A Comparative Study of Telephone versus Onsite Completion of the WOMAC 3.0 Osteoarthritis Index

NICHOLAS BELLAMY, JANE CAMPBELL, JONATHAN HILL, and PHILIP BAND

ABSTRACT. ***Objective.*** Outcome assessment in clinical trials using the Western Ontario and McMaster University (WOMAC 3.0) Osteoarthritis Index is traditionally achieved through self-administration of the Index. However, in other areas of clinical measurement, telephone administration has been shown to be a reliable method of acquiring data that are both accurate and complete. To address this issue in knee osteoarthritis (OA), we conducted a comparative study of telephone administration by interviewer of WOMAC LK3.0 versus onsite self-completion at the hospital.

Methods. Fifty consenting patients with knee OA were randomized to complete the WOMAC LK3.0 Index by telephone interview one day, followed by onsite completion the following day, or vice versa. Neither patients nor interviewers had access to any prior scores.

Results. The mean age of the 50 patients was 66.3 years (range 44–82); 34 (68%) were female and 16 (32%) male. There was excellent agreement between the mean office and telephone scores, with mean differences for the WOMAC LK3.0 pain, stiffness, and function subscale scores and total score of 0.09, 0.12, 0.78, and 0.98, respectively. These differences were well within the respective protocol defined equivalence criteria of ± 1.7, ± 0.9, ± 6.4, and ± 9.1, and represented differences from office scores of 0.9, 2.6, 2.4, and 2.2%, respectively.

Conclusion. The use of telephone interviews for the WOMAC LK3.0 Index is a valid method of obtaining OA outcome measurements. These observations have important implications for designing data acquisition strategies for future OA clinical trials and for longterm observational studies. (J Rheumatol 2002;29:783–6)

Key Indexing Terms:
OUTCOME ASSESSMENT OSTEOARTHRITIS
WOMAC SURVEY TELEPHONE

The use of telephone contact to collect information on patient outcomes is well founded in the literature. This method has been used in several clinical studies and has been found to be a valid tool for patient outcome assessment[1-5]. However, searching under the MeSH headings "telephone," "survey," "VAS," "WOMAC," "osteoarthritis," and "outcomes," no published data could be found specifically validating the use of data collected by telephone in osteoarthritis (OA) outcome studies. The Western Ontario and McMaster University (WOMAC) Osteoarthritis Index has been in use for 19 years[6-8], but to date there has been no formal assessment of the Index comparing telephone administration by interviewer against the usual approach of self-completion onsite in the clinic or doctor's office. Verification of the validity of telephone data, in comparison to visit-collected data, would be useful for designing data acquisition strategies for future OA clinical trials, as well as for longterm longitudinal studies. This study was undertaken to validate telephone assessment of the WOMAC LK3.0 Index by comparing the results of telephone interviews versus onsite assessments.

From the Centre of National Research on Disability and Rehabilitation Medicine (CONROD), Brisbane, Australia; the University of Western Ontario, London, Ontario, Canada; and Matrix Biology Institute, Ridgefield, New Jersey, USA.

Sponsored by Biomatrix, Inc.

N. Bellamy, MD, MSc, FRCPC, FRCP (Glas and Edin), FACP, Professor, Director, Centre of National Research on Disability and Rehabilitation Medicine (CONROD), University of Queensland; J. Campbell, BA, CCRC, Department of Surgery, University of Western Ontario; J. Hill, BSc, Department of Microbiology and Immunology, University of Western Ontario; P. Band, PhD, Matrix Biology Institute.

Address reprint requests to Prof. N. Bellamy, Centre of National Research on Disability and Rehabilitation Medicine, University of Queensland, Department of Medicine, Level 3, Mayne Medical School, Herston Road, Brisbane, Queensland 4006, Australia.

Submitted February 15, 2001; revision accepted October 17, 2001.

