REPORT

Should contemporary rheumatoid arthritis clinical trials be more like standard patient care and vice versa?

T Pincus, T Sokka

....

The information used by rheumatologists when delivering care to patients with rheumatoid arthritis (RA) is derived mainly from two sources: randomised controlled clinical trials and experience in clinical care. However, these two sources differ significantly because (*a*) the extensive inclusion and exclusion criteria result in clinical trial participants being recruited from only a minority of patients seen in standard clinical care; (*b*) assessments in clinical trials are conducted according to standard quantitative measures and indices, while standard clinical care of most patients with RA is generally conducted empirically, without collection of any quantitative data other than laboratory tests to estimate prognosis and document change in status; and (*c*) although baseline databases of various clinical trials (and observational studies) are 60–90% identical in content, they are not standardised and therefore not amenable to direct comparisons. Strategies to promote similarities between clinical trials and standard clinical care in patients with RA may include: more generalised inclusion criteria; incorporation of quantitative measurement into standard care, easily accomplished by asking each patient to complete a simple questionnaire at each visit to a rheumatologist; and consensus among rheumatologists for databases with standard content and format in clinical care and research involving patients with RA.

T he two major sources of information for rheumatologists concerning management of patients with rheumatoid arthritis (RA) are randomised controlled clinical trials and experience in clinical care. Therefore, clinical trials might be expected to be as similar as possible to clinical care in order to enhance their value. However, at this time in the early twenty first century, randomised trials and clinical care of patients with RA are characterised by at least three substantial differences:

- Clinical trial participants tend to be only a minority of patients seen in standard clinical care due to extensive inclusion and exclusion criteria. This reduces the generalisability of the results of clinical trials to patients in standard care.
- Assessments in clinical trials are now conducted according to standard quantitative measures and indices. However, clinical care of most patients with RA continues in the tradition of being conducted empirically, without collection of any quantitative data or specific quantitative goals of therapy, other than laboratory tests to estimate prognosis and document change in status, with the exception of some patients who are treated with new biological therapies in certain clinical settings.
- Baseline databases of patients in clinical trials as well as in observational studies are not standardised and therefore

not amenable to direct comparison, even though 60–90% identical in content.

In this report, we review published data to illustrate the above three points and propose the following strategies to promote similarities between clinical trials and standard clinical care:

- more generalisable inclusion criteria for RA clinical trials, which reflect the current clinical status of a large proportion of patients
- inclusion of quantitative measurements in standard care, by asking all patients to complete a simple questionnaire at each visit to a rheumatologist
- a standard protocol for evaluation of RA (SPERA), to provide a common, standard database for patients with RA in clinical trials and in clinical care.

CLINICAL TRIAL PARTICIPANTS ARE ONLY A MINORITY OF THE PATIENTS SEEN IN STANDARD CLINICAL CARE

The randomised controlled clinical trial is the "gold standard" to compare one treatment with another or with a placebo,[1] mimicking a laboratory "scientific" experiment by isolating a single variable, the therapy, and using randomisation to adjust for additional variables which might affect the results.[2] At this time, all new therapies designed for standard care require documentation of efficacy and acceptable toxicity based on randomised controlled clinical trials. Furthermore, the term "evidence based medicine" has largely come to mean "evidence from clinical trials, rather than from clinical observational studies and case reports".[3] Nonetheless, as is true of all scientific methods, limitations are seen with randomised controlled clinical trials, as described extensively in reports by many observers,[2] [4–20] including previous commentaries by the present authors.[21–27]

Ten limitations of clinical trials in patients with RA (table 1) can be classified broadly as "pragmatic" or "intrinsic". Pragmatic limitations result from logistical complexities and costs of clinical trials and could be overcome, in theory, if sufficient resources were available to include all patients over long periods of time. Intrinsic limitations are inherent in the clinical trial methodology and cannot be overcome even if unlimited resources were available.

The limitations of clinical trials can be illustrated by contrasting the impressions gained of the efficacy of methotrexate therapy based on a meta-analysis of many clinical trials and a long term clinical observational study. The meta-analysis (fig 1) reported that sulfasalazine, D-penicillamine,

....

