

Original Article

Study on the consistency of doctor-report outcomes and patient-report outcomes in symptoms of Traditional Chinese Medicine: a small sample size trial in diabetics

Jun Sun^{1,2}, Wei-An Yuan³, Hao Lu⁴, Zhen Li³, Jian Jiang³

¹Department of Pharmacy, General Hospital of Jinan Military Command, Jinan, Shandong Province, China; ²Post-Doctoral Research Center, HuiRen Group, Nanchang, Jiangxi Province, China; ³Department of Clinical Pharmacology, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China; ⁴Department of Endocrinology, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China

Received January 26, 2015; Accepted April 23, 2015; Epub May 15, 2015; Published May 30, 2015

Abstract: Patient-Reported Outcomes (PROs) data are increasingly common and widely accepted in clinical investigations of symptoms. Agreements or differences regarding Doctor-Report Outcomes (DROs) VS PROs data are unclear. In this study, we conducted a clinical trial and investigated the agreement levels of DROs VS PROs in symptoms of diabetics. This study had a parallel self-controlled and double blind design. In total, 90 diabetics who presented with the required symptoms were enrolled in this study, and 83 patients completed the study. The severity (none, mild, moderate, and severe) of each symptom was consistently compared with the Visual Analogue Scale (VAS) for the PROs measure and the numerical rating scales (NRS) for the DROs measure. The Kappa index was used to test the agreement in the variables. Our results show that most of the agreements regarding DROs VS PROs were moderate; few were modest and substantial. However, the DROs data failed to agree with the PROs data well in which symptoms effective evaluation criteria overlapped with the inclusion criteria. We believe this results are caused by distortion of information during the doctor's decision process. We suggest that the effective evaluation criteria should not be overlapped with the inclusion criteria in the DROs design. And if it is unavoidable, PROs design should be worth considering.

Keywords: Agreement, symptom, trial, patient-report outcomes, kappa index

Introduction

Patients with type 2 diabetes mellitus (T2DM) have multiple T2DM-related, treatment-related, and non-T2DM-related symptoms. These symptoms frequently occur in clusters of three or more, presenting challenges for the assessment and subsequent treatment. In diagnosis and research of Traditional Chinese Medicine (TCM), symptoms are typically described by clinicians - a type of Doctor-Report Outcomes (DROs). This type of symptom assessment, particularly of subjective symptoms, is based on the doctor's understanding of the complaints from the patient. Patient-Reported Outcomes (PROs) was defined more than a decade ago by the US Food and Drug administration (FDA) as "a patient's report of a health condition and its treatment" [1]. PROs measures are frequently used to obtain self-reported health information, e.g., the symptoms, functional status, psy-

chological well-being, quality of life, preferences, perceptions, and satisfaction with care [2]. TCM includes a range of traditional medicine practices that originated in China, and always is adept in symptomatic treatment. Currently, PROs data are increasingly common and widely accepted in TCM clinical trials. Unfortunately, the agreement and difference of DROs VS PROs data are still uncertain.

This study was designed to investigate the agreement of DROs VS PROs in subjective symptom research. We hope it will help us to improve the design skill of future subjective symptom research.

Materials and methods

Participants

We recruited 90 outpatients with T2DM from the Shanghai area to participate in our clinical

The consistency of PROs and PROs data of symptoms

trial. Eligibility for participation required clinical evidence of T2DM, with presentation of the required symptoms. After informed consent was provided, patients who were willing to be randomized underwent an examination of their symptoms and screening process to determine whether they met the eligibility criteria of the study.

Ethical approval

This study was conducted in accordance with the declaration of Helsinki and the ethical regulations, policies and guidelines of the National Institutes of Health. The Shuguang Hospital medical ethics committee at Shanghai University of Traditional Chinese Medicine reviewed and approved this study (Approval No.: 2013-298-67-01). Written informed consent was obtained from the patients included in the study. The patients were informed that they were free to leave the study, without explanation and without any negative consequences to their future treatment. Every precaution was taken to protect the privacy of the research subjects and the confidentiality of their personal information. All the personal patient details were rendered anonymous before the data entry by referring to all the patient records and data only by their assigned research number. There are no known additional risks associated with patient participation in the study, other than the normal risks associated with these common treatments.

Diagnostic criteria of "Deficiency of Liver-yin and Kidney-yin"

The Yin-Yang theory originated from ancient Chinese philosophy, integrated into TCM theory later to help people understand how the human body works. The concept of TCM zang-organs (including heart, liver, spleen, lung, and kidney) is not equal to organs concept of modern medicine, but is similar to abstract term of human body functional entities. Liver-yin and kidney-yin are stipulated by TCM concept of zang-organs, mean one side of equilibrium state of human body function corresponding to zang-organs (Liver and kidney).

