

EXTENDED REPORT

How reliably do rheumatologists measure shoulder movement?

J L Hoving, R Buchbinder, S Green, A Forbes, N Bellamy, C Brand, R Buchanan, S Hall, M Patrick, P Ryan, A Stockman

Ann Rheum Dis 2002;**61**:612–616

See end of article for authors' affiliations

Correspondence to:
Dr J L Hoving, Department of Clinical Epidemiology, Cabrini Hospital and Monash, University Department of Epidemiology and Preventive Medicine, Suite 41 Cabrini Medical Centre, 183 Wattleree Rd, Malvern, Victoria 3144, Australia;
Jan.Hoving@med.monash.edu.au

Accepted
11 January 2002

Objective: To assess the intrarater and interrater reliability among rheumatologists of a standardised protocol for measurement of shoulder movements using a gravity inclinometer.

Methods: After instruction, six rheumatologists independently assessed eight movements of the shoulder, including total and glenohumeral flexion, total and glenohumeral abduction, external rotation in neutral and in abduction, internal rotation in abduction and hand behind back, in random order in six patients with shoulder pain and stiffness according to a 6×6 Latin square design using a standardised protocol. These assessments were then repeated. Analysis of variance was used to partition total variability into components of variance in order to calculate intraclass correlation coefficients (ICCs).

Results: The intrarater and interrater reliability of different shoulder movements varied widely. The movement of hand behind back and total shoulder flexion yielded the highest ICC scores for both intrarater reliability (0.91 and 0.83, respectively) and interrater reliability (0.80 and 0.72, respectively). Low ICC scores were found for the movements of glenohumeral abduction, external rotation in abduction, and internal rotation in abduction (intrarater ICCs 0.35, 0.43, and 0.32, respectively), and external rotation in neutral, external rotation in abduction, and internal rotation in abduction (interrater ICCs 0.29, 0.11, and 0.06, respectively).

Conclusions: The measurement of shoulder movements using a standardised protocol by rheumatologists produced variable intrarater and interrater reliability. Reasonable reliability was obtained only for the movement of hand behind back and total shoulder flexion.

Shoulder pain is common in the general population, its point prevalence averages between 7% and 51% and it is known to increase with age. Restricted range of motion and shoulder pain can interfere with activities in daily life and is associated with work absenteeism and use of medical services.^{1–5} Many patients receive some evaluation by a family doctor, rheumatologist, orthopaedic specialist, or physical therapist.³

A physical examination is often used for both diagnosis and evaluation of treatment success in patients with shoulder pain. One aspect of physical assessment of the shoulder is the evaluation of range of motion. No “gold standard” for the measurement of shoulder range of motion is yet available. Clinical trials that have assessed the efficacy of interventions for shoulder pain have commonly used range of motion of the shoulder as a measurement tool.⁶ To be of value in clinical trials or routine care its reliability (that is, the repeated administration of an instrument to a stable population yielding the same results) should be established.

Multiple methods for estimating shoulder range of motion have been used in the past, including visual estimation, the two armed goniometer, or a gravity referenced goniometer.^{7–16} In many of these studies the methods are poorly described and most looked at passive range of motion only, which may reflect a less functional outcome.¹⁴ The results may not be applicable to symptomatic patients with varying degrees of pain and stiffness. Only a few studies have included subjects with shoulder pain.^{7, 8, 13–16}

We previously developed a standardised protocol, which we assessed in a similar group, for the measurement of active range of motion in clinical trials for shoulder pain.¹³ The intrarater and interrater reliability of our standardised protocol was found to be acceptable when physiotherapists performed the range of motion assessments. We concluded that it would be appropriate for use both in research and clinical practice by

physiotherapists. The gravity inclinometer chosen is fast and easy to use but has not been evaluated in observers with a different professional background.

To determine whether our protocol may also be of value when rheumatologists perform the measurements, this study aimed at examining the interrater reliability (that is reliability when performed by multiple raters) and intrarater reliability (that is reliability when performed again and again by the same rater or examiner) of the standardised protocol when performed by six rheumatologists. Although this study replicates a previous study, which looked at different subjects with similar complaints, differences in professional training and daily use of manual examination techniques between physiotherapists and rheumatologists might influence the reliability of the measurements. For this reason, we compared the results with those obtained by six physiotherapists.