Abbreviations: ACR, American College of Rheumatology; CRP, C-reactive protein; DAS, Disease Activity Score; DMARD, disease modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; HAQ, Health Assessment Questionnaire; RA, rheumatoid arthritis; SPERA, standard protocol for evaluation of RA

Table 1 Some pragmatic and intrinsic limitations of randomised controlled clinical trials in rheumatoid arthritis

Pragmatic limitations
Relatively short observation period in most clinical trials in chronic diseases
Inclusion and exclusion criteria in clinical trials often restrict eligibility to a small minority of patients who will receive the intervention
Inflexible dosage schedules and concomitant drug therapies may limit the apparent efficacy of certain drugs such as methotrexate
Surrogate markers used in clinical trials are reversible measures of inflammatory activity, which may be suboptimal indicators of long term irreversible damage and poor outcomes—for example, tender joints may be poor predictors of joint deformity and work disability
Statistically significant results are not necessarily clinically important and vice versa
Rare adverse events cannot be captured in clinical trials of fewer than 10 000 subjects
Intrinsic limitations
The design of a clinical trial may greatly influence the results, despite inclusion of a control group
Clinical trial reports generally ignore individual variation in responses to treatments
Interpretation of results and adverse events is not standardised and may introduce bias into the reporting of the results of any clinical trial
The format of a clinical trial may distort the ''placebo effect'' in either direction by informing patients that they are to receive one of several regimens rather than the ''best'' therapy
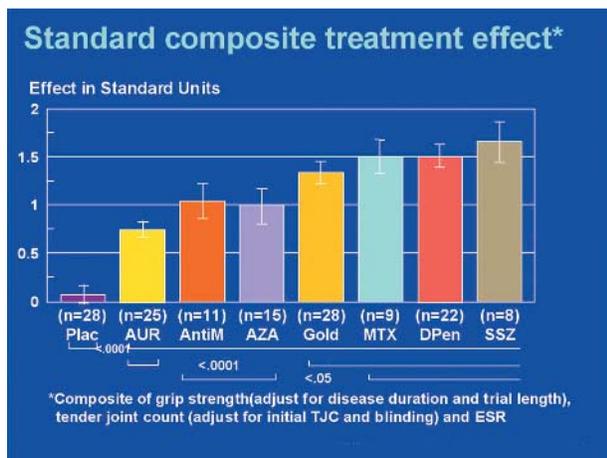
methotrexate, and injectable gold have substantial and roughly equivalent efficacy compared with placebo, with lower efficacy for antimalarial drugs and oral gold.[28] By contrast, in an observational study conducted in seven private practice settings (fig 2 left panel), methotrexate courses were continued over five years in 70% of 538 patients monitored compared with only about 20% of courses of azathioprine, injectable (parenteral) gold salts, penicillamine, and hydroxychloroquine, and fewer than 10% of auranofin courses.[29] These apparently different impressions of methotrexate efficacy can be explained by analyses of a subset of drug courses from the long term clinical study in which only the *first* disease modifying antirheumatic drug (DMARD) courses were analysed over *one* year (fig 2, right panel) (as in many clinical trials), instead of *any* DMARD course over *five* years (fig 2 left panel). No significant differences were seen between courses of any of the DMARDs over one year, similar to results of short term clinical trials and differing considerably from the long term results.[29]

Figure 1 Meta-analyses of clinical trials documenting the efficacy of drugs in the treatment of rheumatoid arthritis (RA). The number of trials for each intervention are given below the respective bars. All drugs had greater efficacy than placebo in the management of RA, determined according to a composite of grip strength (a measure of effectiveness of grip), tender joint count, and erythrocyte sedimentation rate (ESR), adjusted for disease duration, trial length, initial tender joint count, and blinding. In these analyses, no significant differences were seen between sulfasalazine (SSZ), D-penicillamine (DPen), methotrexate (MTX), and injectable gold (Gold). AUR, auranofin; AZA, azathioprine; antiM, antimalarial drugs such as hydroxychloroquine; plac, placebo. Reprinted with permission from Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second line drugs in rheumatoid arthritis: results of 2 meta-analyses. *Arthritis Rheum* 1990;**33**:1449–61, Wiley-Liss, Inc., a subsidiary of John Wiley and Sons, Inc.[28]

