid, 1985-Nov 2003)	25	11	3
ProQuest (Health & Medical, Nov 2003)	10	5	0
Web of Knowledge (ISI, v2.0)	32	10	1
ProQuest (Multiple Databases, Nov 2003)	8	3	0
ScienceDirect (Nov 2003)	7	1	1
SportDiscus (Ovid, 1830-Nov 2003)	0	0	0
EBSCOhost (Nov 2003)	54	19	5
PEDro (Nov 2003)	11	5	2
PubMed (Nov 2003)	55	27	4

Note. Gross number of citations meeting stage one criteria is 116; Gross number of citations meeting stage two criteria is 24.

Table B.2 Summary of hand searches

Journals/Review Articles	Citations Meeting Stage One Criteria	Citations Meeting Stage Two Criteria
Archives of Internal Medicine	0	0
British Medical Journal	0	0
American Journal of Medicine	0	0
Arthritis and Rheumatism	1	0
Seminars in Arthritis and Rheumatism	0	0
Rheumatology	1	0
Acupuncture in Medicine	3	1
American Journal of Acupuncture	1	0
Chinese Acupuncture and Moxibustion	2	0
Shanghai Journal of Acupuncture and Moxibustion	3	0
Complementary Therapies in Medicine	2	1
Alternative Therapies in Health and Medicine	3	0
Ernst (1997)	10	2
Ezzo, et al (2001)	14	2
Singh, Zarow, Mishra, & Dagenais (2000)	3	2

Note. Gross number of citations meeting stage one criteria is 43; Gross number of citations meeting stage two criteria is 8.

Table B.3 Summary of excluded articles (n = 34)

Author, year	Reason for Exclusion
(Ammer & Petschning, 1988)	Neither English or Chinese
(Arichi, Arichi, & Toda, 1983)	Not OA of the hip or knee exclusively
(Berman et al., 1995)	Electro-acupuncture
(Berman et al., 1999)	Electro-acupuncture
(Chen, Li, & Zhang, 1994)	Case report, acupressure
(Christensen et al., 1993)	Neither English or Chinese
(Creamer, Singh, Hochberg, & Berman, 1999)	Retrospective, uncontrolled study
(Ernst, 1997)	Systematic review
(Ezzo et al., 2001)	Systematic review
(Fargas-Babjak & Pomeranz, 1992)	Not OA of the hip or knee exclusively
(Ferrández, García, González, Meis, & Sánchez, 2002)	Neither English or Chinese
(Fink, Künsebeck, & Wippermann, 2000)	Neither English or Chinese
(Gaw, Chang, & Shaw, 1975)	Not OA of the hip or knee exclusively
(Hochberg, 1997)	Published as abstract only
(Jiang, Zhang, Zhao, & Yang, 2001)	Non-randomising Uncontrolled studies
(Junnilla, 1982)	Not OA of the hip or knee exclusively
(Kumar & Wen, 2002)	Non-randomising Uncontrolled studies
(Li & Yang, 2002)	Acupuncture with moxa, uncontrolled trial
(Lundeberg, Eriksson, Lundeberg, & Tomas, 1991)	Not OA of the hip or knee exclusively
(Milligan, Glennie-Smith, Dowson, & Harris, 1981)	Published as abstract only
(Molsberger, Bowing, Jensen, & Lorek, 1994)	Neither English or Chinese
(Petrou, Winkler, Genti, & Balint, 1988)	Published as abstract only

(table continues)

Table B.3 (continued)

Author, year	Reason for Exclusion
(Sangdee et al., 2002)	Electro-acupuncture
(She, 2002)	Acupuncture with moxa, uncontrolled trial
(Singh, Zarow, Mishra, & Dagenais, 2000)	Unsystematic review
(Singh et al., 2001)	Electro-acupuncture
(Song, Tao, Liu, & Gao, 2001)	Acupuncture with sepectrograp irradiation
(Tang, 2000)	Acupuncture with moxa, uncontrolled trial
(Thomas, Eriksson, & Lundeberg, 1991)	Not OA of the hip or knee exclusively
(Tillu, Roberts, & Tillu, 2001)	Two forms of acupuncture was compared
(Tillu, Tillu, & Vowler, 2002)	Non-randomising controlled studies
(Yurtkuran & Kocagil, 1999)	Electro-acupuncture
(Zhang, Gao, & Teng, 2001)	Acupuncture with water injection and moxa
(Zherebkin, 1998)	Neither English or Chinese

APPENDIX C TCM QUESTIONNAIRE FOR OSTEOARTHRITIS OF THE HIP OR KNEE (SAMPLE)

Instructions

This Questionnaire asks for your views about your osteoarthritic hip or knee ***during the past four weeks***. This information will help us with your Traditional Chinese Medicine diagnosis and treatment. Please indicate your responses by ***placing a cross (x)*** in one box per line.

If you have osteoarthritis (OA) in more than one joint (i.e. both hip and/or knee), please choose the most troublesome joint when answering the questions. If you have no opinion on a particular question, leave it blank and move on to the next question.

General information:

Age:…….... Male / Female……

Do you have osteoarthritis in more than one joint (i.e. both hip and/or knee)? **Y**es/**N**o

Do you have on-going medication for osteoarthritis in the hip or knee?…….**Y**es/**N**o

Do you have osteoarthritis in the hip or knee as a primary problem?……**Y**es/**N**o

A. <u>How much pain</u> have you had in your OA hip or knee in the following situations during the past four weeks?

		None	Mild	Moderate	Severe	Extreme
1.	When walking on flat ground...	☐	☐	☐	☐	☐
2.	When lying in bed..……..	☐	☐	☐	☐	☐
3.	During the day.. ……..…..	☐	☐	☐	☐	☐
4.	During the night……..….	☐	☐	☐	☐	☐

B. <u>What kind of sensation</u> have you had in your OA hip or knee during the past four weeks?

	None	Mild	Moderate	Severe	Extreme
5. Stiffness……	☐	☐	☐	☐	☐
6. Weakness……..	☐	☐	☐	☐	☐
7. Cold ……..	☐	☐	☐	☐	☐
8. Hot……..	☐	☐	☐	☐	☐

C. <u>What other symptoms</u> have you had during the past four weeks?

	None	Mild	Moderate	Severe	Extreme
9. Sweating during the day……	☐	☐	☐	☐	☐
10. Sweating during the night……	☐	☐	☐	☐	☐
11. Feeling hot in the afternoon….	☐	☐	☐	☐	☐
12. Poor appetite……	☐	☐	☐	☐	☐
13. Thirsty for cold drinks……	☐	☐	☐	☐	☐
14. Thirsty for hot drinks……	☐	☐	☐	☐	☐
15. Hard bowel motion……	☐	☐	☐	☐	☐
16. Loose bowel motion……	☐	☐	☐	☐	☐
17. Difficulty falling asleep……..	☐	☐	☐	☐	☐
18. Easily woken up from sleep….	☐	☐	☐	☐	☐
19. Ringing in the ears……	☐	☐	☐	☐	☐
20. Dizziness..……..	☐	☐	☐	☐	☐
21. Frequent urination……	☐	☐	☐	☐	☐
22. Feeling “down” (related to OA)…	☐	☐	☐	☐	☐
23. Tiredness……	☐	☐	☐	☐	☐

Thank you for completing these questions!

Participant Code:

This code (unique number) allows us to keep track of the number of respondents to the questionnaire. In the report, information you provide will be referred to only by this code, as your own name will **not** be attached to this code. Please see the Participant Information Sheet for details on how your privacy and confidentiality are protected.

APPENDIX D SF-36 HEALTH SURVEY (SAMPLE)

{PRIVATE **}**
SF-36 HEALTH SURVEY{PRIVATE }

INSTRUCTIONS: This questionnaire asks for your views about your health, how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

circle one)

Excellent……1

Very good……2

Good……3

Fair……4

Poor……5

2. Compared to one year ago, how would you rate your health in general now?

(circle one)

Much better now than one year ago……1

Somewhat better now than one year ago……2

About the same as one year ago……3

Somewhat worse now than one year ago……4

Much worse now than one year ago……5

3. The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

(circle one number on each line)

{PRIVATE } **ACTIVITIES**	**Yes, Limited A Lot**	**Yes, Limited A Little**	**No, Not Limited At All**
a. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
b. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing **several** flights of stairs	1	2	3
e. Climbing **one** flight of stairs	1	2	3
f. Bending, kneeling or stooping	1	2	3
g. Walking **more than one kilometre**	1	2	3
h. Walking **half a kilometre**	1	2	3
i. Walking **100 metres**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

{PRIVATE }	**YES**	**NO**
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Were limited in the **kind** of work or other activities	1	2
d. Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

{PRIVATE **}**	**YES**	**NO**
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. **Accomplished less** than you would like	1	2
c. Didn't do work or other activities as **carefully** as usual	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?

(circle one)

Not at all……..1

Slightly……2

Moderately……..3

Quite a bit……4

Extremely……5

7. How much bodily pain have you had during the past 4 weeks?

(circle one)

No bodily pain……1

Very mild……2

Mild……..3

Moderate……4

Severe……5

Very severe……6

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all……..1

A little bit……..2

Moderately……..3

Quite a bit……4

Extremely……5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks

(circle one number on each line)

{PRIVATE }	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
a. Did you feel full of life?	1	2	3	4	5	6
b. Have you been a very nervous person?	1	2	3	4	5	6
c. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d. Have you felt calm and peaceful?	1	2	3	4	5	6
e. Did you have a lot of energy?	1	2	3	4	5	6
f. Have you felt down?	1	2	3	4	5	6
g. Did you feel worn out?	1	2	3	4	5	6
h. Have you been a happy person?	1	2	3	4	5	6
i. Did you feel tired?	1	2	3	4	5	6

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

(circle one)

All of the time……..1

Most of the time……..2

Some of the time……3

A little of the time……4

None of the time……5

11. How TRUE or FALSE is each of the following statements for you?

(circle one number on each line)

{PRIVATE }		Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a.	I seem to get sick a little easier than other people	1	2	3	4	5
b.	I am as healthy as anybody I know	1	2	3	4	5
c.	I expect my health to get worse	1	2	3	4	5
d.	My health is excellent	1	2	3	4	5

APPENDIX E ADVERTISEMENT, PARTICIPANT INFORMATION, CONSENT FORM, INSTRUCTIONS, AND REMINDER LETTERS FOR THE PILOT STUDY

Appendix E. 1 Advertisement (in practitioner's waiting room) for the pilot study

Questionnaires and Osteoarthritis (OA) Study

Research suggests acupuncture is potentially an effective treatment for OA of the hip or knee

A questionnaire is useful for acupuncture practice and research.

