Assessment of different instruments used as outcome measures in patients with fibromyalgia

Adriana Martins Barros Alves[1], Jamil Natour[2], Marcos Renato Assis[3], Daniel Feldman[2]

ABSTRACT

Objective: To assess the different measure instruments used for patients with fibromyalgia. **Patients and methods:** This study assessed 60 individuals participating in a clinical trial of cross-sectional cohort comparing the effects of exercises performed in water and on land. The following instruments were used: the Fibromyalgia Impact Questionnaire (FIQ) to assess the impact of the disease; the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36) to assess quality of life; the Beck Depression Inventory to assess depression; and the Visual Analogue Scale (VAS) of pain. Those questionnaires were compared with the results obtained in a transitional Likert-type scale, the verbal scale for assessing change (VSAC), considered as a criterion of change in the assessment of other instruments. **Results:** The Spearman coefficient was used to study the correlation between the VSAC measure and the other instruments at two occasions (T1 and T2). At T1, a moderate correlation was observed between VSAC and VAS (r = 0.49), and between VSAC and FIQ (r = 0.41), and a negative correlation was observed between VSAC and the SF-36 domains pain (r = −0.49) and general health perception (r = −0.55), and the SF-36 physical component (r = −0.42). At T2, only the SF-36 domain vitality showed a weak negative correlation with VSAC (r = −0.27). **Conclusion:** Considering VSAC as gold standard, none of the instruments assessed could optimally identify a change in the health status of patients with fibromyalgia.

Keywords: fibromyalgia, questionnaires, quality of life.

INTRODUCTION

Fibromyalgia (FM) is a syndrome characterized by chronic diffuse[1] pain of unknown, probably multifactorial etiology,[2–5] sleep disorders, fatigue, and mood swings.[6,7] Laboratory and imaging tests show neither tissue injury nor changes.[8] Thus, the intensity, the impact on the patients' quality of life, and the temporal or therapeutic intervention variations are subjective, difficult and imprecise.[9]

Measure instruments to quantify those clinical parameters should be carefully selected, because outcome measures should have adequate psychometric properties. To be significant, the instrument should be sensitive to changes and clinically measurable, in addition to having high reliability and validity. Other aspects of those instruments, such as applicability, practicality and clarity, are also important.[10] The only questionnaire specifically developed for FM, the Fibromyalgia Impact Questionnaire (FIQ), although validated only in a limited way,[11–16] is widely used in several countries. In 2009, a study,[17] estimating the minimal clinically important difference (MCID) for the FIQ, concluded that a 14% change in the FIQ total score is clinically relevant, reinforcing its use in research and clinical settings.

Selecting adequate outcome measures for evaluating changes due to interventions in FM in clinical trials on that disease is extremely difficult, because of the subjectivity and heterogeneity of the FM symptoms.[18] In addition, physiological, cognitive-verbal, and behavioral variables need to be

Received on 05/30/2011. Approved on 05/08/2012. The authors declare no conflict of interest. Ethics Committee: 0580/01.
Escola Paulista de Medicina, Universidade Federal de São Paulo – EPM/Unifesp.
1. Physiotherapist
2. Professor of Rheumatology, Escola Paulista de Medicina, Universidade Federal de São Paulo – Unifesp
3. Professor of Rheumatology, Medical School of Marília
Correspondence to: Jamil Natour. Disciplina de Reumatologia, Unifesp/EPM. Rua Botucatu, 740. CEP: 04023-900. São Paulo, SP, Brazil.
E-mail: jnatour@unifesp.br

investigated. In a review of 24 clinical trials involving patients with FM, a great diversity of parameters was used, but the evaluation criteria were not consistently used in any of the trials.[19] A more recent systematic review,[20] comparing the most used variables in clinical trials with the OMERACT (Outcome Measures in Rheumatoid Arthritis Clinical Trials) proceedings for FM,[21] has concluded that each OMERACT domain has an instrument that seems sensitive to change.