One particular limitation of clinical trials in RA at this time concerns their generalisability with respect to the inclusion criteria used for patient selection. In theory, all patients with a given diagnosis such as RA should be eligible to participate in a clinical trial of patients with that diagnosis. However, many variables besides the therapies may affect outcomes, and randomisation cannot be expected to adjust adequately for all variables. Therefore, almost all clinical trials have explicit inclusion criteria, such as a high level of disease activity, and exclusion criteria, such as extensive comorbidities, so that the randomisation groups may be similar.
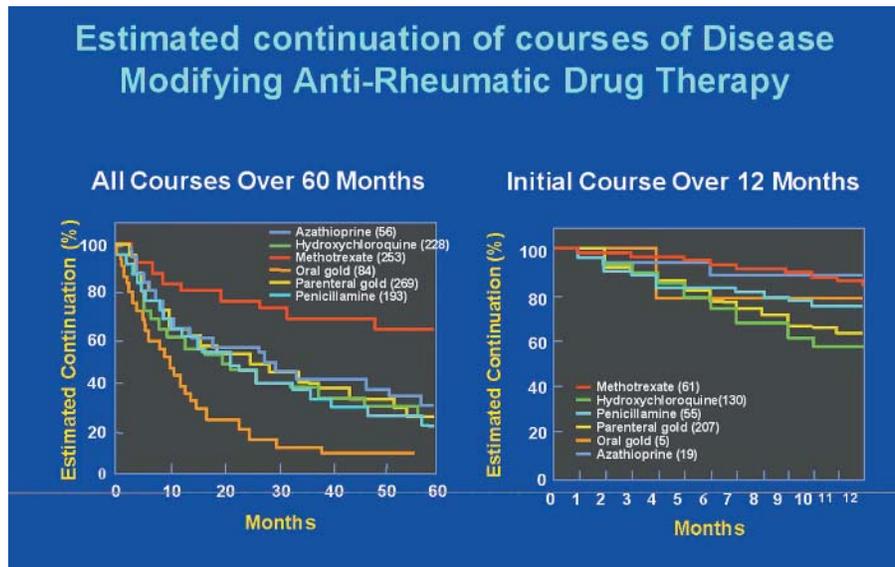
Inclusion and exclusion criteria enhance the comparability of various groups in clinical trials. However, inclusion criteria in RA clinical trials have remained similar over the past 20 years, introducing extensive selection and compromising generalisability. This phenomenon can be illustrated by examining data from cross-sectional studies of two cohorts of consecutive patients with RA identified in 1999–2001 in Nashville, TN, USA.[30 31] An early RA (ERA) cohort included 232 patients with RA of less than three years identified in a large multi-rheumatologist private practice setting. A late RA cohort included 138 patients with long standing RA monitored over 1–18 years in a weekly academic rheumatology clinic. We carried out analyses to find out which patients from these cohorts would be eligible to participate in recent clinical trials of biological agents.

The inclusion criteria for the early RA (ERA) clinical trial[32 33] listed 12 or more tender joints, 10 or more swollen joints, positive rheumatoid factor or erosions, and morning stiffness of 45 minutes or more or erythrocyte sedimentation rate (ESR) of 28 mm/hr or more. Overall, 37 of 232 patients seen in a US multi-rheumatology practice (16%) met the inclusion criteria (table 2). However, 196 of the 232 patients had taken prior methotrexate, and might have been eligible before they took methotrexate. Examination of only the 36 patients who had no prior methotrexate indicated that 11 (31%) would have been eligible for the clinical trial. However, some were already under treatment with other DMARDs and in satisfactory clinical status. Analysis of the 19 patients who were at their first visit and had no prior methotrexate, the most inclusive analysis, indicated that 8 (42%) would have been eligible to participate in this clinical trial (see table 2). This would be the highest proportion of eligible patients according to inclusion criteria for the ERA trial from this cohort.

The inclusion criteria for the Anti-Tumor necrosis factor Trial in Rheumatoid Arthritis with Concomitant Therapy (ATTRACT) study[34 35] comprised six or more tender joints and swollen joints, and two of the three criteria of morning stiffness, ESR, and C-reactive protein (CRP), and the methotrexate dose (see fig 3 for details of criteria). Only

Figure 2 Estimated continuation of courses of six disease modifying antirheumatic drugs (DMARDs) in 532 patients with rheumatoid arthritis. Left panel: estimated continuation of 1083 courses of six DMARDs over 60 months. After two years, about 50% of the courses with four of these drugs, hydroxychloroquine, penicillamine, parenteral gold, and azathioprine, were continued in contrast with 80% of the courses of methotrexate, and only 20% of the courses of oral gold. After five years, about 60% of the methotrexate courses were continued, compared with about 20% of the hydroxychloroquine, penicillamine, parenteral gold, and azathioprine courses and virtually no course of oral gold. Right panel: in a subset of the cohort, estimated continuation of 477 courses of the *initial* (rather than *any*) DMARD in the same patients over 12 months (rather than 60 months). In these analyses, none of the differences was statistically significant, in particular the differences between methotrexate and oral gold (auranofin). These results mimic the results of clinical trials shown in fig 1, which differ considerably from the results seen in left panel. These similarities/differences are explained by a number of limitations connected with randomised controlled clinical trials, including a short time frame of observation, exclusion criteria, the fixed dosage schedule, and the emphasis on statistically significant rather than clinically significant differences. Reprinted with permission from Pincus T, Marcum SB, Callahan LF. Long-term drug therapy for rheumatoid arthritis in seven rheumatology private practices. II. Second-line drugs and prednisone. *J Rheumatol* 1992;**19**:1885–84.[29]

7/138 patients (5%) in the late RA cohort monitored over the long term in a weekly academic rheumatology clinic met the ATTRACT inclusion criteria (see fig 3).

The types of pragmatic limitation illustrated above are potentially surmountable by changes in study design, including sufficient resources and time available to include all patients over long periods. However, certain intrinsic limitations remain. For example, the efficacy of a new drug would likely be greater if enrolment were limited to patients who had no previous experience with DMARDs than to those who had "failed" two previous DMARDs. The inclusion of a "control" group does not eliminate biases, which are intrinsic to the design of any research study.