If you wish to participate in this study to assist the development of a questionnaire for OA of the hip or knee, please ask the receptionist, or contact Ping on 09 630 3168.

Appendix E. 2 Participant information sheet

Project Title

A pilot study to develop and validate a Traditional Chinese Medicine (TCM) questionnaire: A health status instrument for TCM assessment in patients with osteoarthritis of the hip or knee

Invitation

Patients from general practices with symptomatic OA of the hip or knee are invited to be part of the study. If you fulfil the criteria for this study and complete an informed consent document, you can join this study.

What is the purpose of the study?

This study will develop a TCM Questionnaire (TCMQ) for use with patients with OA of the hip or knee.

What happens in the study?

The data of the TCMQ and the MOS 36-Item Short Form Health Survey (SF-36) will be collected at the beginning of the study and two weeks later.

What are the discomforts and risks?

There are no known risks associated with this study.

What are the benefits?

Currently there are no questionnaires for assessing the usefulness of TCM in treating osteoarthritic joints. This study seeks to determine if the questionnaire being developed will be useful in acupuncture research.

How is my privacy and confidentiality protected?

Your information is collected and used in accordance with the Privacy Act 1993. The following steps are also taken to protect your privacy and confidentiality:

- Questionnaire forms are prepared with unique number only

- Mail-out is packaged and posted
- All names and contact details are deleted from data – leaving only unique number and demographic information

Costs of Participating

Participation in the study is voluntary. There is no cost to participants.

Opportunity to consider invitation

You are free to contact Ping Wang, everwell@xtra.co.nz phone 09 630 3168. and have your questions answered. You have the right to withdraw from the research at any time, or to ask for your information to be withdrawn.

Participant Concerns

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Peter Larmer, peter.larmer@aut.ac.nz phone 09 917 9999 ext 7322, Joan Campbell, joancampbell@xtra.co.nz phone 09 520 1919. Concerns regarding the conduct of the researcher should be notified to the Executive Secretary, AUTEC, Madeline Banda, madeline.banda@aut.ac.nz phone 09 917 9999 ext 8044.

Approved by the Auckland University of Technology Ethics Committee on 15 July 2003 **AUTEC Reference number** 03/63 TCM questionnaire for osteoarthritis.

Appendix E. 3 Consent to participation in research

Participant Code:

Title of Project: A pilot study to develop and validate a Traditional Chinese Medicine (TCM) questionnaire: A health status instrument for TCM assessment in patients with osteoarthritis of the hip or knee

Project Supervisors: Peter Larmer Position: Primary Supervisor

Joan Campbell Position: Secondary Supervisor

Researcher: Ping Wang Student of Master of Health Science of AUT

- I have read and understood the information provided about this research project.
- I have had an opportunity to ask questions and to have them answered.
- I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way. If I withdraw, I understand that all relevant information or parts thereof, will be destroyed.
- I agree to take part in this research.

PARTICIPANT

signature:.....Date:……

Confidential information:

Participant name:…….

Contact Address:…….

Contact Phone:……....

Researcher Contact Details:
Ping Wang, everwell@xtra.co.nz phone 09 630 3168

Project Supervisors Contact Details:
Peter Larmer, peter.larmer@aut.ac.nz phone 09 917 9999 ext 7322
Joan Campbell, joancampbell@xtra.co.nz phone 09 520 1919

Approved by the Auckland University of Technology Ethics Committee (AUTEC) on 15 July 2003 **AUTEC Reference number** 03/63

Appendix E. 4 Instructions for participant at entry

Ping Wang
AUT Master of Health Science Student
421 Mt Eden Road
Mt Eden, Auckland
Phone 09 630 3168

Dear Participant

Research suggests acupuncture is potentially an effective treatment for osteoarthritis (OA) of the hip or knee. A questionnaire is useful for acupuncture practice and research.

If you fulfil the American College of Rheumatology (ACR) combined clinical and radiographic criteria for OA of the knee or hip, and you are proficient in English, you are invited to join this study to assist in the development of a questionnaire for OA.

ACR criteria for OA of the knee

Knee pain
Bony growths (osteophytes) shown on the knee joint on X-ray
And at least **1** of the following 3 features:

- **Age** greater than 50 years
- **Stiffness** of the knee joint within 30 minutes of activity
- A **noise** from the knee joint rubbing together

ACR criteria for OA of the hip

Hip pain
And at least **2** of the following 3 features:

- A **normal ESR** blood test (erythrocyte sedimentation rate less than 20 mm/hour)
- X-ray evidence of **bony growths (osteophytes)** on the thigh bone or hip joint
- X-ray evidence of hip joint **narrowing**

It's as easy as 1, 2, 3, 4

1. Complete the enclosed **consent form**.
2. **Leave the consent form with reception**.
3. Complete the enclosed questionnaires (**the SF-36** and **the TCMQ**).
4. **Return the questionnaires** in the enclosed pre-paid envelope.

In order to re-test the TCMQ questionnaire, the same questionnaires (**the SF-36** and **the TCMQ**) will be sent to you to complete **again 2 weeks later**.

Please read the "Participant Information Sheet" for more information.

Thank you very much for your assistance.

Best regards

Ping Wang (Student of MHSc of AUT)

Appendix E. 5 Instructions for participant at 2 weeks

Ping Wang
Student of MHSc of AUT
421 Mt Eden Road
Mt Eden, Auckland
Phone 09 630 3168

Dear Participant

Re: Re-testing questionnaires

In order to re-test the TCMQ questionnaire and complete this study, please take the time to fill out again the same questionnaires (**the SF-36** and **the TCMQ**) and **return** them in the enclosed pre-paid envelope.

Thank you very much for your assistance.

Best regards

Ping Wang
Student of MHSc of AUT

Appendix E. 6 Reminder letter for non-respondent

Ping Wang
Student of MHSc of AUT
421 Mt Eden Road
Mt Eden, Auckland
Phone 09 630 3168

Dear Participant

Re: Reminder letter

About two weeks ago you should have received the **re-testing** questionnaires (**the SF-36** and **the TCMQ**) from us. However at the time of this letter we note that we have not yet received your completed questionnaires.

Perhaps you have been very busy or perhaps have just forgotten about us, if this is the case, then we would be most grateful if you could take the time to complete the enclosed questionnaires and return them as soon as possible.

Thank you very much for your assistance.

Best regards

Ping Wang
Student of MHSc of AUT

APPENDIX F QUESTIONNAIRES (Q) SENT/RECEIVED CHECK LIST

Subject No.	Consent Form Signed (Date)	Q Received (Date)	Retest-Q Sent (Date)	Retest-Q Received (Date)	Reminder Sent (Date)	Reminder Received (Date) or Reason

APPENDIX G STATISTICS ANALYSIS FOR THE TCM QUESTIONNAIRE

Appendix G. 1 Distribution of responses and missing data for items of the TCM questionnaire

Table G.1 Distribution of responses (in each category[a]) and missing data for items of the TCM questionnaire: results from the first test (all in %, n=10)

		Response Categories (Raw Values)					
Scale	Item[b]	1	2	3	4	5	Missing
Exterior (EXT)	1			40	50	10	0
	2			30	50	20	0
	3			50	40	10	0
	4			30	60	10	0
	5		20	30	50		0
	6		10	60	20	10	0
	7			10	50	30	10
	8			20	20	50	10
Interior (INT)	9		10		20	70	0
	10		10	10	10	70	0
	11		10	10	30	50	0
	12	10			10	80	0
	13	10		20	20	50	0
	14			20	10	70	0
	15		20		30	50	0
	16		10	10	20	60	0
	17	10	10	10	30	40	0
	18	10	20	10	30	30	0
	19			50		50	0
	20	10		20	20	50	0
	21		20	20	10	50	0
	22		10	30	30	30	0
	23	10		40	30	20	0

Note. [a]A higher value reflects better health for all items (5=None, 4=Mild, 3=Moderate, 2=Severe, 1=Extreme)

[b]Item numbers correspond to TCM questionnaire in Appendix C.

Table G.2 Distribution of responses (in each category[a]) and missing data for items of the TCM questionnaire: results from the second test (all in %, n=10)

		Response Categories (Raw Values)					
Scale	Item[b]	1	2	3	4	5	Missing
Exterior (EXT)	1	10	10	30	40		10
	2	10	10	20	40	10	10
	3	10		60	30		0
	4	10		30	40	20	0
	5			50	20	10	20
	6	10		30	60		0
	7				40	30	30
	8			30	20	40	10
Interior (INT)	9		10	10	10	70	0
	10			20	20	60	0
	11		20	10	10	60	0
	12			10	20	70	0
	13		20	30	30	10	10
	14			20	20	60	0
	15		10		10	70	10
	16			40	10	50	0
	17		30	10	20	30	10
	18		10	50	20	20	0
	19		20	20	10	40	10
	20		20		20	50	10
	21		10	20	20	50	0
	22		10	30	10	50	0
	23	10	10	20	20	30	10

Note. [a]A higher value reflects better health for all items (5 = None, 4 = Mild, 3 = Moderate, 2 = Severe, 1 = Extreme)

[b]Item numbers correspond to TCM questionnaire in Appendix C.

Appendix G. 2 Descriptive statistics for item and scale values of the TCM questionnaire

Table G.3 Descriptive statistics for item values of the TCM questionnaire: results from the first test (n=10)

Item[a]	Maximum	Minimum	Range	Mean	Standard Deviation	Skewness
1	5.00	3.00	2.00	3.70	.67	.43
2	5.00	3.00	2.00	3.90	.74	.17
3	5.00	3.00	2.00	3.60	.70	.78
4	5.00	3.00	2.00	3.80	.63	.13
5	4.00	2.00	2.00	3.30	.82	-.69
6	5.00	2.00	3.00	3.30	.82	.81
7	5.00	3.00	2.00	4.22	.63	-.25
8	5.00	3.00	2.00	4.33	.82	-.84
9	5.00	2.00	3.00	4.50	.97	-2.27
10	5.00	2.00	3.00	4.40	1.07	-1.69
11	5.00	2.00	3.00	4.20	1.03	-1.24
12	5.00	1.00	4.00	4.50	1.27	-2.85
13	5.00	1.00	4.00	4.00	1.33	-1.41
14	5.00	3.00	2.00	4.50	.85	-1.36
15	5.00	2.00	3.00	4.10	1.20	-1.20
16	5.00	2.00	3.00	4.30	1.06	-1.44
17	5.00	1.00	4.00	3.80	1.40	-1.08
18	5.00	1.00	4.00	3.50	1.43	-.57
19	5.00	3.00	2.00	4.00	1.05	.00
20	5.00	1.00	4.00	4.00	1.33	-1.41
21	5.00	2.00	3.00	3.90	1.29	-.56
22	5.00	2.00	3.00	3.80	1.03	-.27
23	5.00	1.00	4.00	3.50	1.18	-.76

Note. [a]Item numbers correspond to TCM questionnaire in Appendix C.