A consensus has not been achieved on the adequate gold standard for assessing clinical improvement resulting from different therapeutic interventions in FM, especially in the Brazilian population.[22] Considering that there are no objective measures identifying patients' improvement, subjective measures such as questionnaires assessing the quality of life, impact of the disease, and pain scales have been used. Regarding subjective symptoms, the patient's perception is extremely important, because it comprises the complex assessment of the multiple domains that affect the individual's biopsychosocial integrity.[23] The information provided by patients regarding their health condition should be considered the gold standard to guide their treatment.[24–27] Thus, this study aimed at assessing the correlation between measure instruments used in the FM treatment and the objective questionnaire provided by the patient, assuming that the latter is a more sensitive parameter.

PATIENTS AND METHODS

Sample

The patients of this study were recruited from a clinical trial assessing the effects of exercise performed in water and that performed on land by women diagnosed with FM. The 60 female patients participating in this study met the American College of Rheumatology (ACR) criteria for FM. They were systematically selected from the Rheumatology Outpatient Clinic of the Federal University of São Paulo (Unifesp). Of the 60 female patients included, only 51 completed all assessments, comprising the object of this analysis.

Procedures

The project was approved by the Ethics Committee of the Unifesp, as participants provided written informed consent. The patients selected were assessed by use of the following instruments: (a) the FIQ,[9] a questionnaire assessing the impact of the disease, and whose scoring is directly proportional to that impact – the higher the score, the worse the health condition. That questionnaire comprises 10 questions and quantifies functional disability, pain intensity, sleep disorders, anxiety, depression, and well-being over the past week; (b) the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36), an instrument validated for the Brazilian population,[28] assessing quality of life. The SF-36 is divided into two components: the physical component, which determines the patient's physical status by assessing the domains physical functioning, physical role, pain, and general health perceptions; and the mental component, which reveals the patient's psychoemotional status by assessing the domains vitality, emotional role, social functioning, and mental health. In the SF-36, a higher score indicates better health status; (c) the Beck Depression Inventory, comprising 21 questions that assess the patient's depression status – the higher the score, the worse the depression; and (d) the Visual Analogue Scale (VAS) of pain, according to which patients classify their pain in a 0–10 numeric scale, the highest score corresponding to the worst pain possible. The assessments were performed at the beginning of treatment (T0), at the eighth week (T1), and at the fifteenth week of treatment (T2).

Patient's perspective was assessed based on a transitional five-point Likert scale, the verbal scale for assessing change (VSAC), as follows: 1, significant improvement; 2, moderate improvement; 3, mild improvement; 4, no improvement; and 5, worsening. It served as reference (gold standard) for global perception of change.

All instruments were applied by an examiner blinded to the patient's therapeutic group.

Statistical analysis

The following statistical methods were used to assess this study's results: (a) descriptive statistics, to analyze demographic and clinical variables (mean and standard deviation); (b) Spearman coefficient, to assess the correlation between the scores of change of the different instruments used and VSAC. The values used in such comparisons originated from the difference between the scorings at T0 and T1 for each questionnaire and the VAS – that is, between the first (T0) and second assessment (T1). Then, those final values were compared with VSAC at T1. The VSAC at T2 was compared with the difference between the scorings at T1 and T2 for the questionnaires and the VAS.

In addition, (c) linear regression analysis was used to assess which of the measures better relates to the change perceived by patients at T1 and T2. In addition, (d) the magnitude of the effect of each instrument was calculated by dividing the mean of the baseline scores until the eighth week (T1) by the standard deviation of the baseline scores. This method was used to evaluate the intensity of change, indicating the MCID. The analyses

were performed according to a protocol – thus, the size of the sample used for the statistical calculations was the number of patients completing all measure instruments at all assessments.

RESULTS

Of the 60 female patients selected for the study, 51 responded to all instruments at all assessments. Figure 1 shows the classification score of the VSAC at T1 and T2. Table 1 shows the mean values and the respective standard deviations obtained in the FIQ, Beck Depression Inventory, VAS and SF-36 at T0, T1 and T2.

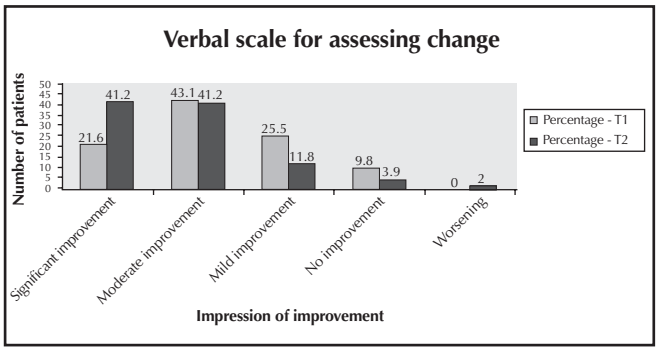

Figure 1
Performance of the verbal scale for assessing change at T1 and T2.