Clinical trial results are generally reported as identifying the "preferred" or "best" therapy for *all* patients, rather than for *most* patients, as almost all trial results include a few individuals with good results in all treatment groups. However, one interpretation of the results may be that hospital formularies in the USA should include only a single drug or limited number of drugs with the same mechanism of action.[36] Furthermore, interpretation of data on adverse

Table 2 Analysis of patients with rheumatoid arthritis (RA) of less than 3 years' duration who met criteria for inclusion in the early RA (ERA) clinical trial according to therapies taken

Prior methotrexate status	No. of patients	No. (%) who met inclusion criteria for ERA clinical trial*
No prior methotrexate: first visit	19	8 (42%)
No prior methotrexate: not first visit	17	3 (18%)
All no prior methotrexate	36	11 (31%)
Prior methotrexate, leflunomide, etanercept or infliximab	196	26 (13%)
Total—all patients	232	37 (16%)

* Inclusion criteria: 12 or more tender joints and 10 or more swollen joints; positive rheumatoid factor or radiographic erosions; morning stiffness of 45 minutes or more or ESR of 28 mm/hr or more.
Source: Sokka *et al. Arthritis Rheum* 2003;**48**:313–18.[31]
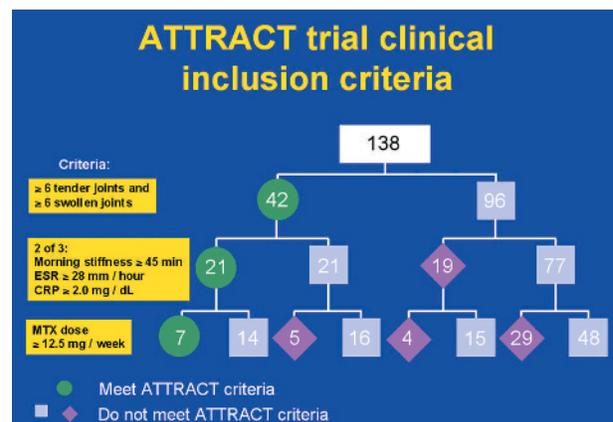
Figure 3 Patients with rheumatoid arthritis who were potential participants in the Anti-Tumor necrosis factor Trial in Rheumatoid Arthritis with Concomitant Therapy (ATTRACT) study of infliximab plus methotrexate versus methotrexate. We analysed whether the 138 patients would meet the inclusion criteria of the ATTRACT study (see figure and text for details). Reprinted with permission from: Sokka T, Pincus T. Eligibility of patients in routine care for major clinical trials of anti-tumour necrosis factor alpha agents in rheumatoid arthritis. *Arthritis Rheum* 2003;**48**:313–18. John Wiley and Sons, Inc., New Jersey.[31]

events remains a subjective consideration in the reporting of clinical trials, particularly in those where no clear cut advantage or disadvantage of one therapy over another is seen. All interventions are associated with some type of adverse event, in at least some patients, whose interpretation is not standardised prospectively, and may introduce considerable disagreement when analysing the results.

Conclusion

These observations strongly suggest that current inclusion criteria for clinical trials in RA are too restrictive and often do not apply to the majority of patients seen in contemporary clinical settings. It may be appropriate to consider inclusion of patients with fewer than six swollen joints or six tender joints, perhaps patients who have as few as two swollen joints, particularly with a goal of remission as the objective of current clinical care. Furthermore, 40–70% of patients with RA have normal ESR,[30] [37] and therefore would be ineligible for clinical trials according to this criterion. Why not consider inclusion criteria based on the American College of Rheumatology (ACR) Core Data Set or the Disease Activity Score (DAS), the primary outcome criteria?

CLINICAL TRIALS ARE NOW CONDUCTED ACCORDING TO STANDARD QUANTITATIVE DATA FROM FORMAL JOINT EXAMINATIONS, LABORATORY TESTS, AND PATIENT QUESTIONNAIRES, BUT STANDARD CLINICAL CARE USUALLY INCLUDES ONLY LABORATORY TESTS

Clinical trials are now conducted according to standard quantitative data included in the American College of Rheumatology (ACR) Core Data Set[38–40] and the Disease Activity Score (DAS).[41] [42] These indices include measures from three types of sources—the joint examination, laboratory, and patient questionnaires. However, as noted above standard clinical care of most patients with RA usually does not include formal joint examination and patient questionnaire measures. Therefore, laboratory tests are the only quantitative data collected in standard clinical care, other than for some patients who are treated with biological agents in certain settings, who may have formal assessment of joint counts and sometimes questionnaires as well. Standard rheumatology care of most patients generally does not include quantitative goals or benchmarks, analogous to a certain level of blood pressure or serum cholesterol, to estimate prognosis and document change in status.