MATERIALS AND METHODS

This was a single center, outpatient study, designed to compare patient responses to the WOMAC LK3.0 instrument administered by telephone and during an office visit. It was conducted as part of a larger study, comparing telephone vs office administration of 3 visual analog measures of pain and function (TPQ-VAS) and a 5 point Likert scaled patient global assessment (PGA) question whose performance we are evaluating. In each case, the WOMAC LK3.0 was the first questionnaire completed, since we wished to avoid reactivity on WOMAC LK3.0 scores. The WOMAC LK3.0 is a tridimensional joint targeted, patient centered questionnaire containing 5 pain, 2 stiffness, and 17 physical function questions, responses to which are scaled on 5 point (none, mild, moderate, severe, extreme) adjectival scales resulting in subscale score sizes of 0–20 for pain, 0–8 for stiffness, 0–68 for physical function, and 0–96 for total score. Male and female outpatients from the investigator's practice were initially contacted by one of the following methods: (1) by letter requesting their participation in the study; (2) by telephone call; or (3) during their visit to an outpatient clinic. Screening procedures included a review of the inclusion criteria. Patients meeting the following criteria were eligible for entry into the

study: over 35 years of age, inclusive; radiographic evidence of knee OA[9]; ability to understand English; and willingness to participate in the study according to the protocol. Because of the noninvasive nature of this study, there were no explicit exclusion criteria. Patients who wished to enroll received a blank copy of the WOMAC LK3.0 questionnaire for reference during the telephone interview. The first 50 patients from the investigator's patient population who consented to participate were selected for study. After enrolling in the study, patients were randomized to one of 2 groups. Group 1 completed questionnaires by telephone on Day 1, followed on Day 2 by questionnaire completion in the office. Group 2 completed questionnaires in the office on Day 1, followed by telephone completion on Day 2. All interviews, whether by telephone or in the office, were supervised by an evaluator trained in the administration and evaluation of the WOMAC LK3.0 Index. At the beginning of the study, patients were asked to select the knee with the worst OA symptoms. This was then designated their study knee. The WOMAC LK3.0 questionnaire contained written instructions for completion of the form. Patients were instructed to rate the severity of the pain, stiffness, and physical disability they had experienced in their study knee over the previous 4 weeks. The WOMAC LK3.0 was self-administered in the office. During the telephone assessment, the evaluator read each WOMAC LK3.0 question, and then recorded the patient's response on the blank questionnaire. Neither the patient nor the evaluator had access to any prior WOMAC LK3.0 scores.

Patients were free to withdraw from the study at any time. However, because of the short duration and noninvasive nature of the trial, no explicit provisions for removal of patients from assessment were included in the study protocol. This study did not involve drug administration or other invasive procedures, and patients were not requested to make any changes to their existing treatment regimen prior to or during the assessment period. Prior and concomitant therapies were not expected to have a material effect on the results because of the short retest interval, and therefore concomitant therapies were neither recorded nor controlled.

Statistical issues. As noted, this study was conducted within the framework of a larger study. As a consequence, sample size calculations were performed to estimate the minimum sample size required to establish with 95% confidence that TPQ-VAS telephone values did not differ by more than 15 mm from office administered values. Statistical analyses for equivalence were performed using analysis of variance (ANOVA) (alpha = 0.05) to establish the upper and lower limits of the relevant confidence interval[10]. Twenty percent has been used in the arthritis literature as a threshold value for declaring a clinically important difference[11], and earlier studies have established standard deviations for VAS measures of pain of about 15 mm[12,13]. Equivalence was to be inferred if the 95% confidence limits for the differences between the office and telephone scores were within ± 20% of the mean office scores. Based on these assumptions, a sample size of 50 patients was determined to be adequate for the required statistical power. This sample size was also considered adequate for evaluating similar issues for the WOMAC LK3.0 and the PGA (i.e., equivalence was to be inferred if the 95% confidence limits for the differences between the office and telephone scores were within ± 20% of the mean office scores).

After completion of the questionnaire stage of the study, data were entered from the patient questionnaires into a database and subjected to quality assurance procedures that were double verified, providing 100% verification for the key OA outcome measurements. There was no a priori reason to suspect that performing the assessment in the office first would influence the outcome of the subsequent telephone interview and vice versa. However, because of this, it was necessary to ensure that results from the office visit and the telephone interview were consistent, regardless of order of presentation.