One type of TCM symptom cluster, "Deficiency of Liver-yin and Kidney-yin" is composed of the main symptoms as follows: dizziness, blurred vision, soreness and weakness of the waist or

knees (SWWK), dry eyes, dysphoria, fever in the chest palms and soles (DFCS); and minor symptoms: constipation, amnesia, disturbed sleep, tinnitus, and fatigue. Patients who have at least 3 of the main symptoms and 2 minor symptoms should be diagnosed with "Deficiency of Liver-yin and Kidney-yin."

Inclusion criteria

- Aged 40 to 75 years old, male or female.
- Meeting the TCM diagnosis criteria of "Deficiency of Liver-yin and Kidney-yin".
- At least one of the main symptoms reaches score of 3 (DROs).
- Being diagnosed with T2DM.
- Understanding the entire study process well.
- Volunteering to participate and signed informed consent forms.

Research design

We have carried out a clinical trial of the "investigation of Ganshenan granules treatment in deficiency of liver-yin and kidney-yin of type 2 diabetes patients" (clinical trial registration number: ChiCTR-TRC-14004179). This study was a parallel self-controlled and double-blind clinical study. Endocrinologists invited their patients to participate in the study and were responsible for managing patients with no specific instructions given (in particular, the participants were not aware of the primary objective of the trial). This study intended to investigate the agreement of DROs VS PROs before and after treatment in patients with the TCM symptom cluster of "Deficiency of Liver-yin and Kidney-yin".

The study began in March 2014 and continued for 8 months until October 2014. Individuals interested in our study was first screened by an interviewer. Potentially eligible persons were scheduled for a baseline assessment visit for confirmation of eligibility. After obtaining informed consent and confirmation of meeting inclusion criteria, 90 participants were randomized with 2:1 probability to one of two intervention arms: the Ganshenan granules treatment group; and the placebo granules treatment group. The approved indications of "Ganshenan

The consistency of PROs and PROs data of symptoms

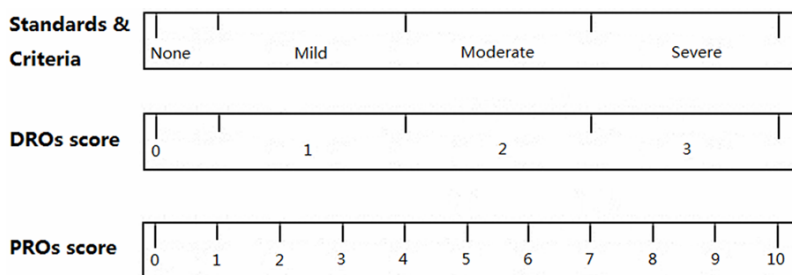


Figure 1. Symptom measure of the PROs VS DROs. The PROs score: At 0, 2, and 4 weeks of the trial, the patient self-assessment of the symptom severity was based on a self-administered symptom scoring, using an anchored or categorized VAS ranging from 0 (representing the best imaginable state) to 10 (the worst imaginable state). The DROs score: The investigator assessment of the symptom severity was based on a symptom NRS, ranging from 0 (representing the best state) to 3 (the worst state). The investigator assessments were performed synchronously at 0, 2, and 4 weeks with the patient self-assessment.

granules” are dizziness, blurred vision, soreness and weakness at the waist and knees, amnesia, disturbed sleep, tinnitus, and fatigue. The Ganshenan granules and placebo granules were provided by Hui ren Pharmaceutical Co., Ltd (Nanchang, China). The intervention period lasted 4 weeks in each arm. Ganshenan is a prescription in TCM and pharmacy to treat symptoms of deficiency of liver-yin and kidney-yin in china. This Chinese herbs formula was created in Ming dynasty and named as Yanshou pill, which was composed of cuscuta, black sesame, eclipta, mulberry, etc.

Outcome measures

The Guideline of “Traditional Chinese Medicine Clinical Research (2002)”, which was published by the China Food and Drug Administration was referred in this study [3]. Most of the TCM symptom cluster diagnosis and assessment were well defined and standardized by this guideline. In this study, the clinical criteria of symptom severity (none, mild, moderate, severe) was referred to the standards recommended in the guideline [3]. The method of symptoms measures was followed the “FDA guidelines for industry patient-reported outcome measures” [4]. The severity (none, mild, moderate, and severe) of every symptom was anchored consistently to the Visual Analogue Scale (VAS) for the PROs measure and numerical rating scale (NRS) for the DROs measure, as shown in **Figure 1**. The measures were performed at three time points - 0, 2, and 4 weeks after enrollment.