In RA, no single measure can serve as a "gold standard" to assess clinical status in all individual patients, unlike in hypertension or hyperlipidaemia. Therefore pooled indices[43] have been developed to assess patients with RA. Early pooled
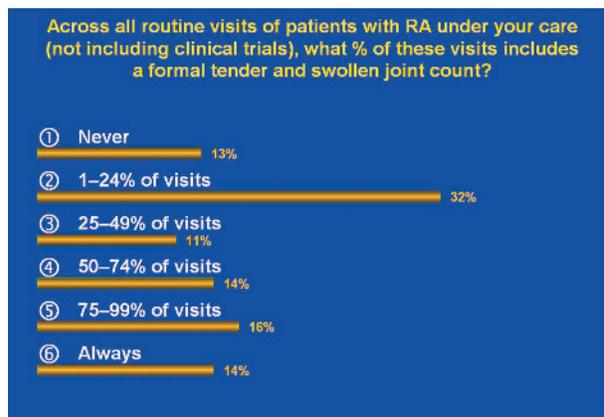
Figure 4 Responses of ~500 European rheumatologists to a question presented in September 2003 at the European launch of adalimumab (see figure). Note that 13% of rheumatologists indicated that no visit included a formal joint count, and only 14% indicated that all visits of patients with RA included formal joint counts. Overall, 45% of rheumatologists indicated that fewer than 25% visits included a formal joint count, 56% indicated that fewer than 50% of visits included a formal joint count, and 70% indicated that fewer than 75% of visits included a formal joint count. By permission of Abbott Immunology, Chicago, IL.

indices (table 3) such as "therapeutic scorecard in rheumatoid arthritis" of Steinbrocker and Blazer,[44] the Lansbury index,[45] and the Paulus criteria,[46] have largely been supplanted by the ACR Core Data Set[38–40] and DAS[41] [42] (see table 3). Recently, a simplified disease activity index (SDAI), which includes measures in the DAS and a physician global assessment,[47] as well as an index consisting of only the three patient self-report measures in the Core Data Set, physical funtion, pain, and global status[48] have been reported. Among these measures, however, only CRP or ESR are measured regularly (in some settings) in standard care, while formal quantitative counts of swollen and tender joints, patient questionnaire scores for physical function, pain and global status, and quantitative physician or assessor global scores are usually not performed.

At a recent meeting of European rheumatologists, only 14% indicated that they performed formal quantitative joint counts at all visits of patients with RA (fig 4). Indeed, 56% of rheumatologists acknowledged performing formal quantitative joint counts in fewer than half of the visits of patients with RA (see fig 4). Most likely joint count performance is over estimated as physicians tend to exaggerate what they think they should do. Furthermore, tender and swollen joint counts have been found to improve more with placebo

Table 3 Indices of various measures used to analyse rheumatoid arthritis

Measure	Steinbrocker[44]	Lansbury[45]	Paulus[46]	ACR Core Data Set[38–40]	DAS[41] [42]	SDAI[47]	Patient Only index[48]
Joint swelling	+		+	+	+	+	
Joint pain/tenderness	+	+	+	+	+	+	
Joint motion	+						
ESR	+	+	+	+	+	+	
Haemoglobin	+	+					
Functional status	+			+			+
Pain	+	+		+			+
Patient global	+		+	+	+	+	+
Weight	+						
Fever		+					
Muscle weakness		+					
Morning stiffness		+	+				
Fatigue		+					
Assessor global			+	+		+	

ACR, American College of Rheumatology; ESR, erythrocyte sedimentation rate; DAS, Disease Activity Score; SDAI, Simplified Disease Activity Index.

Table 4 Change in American College of Rheumatology (ACR) Core Data Set measures over 12 months in rheumatoid arthritis clinical trial of leflunomide versus methotrexate versus placebo

Measure	Leflunomide	Placebo	Methotrexate	Effect size	Relative efficiency
Tender joints	−7.7	−3.0	−6.6	−0.59	1.00
Swollen joints	−5.7	−2.9	−5.4	−0.44	0.56
Assessor global	−2.8	−1.0	−2.4	−0.68	1.33
ESR	−6.3	+2.6	−6.5	−0.41	0.48
HAQ	−0.45	+0.03	−0.26	−0.80	1.84
MHAQ	−0.29	+0.07	−0.15	−0.69	1.37
Pain	−2.2	−0.4	−1.7	−0.65	1.21
Patient global	−2.1	+0.1	−1.5	−0.81	1.88

Sources: Tugwell *et al. Arthritis Rheum* 2000;43:506–14[49] and Strand *et al. Arch Int Med* 1999;159:2542–50.[50] (M)HAQ, (modified) Health Assessment Questionnaire, ESR, erythrocyte sedimentation rate.

treatment than any of the other measures included in the ACR Core Data Set (see table 4). These phenomena suggest that there is more variation and even bias in the assessment of joint count measures than in the assessment of patient measures.