METHODS

The standardised protocol for the measurement of shoulder movement with the Plurimeter-V gravity inclinometer (Dr Rippstein, Zurich, Switzerland) was described in detail previously.¹³ To minimise potential sources of variation, for each movement, we specified the position of the patient, positioning of the instrument on the patient, stabilisation of the joint to be measured, and determination of the end point of movement. The gravity referenced inclinometer was used to measure in total eight movements: total shoulder flexion; glenohumeral flexion; total and glenohumeral abduction; internal rotation; and external rotation in both neutral and abduction. Internal rotation was performed in abduction only. Hand behind back, assessed by palpation to the nearest spinal process, was also included.

Table 1 Summary statistics for each movement measured in six patients in rounds 1 and 2 by the six rheumatologists

	Round	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
TSF	1	102.7	14.2	88.8	18.9	136.5	12.1	131.7	21.7	66.7	22.3	131.3	5.0
	2	89.2	8.5	86.5	11.1	134.7	12.2	136.0	13.7	80.5	29.4	133.0	12.4
GHF	1	71.0	7.0	68.5	10.5	100.3	13.1	97.3	18.4	57.7	12.2	111.5	36.6
	2	82.3	16.4	61.3	12.2	95.66	21.0	123.3	16.0	45.5	13.1	91.0	14.8
TSA	1	68.2	17.8	75.3	14.6	118.2	29.7	99.7	15.6	77.5	16.2	85.0	29.2
	2	74.8	14.8	81.2	21.0	127.0	18.3	140.0	29.7	73.3	25.7	85.0	31.6
GHA	1	46.7	15.4	54.0	14.2	52.8	16.0	74.6	27.0	37.7	11.7	57.3	10.0
	2	67.0	6.4	53.5	14.3	82.3	21.1	107.7	28.2	49.7	25.3	62.7	15.3
ERN	1	40.0	16.6	66.3	19.7	31.8	9.8	46.6	23.1	46.2	12.4	39.8	24.1
	2	38.2	20.4	66.7	10.6	35.0	13.7	42.2	21.3	52.3	16.2	35.3	18.1
ERA	1	35.7	14.7	40.5	23.1	32.7	19.8	34.3	21.1	62.5	41.5	35.2	25.9
	2	41.0	17.7	54.5	26.7	42.3	8.1	46.8	26.7	49.5	26.8	21.3	10.6
IRA	1	54.3	16.5	65.0	21.1	64.8	24.9	70.0	31.9	78.5	21.4	65.3	22.1
	2	54.7	22.7	58.7	27.1	66.2	16.6	80.7	23.0	60.5	25.7	57.8	26.1
HBB	1	11.0	8.0–15.0	7.0	5.0–8.0	9.0	4.0–11.0	9.5	3.0–12.0	10.0	9.0–14.0	19.0	17.0–19.0
	2	11.0	10.0–13.0	6.5	4.0–9.0	7.0	4.0–11.0	8.0	6.0–10.0	11.0	7.0–13.0	19.0	16.0–19.0

TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back; SD, standard deviation. For HBB median and range are presented instead of mean and SD.

Six patients with varying degrees of pain and stiffness in the shoulder, and varying diseases, were recruited from a private rheumatology practice. Subjects were excluded if repeated testing was likely to aggravate the condition.

Six rheumatologists undertook a one hour training session to familiarise themselves with the process and standardise the measurement methods between the examiners. All were given a detailed manual. All had experience in measuring shoulder range of motion, but none had used an inclinometer. The examiners in the study were instructed to demonstrate each movement to the patient and ask the patient to move the arm until the first sensation of pain so that each examiner agreed on the definition of the first onset of pain. Care was taken in aligning the inclinometer on the patient's arm, ensuring consistent positioning of adjacent joints. The participant baseline data collected included age, sex, duration of the shoulder pain, and a history of trauma or shoulder surgery. In addition, the patients rated their severity on an ordinal five point scale (1, not at all severe; 5, maximum severity).