Table 1
Mean (standard deviation) of the FIQ, Beck Depression Inventory, VAS and SF-36 at T0, T1 and T2

	T0	T1	T2
FIQ	63.29 (13.86)	43.28 (19.36)	38.07 (19.46)
Beck	18.60 (9.11)	11.25 (10.25)	9.58 (9.49)
VAS	8.27 (1.55)	5.85 (2.32)	5.05 (2.42)
SF-36/PF	57.41 (21.28)	62.91 (25.59)	66.00 (29.52)
SF-36/PR	18.75 (30.05)	43.33 (41.90)	53.75 (45.29)
SF-36/PA	31.66 (15.94)	42.91 (21.50)	49.63 (27.48)
SF-36/VT	30.91 (18.67)	47.41 (23.17)	49.16 (28.24)
SF-36/GH	45.81 (19.64)	53.21 (25.34)	54.63 (28.36)
SF-36/SF	54.37 (30.77)	71.82 (33.54)	69.16 (37.42)
SF-36/ER	38.33 (41.54)	55.00 (42.88)	56.66 (45.22)
SF-36/MH	45.40 (22.29)	56.93 (26.90)	57.60 (30.80)
SF-36/PCOMP	35.08 (6.93)	41.16 (7.68)	44.72 (8.59)
SF-36/MCOMP	38.27 (12.84)	46.22 (11.85)	47.25 (12.62)

PF: physical functioning; PR: physical role; PA: pain; VT: vitality; GH: general health perception; SF: social functioning; ER: emotional role; MH: mental health; PCOMP: physical component; MCOMP: mental component.

The Spearman correlation coefficient revealed a significant correspondence of the VSAC with the following at T1: the pain VAS ($P < 0.001$); the FIQ; and SF-36 domains pain and general health perception; and the SF-36 physical component (Table 2). At T2, the only variable showing statistically significant correlation with VSAC was the SF-36 domain vitality ($P = 0.04$) (Table 3).

The simple linear regression analysis revealed, at T1, a statistically significant correlation of VSAC with the VAS for

Table 2
Correlation of the VSAC1 measure with the difference between the first and second assessments of FIQ, Beck Depression Inventory, VAS and SF-36 domains and components

VSAC1	r	P
FIQ	0.41	0.002**
Beck	0.32	0.02*
VAS	0.49	0.001**
PF/SF-36	–0.23	0.09
PR/SF-36	–0.28	0.04*
PA/SF-36	–0.49	0.001**
VT/SF-36	–0.29	0.03*
GH/SF-36	–0.55	0.001**
SF/SF-36	–0.26	0.06
ER/SF-36	–0.17	0.06
MH/SF-36	–0.31	0.02*
PCOMP/SF-36	–0.42	0.002**
MCOMP/SF-36	–0.25	0.06

**Significant correlation P > 0.01; *Significant correlation P > 0.05. r: Spearman correlation coefficient.

Table 3
Correlation of the VSAC2 measure with the difference between the second and third assessments of FIQ, Beck Depression Inventory, VAS and SF-36 domains and components

VSAC2	r	P
FIQ	0.07	0.62
Beck	–0.18	0.19
VAS	0.18	0.18
PF/SF-36	–0.1	0.47
PR/SF-36	–0.13	0.35
PA/SF-36	–0.17	0.23
VT/SF-36	–0.27	0.04*
GH/SF-36	–0.03	0.8
SF/SF-36	–0.76	0.59
ER/SF-36	–0.08	0.55
MH/SF-36	0.02	0.86
PCOMP/SF-36	–0.12	0.38
MCOMP/SF-36	–0.09	0.52

**Significant correlation P > 0.01; *Significant correlation P > 0.05. r: Spearman correlation coefficient.

pain (P = 0.001) and the SF-36 domain general health perception (P < 0.001) (Table 4). Table 4 also shows that, at T2, the only statistically significant variable was the SF-36 domain vitality (P = 0.023).