For each outcome measure, ANOVA statistics were computed for the telephone and office scores, classifying them with respect to those obtained first and those obtained second. The first and second office scores and the first and second telephone scores were to be pooled if p values for a given outcome measure were greater than 0.05. Comparisons were made using the pooled error estimate obtained from the ANOVA. Scores for outcome measures that satisfied the pooling criterion were pooled and evaluated for equivalence as follows. WOMAC LK3.0 data were treated as continuous normally distributed data for the purposes of the analyses. Difference scores for each WOMAC LK3.0 component were calculated on a by-patient basis by subtracting the value determined on the telephone from the corresponding office visit value. These differences were then subjected to statistical analysis based on the principles advanced by Bland and Altman[10], and using the SAS program PROC Means.

RESULTS

An initial cohort of 50 patients was selected for the study. One patient, who did not have a knee radiograph on file to provide radiographic evidence of OA, was subsequently dropped from the study. One other patient, randomized to telephone assessment on Day 1, was assessed on Day 2 at her home instead of in the office and was also dropped from the study. Data from these patients were not included in the database. Two additional patients were recruited to achieve the desired study cohort of 50 patients. Of the final 50 patients, 22 were randomized to Group 1 (telephone first) and 28 were randomized to Group 2 (office first).

In violation of the protocol, 5 patients did not complete questionnaires on consecutive days. Four of these patients completed the second assessment on Day 3, rather than Day 2. One patient completed the second assessment on Day 5 instead of Day 2. However, because patients were rating the condition of their study knee over a 4 week timeframe, these deviations were not considered to be serious violations. Further, any effect of these more extensive intervals would be expected to decrease rather than increase observed levels of agreement. Therefore, the scores for these patients were included in the analysis.

The mean age of the 50 randomized patients was 66.3 years (range 44 to 82). Of these 50 patients, 34 (68%) were female and 16 (32%) male. The severity of the patients' OA ranged from Grade I to IV, as graded by Kellgren-Lawrence radiographs[9]: Grade I 20%, Grade II 38%, Grade III 40%, Grade IV 2%.

The significance of the effect of order (first or second) on assessment method (office or telephone interview) was analyzed by ANOVA. The results are provided in Table 1. The results indicate there were no statistically significant differences attributed to the order of questionnaire administration for any of the outcome measures. This means that office and telephone scores from those patients randomized to the telephone-first questionnaire were statistically equivalent to the corresponding office and telephone scores for patients in the office-first group. Therefore, the results from the 2 groups of patients were combined for the analysis of success criteria.

As described above, the WOMAC LK3.0 telephone and office visit methods were considered equivalent if the combined 2 sided 95% CI for the differences between methods were within ± 20% of the mean office scores. Although all patients were symptomatic, this study included patients with a wide range of OA symptoms. As a result,

Table 1. Analysis of sequence effects.

Outcome Measure	Mean Office Scores (SD)			Mean Telephone Scores (SD)		
	First, n = 28	Second, n = 22	p	First, n = 22	Second, n = 28	p
WOMAC Pain	9.11 (3.17)	8.16 (3.15)	0.424	8.15 (3.19)	8.96 (2.77)	0.491
WOMAC Stiffness	5.00 (1.70)	3.95 (1.68)	0.120	3.91 (1.72)	4.82 (1.72)	0.174
WOMAC Function	34.71 (11.46)	28.96 (12.37)	0.212	27.65 (11.53)	34.36 (10.52)	0.147
WOMAC Total	48.82 (14.70)	41.07 (16.41)	0.205	39.71 (15.52)	48.14 (14.05)	0.168

zero scores on some, but never on all outcome measures, did occur for the outcome measures of specific patients. The primary success criteria evaluated in this study were analyzed by subtracting the telephone score from the corresponding office visit score for each WOMAC component. The differences were then subjected to a statistical analysis using the SAS program PROC Means. The results of this analysis are provided in Table 2. There was excellent agreement between the mean office and telephone scores, with mean differences for the WOMAC LK3.0 pain, stiffness, function, and total scores of 0.09, 0.12, 0.78, and 0.98, respectively. These differences are also well within the protocol-defined equivalence criteria of ± 1.7, ± 0.9, ± 6.4, and ± 9.1, respectively, for pain, stiffness, physical function, and total WOMAC LK3.0 scores, and represent differences from office scores of 0.9, 2.6, 2.4, and 2.2%, respectively.