Before the measurements the patients were taken through a series of warm up exercises to reduce the risk of a mobilisation effect from the repeated movement assessment. According to a 6x6 Latin square design each patient was examined twice by the same examiner (in round 1 and round 2). Each examiner spent about five minutes with each patient in each

round. Between rounds 1 and 2 there was a one hour break for both patients and examiners. The order of the examiners and the order of the eight movements were randomly assigned for each round separately. In addition, the examiners were in separate rooms, and had no contact with the other examiners during the measurement procedures. Thus the examiners were unaware of each other's scores and their own scores of the previous round.

Statistical analyses

Analysis of variance was used to partition the total variability into components of variation due to patients, examiners, round, two-way interactions of individual components, and residual error (using the SAS VARCOMP procedure).^{17,18} All sources were considered as random effects.

Results are presented as the total variability and the percentage of the total variability attributable to each variance component. Intrarater and interrater reliability were calculated from the variance (var) components through appropriate intraclass correlation coefficients (ICCs)¹⁹ according to the following formulae:

$$\text{Intrarater reliability} = (\text{var}(\text{patient}) + \text{var}(\text{observer}) + \text{var}(\text{patient} \times \text{observer})) / \text{var}(\text{total})$$

Table 2 Interrater and intrarater reliability of rheumatologists

	Intrarater ICC (95% CI)*	Interrater ICC (95% CI)	Interrater ICC	
			Round 1	Round 2
Total shoulder flexion	0.83 (0.52 to 0.94)	0.72 (0.33 to 0.93)	0.72	0.73
Glenohumeral flexion	0.62 (0.18 to 0.83)	0.65 (0.32 to 0.86)	0.56	0.72
Total shoulder abduction	0.56 (0.21 to 0.73)	0.49 (0.03 to 0.80)	0.32	0.56
Glenohumeral abduction	0.35 (0.04 to 0.65)	0.51 (0.22 to 0.77)	0.28	0.51
External rotation in neutral	0.75 (0.42 to 0.89)	0.29 (0.02 to 0.68)	0.24	0.30
External rotation in abduction	0.43 (0.12 to 0.80)	0.11 (0.02 to 0.58)	0.06	0.19
Internal rotation in abduction	0.32 (0.02 to 0.67)	0.06 (0.01 to 0.45)	0.00	0.02
Hand behind back	0.91 (0.41 to 0.97)	0.80 (0.19 to 0.95)	0.80	0.80

ICC, intraclass correlation coefficient; 95% CI, 95% confidence interval.

Table 3 Percentage of total variance attributable to patient, observer, round, two-way interactions, residual error and total variance for shoulder movements

	Patient (%)	Observer (%)	Round (%)	Patient by round (%)	Round by observer (%)	Patient by observer (%)	Residual error (%)	Total variance
TSF	71	2	0	1	0	10	16	1007.6
GHF	52	4	0	13	4	7	21	867.6
TSA	41	14	3	5	0	0	37	1138.1
GHA	30	5	17	5	0	0	44	743.7
ERN	29	38	0	0	4	8	21	446.8
ERA	8	36	0	3	0	0	54	667.4
IRA	6	26	0	0	0	0	68	640.2
HBB	79	1	0	0	0	10	9	21.7

TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back.

$$\text{Interrater reliability} = (\text{var}(\text{patient}) + \text{var}(\text{round}) + \text{var}(\text{patient} \times \text{round})) / \text{var}(\text{total})$$

The closer the ICC is to 1, the better the reliability. Statistical uncertainty in the computed ICCs is expressed through approximate 95% confidence intervals obtained from 5000 bootstrap resamples of both subjects and rheumatologists.

Interexaminer reliabilities were also calculated for each round separately:

$$\text{Interrater reliability} = \text{var}(\text{patient}) / \text{var}(\text{total})$$

RESULTS

Six patients (two male, four female), with a mean age of 72 years (range 54–82), took part in the study. The right shoulder was assessed in four patients and the left shoulder in two. The mean duration of symptoms was 16 months (range 1–48). Two patients had a history of shoulder trauma and one patient had had previous shoulder surgery. The severity rating by each patient ranged from 2 (mild severity) to 3 (moderate severity) on the ordinal five point scale.