Based on the calculation of the magnitude of the effect (ME), VAS proved to be the most statistically significant parameter to measure change (ME = –1.60), followed by FIQ (ME = –1.44). Table 5 shows the other important statistically significant parameters to measure change.

Table 4
Linear regression analysis showing the correlation between VSAC1 and VSAC2 and the other instruments

VSAC1	B[a]	Standard error	Beta[o]	P
VAS	0.135	0.038	0.4	0.001
GH/SF-36	–2.27	0.006	–0.45	0.001
VSAC2	**B[a]**	**Standard error**	**Beta[o]**	**P**
VIT/SF-36	–1.87	0.008	–0.32	0.02

[a]Non-standardized correlation coefficient; [o]Standardized correlation coefficient.

Table 5
Magnitude of the effect (ME) calculated for all instruments

	Mean	Standard deviation	ME
VAS1	8.27	1.55	–1.6*
VAS2	5.78	2.32	0.31
PF1/SF-36	57.05	20.42	0.57
PF2/SF-36	68.72	17.91	0.4
PR1/SF-36	21.07	31.76	0.92*
PR2/SF-36	50.49	41.37	0.29
PA1/SF-36	33.11	15.67	0.95*
PA2/SF-36	48.01	16.69	0.56
GH1/SF-36	48.13	18.74	0.55
GH2/SF-36	58.54	20.66	0.19
VIT1/SF-36	31.96	18.74	1.09*
VIT2/SF-36	52.25	18.68	0.27
SF1/SF-36	56.37	30.85	0.73
SF2/SF-36	79.16	25.33	0.02
ER1/SF-36	41.83	42.6	0.41
ER2/SF-36	59.47	42.32	0.15
MH1/SF-36	49.33	21.08	0.62
MH2/SF-36	62.5	22.25	0.18
FIQ1	63.29	13.86	–1.44*
FIQ2	43.28	19.36	–0.26
Beck1	18.6	9.11	–0.8
Beck2	11.25	10.25	0.16

*Statistically significant values (ME > 0.8).

DISCUSSION

Studies on FM have used different measures and instruments to assess therapeutic performance, hindering the attempts of extrapolation or comparison between treatments. At the same time, the large number of parameters studied makes the assessments exhausting and redundant.

This study compared the results of the measuring instruments FIQ, Beck Depression Inventory, SF-36 and pain VAS with the five-point Likert-like comparative transitional scale, the VSAC perceived by the patient and used as gold standard. Global classifications of symptom change provided by patients are considered a valid external criterion[23–27] and have been recently applied to populations with FM.[21,29,30] While completing the VSAC, the patient translated her impression of improvement as a general aspect; the relationship of that answer with other instruments could indicate which aspects influenced the impression of improvement. At the first assessment, the VSAC correlated with the SF-36 domains of pain and general health perception, the SF-36 physical component, in addition to the FIQ and the pain VAS.

Dunkl et al.[25] have reported results similar to ours, with correlation between the VSAC and FIQ. In our study, although to a lesser extent, a correlation was also observed with the SF-36 domains physical role, mental health, and vitality, and the Beck Depression Inventory. The result of the regression analysis confirms the SF-36 domain general health perception and the pain VAS as important variables. Thus, the change in pain intensity, in general health status and in physical well-being are fundamental aspects for the patient's impression of improvement, especially in the initial treatment. In the second assessment provided by the patient, only the SF-36 domain vitality proved to have any relationship with the subjectivity of improvement. This can show that, after the initial improvement, aspects other than pain and general health status begin to have more influence on the subjective impression of improvement. Then, vitality begins to have more importance.

One possible explanation for the difference found in the relationship between the instruments in the first and second data comparison would be the intensity of changes. Thus, patients would only perceive changes in the aspects they consider important to contribute to the impression of improvement if that had a greater intensity. That statement is based on the fact that, most of the instruments used showed an improvement, sometimes small, not only between the first and second assessments, but also between the second and third assessments. The time interval between the application of the instruments and the minimum difference necessary

to identify changes might also have influenced the results. According to Stratford,[31] the lack of a gold standard for attributes such as functional disability generates methodological dilemma. Beaton[27] has reported that , in addition to the already established psychometric properties, the challenge of interpretability should also be faced. Thus, using the MCID is required. For determining MCID, not only the patient's and physician's perspectives should be considered, but also the methodological approach and the patient's health status at the beginning of treatment. However, a greater number of methodological studies are required to determine the best way to quantify MCID. Regarding FIQ, a study[29] has concluded that a 14% change in the final score would determine a MCID. Specially for FM, that can be of great clinical value, considering the subjectivity and variability of the symptoms.