The exact time taken to complete the questionnaires was not measured. However, completion times estimated by evaluators were between 5 and 10 minutes.

DISCUSSION

Telephone contact to collect information on patient outcomes has been used extensively in clinical studies and has been found to be a valid tool for patient outcome assessment. The ability to use telephone contacts to obtain OA patient health assessments would be of great benefit to those patients for whom office visits can be both difficult and inconvenient.

This study was designed to validate the use of WOMAC LK3.0 telephone interviews as a measure of OA study outcome by comparing the results of telephone versus office visit assessments. The WOMAC LK3.0 demonstrated excellent agreement between the mean office and telephone scores, with mean differences for the WOMAC LK3.0 outcome measures ranging from 0.09 to 0.98. The 95% CI were well within established equivalence criteria of ± 20% of the mean office scores for the respective outcome measures. The telephone administered WOMAC can therefore be considered as validated by the criteria established in the protocol. This demonstration of clinical and statistical equivalence provides the first evidence that the use of telephone interviews using the WOMAC LK3.0 Index is a valid method of obtaining OA outcome measurements.

In recognizing the following potential limitations, it is acknowledged that the results of this study are directly generalizable only to individuals and groups having similar characteristics to this group of patients. We have not specifically addressed issues peculiar to the elderly, those not fluent in English, institutionalized patients, those not under medical care, or those of low education or socioeconomic status. Furthermore, the study population was not constituted to permit such an analysis. It is notable that Bombardier, *et al* have successfully administered the WOMAC Index to OA patients seen by family physicians, supporting the contention that this method of administration is feasible in a community based setting[14]. With respect to memory effects, we have experience with varying the timeframe of the WOMAC Index[15] and the interval between pain assessments[16,17], as well as performing repeated assessments

Table 2. WOMAC LK3.0 outcome measures: analysis of differences between telephone and office assessments (n = 50).

	WOMAC Pain	WOMAC Stiffness	WOMAC Function	WOMAC Total
Mean office scores (SD)	8.69 (3.16)	4.54 (1.75)	32.18 (12.09)	45.41 (15.80)
Mean telephone scores (SD)	8.61 (2.96)	4.42 (1.76)	31.41 (11.47)	44.43 (15.16)
Mean difference	0.09	0.12	0.78	0.98
SD, difference	1.83	0.92	3.43	4.05
Paired t, difference	0.33	0.92	1.60	1.71
Prob > T, difference	0.74	0.36	0.12	0.09
Lower 95%, difference	–0.44	–0.14	–0.20	–0.17
Upper 95%, difference	0.61	0.38	1.75	2.13
Protocol defined equivalence criteria ± 20% of mean office score	±1.74	±0.91	±6.44	±9.08

of health status on the same patients without time/patient interactions[18]. While the potential for memory effects deserves recognition, our experience in dissecting circadian rhythmicity in knee OA[18], hand OA[19], and rheumatoid arthritis[20] strongly suggests that patients can detect even small changes in symptom intensity, and memory effects are negligible or absent.

We did not collect quantitative data on patient and interviewer experience evaluations, or the costs and time of administration. The costs of conducting telephone interviews are mostly attributable to the costs of recruiting, training, and retaining interviewers and charges for telephone service and line usage. The former does not impose a high skill requirement, while the latter offers opportunities to negotiate favorable rates if long distance or high volume usage is contemplated. Offsetting this is the greater convenience and cost savings to the patient of not having to leave home, and the opportunity to collect community based rather than specialist based information.

It should be noted that the administration mode-dependent differences detected are very small and in magnitude fall below the values of published definitions of minimum perceptible clinical improvement[21], minimum clinically important difference[22], and responder criteria for OA clinical trials[23]. This suggests that the WOMAC Index is capable of detecting meaningful alterations in health status, when administered by telephone. It is worth reiterating that in this study patients were provided with a blank copy of the WOMAC Index for reference during the telephone interview.