Table 1 shows the means and standard deviations for each of the six patients in rounds 1 and 2, combining the scores of the six rheumatologists for each movement. The range of motion of patients varied considerably. For example, the mean total shoulder flexion in round 1 varied between 66.7° (SD 22.3) and 136.5° (12.1). The differences between rounds 1 and 2 were smaller.

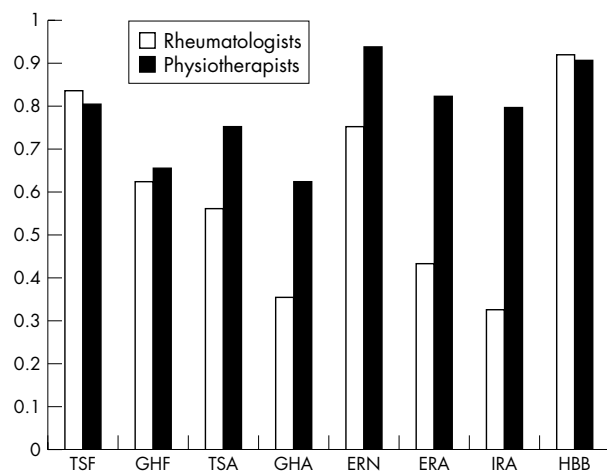


Figure 1 Intrarater reliability of rheumatologists and physiotherapists for each movement. TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back.

Table 2 shows the intrarater and interrater ICCs calculated for each of the eight shoulder movements together with their approximate 95% confidence intervals. The movement of hand behind back and total shoulder flexion gave the highest ICC scores for both intrarater reliability (0.91 and 0.83, respectively) and interrater reliability (0.80 and 0.72, respectively). Low ICC scores were found for the movements of glenohumeral abduction, external rotation in abduction, and internal rotation in abduction (intrarater ICCs 0.35, 0.43, and 0.32, respectively); and external rotation in neutral, external rotation in abduction, and internal rotation in abduction (interrater ICCs 0.29, 0.11 and 0.06, respectively).

Table 3 gives the individual components of variation as a percentage of the total variance. The proportion of the variance between examiners for the movements of external rotation in neutral, external rotation in abduction, and internal rotation in abduction was considerably greater than that of the explained variance of the patients, thus yielding very low interrater ICCs (see respective formula) for these movements.

In a previous study, we evaluated the reliability of shoulder measurement between physiotherapists using the same design and a similar patient group with shoulder pain.¹³ Figures 1 and 2 display graphically the ICCs of the rheumatologists compared with those of the physiotherapists. Overall, the physiotherapists achieved higher intrarater and interrater reliability ratings than rheumatologists, especially for the movements of external rotation and internal rotation.

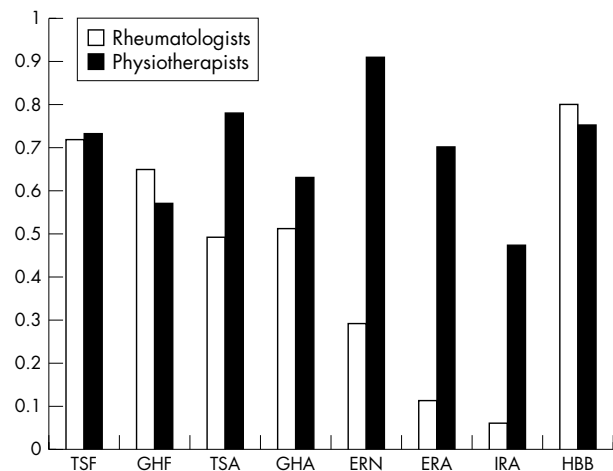


Figure 2 Interrater reliability of rheumatologists and physiotherapists for each movement. TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back.