Table G.4 Descriptive statistics for item values of the TCM questionnaire: results from the second test (n=10)

Item[a]	Maximum	Minimum	Range	Mean	Standard Deviation	Skewness
1	4.00	1.00	3.00	3.11	.99	-1.12
2	5.00	1.00	4.00	3.33	1.15	-.83
3	4.00	1.00	3.00	3.10	.88	-1.46
4	5.00	1.00	4.00	3.60	1.17	-1.07
5	5.00	3.00	2.00	3.50	.67	1.41
6	4.00	1.00	3.00	3.40	.97	-1.96
7	5.00	4.00	1.00	4.42	.44	.48
8	5.00	3.00	2.00	4.11	.87	-.27
9	5.00	2.00	3.00	4.40	1.07	-1.69
10	5.00	3.00	2.00	4.40	.84	-1.00
11	5.00	2.00	3.00	4.10	1.29	-1.01
12	5.00	3.00	2.00	4.60	.70	-1.66
13	5.00	2.00	3.00	3.33	.94	.12
14	5.00	3.00	2.00	4.40	.84	-1.00
15	5.00	2.00	3.00	4.56	.96	-2.60
16	5.00	3.00	2.00	4.10	.99	-.24
17	5.00	2.00	3.00	3.56	1.26	-.17
18	5.00	2.00	3.00	3.50	.97	.45
19	5.00	2.00	3.00	3.78	1.23	-.37
20	5.00	2.00	3.00	4.11	1.20	-1.24
21	5.00	2.00	3.00	4.10	1.10	-.86
22	5.00	2.00	3.00	4.00	1.15	-.54
23	5.00	1.00	4.00	3.56	1.34	-.68

Note. [a]Item numbers correspond to TCM questionnaire in Appendix C.

Table G.5 Descriptive statistics for scale value of TCM questionnaire: results from the first test (n=10)

Test 1 Scale	Maximum	Minimum	Range	Mean	Standard Deviation	Skewness
EXT	87.50	46.88	40.63	69.22	11.07	-.51
INT	95.00	36.67	58.33	76.67	18.86	-1.30
SUM	88.04	48.37	39.67	74.08	13.47	-.89

Table G.6 Descriptive statistics for scale values of TCM questionnaire: results from the second test (n=10)

Test 2 Scale	Maximum	Minimum	Range	Mean	Standard Deviation	Skewness
EXT	78.13	24.69	53.44	64.28	17.26	-1.47
INT	98.33	36.67	61.67	75.83	19.18	-1.02
SUM	91.30	32.50	58.80	71.82	17.26	-1.34

Appendix G. 3 Item to scale correlations of the TCM questionnaire

Table G.7 Item to scale correlations[a] of the TCM questionnaire: results from the first test (n=10)

Test 1 Item[b]	EXT	INT	SUM
1	.43[a]	.15	.25[a]
2	.47[a]	-.19	-.05[a]
3	.10[a]	.06	.08[a]
4	.60[a]	.16	.30[a]
5	.61[a]	.11	.25[a]
6	.25[a]	.35	.39[a]
7	.84[a]	.16	.36[a]
8	.42[a]	.01	.12[a]
9	.10	.59[a]	.57[a]
10	.01	.49[a]	.44[a]
11	.16	.31[a]	.33[a]
12	-.08	.63[a]	.54[a]
13	.06	.94[a]	.86[a]
14	.25	.35[a]	.39[a]
15	-.50	.15[a]	-.01[a]
16	.19	.65[a]	.65[a]
17	.43	.73[a]	.80[a]
18	.43	.61[a]	.69[a]
19	.19	.35[a]	.37[a]
20	-.01	.91[a]	.81[a]
21	.19	.53[a]	.54[a]
22	.12	.56[a]	.54[a]
23	.01	.75[a]	.67[a]

Note. EXT = Exterior; INT = Interior; SUM = Summary

[a]Item to scale correlation corrected for overlap (i.e. correction between an item and the sum score of all other items in the same scale) (Howard & Forehand, 1962; Ware, Kosinski, & Gandek, 2002)

[b]Item numbers correspond to TCM questionnaire in Appendix C.

Table G.8 Item to scale correlations[a] of TCM questionnaire: results from the second test (n=10)

Test 2 Item[b]	EXT	INT	SUM
1	.82[a]	.53	.65[a]
2	.89[a]	.64	.75[a]
3	.82[a]	.84	.90[a]
4	.70[a]	.53	.62[a]
5	.12[a]	-.34	-.21[a]
6	.98[a]	.68	.82[a]
7	.30[a]	.46	.44[a]
8	.61[a]	.71	.73[a]
9	.40	.75[a]	.68[a]
10	.47	.73[a]	.69[a]
11	.73	.74[a]	.79[a]
12	.30	.75[a]	.64[a]
13	.49	.67[a]	.66[a]
14	.43	.51[a]	.52[a]
15	.68	.59[a]	.67[a]
16	.56	.46[a]	.53[a]
17	.39	.72[a]	.65[a]
18	.41	.60[a]	.57[a]
19	.35	.29[a]	.33[a]
20	.46	.86[a]	.77[a]
21	.72	.96[a]	.94[a]
22	.72	.62[a]	.71[a]
23	.43	.86[a]	.76[a]

Note. EXT = Exterior; INT = Interior; SUM = Summary

[a]Item to scale correlation corrected for overlap (i.e. correction between an item and the sum score of all other items in the same scale) (Howard & Forehand, 1962; Ware, Kosinski, & Gandek, 2002)

[b]Item numbers correspond to TCM questionnaire in Appendix C.

Appendix G. 4 One-way ANOVA of the TCM questionnaire scales by different factors

Table G.9 One-way ANOVA of the TCM questionnaire scales by different factors

Scale	Factor	N	Mean	Std. Deviation	F Statistics	Sig.
EXT	Male	5	67.19 (61.44)	12.30 (22.73)		
	Female	5	71.25 (67.13)	10.69 (11.56)	.311 (.249)	.593 (.631)
INT	Male	5	70.67 (73.33)	24.08 (26.80)		
	Female	5	82.67 (78.33)	11.40 (9.72)	1.014 (.154)	.343 (.705)
SUM	Male	5	69.46 (69.20)	16.06 (23.65)		
	Female	5	78.70 (74.43)	9.86 (9.70)	1.202 (.210)	.305 (.659)
EXT	More Joints	5	71.88 (57.00)	11.05 (20.95)		
	One Joint	4	65.63 (71.48)	13.26 (11.48)	.598 (1.517)	.465 (.258)
INT	More Joints	5	74.00 (68.00)	13.97 (19.20)		
	One Joint	4	90.00 (91.25)	4.91 (4.98)	4.667 (5.429)	.068 (.053)
SUM	More Joints	5	73.26 (64.17)	12.29 (19.49)		
	One Joint	4	81.52 (84.38)	6.93 (5.61)	1.419 (3.931)	.272 (.088)
EXT	Use Drugs	5	77.19 (75.25)	6.40 (2.91)		
	No Drugs	4	64.84 (53.05)	3.93 (21.94)	11.262 (5.191)	.012 (.057)
INT	Use Drugs	5	77.33 (80.67)	23.23 (17.74)		
	No Drugs	4	73.75 (66.25)	17.66 (21.36)	.065 (1.230)	.807 (.304)
SUM	Use Drugs	5	77.28 (78.78)	16.46 (12.48)		
	No Drugs	4	70.65 (61.66)	12.20 (21.37)	.447 (2.288)	.525 (.174)
EXT	Primary	5	68.75 (62.50)	7.65 923.03)		
	Not Primary	3	77.08 (68.02)	9.55 (13.17)	1.875 (.139)	.220 (.722)
INT	Primary	5	76.00 (73.67)	16.18 (22.74)		
	Not Primary	3	88.33 (82.22)	6.01 (13.98)	1.528 (.335)	.263 (.584)
SUM	Primary	5	73.48 (69.78)	12.12 (22.82)		
	Not Primary	3	84.42 (77.28)	5.36 (12.48)	2.089 (.264)	.199 (.626)

Note. EXT = Exterior; INT = Interior; SUM = Summary

Result from the second test is presented in brackets.

Appendix G. 5 Mean differences against means for the TCM questionnaire scale scores between test and re-test

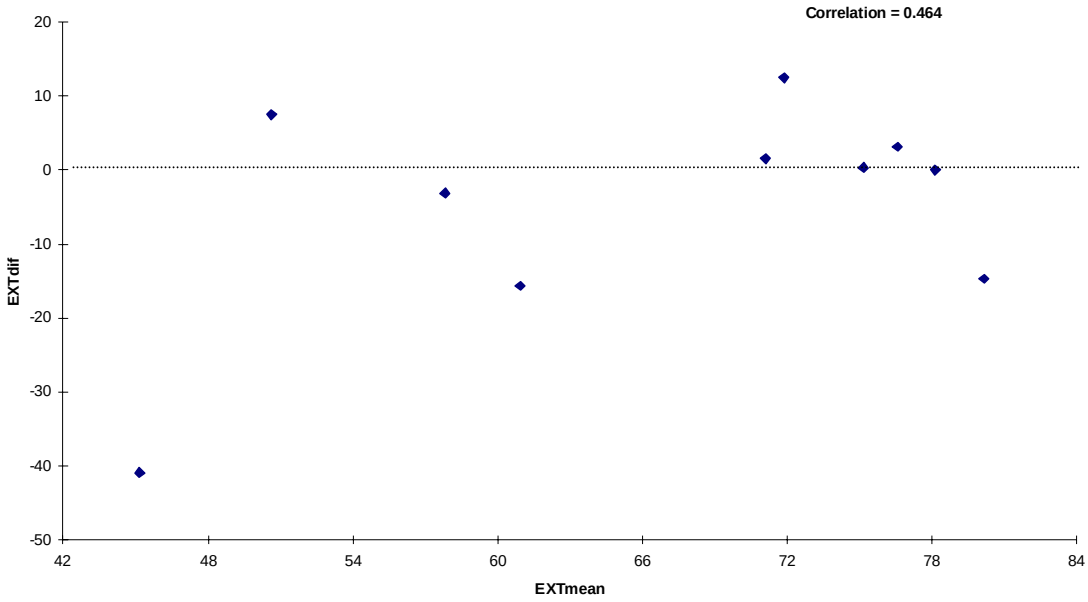

Figure G.1 Mean differences (EXTdif) against means (EXTmean) for exterior scale (EXT) scores between test and re-test (n=10). Dashed line represents zero difference