The Health Assessment Questionnaire (HAQ)[51] and Arthritis Impact Measurement Scales (AIMS)[52] were published in 1980, and evidence of patient questionnaires being better clinical measures for predicting morbidity and mortality in RA than laboratory tests or radiographs was published in 1984.[53] Yet most rheumatologists have not incorporated patient questionnaires into standard clinical care. One reason may be that the traditional ''biomedical'' paradigm, the basis for the most spectacular advances of twentieth century medicine, regards data derived from a physician and/or high technology imaging and laboratory source as critical, while data from a patient are seen as unimportant, except to provide clues to obtain definitive high technology data.

A second reason may be that most rheumatologists have little experience with simple questionnaires in clinical care; most of the experience with patient questionnaires involves cumbersome and lengthy documents designed for clinical research and clinical trials (table 5). In formal clinical research, an extensive questionnaire is appropriate to have as complete a database as might be needed to address questions in a study. Patients and clinicians recognise and accept the inconveniences of lengthy questionnaires with complex scoring. Indeed, the clinician is not expected to review the patient questionnaire data, which are not interpreted at the clinical site but rather entered into a database at a data centre (see table 5).

In contrast, a questionnaire for standard care must be feasible and practical (see table 5), completed by a patient within 5–10 minutes, scanned (''eyeballed'') by a health professional in less than 5 seconds, and scored formally and entered into a flowsheet to compare with previous visits in

less than 30 seconds. Furthermore, a questionnaire for standard patient care should be clinically applicable to patients with all diagnoses, and provide time-saving information to the physician by enhancing a patient's capacity to describe his or her concerns in the limited time allotted for a clinical visit.

A recent report presented a pooled index comprised only of the three patient measures from the ACR Core Data Set (physical function, pain, and global status)[48] (see table 3) to distinguish active leflunomide or methotrexate treatment from placebo treatment.[50] The results according to this ''patient only'' index were compared with those according to indices of ''assessor only'' measures, patient measures plus ESR, assessor measures plus ESR, and all seven ACR Core Data Set measures.[48] The capacity of each index to distinguish active treatment from placebo treatment was similar to that of all the others and to the ACR 20 and DAS (table 6).[48]

The strong likelihood that most patients with rheumatic diseases will not be included in clinical trials, as well as the limitations of clinical trials noted above, indicate that long term observational studies are required to obtain optimal information concerning results of clinical care in RA, including the new biological agents. Many current concepts about RA have been derived primarily from long term observational studies, including evidence for substantial work disability,[54] radiographic damage in early RA[55–57] and over time,[58] functional decline,[53] premature mortality in RA,[53 59] and gastropathy associated with non-steroidal anti-inflammatory drugs.[60 61] The concepts are based on data that cannot be obtained from clinical trials.

These considerations suggest that a simple approach to the further development of clinical rheumatology as a quantitative scientific field would be for each patient to complete a questionnaire at each visit. Patient questionnaires provide the most valuable data to predict and monitor functional losses, work disability, and death, and have more predictive value than any known laboratory test or radiograph in RA.[62–65] The questionnaires address the primary concerns of patients with RA—functional limitations, pain, fatigue, psychological distress. Patient questionnaires are also of documented clinical value in lupus, scleroderma, osteoarthritis, fibromyalgia, and ankylosing spondylitis.[66] The patient questionnaire data are correlated significantly with traditional data—that is, ESR, joint count, walk time—and can explain other measures such as global assessment and joint count more effectively than any other measure of clinical status.[67] Furthermore, patient questionnaires are as sensitive to changes in clinical status as traditional measures in clinical trials.[63]

Clinical rheumatologists have expressed a number of misconceptions concerning patient questionnaires (table 7), which may be broadly classified into three categories: validity, consequences, and logistics. Rheumatologists have

Table 5 Features of patient questionnaires in clinical research versus clinical care

Clinical research	Clinical care
Complete, long	Patient friendly, completed in 5–10 minutes
Take time for office routine	Saves time for health professionals
Results unknown to clinician	Scan ''eyeball'' results in 5–10 seconds
Scoring too complex in clinic	Score results in 15–30 seconds
Results not interpreted to add to clinical decisions	Results compared to previous scores to assess trends
Send to data centre	Document in medical record, flowsheet

Table 6 Improvement with leflunomide, methotrexate or placebo for three pooled indices, American College of Rheumatology (ACR) 20 and the Disease Activity Score (DAS)