DISCUSSION

In this study we found that using a standardised protocol for the measurement of shoulder movements produced variable intrarater and interrater reliability. Reasonable reliability (ICC > 0.7) was obtained for the movement of hand behind back and total shoulder flexion only. The other movements in our protocol produced fair to poor reliability, especially for those movements demanding more complex stabilisation and handling skills, such as external rotation and internal rotation. There was a learning effect as demonstrated by the increase in interrater reliability in round 2, though we did not formally test for statistical significance of this apparent increase in reliability. However, we think that the reliability can be improved with practice.

As far as we know, no previous published studies have investigated the differences in reliability between different groups of raters for measuring shoulder movements. Physiotherapists and rheumatologists may differ in their use of manual techniques, especially those movements requiring fixation, which are performed more often by physiotherapists. There may also be other determinants of the differences other than the professional background of the raters, such as experience in examining shoulders, different verbal and non-verbal cues to patients, and level of confidence, which may also have a role. The differences in reliability might also depend on the selection of the individual raters. For example, the physiotherapists selected in our previous study might have been a more homogeneous group. In addition, we cannot exclude the possibility that the observed differences between the ICCs of the physiotherapists and rheumatologists were due to real differences between the patients. Although both studies recruited patients with stiff painful shoulders with similar duration of symptoms and pain scores, some minor differences were noted, including the fact that patients in this study had more homogeneous pain scores. Unfortunately, the small number of patients in our study does not allow for more analyses to examine the sources of disagreement between the rheumatologists and physiotherapists.

Other studies have reported interrater ICC values for patients with shoulder disorders ranging between 0.26 and 0.95.^{7 13-15} The reliability scores in our study are consistent with the overall conclusion of previous studies that reliability is highly variable between the various shoulder movements. The measurement instruments and the methods that were used in previous studies were also different, as were the reliability statistics (such as the limits of agreement¹⁶). Previous publications have discussed the use of appropriate ICCs. Studies that have reported on the reliability of shoulder movements in healthy subjects^{9 10 12 20} may either overestimate or underestimate the reliability of shoulder movements when applied in a clinical setting. For reasons of clarity we provide the formulas of the ICCs that we used. ICC values are known to depend on the variation in the study group.²¹ More homogeneous study groups may produce lower ICCs as the main contributor of the variance (patient) is negligible. As the end point of range of movement was determined by the patients' subjective reporting of pain the results of this study are only applicable to patients with shoulder pain.

With further training, and with standardisation of methods of measurement, the rheumatologists may reach higher levels of reliability. Considering the variability in reliability, it may be questioned whether it is necessary to have patients perform all isolated shoulder movements, especially if there are time constraints. The scapular fixation may require extra training to perform in a standardised way. One way to overcome this would be to eliminate complex shoulder movements, and choose the most reliable movements—for example, hand behind back instead of internal rotation in abduction, which both assess internal rotation. Although we had the examiners

do a single movement only, the ICCs may also improve with two readings.

In conclusion, we recommend that for clinical trials of painful shoulder disorders where shoulder movements are an end point the assessments should be performed by the same examiner whenever possible. The examiner requires training and practice in a standardised protocol. For rheumatologists, one hour's training may not be enough. Consideration should be given to choosing only those measures with known reliability.

ACKNOWLEDGEMENTS

Electro-Med, Melbourne for the supply of the Plurimeter-V inclinometer.

Authors' affiliations

J L Hoving, R Buchbinder, Department of Clinical Epidemiology, Cabrini Hospital, Australia
J L Hoving, R Buchbinder, A Forbes, Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia
S Green, Monash Institute of Health Services Research, Monash Medical Centre, Clayton, Australia
N Bellamy, Centre of National Research on Disability and Rehabilitation Medicine, Royal Brisbane Hospital, Herston, Australia
C Brand, A Stockman, Centre for Rheumatic Disease, Royal Melbourne Hospital, Parkville, Australia
R Buchanan, Department of Rheumatology, Austin and Repatriation Medical Centre, Heidelberg, Australia
S Hall, Cabrini Hospital, Melbourne, Australia
M Patrick, Box Hill Hospital, Box Hill, Australia
P Ryan, Alfred Hospital, Prahran, Australia