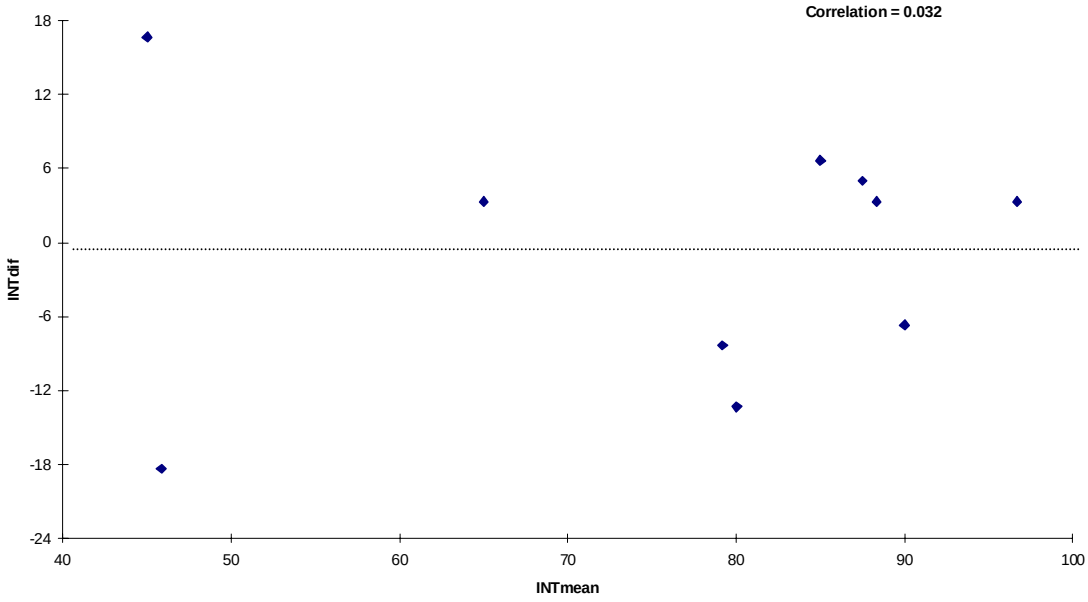

Figure G.2 Mean differences (INTdif) against means (INTmean) for interior scale (INT) scores between test and re-test (n=10). Dashed line represents zero difference

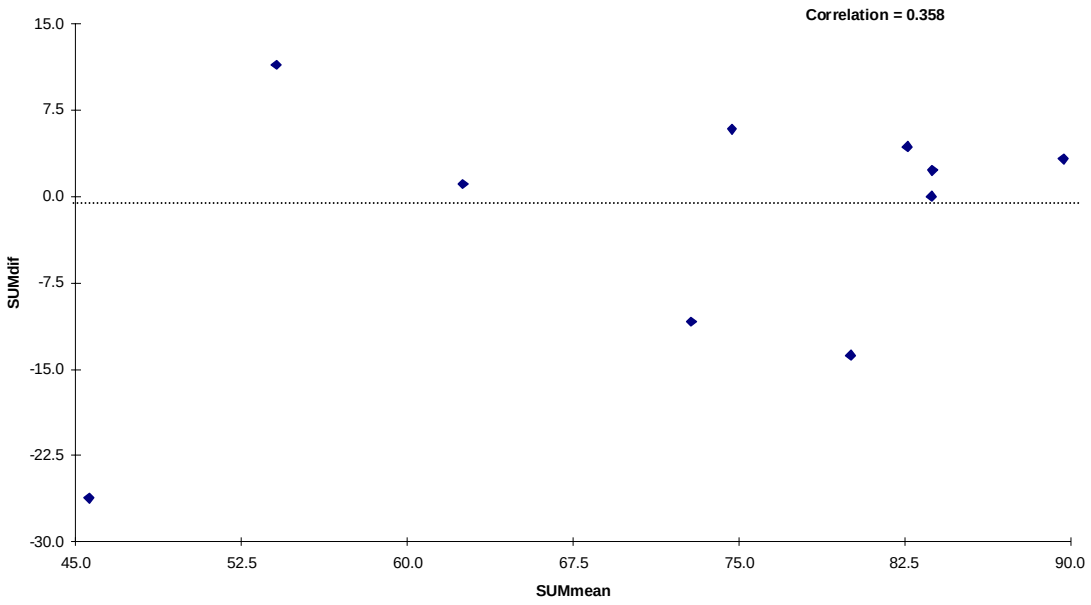

Figure G.3 Mean differences (SUMdif) against means (SUMmean) for summary scale (SUM) scores between test and re-test (n=10). Dashed line represents zero difference

APPENDIX H TABLES OF CHECKING CONTENT VALIDITY OF THE TCM QUESTIONNAIRE

Table H.1 Checking content validity and abbreviated content for items in the TCM questionnaire scales

Scale	Label	Subscale[a]	Item[b]	Abbreviated Item Content
Exterior (Item1-8)	EXT	Cold	7	Feeling cold in OA joint
		Hot	8	Feeling hot in OA joint
		Deficiency	2	Pain in OA joint when lying (except phlegm and blood stagnation)
			6	Weakness in OA joint
		Excess	1	Pain in OA joint when walking
			5	Stiffness in OA joint
		Yin	4 + Items of Cold and Deficiency	Pain in OA joint during the night + Items of Cold and Deficiency
		Yang	3 + Items of Hot and Excess	Pain in OA joint during the day + Items of Hot and Excess

(table continues)

Table H.1 (continued)

Scale	Label	Subscale[a]	Item[b]	Abbreviated Item Content
Interior (Item 9-23)	INT	Cold	14	Thirsty for hot drinks
		Hot	13	Thirsty for cold drinks
		Deficiency	9	Sweating during the day
			10	Sweating during the night
			11	Feeling hot in the afternoon (may also be clarified into the "hot" subscale)
			12	Poor appetite
			16	Loose bowel motion (may also be clarified into the "cold" subscale)
			17	Difficulty falling asleep
			18	Easily woken up from sleep
			19	Ringing in the ears
			20	Dizziness
			21	Frequent urination
			22	Feeling "down" (related to OA, and except phlegm and blood stagnation)
			23	Tiredness
		Excess	15	Hard bowel motion (except yin deficient)
		Yin	Items of Cold and Deficiency	Items of Cold and Deficiency
		Yang	Items of Hot and Excess	Items of Hot and Excess
Summary (Item 1-23)	SUM	Exterior	1-4	Pain in OA joint (walking, lying, day, night)
			5-8	Other sensations in OA joint (stiff, weak, cold, hot)
		Interior	9-23	Other symptoms in general body

Note. [a]Subscale in this table is guideline only; the subscales referring to the "eight principle syndromes" often occur in combinations; some items within subscale may also vary in the combinations with the "four examination methods" of TCM

[b]Item numbers correspond to TCM questionnaire in Appendix C.

Table H.2 Summary of the eight patterns of OA in TCM captured by the TCM questionnaire scales and subscales

Pattern of OA	TCM Questionnaire Scales and Subscales							
	Yin	Yang	Cold	Hot	Exterior	Interior	Deficiency	Excess
1. Wind Bi or wandering Bi		●			●			●
2. Cold Bi or painful Bi	●		●		●			●
3. Damp Bi or fixed Bi	●				●			●
4. Febrile Bi		●		●	●			●
5. Phlegm and blood stagnation Bi	●					●		●
6. Qi and blood deficiency Bi	●					●	●	
7. Yang deficiency Bi	●		●			●	●	
8. Yin deficiency Bi				●		●	●	

Note. The main scales of exterior and interior are in shade

The combination of both scales often occurs in chronic condition of OA.

Table H.3 Items in the TCM questionnaire scales correspond to the main "ten questions" and TCM inquiry

Item[a]	Hot and Cold	Sweat	Head and Body	Stools and Urine	Food	Drink	Hearing	Sleep	Emotion	Energy
1			●							
2			●							
3			●							
4			●							
5			●							
6			●							
7			●							
8			●							
9		●								
10		●								
11	●									
12					●					
13						●				
14						●				
15				●						
16				●						
17								●		
18								●		
19							●			
20			●							
21				●						
22									●	
23										●

Note. [a]Item numbers correspond to TCM questionnaire in Appendix C.

REFERENCES

Allen, J. J., Schnyer, R. N., & Hitt, S. K. (1998). The efficacy of acupuncture in the treatment of major depression in women. *Psychological Science, 9*, 397-401.

Altman, D. (1991). The nomogram for calculating sample size or power. In D. Altman (Ed.), *Practical statistics for medical research* (pp. 456). London: Chapman and Hall.

Altman, R., Alarcon, G., & Appelrouth, D. (1991). The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis and Rheumatism, 34*, 505-514.

Altman, R., Asch, E., & Bloch, G. (1986). Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. *Arthritis and Rheumatism, 29*, 1039-1049.

Ammer, K., & Petschning, R. (1988). Vergleich der Wirksamkeit von Akupunktur and Physikotherapie bei ambulanten Gonarthrosepatienten. *Wiener Medizinische Wochenschrift, 138*(22), 566-569.

Andersen, J., & Felson, D. T. (1988). Factors associated with knee osteoarthritis in the HANES survey: Evidence for an association with overweight, race and physical demands of work. *American Journal of Epidemiology, 128*, 179-189.

Andrich, D. (1988). *Rash models of measurement*. Beverly Hills, CA: Sage Publications.

Arichi, S., Arichi, H., & Toda, S. (1983). Acupuncture and rehabilitation (III) effects of acupuncture applied to the normal side on osteoarthritis deformans and

rheumatoid arthritis of the knee and on disorders in motility of the knee joint. *American Journal of Chinese Medicine, 11*(1-4), 146-149.

Badley, E. M. (1991). Population projections and the effect on rheumatology. *Annals of the Rheumatic Diseases, 50*, 3-6.

Badley, E. M. (1995). The economic burden of musculoskeletal disorders in Canada is similar to that for cancer, and may be higher. *Journal of Rheumatology, 22*, 204-206.

Badley, E. M., Thompson, R. P., & Wood, P. H. N. (1978). The prevalence and severity of major disabling conditions: A reappraisal of the government social survey on the handicapped and impaired in Great Britain. *International Journal of Epidemiology, 7*, 145-151.

Batchlor, E. E., & Paulus, H. E. (1992). Principles of drug therapy. In R. W. Moskowitz, D. S. Howell, V. M. Goldberg, & H. J. Mankin (Eds.), *Osteoarthritis: Diagnosis and medical/surgical management* (pp. 465-491). Philadelphia: W. B. Saunders.