These issues notwithstanding, we believe our findings can be applied to designing data acquisition strategies for future OA clinical trials and longterm observational studies. From a research and regulatory perspective, it facilitates the completeness and speed of data acquisition and transfer, and from the patient's perspective allows an accurate assessment of OA status without the inconvenience and physical demands associated with office visits.

REFERENCES

1. Smith PM, Illig SB, Fiedler RC, Hamilton BB, Ottenbacher KJ. Intermodal agreement of follow-up telephone function assessment using the Functional Independence Measure in patients with stroke. Arch Phys Med Rehabil 1996;77:431-5.
2. Korner-Bitensky N, Wood-Dauphine S. Barthel Index information elicited over the telephone. Is it reliable? Am J Phys Med Rehabil 1995;74:9-18.
3. Rohde P, Lewinsohn PM, Seeley JR. Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. Am J Psychiatry 1997;154:1593-8.
4. Balas EA, Jaffrey F, Kuperman GJ, et al. Electronic communication with patients. Evaluation of distance medicine technology. JAMA 1997;278:152-9.
5. McGrory BJ, Shinar AA, Freiberg AA, Harris WH. Enhancement of the value of hip questionnaires by telephone follow-up evaluation. J Arthroplasty 1997;12:340-3.
6. Bellamy N. Osteoarthritis — An evaluative index for clinical trials [MSc thesis]. Hamilton: McMaster University; 1982.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
8. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988;1:95-108.
9. Kellgren J, Lawrence J. Atlas of standard radiographs: The epidemiology of chronic rheumatism. Vol. 2. Oxford: Blackwell Scientific; 1963.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 8:307-10.
11. Bellamy N. Standardised procedures for outcome measurement and parameters for calculating sample size for antirheumatic drug studies. Musculoskeletal clinical metrology. Dordrecht: Kluwer Academic Publishers; 1993:214-24.
12. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. III. Setting the delta for clinical trials — results of a consensus development (Delphi) exercise. J Rheumatol 1992;19:451-7.
13. Felson D, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
14. Bombardier C, Maguire L, Li L, Maetzel A, Jansz G, The CHAP Team. Characteristics of patients with osteoarthritis seen by family physicians: results of a telephone survey in Ontario [abstract]. J Rheumatol 2001;28:1416.
15. Griffiths G, Bellamy N, Kean WF, Campbell J, Gerecz-Simon E. A study of the time frame dependency of responses to the WOMAC Osteoarthritis Index. Inflammopharmacology 1993;2:85-7.
16. Bellamy N, Goldsmith GH, Buchanan WW, Campbell J, Duku E. Prior score availability: Observations using the WOMAC Osteoarthritis Index. Br J Rheumatol 1991;30:150-1.
17. Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drug trials. I. Effects of standardization procedures on observer dependent outcome measures. J Rheumatol 1992;19:436-43.
18. Bellamy N, Sothern RB, Campbell J. Rhythmic variations in pain perception in osteoarthritis of the knee. J Rheumatol 1990; 17:364-72.
19. Bellamy N, Campbell J, Sothern R, Soucy E, Flynn J, Buchanan WW. Circadian rhythm in pain, stiffness and manual dexterity in osteoarthritis of the hand: relationship between discomfort and disability. International Congress on Chronobiology, Paris, September 7-11, 1997. Paris Chronobiology International 1997;24 Suppl 1:16.
20. Bellamy N, Sothern RB, Campbell J, Buchanan WW. Circadian rhythm in pain, stiffness and manual dexterity in rheumatoid arthritis: relation between discomfort and disability. Ann Rheum Dis 1991;50:243-8.
21. Ehrich EW, Davies GM, Watson DJ, Bolognese JA, Seidenberg BC, Bellamy N. Minimal perceptible clinical improvement with the Western Ontario and McMaster Universities Osteoarthritis Index questionnaire and global assessments in patients with osteoarthritis. J Rheumatol 2000;27:2635-41.
22. Bellamy N, Carr A, Dougados M, Shea B, Wells G. Towards a definition of "difference" in osteoarthritis. J Rheumatol 2001;28:427-30.
23. Dougados M, LeClaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Special article: Response criteria for clinical trials on osteoarthritis of the knee and hip: A report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials Response Criteria Initiative. Osteoarthritis Cartilage 2000;8:395-403.