REFERENCES

- Chakravarty K**, Webley M. Shoulder joint movement and its relationship to disability in the elderly. *J Rheumatol* 1993;20:1359-61.
- Bassey EJ**, Morgan K, Dallosso HM, Ebrahim SB. Flexibility of the shoulder joint measured as range of abduction in a large representative sample of men and women over 65 years of age. *Eur J Appl Physiol Occup Physiol* 1989;58:353-60.
- Van der Windt DA**, Koes BW, de Jong BA, Bouter LM. Shoulder disorders in general practice: incidence, patient characteristics, and management. *Ann Rheum Dis* 1995;54:959-64.
- Pope DP**, Croft PR, Pritchard CM, Macfarlane GJ, Silman AJ. The frequency of restricted range of movement in individuals with self-reported shoulder pain: results from a population-based survey. *Br J Rheumatol* 1996;35:1137-41.
- Croft P**, Pope D, Silman A. The clinical course of shoulder pain: prospective cohort study in primary care. *Primary Care Rheumatology Society Shoulder Study Group. BMJ* 1996;313:601-2.
- Green S**, Buchbinder R, Glazier R, Forbes A. Interventions for shoulder pain. *Cochrane Database Syst Rev* 2000;CD001156.
- Riddle DL**, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting. *Shoulder measurements. Phys Ther* 1987;67:668-73.
- Pandya S**, Florence J, King W, Robison JD, Oxman M, Province M. Reliability of goniometric measurements in patients with Duchenne muscular dystrophy. *Phys Ther* 1985;65:1339-42.
- Bovens AM**, van Baak MA, Vrencken J, Wijnen J, Verstappen F. Variability and reliability of joint measurements. *Am J Sports Med* 1990;18:58-63.
- Clarke GR**, Willis LA, Fish WW, Nichols PJR. Assessment of movement at the glenohumeral joint. *Orthopaedics* 1974;7:55-71.
- Andrews AW**, Bohannon RW. Decreased shoulder range of motion on paretic side after stroke. *Phys Ther* 1989;69:768-72.
- Bower K**. The Hydrogoniometer and assessment of gleno-humeral joint motion. *The Austr J Physiother* 1992;28:12-17.
- Green S**, Buchbinder R, Forbes A, Bellamy N. A standardized protocol for measurement of range of movement of the shoulder using the Plurimeter-V inclinometer and assessment of its intrarater and interrater reliability. *Arthritis Care Res* 1998;11:43-52.
- Croft P**, Pope D, Boswell R, Rigby A, Silman A. Observer variability in measuring elevation and external rotation of the shoulder. *Primary Care Rheumatology Society Shoulder Study Group. Br J Rheumatol* 1994;33:942-6.
- MacDermid JC**, Chesworth BM, Patterson S, Roth JH. Intratester and intertester reliability of goniometric measurement of passive lateral shoulder rotation. *J Hand Ther* 1999;12:187-92.
- Triffitt PD**, Wildin C, Hajjoff D. The reproducibility of measurement of shoulder movement. *Acta Orthop Scand* 1999;70:322-4.
- SAS 6.10**. SAS-Institute. Cary, NC., USA, SAS Institute, 1993.

- 18 **Streiner D**, Norman G. *Health measurement scales: a practical guide to their development and use*. 2nd ed. Oxford: Oxford University Press, 1995.
- 19 **Dunn G**. *Design and analysis of reliability studies: the statistical evaluation of measurement errors*. Oxford: Oxford University Press, 1989.
- 20 **Jordan K**, Dziedzic K, Jones PW, Ong BN, Dawes PT. The reliability of the three-dimensional FASTRAK measurement system in measuring cervical spine and shoulder range of motion in healthy subjects. *Rheumatology (Oxford)* 2000;39:382–8.
- 21 **Fleiss, JF**. Reliability of measurement. *In: The design and analysis of clinical experiments*. New York: Wiley, 1986:1–33.

Annals of the Rheumatic Diseases through the ages

Browse the Archive

Annals of the Rheumatic Diseases online has an archive of content dating back to 1965.
Full text from 1997; abstracts from 1975; table of contents from 1965

www.annrheumdis.com