Statistical analysis

Statistical analysis was done by the SAS 9.2 software. The agreement of doctor assessment (NRS) and patient self-assessment (anchored or categorized VAS) were investigated by Cohen’s *Kappa* index [5]. The inter-group difference of *Kappa* index was analyzed by Wilcoxon scores (rank sums) test. *P* values of less than 0.05 were considered statistically significant.

Results

90 patients were enrolled in this study, and 83 of them finished it. The severity data was collected from all the 3 round symptoms measures (0, 2, 4 week) and divided into 4 levels: none, mild, moderate, and severe. The results of the agreement strength of the DROs VS PROs were summarized in **Table 1** after the statistical analysis. The statistics results of *Kappa* index indicated that most of the agreement strengths of the DROs VS PROs were moderate; few were modest and substantial. We failed to find any statistical difference between the main symptoms and minor symptoms in all the 3 round tests ($P>0.05$).

To mine the detailed information of the data, we reclassified the 1st-round symptoms measures data into 2 levels: non-severe and severe. The statistics results were show in **Table 2**. The results of the *Kappa* index showed that DROs VS PROs agreement strengths were between negligible and modest level for main symptoms; but moderate for minor symptoms. And the difference of DROs VS PROs agreement strength between main and minor symptoms was significant ($P=0.0331$).

Discussion

Unrelieved symptoms, such as those of diabetes, significantly affect the quality of patients life. Many patients are suffering from more than one symptom simultaneously. For symptom management or evaluation, symptom measure is the first and key step. In most cases,

The consistency of PROs and PROs data of symptoms

Table 1. The results of agreement strength of DROs VS PROs in symptoms evaluation (4 levels)

Symptoms	1st-round test (0 week)		2nd-round test (2 week)		3rd-round test (4 week)		
	Kappa index	95% CI	Kappa index	95% CI	Kappa index	95% CI	
main symptoms	Dizziness	0.542	[0.368, 0.715]	0.511	[0.327, 0.695]	0.597	[0.442, 0.751]
	Blurred vision	0.470	[0.326, 0.613]	0.665	[0.521, 0.809]	0.705	[0.562, 0.847]
	SWWK	0.155	[0.048, 0.262]	0.350	[0.189, 0.510]	0.510	[0.336, 0.685]
	Dry eyes	0.449	[0.294, 0.605]	0.531	[0.372, 0.690]	0.533	[0.367, 0.699]
	DFCS	0.509	[0.367, 0.652]	0.291	[0.133, 0.449]	0.380	[0.203, 0.557]
minor symptoms	Constipation	0.499	[0.339, 0.658]	1.000	[1.000, 1.000]	1.000	[1.000, 1.000]
	Amnesia	0.367	[0.197, 0.536]	0.479	[0.295, 0.662]	0.324	[0.104, 0.544]
	Disturbed sleep	0.570	[0.417, 0.723]	0.365	[0.196, 0.535]	0.544	[0.395, 0.693]
	Tinnitus	0.570	[0.430, 0.711]	0.422	[0.262, 0.583]	0.518	[0.369, 0.668]
	Fatigue	0.589	[0.429, 0.749]	0.467	[0.282, 0.653]	0.355	[0.155, 0.555]
Wilcoxon Scores (Pr > Z)		0.2403		1.000		0.6859	

Strength of agreement (Kappa index): Poor (below zero); Negligible (0-0.2); Modest (0.21-0.4); Moderate (0.41-0.6); Substantial (0.61-0.9); Almost perfect (0.81-1).

Table 2. The results of agreement strength of DROs VS PROs in symptoms evaluation (2 levels)

Symptoms	1st-round test (0 week)		
	Kappa index	95% CI	
main symptoms	Dizziness	0.322	[-0.053, 0.697]
	Blurred vision	0.332	[0.053, 0.611]
	SWWK	0.154	[0.048, 0.259]
	Dry eyes	0.183	[-0.186, 0.551]
	DFCS	0.119	[-0.166, 0.404]
minor symptoms	Constipation	0.491	[-0.109, 1.000]
	Amnesia	0.476	[0.041, 0.911]
	Disturbed sleep	0.491	[-0.109, 1.000]
	Tinnitus	0.423	[0.080, 0.766]
	Fatigue	0.654	[0.206, 1.000]
Wilcoxon Scores (Pr > Z)		0.0331	

researchers or clinicians accept patient complaints as the basis of symptom assessment. Most of our clinicians believe themselves are capable of doing impartial symptom assessment: collecting diseases information dispassionately, analyzing the diseases information in unbiased way, and drawing conclusions carefully. In contrast to that view, studies have shown that people are affected by a variety of factors when they make decisions, people are even unaware of that sometimes. What factors might make people distorting the information? The following two potential psychological mechanisms might account for information distortion: the desire to maintain consistency and the desire to reduce effort [6]. The desire to maintain consistency might involve ego-defensive mechanism and preference for a "consistent world". Holyoak and Simon [7] explicitly proposed that pre-decisional distortion of informa-

tion is driven by the goal of consistency. Russo J. Edward [8] reached similar conclusion that a consistent goal appears to drive information distortion.