Beinfield, H., & Korngold, E. (1992). *Between heaven and earth*. New York: Ballantine.

Bellamy, N., Buchannan, W. W., Goldsmith, C. H., Campbell, J., & Stitt, L. W. (1988a). Validation study of WOMAC: A health status instrument for measuring clinically improtant patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *Journal of Orthopedic Rheumatology, 1*, 95-108.

Bellamy, N., Buchannan, W. W., Goldsmith, C. H., Campbell, J., & Stitt, L. W. (1988b). Validation study of WOMAC: A health status instrument for measuring clinically improtant patient relevant outcomes to antirheumatic drug therapy in

patients with osteoarthritis of the hip or knee. *Journal of Rheumatology, 15*(12), 1833-1840.

Bensoussan, A., & Myers, S. P. (1996). *Towards a safer choice: The practice of traditional Chinese medicine in Australia*. Sydney: University of Western Sydney.

Berman, B., Lao, L., Greene, M., Anderson, R., Wong, R., Langenberg, P., & Hochberg, M. (1995). Efficacy of traditional Chinese acupuncture in the treatment of symptomatic knee osteoarthritis: A pilot study. *Osteoarthritis and Cartilage, 3*(2), 139-142.

Berman, B., Singh, B., Lao, L., Langenberg, P., Li, H., Hadhazy, V., Bareta, J., & Hochberg, M. (1999). A randomised trial of acupuncture as an adjunctive therapy in osteoarthritis of the knee. *Rheumatology, 38*(4), 346-354.

Birch, S. (1997). Issues to consider in determining an adequate treatment in a clinical trial of acupuncture. *Complementary Therapies in Medicine, 5*, 8-12.

Bland, J. H. (1983). The reversibility of osteoarthritis: A review. *American Journal of Medicine, 74*, 16-26.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307-310.

Bosi Ferraz, M., Quaresma, M. R., Aquino, L. R., Atra, E., Tugwell, P., & Goldsmith, C. H. (1990). Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis. *Journal of Rheumatology, 17*, 1022-1024.

Bowling, A. (2001a). *Measuring disease: A review of disease-specific quality of life measurement scales* (2nd ed.). Buckingham, UK: Open University Press.

Bowling, A. (2001b). *Measuring health: A review of quality of life measurement scales* (2nd ed.). Buckingham, UK: Open University Press.

Brandt, K. D. (1993). Management of osteoarthritis. In W. N. Kelley, E. D. Harris, S. Ruddy, & C. B. Sledge (Eds.), *Textbook of rheumatology* (14th ed., pp. 1385-1399). Philadelphia: W. B. Saunders.

Brandt, K. D., & Fife, R. S. (1986). Ageing in relation to the pathogenesis of osteoarthritis. *Clinics in Rheumatic Diseases, 12*, 117-130.

Brazier, J. E., Harper, R., Jones, N. M. B., O'Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF-36 health survey questionnaire: New outcome measure for primary care. *British Medical Journal, 305*, 160-164.

Brooks, P. M., & Day, R. O. (1991). Non-steroidal anti-inflammatory drugs: Difference and similarities. *New England Journal of Medicine, 324*, 1716-1725.

Buchner, A., Erdfelder, E., & Faul, F. (1997). *G*Power (version 2.0)*. Retrieved 3 February, 2004, from http://www.psycho.uniduesseldorf.de/aap/projects/gpower

Bullough, P. G. (1992). The pathology of osteoarthritis. In R. W. Moskowitz, D. S. Howell, V. M. Goldberg, & H. J. Mankin (Eds.), *Osteoarthritis* (pp. 39-69). Philadelphia: W. B. Saunders.

Campbell, J. (2002). *Lecture notes for the postgraduate diploma of Traditional Chinese Acupuncture*. Auckland, New Zealand: Auckland University of Technology.

Campbell, J. (2004). *New Zealand Acupuncture Standards Authority presentation to Health Select Committee*. Auckland, New Zealand: New Zealand Acupuncture Standards Authority.

Cassidy, C. M. (2001). Social and cultural context of complementary and alternative medicine systems. In M. S. Micozzi (Ed.), *Fundamentals of complementary and alternative medicine* (pp. 36). Philadelphia: Churchill Livingstone.

Chen, W., Li, Z., & Zhang, J. (1994). Treating osteoarthritis of the knee joint by traditional Chinese medicine. *Journal of Traditional Chinese Medicine, 14*, 279-282.

Cheng, X. (1997). *Chinese acupuncture and moxibustion*. Beijing, China: Foreign Languages Press.

Chitnavis, J., Sinsheimer, J. S., Clipsham, K., Loughlin, J., Sykes, B., Burge, P. D., & Carr, A. J. (1997). Genetic influences in end-stage osteoarthritis: Sibling risks of hip and knee replacement for idiopathic osteoarthritis. *Journal of Bone and Joint Surgery, 79*, 660-664.

Christensen, B., Iuhl, I., Wilbek, H., Bülow, H., Dreijer, N., & Rasmussen, H. (1993). Acupuncture treatment of knee arthrosis: A long-term study. *Ugeskrift for Laeger, 155*(49), 4007-4011.

Christensen, B. V., Iuhl, I. U., Vilbek, H., Bülow, H. H., Dreijer, N. C., & Rasmussen, H. F. (1992). Acupuncture treatment of severe knee osteoarthrosis: A long-term study. *Acta Anaesthesiologica Scandinavian, 36*(6), 519-525.

Cicuttini, F. M., & Spector, T. D. (1997). Evidence for the increasing prevalence of osteoarthritis with aging: Does this pertain to the oldest old? In D. Hamerman (Ed.), *Osteoarthritis* (pp. 49-62). Baltimore, MD: Johns Hopkins University Press.

Clarke, A. K. (1991). Chronic knee and hip pain trials. In P. Schapbach, & N. J. Gerber (Eds.), *Physiotherapy: Controlled trials and facts* (pp. 108-129). Basel, Switzerland: Karger.

Coggon, D., Kellingray, S., & Inskip, H. (1998). Osteoarthritis of the hip and occupational lifting. *American Journal of Epidemiology, 147*, 523-528.

Cooper, C. (1994). Osteoarthritis: Epidemiology. In J. H. Klippel, & P. Dieppe (Eds.), *Rheumatology*. St. Louis, MO: Mosby.

Cooper, C., Campbell, L., & Byng, P. (1996). Occupational activity and the risk of hip osteoarthritis. *Annals of the Rheumatic Diseases, 55*, 680-682.

Cooper, C., McAlindon, T., & Coggon, D. (1994). Occupational activity and osteoarthritis of the knee. *Annals of the Rheumatic Diseases, 53*, 90-93.

Cornbach, L. J., & Warrington, W. G. (1951). Time-limit tests: Estimating their reliability and degree of speeding. *Psychometrika, 16*, 167-188.

Creamer, P., Singh, B., Hochberg, M., & Berman, B. (1999). Are psychosocial factors related to response to acupuncture among patients with knee osteoarthritis? *Alternative Therapies in Health and Medicine, 5*(4), 72-76.

Crichton, N. J. (1993). The importance of statistics in research design. In G. T. Lewith, & D. Aldridge (Eds.), *Clinical research methodology for complementary therapies* (pp. 14-15). London: Hodder and Stoughton.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Phychometrika, 16*, 297-334.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalisability for scores and profiles*. New York: Wiley.

Danielsson, L., & Lindberg, H. (1997). Prevalence of coxarthrosis in an urban population during four decades. *Clinical Orthopaedics and Related Research, 342*, 106-110.

Davis, M. A., Ettinger, W. H., Neuhaus, J. M., Cho, S. A., & Houck, W. W. (1989). The association of knee injury and obesity with unilateral and bilateral osteoarthritis of the knee. *American Journal of Epidemiology, 130*, 278-288.

Deng, T. (1999). *Practical diagnosis in traditional Chinese medicine* (M. Ergil, & S. Yi, Trans.). Edinburgh: Churchill Livingstone.

Dieppe, P. (1990). Some recent approaches to arthritis research. *Seminars in Arthritis and Rheumatism, 20*, 2-11.

Dieppe, P. (1995). The classification and diagnosis of osteoarthritis. In K. E. Kuettner, & V. M. Goldberg (Eds.), *Osteoarthritic disorders* (pp. 5-12). Rosemont: American Academy of Orthopaedic Surgeons.

Dieppe, P. (1999). Osteoarthritis: Time to shift the paradigm. This includes distinguishing between severe disease and common minor disability. *British Medical Journal, 318*, 1299-1300.

Dimitrov, D., Rumrill, P., Fitzgerald, S., & Hennessey, M. (2001). Reliability in rehabilitation measurement. *Work, 16*, 159-164.

Doyle, D. (1986a). Management of osteoarthrosis. In J. T. Scott (Ed.), *Copeman's textbook of the rheumatic diseases* (6th ed., Vol. 2). New York: Churchill Livingstone.

Doyle, D. (1986b). Osteoarthritis. *The Practitioner, 230*, 335-340.

Ergil, K. V. (2001). Chinese medicine. In M. S. Micozzi (Ed.), *Fundamentals of complementary and alternative medicine* (pp. 314). Philadelphia: Churchill Livingstone.

Ernst, E. (1997). Acupuncture as a symptomatic treatment of osteoarthritis: A systematic review. *Scandinavian Journal Rheumatology, 26*(6), 444-447.

Ezzo, J., Hadhazy, V., Birch, S., Lao, L., Kaplan, G., Hochberg, M., & Berman, B. (2001). Acupuncture for osteoarthritis of the knee: A systematic review. *Arthritis and Rheumatism, 44*(4), 819-825.

Fang, Y., Deng, T., Li, K., Cheng, K., Jin, S., Huang, X., & Dong, J. (1985). *Practical internal medicine of traditional Chinese medicine*. Shanghai, China: Shanghai Science and Technology Press (in Chinese).

Fargas-Babjak, A., & Pomeranz, B. (1992). Acupuncture-like stimulation with Codetron for rehabilitation of patients with chronic pain syndrome and osteoarthritis. *Acupuncture & Electro-Therapeutics Research, 17*, 95-105.

Fayers, P. M., & Machin, D. (1998). Factor analysis. In M. J. Staquet, R. D. Hays, & P. M. Fayers (Eds.), *Quality of life assessment in clinical trials: Methods and practice*. Oxford: Oxford University Press.

Felson, D. T. (1988). Epidemiology of hip and knee osteoarthritis. *Epidemiologic Reviews, 10*, 1-28.

Felson, D. T. (1993). The course of osteoarthritis and factors that affect it. *Rheumatic Disease Clinics of North America, 19*, 607-615.