Pooled index	20% Improvement response			z scores*	
	Leflunomide	Methotrexate	Placebo	Leflunomide v placebo	Methotrexate v placebo
ACR 20	52%	46%	26%	4.36	3.32
Patient only	64%	56%	33%	4.85	3.00
Assessor only	74%	69%	43%	5.12	4.86
All Core Data Set	57%	49%	29%	5.89	5.12
DAS	59%	59%	30%	4.84	4.85

*All z scores $p < 0.001$.
Source: Pincus et al. Arthritis Rheum 2003;**48**:625–30.[48]

suggested that patient questionnaires are unrelated to laboratory or radiographic data. However, significant correlations are seen between these measures, although at levels which indicate these measures are not redundant, but rather related to one another. Rheumatologists must continue to take a history and examine joints. The purpose of the questionnaire is to reduce the time required to gather factual information and use the time saved for patient discussion and counselling. Appropriate use of a questionnaire saves physician time, provided it is easily scanned ("eyeballed") in 5–15 seconds. A rheumatologist is not obliged to create a database with patient questionnaire data anymore than with laboratory or radiographic data, although a flowsheet is a very good way to monitor patient status.

Longitudinal databases have been established for post-marketing surveillance of biological agents. These databases are excellent for identifying possible problems with safety of the new agents, but they cannot provide optimal answers concerning efficacy of these agents and capacity to improve long term outcomes. Furthermore, these databases often do not include data concerning patients who were not treated with the biological agents, many of whom may have favourable outcomes, for comparison with results of therapy with biological agents.

Conclusion
The standard medical record generally includes little or no quantitative data documenting baseline status whether patients with RA are better or worse over long periods, and patients often see different physicians over the years. Rheumatologists can monitor consecutive patients in clinical care to identify long term outcomes by using patient self-report questionnaires as a standard component of each visit in clinical care.[68–70] The patient questionnaire, in addition to appropriate laboratory tests and imaging data, would provide quantitative data documenting severity and helping monitor improvement in each individual patient in the care of any rheumatologist.

BASELINE DATABASES ARE NOT STANDARDISED AND THEREFORE NOT AMENABLE TO DIRECT COMPARISON
The advancement of science depends on standardisation of measurement methods. This phenomenon is readily accepted in laboratory science, in which say, ESR or rheumatoid factor are assessed according to a standard protocol. Standardised pooled indices such as the ACR Core Data Set[38–40] and DAS[41 42] have been adopted in almost all RA clinical trials. However, while outcome measures used in clinical trials are now standardised, additional baseline data collected at the onset of a clinical trial, as well as in observational studies, are not standardised at this time, despite the likelihood that 60–90% or more of the data collected are identical. The content and format of databases differ sufficiently so that different databases can be compared only with great difficulty. It is possible that standardisation of databases would contribute substantially to clinical knowledge of RA.

We have used a simple "standard protocol to evaluate rheumatoid arthritis" (SPERA),[71] which provides a pragmatic assessment on five pages that can be completed in 15–30 minutes, to capture most of the information that should be collected on the initial visit of a patient with RA or any early arthritis which might be RA. This protocol, which has been used for more than 20 years, incorporates the five core domains, health status, disease process, damage, mortality,

Table 7 Some misconceptions concerning use of patient questionnaires in clinical care

Misconception	More accurate conception
Validity	
Unrelated to lab, radiographs	Correlated with lab, radiograph, joint count
Less predictive than …	As or more predictive than …
Less sensitive than …	As or more sensitive in clinical trials and care
Less reproducible than …	As or more reproducible as lab tests, radiographs
Consequences	
No need to examine joints	All joints should be examined, but formal joint count does not add important data
Replace patient history	Enhance patient history
Try to be as complete as possible—extensive patient data needed	Try to be as pragmatic as possible—better to obtain 80% of ideal data in 100% of patients than ideal data in 5%
MDHAQ score dictates stop or start drug	One of many measures in clinical decisions, ESR, joint examination, etc.
Logistics	
Takes extra time	Saves time for rheumatologist
Burden for patients	5–10 minutes in waiting room
Must score responses	"Eyeball" data in 5–10 seconds
Need software, database	Why? Are labs in the database?
Focus on instrument	Focus on patient

and toxicity/adverse reactions, from a consensus recommendation of an Outcome MEasures in Rheumatoid Arthritis Clinical Trials (OMERACT) conference in 1998.[72] The SPERA approach may be viewed as an effort to extend the concept of a uniform clinical database for rheumatic diseases, initially proposed by Fries in the 1970s,[73 74] beyond clinical trials to routine clinical care.