The question is whether clinicians distort PROs information to insure that the information agrees with the inclusion criterion that, "At least one of the main symptoms scored 3 (DROs)". The results of this study (as shown in **Table 2**) strongly supported that assumption. After reviewing the raw data, we found that doctors, compared with patients, made much more "severe" judgments in the 1st-round symptoms measures (data not shown). The DROs data failed to fully agree with the PROs data in "the main symptoms" assessment because of the distortions of information in the doctor's decision process. In view of that finding, we suggested the following principle for the design of clinical research of symptoms that the effective evaluation criteria should not be overlapped with the inclusion criteria in the DROs design. If this situation is inevitable, the design of the final evaluation criteria should be based on the PROs data.

If the patients could provide complete symptom assessment with anchored or categorized VAS scale, should we consider that to be the gold standard of symptom assessment? The answer of this question is more complicated. Although the VAS appeared to be a very simple metric ruler, however it led to increasing dispute, including dispute on the following fields:

The consistency of PROs and PROs data of symptoms

the continuity of the data [9]; practical, metric and ontological limitations [10]; problematic correlation with medical treatment [11]. Even more it was suggested that raw pain VAS data should not be used as a primary outcome measure or to inform parametric-based randomized controlled trial power calculations in research studies [12].

It is obvious that patients are able to accurately recognize and timely collect subjective symptoms information. If we develop novel designs in symptom research, we could gain more detailed information from PROs design such as the duration of the symptoms, from which DROs design is unable to provide. In conclusion, PROs design can provide more reliable and detailed data in symptoms research compared with DROs design.

Acknowledgements

This study was financially supported by Jiangxi HuiRen Pharmaceutical Co., Limited and the National Key New Drugs Creation Project Innovative drug research and development platform (NO. 2012ZX09303009-001) of China.

Disclosure of conflict of interest

None.

Address correspondence to: Jian Jiang, Department of Clinical Pharmacology, Shuguang Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China. Tel: +86 2120256052; Fax: +86 2120256051; E-mail: 13817955085m@sina.cn

References

- [1] Acquadro C, Berzon R, Dubois D, Leidy NK, Marquis P, Revicki D and Rothman M. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health* 2003; 6: 522-531.
- [2] Martinez-Martin P, Rodriguez-Blazquez C and Frades-Payo B. Specific patient-reported outcome measures for Parkinson's disease: analysis and applications. *Expert Rev Pharmacoecon Outcomes Res* 2008; 8: 401-418.
- [3] CFDA. The Guideline of Traditional Chinese Medicine Clinical Research 2002; 233-236.
- [4] FDA. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009; 1-17.
- [5] Cohen J. A coefficient of agreement for nominal scales. *Psychological Bulletin* 1960; 20: 37-46.
- [6] Russo JE, Medvec VH and Meloy MG. The Distortion of Information during Decisions. *Organizational Behavior and Human Decision Processes* 1996; 66: 102-110.
- [7] Keith J, Holyoak DS. Bidirectional Reasoning in Decision Making by Constraint Satisfaction. *Journal of Experimental Psychology: General* 1999; 128: 3-31.
- [8] Russo JE, Carlson KA, Meloy MG and Yong K. The goal of consistency as a cause of information distortion. *J Exp Psychol Gen* 2008; 137: 456-470.
- [9] Kersten P, Kucukdeveci AA and Tennant A. The use of the Visual Analogue Scale (VAS) in rehabilitation outcomes. *J Rehabil Med* 2012; 44: 609-610.
- [10] Franchignoni F, Salaffi F and Tesio L. How should we use the visual analogue scale (VAS) in rehabilitation outcomes? I: How much of what? The seductive VAS numbers are not true measures. *J Rehabil Med* 2012; 44: 798-799; discussion 803-794.
- [11] Kelly AM. Patient satisfaction with pain management does not correlate with initial or discharge VAS pain score, verbal pain rating at discharge, or change in VAS score in the Emergency Department. *J Emerg Med* 2000; 19: 113-116.
- [12] Kersten P, White PJ and Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. *PLoS One* 2014; 9: e99485.