Felson, D. T., Naimark, A., Anderson, J., Kazis, L., Castelli, W., & Meenan, R. F. (1987). The prevalence of knee osteoarthritis in the elderly: The Framingham osteoarthritis study. *Arthritis and Rheumatism, 30*, 914-918.

Felson, D. T., Zhang, Y., Anthony, J. M., Naimark, A., & Anderson, J. J. (1992). Weight loss reduces the risk for symptomatic knee osteoarthritis in women: The Framingham study. *Annals of Internal Medicine, 116*, 535-539.

Felson, D. T., Zhang, Y., Hannan, M. T., Naimark, A., Weissman, B., Aliabadi, P., & Levy, D. (1997). Risk factors for incident radiographic knee osteoarthritis in the elderly: The Framingham study. *Arthritis and Rheumatism, 40*, 728-733.

Ferrández, I. A., García, O. L., González, G. A., Meis, M. M. J., & Sánchez, R. B. M. (2002). Effectiveness of acupuncture in the treatment of pain from osteoarthritis of the knee. *Atencion Primaria, 30*(10), 602-608.

Fink, M., Kunsebeck, H., Wipperman, B., & Gehrke, A. (2001). Non-specific effects of traditional Chinese acupuncture in osteoarthritis of the hip. *Complementary Therapies in Medicine, 9*(2), 82-89.

Fink, M., Künsebeck, H., & Wippermann, B. (2000). Effect of needle acupuncture on pain perception and functional impairment of patients with coxarthrosis. *Zeitschrift fur Rheumatologie, 59*(3), 191-199.

Finocchiaro, C., Abramson, M. J., & Ryan, P. F. (1996). Arthritis in Australia: An emerging public health problem. *Medical Journal of Australia, 165*, 352.

Forbes, W. F. (1997). General concepts of the association of aging and disease. In D. Hamerman (Ed.), *Osteoarthritis* (pp. 3-14). Baltimore, MD: Johns Hopkins University Press.

Fries, J. F., Spitz, P. W., & Kraines, R. G. (1980). Measurement of patient outcome in arthritis. *Arthritis and Rheumatism, 23*, 137-145.

Fries, J. F., Spitz, P. W., & Young, D. Y. (1982). The dimensions of health outcomes: The health assessment questionnaire, disability and pain scales. *Journal of Rheumatology, 9*, 789-793.

Froberg, D., & Kane, R. L. (1989). Methodology for measuring health-state preference 1: Measurement strategies. *Journal of Clinical Epidemiology, 42*, 345-354.

Gabriel, S. E., Crowson, C. S., & O'Fallon, W. M. (1995). Costs of osteoarthritis: Estimates from a geographically defined population. *Journal of Rheumatology, 43*(Supp.), 23-25.

Gascoigne, S. (2001). *The clinical medicine guide: A holistic perspective*. Clonakilty, Ireland: Jigme Press.

Gaw, A., Chang, L., & Shaw, L. (1975). Efficacy of acupuncture on osteoarthritic pain: A controlled, double-blind study. *The New England Journal of Medicine, 293*(8), 375-378.

Godfrey, C. M., & Morgan, P. (1978). A controlled trial of the theory of acupuncture in musculoskeletal pain. *Journal of Rheumatology, 5*(2), 121-124.

Griffin, M. R., Brandt, K. D., Liang, M. H., Pincus, T., & Ray, W. A. (1995). Practical management of osteoarthritis: Integration of pharmacologic and nonpharmacologic measures. *Archives of Family Medicine, 4*, 1049-1055.

Griffin, M. R., Piper, J. M., Daugherty, J. R., Snowden, M., & Ray, W. A. (1991). Non-steroidal anti-inflammatory drug use and increased risk for peptic ulcer disease in elderly persons. *Annals of Internal Medicine, 114*(63), 257-263.

Griffin, M. R., Ray, W. A., & Schaffner, W. (1988). Non-steroidal anti-inflammatory drug use and death from peptic ulcer in elderly persons. *Annals of Internal Medicine, 109*, 359-363.

Guccione, A. A. (1997). Osteoarthritis, co-morbidity, and physical disability. In D. Hamerman (Ed.), *Osteoarthritis* (pp. 84-98). Baltimore, MD: Johns Hopkins University Press.

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Haley, S. M., McHorney, C. A., & Ware, J. E. (1994). Evaluation of the MOS SF-36 physical functioning scale (PF-10): 1. Unidimensionality and reproducibility of the Rasch item scale. *Journal of Clinical Epidemiology, 47*, 671-684.

Haslam, R. (2001). A comparison of acupuncture with advice and exercises on the symptomatic treatment of osteoarthritis of the hip: A randomised controlled trial. *Acupuncture in Medicine, 19*(1), 19-26.

Hays, R. D., Anderson, R. T., & Revicki, D. (1998). Assessing reliability and validity of measurement in clinical trials. In M. J. Staquet, R. D. Hays, & P. M. Fayers (Eds.), *Quality of life assessment in clinical trials* (pp. 169-182). New York: Oxford University Press.

Hays, R. D., & Hadoron, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research, 1*, 73-75.

Hill, C. L., LaValley, M. P., & Felson, D. T. (2002). Discrepancy between published report and actual conduct of randomised clinical trials. *Journal of Clinical Epidemiology, 55*, 783-786.

Hochberg, M. (1997). Acupuncture useful for treating osteoarthritis of the knee. *Geriatrics, 52*(9), 98.

Hochberg, M. C. (2001). Osteoarthritis. In A. J. Silman, & M. C. Hochberg (Eds.), *Epidemiology of the rheumatic diseases* (2nd ed., pp. 205-229). New York: Oxford University Press.

Hochberg, M. C., Altman, R. D., Brandt, K. D., Clark, B. M., Dieppe, P. A., & Griffin, M. R. (1995a). Guidelines for the medical management of osteoarthritis: Part 1. osteoarthritis of the hip. *Arthritis and Rheumatism, 38*, 1535-1540.

Hochberg, M. C., Altman, R. D., Brandt, K. D., Clark, B. M., Dieppe, P. A., & Griffin, M. R. (1995b). Guidelines for the medical management of osteoarthritis: Part 2. osteoarthritis of the knee. *Arthritis and Rheumatism, 38*, 1541-1546.

Hochberg, M. C., & Lethbridge-Cejku, M. (1997). Epidemiologic considerations in the primary prevention of osteoarthritis. In D. Hamerman (Ed.), *Osteoarthritis* (pp. 169-186). Baltimore, MD: Johns Hopkins University Press.

Holden, R. R., Fekken, G. C., & Jackson, D. N. (1985). Structured personality test item characteristics and validity. *Journal of Research in Personality, 19*, 386-394.

Hopwood, V., & Lewith, G. (2002). Acupuncture trials and methodological considerations. *Clinical Acupuncture and Oriental Medicine, 3*(4), 192-199.

Howard, K. I., & Forehand, G. G. (1962). A method for correcting item-total correlations for the effect of relevant item inclusion. *Educational and Psychological Measurement, 22*, 731-735.

Huang, X. (1999). Preliminary observation of relationship between deqi phenomenon and acupuncture effects. *Chinese Acupuncture and Moxibustion, 19*(1), 19-21.

Hunt, S. M., McKenna, S. P., McEwen, J., Backett, E. M., Williams, J., & Papp, E. (1981). The Nottingham Health Profile: Subjective health status and medical consultations. *Social Science and Medicine, 15A*, 221-229.

Huskisson, E. C. (1974). Measurement of pain. *Lancet, 2*, 1127-1131.

Jadad, A. R., Moore, R. A., Carrol, D., Jenkinson, C., Reynolds, D. J. M., Gavaghan, D. J., & McQuay, H. J. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials, 17*, 1-12.

Jenkinson, C. (1991). Why are we weighting? A critical examination of the use of item weights in a health status measure. *Social Science and Medicine, 32*, 1413-1416.

Jenkinson, C., & McGee, H. (1998). *Health status measurement: A brief but critical introduction*. Oxon, UK: Radcliffe Medical Press.

Jiang, A., Zhang, L., Zhao, C., & Yang, F. (2001). Clinical effect of acupuncture treatment in 109 cases of knee osteoarthritis. *Journal of Traditional Chinese Medicine, 21*(4), 282-285.

Junnilla, S. (1982). Acupuncture superior to piroxicam in the treatment of osteoarosis. *American Journal of Acupuncture, 10*, 341-346.

Kaplan, R. M. (1998). Profile versus utility based measures of outcome for clinical trials. In M. J. Staquet, R. D. Hays, & P. M. Fayers (Eds.), *Quality of life assessment in clinical trials: Methods and practice* (pp. 72). New York: Oxford University Press.

Kaptchuk, T. J. (2000). *Chinese medicine: The web that has no weaver* (Rev. ed.). London: Rider.

Keen, R. W., Hart, D. J., Lanchbury, J. S., & Spector, T. D. (1997). Association of early osteoarthritis of the knee with a Taq I polymorphism of the vitamin D receptor gene. *Arthritis and Rheumatism, 40*, 1444-1449.

Kellgren, J. H., & Lawrence, J. S. (1958). Osteo-arthrosis and disk degeneration in an urban population. *Annals of the Rheumatic Diseases, 17*, 388-397.

Kellgren, J. H., & Lawrence, J. S. (1963). *The epidemiology of chronic rheumatism: Atlas of standard radiographs*. Oxford: Blackwell Scientific.

Kelsey, J. L., & Hochberg, M. C. (1988). Epidemiology of chronic musculoskeletal disorders. *Annual Review of Public Health, 9*, 379-401.

Koller, M., Kussmann, J., & Lorenz, W. (1996). Symptom reporting in cancer patients: the role of negative affect and experienced social stigma. *Cancer, 77*, 983-995.

Kumar, A., & Wen, X. (2002). Acupuncture treatment for osteoarthritic pain and Inflammation of the knee. *Alternative Therapies in Health and Medicine, 8*(6), 128.

Lawrence, J. S., Bremner, J. M., & Bier, F. (1966). Osteoarthrosis, prevalence in the population and relationship between symptoms and X-ray changes. *Annals of the Rheumatic Diseases, 25*, 1-24.

Lawrence, R. C., Hochberg, M. C., Kelsey, J. L., McDuffie, F. C., & Medsger, T. A. (1989). Estimates of the prevalence of selected arthritic and musculoskeletal diseases in the United States. *Journal of Rheumatology, 16*, 427-441.

Lee, K. P., Schotland, M., Bacchetti, P., & Bero, L. A. (2002). Association of journal quality indicators with methodological quality of clinical research articles. *The Journal of the American Medical Association, 287*, 2805-2808.