In the present study, the ME was calculated for each instrument as a way to determine MCID. On the first assessment, the ME was clinically important for the SF-36 domains vitality, pain and physical role, and the Beck Depression Inventory, in addition to the FIQ and pain VAS. None of these were clinically important on the second assessment. Thus, the validity of the numerical interpretation of the ME in FM should be questioned, because that effect not always represents a true MCID.

Based on our data, pain remained as a central aspect for the impression of change on the health status. Pain is basically a subjective symptom, associated with the interaction of the physical, psychic and cultural dimensions involved in its manifestation, what makes its measurement difficult. However, our study revealed that, when monitoring the patient in a clinical setting, the use of a VSAC proved to suffice. In clinical trials, other instruments can be used depending on the need for specific data in different aspects of the spectrum of the patient/disease relationship. It is worth noting that each instrument assesses an individual's different dimension, which might explain the lack of more correlation between the instruments, generating the need to choose not only an instrument to assess the therapeutic response, but, depending on the objective, to select the most adequate instrument.

It is worth emphasizing that, although the several instruments, such as pain VAS, FIQ, SF-36 and Beck Depression Inventory, showed a ME over 0.8, indicating effective magnitude, none of them could detect changes in the patients' health status at T2 as compared with the VSAC. Thus, the psychometric properties of the instruments are not ideal for FM. Wolfe[32] has proposed a version of the Health Assessment Questionnaire (FHAQ), which should be further studied and validated to be used in FM protocols.

CONCLUSION

Considering the VSAC as gold standard, none of the instruments assessed was able to ideally capture a change in the FM patient's health status. We emphasize the importance of assessing the psychometric properties of those instruments, in addition to studying the use of other instruments in clinical trials involving patients with FM.

REFERENCES

1. Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: Report of multicentre criteria committee. Arthritis Rheum 1990; 33:160–72.
2. Gibson SJ, Littlejohn GO, Gorman MM, Helme RD, Granges G. Altered heat pain thresholds and cerebral event-related potentials follow CO2 laser stimulation in subjects with fibromyalgia syndrome. Pain 1994; 58:185–93.
3. Crofford LJ, Pillemer SR, Kalogeras KT. Hypothalamic-pituitary adrenal axis perturbations in patients with fibromyalgia. Rev Rheum Mal Osteartic 1992; 59:497–500.
4. Perlis ML, Giles DE, Bootzin RR, Dikman ZV, Fleming GM, Drummond SP. Alpha-sleep and information processing, perception of sleep, pain, and arousability in fibromyalgia. Int J Neurosci 1997; 89:265–80.
5. Branco J, Atalaia A, Paiva T. Sleep cycles and alpha-delta sleep in fibromyalgia syndrome. J Rheumatol 1994; 21:1132–6.
6. Yunus MB. Psychological aspects of fibromyalgia syndrome: a component of the dysfunctional spectrum syndrome. Bailliers Clin Rheum 1992; 8:90–4.
7. Kurtze N, Gundersen KT, Svebak S. The role of anxiety and depression in fatigue and patterns of pain among subgroups of fibromyalgia patients. British L Medical Psychol 1998; 71:185–94.
8. Russell IJ. Fibromyalgia Syndromes. Phys Med Rehabil Clin North Am 1997; 8:213–26.
9. Buckhardt CS, Clark SR, Bennett RM. The Fibromyalgia Impact questionnaire: development and validation. J Rheumatol 1991; 18:728–33.
10. Wright J, Young N. A comparison of different indices of responsiveness. J Clin Epidemiol 1997; 50:239–46.
11. Hedin PJ, Hamme H, Burckhardt CS. The Fibromyalgia Impact Questionnaire, a Swedish translation of a new tool for evaluation of the fibromyalgia patient. Scan J Rheumatol 1995; 24:69–75.
12. Buskila D, Neumann L. Assessing functional disability and health status of women with fibromyalgia: validation of a Hebrew version of the Fibromyalgia Impact Questionnaire. J Rheumatol 1996; 23:903–6.
13. Kim YA, Lee SS. Validation of a Korean version of the Fibromyalgia Impact Questionnaire. J Korean Med Sci 2002; 17(2):220–4.
14. Offenbaecher M, Waltz M. Validation of a German version of the Fibromyalgia Impact Questionnaire. J Rheumatol 2000; 27:1984–8.