Lei, H., & Skinner, H. A. (1980). A psychometric study of life evens and social readjustment. *Journal of Psychosomatic Research, 24*, 57-65.

Lewith, G., Jonas, W. B., & Walach, H. (2002). *Clinical research in complementary therapies: Principles, problems and solutions*. London: Churchill Livingstone.

Lewith, G., Walach, H., & Jonas, W. B. (2002). Balanced research strategies for complementary and alternative medicine. In G. Lewith, W. Jonas, & H. Walach (Eds.), *Clinical research in complementary therapies: principles, problems and solutions* (pp. 19). London: Churchill Livingstone.

Li, B., & Yang, Q. (2002). Acupuncture combined with herbs for treatment in 120 cases of osteoarthritic knee. *Shanghai Journal of Acupuncture and Moxibustion, 21*(3), 31.

Liang, M. H., & Fortin, P. (1991). Management of osteoarthritis of the hip and knee. *New England Journal of Medicine, 325*, 125-127.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-55.

Lozada, C. J., & Altman, R. D. (2001). Management of osteoarthritis. In W. J. Koopman (Ed.), *Arthritis and allied conditions: A textbook of rheumatology* (14th ed., pp. 2251). Philadelphia: Lippincott Williams & Wilkins.

Lu, A. A. (1996). *Designing and conducting health surveys: A comprehensive guide* (2nd ed.). San Francisco: Jossey-Bass.

Lundeberg, T., Eriksson, S., Lundeberg, S., & Tomas, M. (1991). Effect of acupuncture and naloxone in patients with osteoarthritis pain: A sham acupuncture controlled study. *Pain Clinic, 4*, 155-161.

MacFarlane, D. (1992). Osteoarthritis. *The practitioner, 236*, 1061-1065.

Maciocia, G. (1996). *The foundations of Chinese medicine*. New York: Churchill Livingstone.

Maciocia, G. (1997). *The practice of Chinese medicine*. New York: Churchill Livingstone.

McAlindon, T. E., Felson, D. T., & Zhang, Y. (1996). Relation of dietary intake and serum levels of vitamin D to progression of osteoarthritis of the knee among participants in the Framingham Study. *Annals of Internal Medicine, 125*, 353-359.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46.

McHorney, C. A., & Ware, J. E. (1995). Construction and validation of an alternate form general mental health scale for the medical outcomes study short form 36-item health survey. *Medical Care, 33*, 15-28.

McHorney, C. A., Ware, J. E., & Raczek, A. E. (1993). The MOS 36-item short form health survey (SF-36): 2. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care, 31*, 247-263.

Mclndoe, A. K., Young, K., & Bone, M. E. (1995). A comparison of acupuncture with intra-articular steroid injection as analgesia for osteoarthritis of the hip. *Acupuncture in Medicine, 13*(2), 67-70.

Meenan, R. F., Gertman, P. M., & Mason, J. H. (1980). Measuring health status in arthritis: The arthritis impact measurement scales. *Arthritis and Rheumatism, 23*, 146-152.

Meenan, R. F., Mason, J. H., Anderson, J. J., Guccione, A. A., & Kazis, L. E. (1992). AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. *Arthritis and Rheumatism, 35*, 1-10.

Milligan, J., Glennie-Smith, K., Dowson, D., & Harris, J. (1981). *Comparison of acupuncture with physiotherapy in the treatment of osteoarthritis of the knees.* Paper presented at the 15th International Congress of Rheumatology, Paris.

Molsberger, A., Bowing, G., Jensen, K., & Lorek, M. (1994). Schmerztherapie mit Akupunktur bei Gonarthrose: Eine kontrollierte studie zur analgetischen Wirkung der akupunktur bei Gonarthrose. *Der Schmerz, 8*, 37-42.

Moskowitz, R. W. (1992). Osteoarthritis: Symptoms and signs. In R. W. Moskowitz, D. S. Howell, V. M. Goldberg, & H. J. Mankin (Eds.), *Osteoarthritis: Diagnosis and medical/surgical management* (2nd ed., pp. 255-261). Philadelphia: W. B. Saunders.

Moskowitz, R. W., Davis, W., Sammarco, J., Mast, W., & Chase, S. W. (1970). Experimentally included corticosteroid arthropathy. *Arthritis and Rheumatism, 134*, 236.

Moskowitz, R. W., & Holderbaum, D. (2001). Clinical and laboratory findings in osteoarthritis. In W. J. Koopman (Ed.), *Arthritis and allied conditions: A textbook of rheumatology* (14th ed., pp. 2216-2245). Philadelphia: Lippincott Williams & Wilkins.

Mulrow, C. D., & Oxman, A. (1997). Critical appraisal of studies: Cochrane collaboration handbook. In *The Cochrane Library database on disk and CD-ROM. Issue 4*. Oxford: Cochrane Collaboration.

NHS Centre for Reviews and Dissemination. (1996). *Undertaking systematic reviews of research on effectiveness: CRD guidelines for those carrying out or commissioning reviews*. New York: Author.

Ni, M. (1995). *The Yellow Emperor's classic of medicine: A new translation of the Neijing Suwen with commentary*. Boston: Shambhala.

Nishisato, N., & Torii, Y. (1970). Effects of categorizing continuous normal distributions on the product-moment correlation. *Japanese Psychological Research, 13*, 45-49.

Norheim, A. J. (1996). Adverse effects of acupuncture: A study of the literature for the years 1981-1994. *Journal of Alternative and Complementary Medicine, 2*, 291-297.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Olshansky, S. J., & Cassel, C. K. (1997). Implications of the accrual of chronic nonfatal conditions in very elderly persons. In D. Hamerman (Ed.), *Osteoarthritis* (pp. 15-29). Baltimore, MD: Johns Hopkins University Press.

Osoba, D. (1998). Guidelines for measuring health-related quality of life in clinical trials. In M. J. Staquet, R. D. Hays, & P. M. Fayers (Eds.), *Quality of life assessment in clinical trials: Methods and practice* (pp. 28). New York: Oxford University Press.

Palmoski, M., & Brandt, K. (1980). Effects of some nonsteroidal antiinflamatory drugs on articular cartilage proteoglycan metabolism and organization. *Arthritis and Rheumatism, 23*, 1010.

Palmoski, M., & Brandt, K. (1982). Aspirin aggravates the degeneration of canine joint cartilage caused by immobilization. *Arthritis and Rheumatism, 25*, 1333.

Palmoski, M., & Brandt, K. (1983). In vivo effect of aspirin on canine osteoarthritic cartilage. *Arthritis and Rheumatism, 26*, 994.

Petrou, P., Winkler, V., Genti, G., & Balint, G. (1988). Double blind trial to evaluate the effect of acupuncture treatment on knee osteoarthritis. *Scandinavian Journal Acupuncture, 3*, 112-115.

Peyron, J. G., & Altman, R. D. (1992). The epidemiology of osteoarthritis. In R. W. Moskowitz, D. S. Howell, V. M. Goldberg, & H. J. Mankin (Eds.), *Osteoarthritis* (pp. 15-37). Philadelphia: W. B. Saunders.

Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research: Applications to practice* (2nd ed.). New Jersey: Prentice Hall Health.

Prescott, R. J., Counsell, C. E., & Gillespie, W. J. (1999). Factors that limit the quality, number and progress of randomisied controlled trials. *Health Technology Assessment, 3*, 1-143.

Qiu, L. L., Kong, Z., & Tu, J. S. (1988). *Therapeutics of Chinese acupuncture and moxibustion*. Nanjing, China: Jiangsu Science and Technology Press.

Radin, E. L., Schaffler, M., Gibson, G., & Tashman, S. (1995). Osteoarthrosis as the result of repetitive trauma. In K. E. Kuettner, & V. M. Goldberg (Eds.), *Osteoarthritc disorders* (pp. 197-203). Rosemont: American Academy of Orthopaedic Surgeons.

Rash, G. (1980). *Probablistic models for some intelligence and attainment tests*. Chicago: University of Chicago.

Ren, X. S., Amick, B., Zhou, L., & Gandek, B. (1998). Translation and psychometric evaluation of a Chinese version of the SF-36 health survey in the United States. *Journal of Clinical Epidemiology, 51*(11), 1229-1138.

Roos, H., Lindberg, H., & Gardsell, P. (1994). The prevalence of gonarthrosis and its relation to meniscectomy in former soccer players. *American Journal of Sports Medicine, 22*, 219-222.

Salaffi, F., Cavalieri, F., Nolli, M., & Ferraccioli, G. (1991). Analysis of disability in knee osteoarthritis: Relationship with age and psychological variables but not with radiographic score. *Journal of Rheumatology, 18*, 1581-1586.

Sandell, L. J. (1995). Molecular biology of collagens in normal and osteoarthritic cartilage. In K. E. Kuettner, & V. M. Goldberg (Eds.), *Osteoarthritic disorders* (pp. 131-146). Rosemont: American Academy of Orthopaedic Surgeons.

Sangdee, C., Teekachunhatean, S., Sananpanich, K., Sugandhavesa, N., Chiewchantanakit, S., Pojchamarnwiputh, S., & Jayasvasti, S. (2002). *Electroacupuncture versus Diclofenac in symptomatic treatment of osteoarthritis of the knee: A randomized controlled trial*. Retrieved 19 November, 2003, from http://www.biomedcentral.com/1472-6882/2/3

Sanson-Fisher, R. W., & Perkins, J. J. (1998). Adaptation and validation of the SF-36 health survey for use in Australia. *Journal of Clinical Epidemiology, 51*(11), 961-967.

Schlegel, S., & Paulus, H. (1986). Update on NSAID use in rheumatic diseases. *Bulletin on Rheumatic Diseases, 36*, 1-8.

Schnitzer, T. J. (1993). Osteoarthritis treatment update: Minimizing pain while limiting patient risk. *Postgraduate Medicine, 93*, 89-95.

Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement, 41*, 1101-1114.

Selltiz, C., Wrightsman, L. S., & Cook, S. W. (1976). *Research methods in social relations*. New York: Holt Rinehart & Winston.

Sharma, L. (2001). Epidemiology of osteoarthritis. In R. W. Moskowitz, A. S. Howell, R. D. Altman, J. A. Buckwalter, & V. M. Goldberg (Eds.), *Osteoarthritis: Diagnosis and medical/surgical management* (3rd ed., pp. 3). Philadelphia: W. B. Saunders.

She, Q. (2002). Acupuncture combined with herbs for treatment in 108 cases of osteoarthritic knee. *Chinese Acupuncture and Moxibustion, 22*(6), 404.

Sherman, K. J., Hogeboom, C. J., & Cherkin, D. C. (2001). How traditional Chinese medicine acupuncturists would diagnose and treat chronic low back pain: Results of a survey of licensed acupuncturists in Washington State. *Complementary Therapies in Medicine, 9*, 146-156.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.

Silman, A. J., & Hochberg, M. C. (1993). *Epidemiology of the rheumatic diseases*. Oxford: Oxford University Press.

Simon, L. S. (2000). Osteoarthritis: A review. *Clinical Cornerstone, 2*, 26-37.

Singh, B., Berman, B., Hadhazy, V., Bareta, J., Lao, L., Zarow, F., & Hochberg, M. (2001). Clinical decisions in the use of acupuncture as an adjunctive therapy for osteoarthritis of the knee. *Alternative Therapies in Health and Medicine, 7*(4), 58-65.

Singh, B., Zarow, F., Mishra, L., & Dagenais, S. (2000). Acupuncture as a treatment for osteoarthritis of the knee: Clinical relevance. *Topics in Clinical Chiropractic, 7*(2), 39-42.

Slavin, R. E. (1995). Best evidence synthesis: An intelligent alternative to meta-analysis. *Journal of Clinical Epidemiology, 48*, 9-18.

Sobell, D. S. (1979). *Ways of health*. New York: Harcourt Brace.

Song, W., Tao, W., Liu, M., & Gao, X. (2001). Analysis of the therapeutic of acupuncture plus spectrograp irradiation on treating gonitis. *Shanghai Journal of Acupuncture and Moxibustion, 20*(4), 22-23.

Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS short-form general health survey: Reliability and validity in a patient population. *Medical Care, 26*, 724-736.

Stewart, A. L., & Ware, J. E. (1992). *Measuring functioning and well-being: The medical outcomes study approach*. Durham, NC: Duke University Press.

Streiner, D. L. (1993). A checklist for evaluating the usefulness of rating scales. *Canadian Journal Psychiatry, 38*, 140-148.

Streiner, D. L., Goldberg, J. O., & Miller, H. R. (1993). MCMI-II weights: Their lack of effectiveness. *Journal of Personality Assessment, 60*, 471-476.

Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales: A practical guide to their development and use* (2nd ed.). New York: Oxford University Press.

Summers, M. N., Haley, W. E., Reveille, J. D., & Alarcon, G. S. (1988). Radiographic assessment and psychologic variables as predictors of pain and functional impairment in osteoarthritis of the knee or hip. *Arthritis and Rheumatism, 31*, 204-209.

Takeda, W., & Wessel, J. (1994). Acupuncture for the treatment of pain of osteoarthritic knees. *Arthritis Care and Research, 7*(3), 118-122.

Tang, D. (2000). Acupuncture combined with moxibustion and tuina for treatment in osteoarthritic knee. *Chinese Acupuncture and Moxibustion, 20*(6), 348.

Tepper, S., & Hochberg, M. C. (1993). Factors associated with hip osteoarthritis: Data from the first national health and nutrition examination survey (NHANES-I). *American Journal of Epidemiology, 137*, 1081-1088.

Thomas, K., & Fitter, M. (2002). Possible research strategies for evaluating CAM interventions. In G. Lewith, W. Jonas, & H. Walach (Eds.), *Clinical research in complementary therapies: Principles, problems and solutions* (pp. 73). London: Churchill Livingstone.

Thomas, M., Eriksson, S., & Lundeberg, T. (1991). A comparative study of diazepam and acupuncture in patients with osteoarthritis pain: a placebo controlled study. *American Journal of Chinese Medicine, 19*, 95-100.

Thompson, J. (1988). Ageing of the population: contemporary trends and issues. *Population Trends, 21*, 18-22.

Tillu, A., Roberts, C., & Tillu, S. (2001). Unilateral versus bilateral acupuncture on knee function in advance osteoarthritis of the knee: A prospective randomised trial. *Acupuncture in Medicine, 19*(1), 15-18.

Tillu, A., Tillu, S., & Vowler, S. (2002). Effect of acupuncture on knee function in advanced osteoarthritis of the knee: A prospective, non-randomised controlled study. *Acupuncture in Medicine, 20*(1), 19-21.

Towheed, T., Shea, B., Wells, G., & Hochberg, M. (1998). *Osteoarthritis: A systematic review of randomized controlled trials of analgesia and anti-inflammatory therapy in osteoarthritis (OA) of the hip* (Cochrane Review). Oxford: Cochrane Library.

Uitterlinden, A. G., Burger, H., & Huang, Q. (1997). Vitamin D receptor genotype is associated with radiographic osteoarthritis at the knee. *Journal of Clinical Investigation, 100*, 259-263.

Van Saase, J. L., Van Romunde, L. K., Cats, A., Vandenbroucke, J. P., & Valkenburg, H. A. (1989). Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiologic osteoarthritis in a Dutch population with that in 10 other populations. *Annals of the Rheumatic Diseases, 48*, 271-280.

Van Saase, J. L., Vandenbroucke, J. P., Van Romunde, L. K., & Valkenburg, H. A. (1988). Osteoarthritis and obesity in the general population: A relationship calling for an explanation. *Journal of Rheumatology, 15*, 1152-1158.

Van Tulder, M. W., Koes, B. W., & Bouter, L. M. (1997). Conservative treatment of acute and chronic non-specific low back pain: A systematic review of randomised controlled trials of the most common interventions. *Spine, 22*, 2128-2156.

Verbrugge, L. M., & Patrick, D. L. (1995). Seven chronic conditions: Their impact on US adults' activity levels and use of medical services. *American Journal of Public Health, 85*, 173-182.

Vicent, C. A., & Richardson, P. H. (1986). The evaluation of therapeutic acupuncture: Concepts and methods. *Pain, 24*, 1-13.

Vincent, C., & Furnham, A. (2001). *Complementary medicine: A research perspective*. New York: Wiley.

Vingard, E., Alfredsson, L., Goldie, I., & Hogstedt, C. (1991). Occupation and osteoarthrosis of the hip and knee: A register-based cohort study. *International Journal of Epidemiology, 20*, 1025-1031.

Vingard, E., Alfredsson, L., Goldie, I., & Hogstedt, C. (1993). Sports and osteoarthritis of hip: An epidemiologic study. *American Journal of Sports Medicine, 21*, 195-200.

Ware, J. E. (1993). Measuring patients' views: The optimum outcome measure. *British Medical Journal, 306*, 1429-1430.

Ware, J. E., & Gandek, B. (1998). Methods for testing data quality, scaling assumptions, and reliability: The IQOLA project approach. *Journal of Clinical Epidemiology, 51*(11), 945-952.

Ware, J. E., Kosinski, M., & Gandek, B. (2002). *SF-36 health survey: Manual & interpretation guide*. Lincoln, RI: QualityMetric Incorporated.

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): A conceptual frame work and item selection. *Medicine Care, 30*, 473-483.

Watson, M. C., Brookes, S. T., Kirwan, J. R., & Faulkner, A. (1998). *Osteoarthritis: The comparative efficacy of non-aspirin non-steroidal anti-inflammatory drugs for the management of osteoarthritis of the knee* (Cochrane Review). Oxford: Cochrane Library.

White, A., Hayhoe, S., Hart, A., & Ernst, E. (2001). Survey of adverse events following acupuncture: A prospective study of 32,000 consultations. *Acupuncture in Medicine, 19*(2), 84.

White, A., Trinh, K., & Hammerschlag, R. (2002). Performing systematic reviews of clinical trials of acupuncture: Problems and solutions. *Clinical Acupuncture and Oriental Medicine, 3*, 26-31.

White, A. R. (2002). Acupuncture research methodology. In G. Lewith, W. Jonas, & H. Walach (Eds.), *Clinical research in complementary therapies: Principles, problems and solutions* (pp. 317). London: Churchill Livingstone.

White, A. R., Filshie, J., & Cummings, T. M. (2001). Clinical trials of acupuncture: Consensus recommendations for optimal treatment, sham controls and blinding. *Complementary Therapies in Medicine, 9*, 237-245.

Wittmann, W. W., & Walach, H. (2002). Evaluating complementary medicine: Lessons to be learned from evaluation research. In G. Lewith, W. B. Jonas, & H. Walach (Eds.), *Clinical research in complementary therapies: Principles, problems and solutions* (pp. 103-104). London: Churchill Livingstone.

Wolfe, F. (1995). Health status questionnaires. *Rheumatic Disease Clinics of North America, 21*, 445-464.

World Health Organisation. (1958). *The first ten years of the health organisation*. Geneva: World Health Organisation.

World Health Organization. (1985). *The role of traditional medicine in primary health care* (WPR/RC36/Technical Discussions). Manila: World Health Organization.

Wright, B. D., & Masters, G. (1982). *Rating scale analysis: Rash measurements*. Chicago: MESA.

Xi, Y. (1985). *Acupuncture and moxibustion techniques*. Shanghai, China: Shanghai Science and Technology Press (in Chinese).

Yamauchi, N. (1976). The results of therapeutic acupuncture in a pain clinic. *Arthritis Care Research, 23*(2), 196-206.

Yang, J. S., Chao, Y. M., Guang, J. D., Yuan, J. L., Gao, Y. Z., & Di, E. G. (1985). *Acupuncture points*. Shanghai, China: Shanghai Science and Technology Press (in Chinese).

Yates, B. T., Wagner, J. L., & Surprenant, L. M. (in press). Recall of health-risky behaviours for the prior two or four weeks via computerized versus printed questionnaire. *Computers and Human Behaviour*.

Ying, H. H., Zhang, B. N., Zhang, Z. Y., Zhang, X. C., Meng, X. M., Liu, Y. C., & Quan, C. H. (1995). *The foundations of traditional Chinese medicine*. Shanghai, China: Shanghai Science and Technology Press (in Chinese).

Yurtkuran, M., & Kocagil, T. (1999). TENS, electroacupuncture and ice massage: Comparison of treatment for osteoarthritis of the knee. *American Journal of Acupuncture, 27*, 133-140.

Zhang, J. B. (1986). *The complete book of Jing Yue (Jing Yue quan shu)*. Shanghai, China: Shanghai Science Publishing House. (Original work published in 1624).

Zhang, Y., Gao, X., & Teng, S. (2001). Water injection acupuncture combined with moxibustion for treatment in 30 cases of osteoarthritic knee. *Shanghai Journal of Acupuncture and Moxibustion, 20*(6), 26.

Zherebkin, V. (1998). The use of acupuncture reflex therapy in the combined treatment of osteoarthrosis patients. *Likars'ka Sprava, 8*, 149-151.

Zhu, J. P., & Rose, K. (2002). Chinese medicine and complexity. *Clinical Acupuncture and Oriental Medicine, 3*, 77-91.