15. Zijlstra TR, Taal E. Validation of a Dutch translation of the fibromyalgia impact questionnaire. *Rheumatology* 2007; 46 (1):131–4.
16. Perrot S, Dumont, D. Quality of life in women with fibromyalgia syndrome: validation of the QIF, the French version of the fibromyalgia impact questionnaire. J Rheumatol 2003; 30(5):1054–9.
17. 17. Bennett R, Bushmakin A. Minimal clinically important difference in Fibromyalgia impact questionnaire. J Rheumatol 2009; 36(6):1304–11.
18. Hewett JE, Buckelew SP, Johnson JC, Shaw SE, Huyser B, Fu YZ. Selection of measures suitable for evaluating changes in fibromyalgia clinical trials. J Rhematol 1995; 22:2307–12.
19. White KP, Harth M. An analytical review of 24 controlled clinical trials for fibromyalgia syndrome (FMS). Pain 1996; 64:211–9.
20. Carville S, Choy E. Systematic review of discriminating power of outcome measures used in clinical trials of fibromyalgia. J Rheumatol 2008; 35(11):2094–105.
21. Mease P, Arnold L. Fibromyalgia syndrome module in OMERACT 9: domain construct. J Rheumatol 2009; 36(11):2318–29.
22. Heyman RE, Helfeinstein M, Feldman D. A double-blind, randomized, controlled study of amitriptyline, nortriptyline and placebo in patients with fibromyalgia. An analysis of outcome measures. Clin and Exp Rheumatol 2002; 19: 353–7.
23. Bellamy N, Kenneth M, Brooks P, Barraclough D, Tellus M. A survey of outcome measures in routine rheumatology outpatient practice in Australia. J Rheumatol 1999; 26:1593–9.
24. Husted J, Gladman D, Cook R, Farewell V. Responsiveness of health status instruments to changes in articular status and perceived health in patients with psoriatic arthritis. J Rheumatol 1998; 25:2146–55.
25. Dunkl P, Taylor G, McConnell G, Alfano A, Conaway M. Responsiveness of Fibromyalgia clinical outcomes measures. J Rheumatol 2000; 27:2683–91.
26. Wright J, Young N. A comparison of different indices of responsiveness. J Clin Epidemiol 1997; 50:239–46.
27. Beaton D, Hogg-Johnson S, Bombadier C. Evaluating changes in health status:reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol 1997; 50:79–93.
28. Ciconelli RM, Ferraz MB, Santos W, Meinao I, Quaresma MR. Tradução para a língua portuguesa e validação do questionário genérico de avaliação de qualidade de vida SF-36 . Rev Bras Reumatol 1999; 39(3):143–50.
29. Hudson J, Arnold L. What makes patients with fibromyalgia feel better Correlations Between Patient Global Impression of Improvement and Changes in Clinical Symptoms and Function: A Pooled Analysis of 4 Randomized Placebo-controlled Trials of Duloxetine. J Rheumatol 2009; 36(11):2517–22.
30. Assis MR, Alves A, Natour J. A randomized controlled trial of deep water running: clinical effectiveness of aquatic exercise to treat fibromyalgia. Arthritis Rheum 2006; 55(1):57–65.
31. Stratford P, Binkley J. comparison study of the back pain scale and Roland Morris questionnaire. J Rheumatol 2000; 27:1928-35.
32. Wolfe F, Hawley D, Goldenberg D, Russel I, Buskila D, Neumann L. The assessment of functional impairment in fibromyalgia (FM): Rasch analyses of 5 functional scales and the development of the FM health assessment questionnaire. J Rheumatol 2000; 27: 1989–99.