The measures collected for the SPERA database have been about 95% identical over the past two decades. This database has proved useful in clinical research concerning prognosis and monitoring of patients, including:

- observation of radiographic damage in most patients within the first two years of disease during the 1980s[56]
- development of a 28 joint count[75]
- documentation of the absence of correlation between joint tenderness scores and radiographic scores in cross-sectional studies[76]
- description of significant correlations between patient questionnaire scores with joint counts, radiographic scores, and laboratory tests[67]
- prediction of work disability[64] and mortality[65 77]
- recognition of a relatively small proportion of patients being eligible for contemporary clinical trials.[30 31]

The five pages in the SPERA protocol include:

- Clinical features of RA: classification criteria, comorbidities, extra-articular manifestations, surgeries, laboratory tests, family history, and work status
- All medications used for RA since disease onset
- A 42 joint count, comprising 10 proximal interphalangeal and 10 metacarpophalangeal joints of the hand, 2 wrists, 2 elbows, 2 shoulders, 2 hips, 2 knees, 2 ankles, and 10 metatarsophalangeal joints (hips and shoulders are not scored for swelling). All joints are scored for four possible problems, tenderness or pain on motion, swelling, limited motion or deformity, and surgery, with a space to indicate that a joint is normal. Inclusion or limited motion or deformity at baseline may help to recognise one basis for non-improvement of swollen joints, as swelling is much less likely to improve in joints with subluxation and other damage than if there is no damage to the joint. Surgery is included to account for a joint that might have no swelling or tenderness, but after a total joint replacement, which is not entirely "normal". A check mark indicating that a joint is "normal" saves time when performing a joint count.
- A patient self-report multidimensional (MD)HAQ,[78] including 10 items for physical function in a standard HAQ format, visual analogue scales to assess pain, global health, and fatigue, minutes of morning stiffness, and symptom checklist. This questionnaire provides data similar to the HAQ, which can be used instead of the MDHAQ.
- Radiographic scoring sheet: we have used the Sharp and Larsen methods, which, as indicated by several reports, yield very similar data.[79 80]

The two pages of clinical features and medications are kept in the standard care patient record for updating.

Conclusion

Although rheumatologists collect data at baseline in clinical trials and clinical care which are 60–90% identical, sufficient differences exist between databases to make direct comparisons quite cumbersome. A simple, standardised database and format which can be completed by a rheumatologist in less than 15 minutes could be used in all clinical trials as well as observational studies, and might greatly advance clinical research in RA. We certainly do not suggest that the SPERA database is exactly what should be used, but rather would advocate efforts at a consensus for a standard database, as suggested by Fries more than two decades ago.[74] Furthermore, individual investigators should always be free to include any additional data that are considered important. It is also suggested that the forms and content of the standard database be updated every two to four years, based on accumulated new information, to further enhance clinical rheumatology as a quantitative science.

. . . .

Authors' affiliations
T Pincus, T Sokka, Division of Rheumatology and Immunology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
T Sokka, Department of Medicine, Jyväskylä Central Hospital, Jyväskylä, Finland

Correspondence to: Dr T Pincus, Professor of Medicine, Division of Rheumatology and Immunology, Vanderbilt University School of Medicine, 203 Oxford House, Box 5, Nashville, TN 37232-4500, USA; t.pincus@vanderbilt.edu

REFERENCES

1 **Friedman LM**, Furberg CD, DeMets DL. Fundamentals of clinical trials. Littleton: John Wright–PSG Inc., 1984.
2 **Feinstein AR**. An additional basic science for clinical medicine: II. The limitations of randomized trials. *Ann Intern Med* 1983;**99**:544–50.
3 **Sackett DL**, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. *BMJ* 1996;**312**:71–2.
4 **Freiman JA**, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: survey of 71 "negative" trials. *N Engl J Med* 1978;**299**:690–4.
5 **Sackett DL**, Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;**301**:1410–12.
6 **Sackett DL**. The competing objectives of randomized trials. *N Engl J Med* 1980;**303**:1059–60.
7 **Huskisson EC**. Important factors in the success and failure of clinical trials [closing remarks]. *Agents Actions Suppl* 1980;**7**:323–4.
8 **Freireich EJ**. The randomized clinical trial as an obstacle to clinical research. In: Varco RL, Delaney JP, eds. *Controversy in Surgery*. Philadelphia: WB Saunders, 1983:5–12.
9 **Diamond GA**, Forrester JS. Clinical trials and statistical verdicts: probable grounds for appeal. *Ann Intern Med* 1983;**98**:385–94.
10 **Chalmers TC**, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;**309**:1358–61.
11 **Bombardier C**, Tugwell P. Controversies in the analysis of longterm clinical trials of slow acting drugs [editorial]. *J Rheumatol* 1985;**12**:403–5.
12 **Guyatt G**, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy—randomized trials in individual patients. *N Engl J Med* 1986;**314**:889–92.
13 **Hawley DJ**, Wolfe F. Are the results of controlled clinical trials and observational studies of second line therapy in rheumatoid arthritis valid and generalizable as measures of rheumatoid arthritis outcome: analysis of 122 studies. *J Rheumatol* 1991;**18**:1008–14.
14 **Felson DT**, Anderson JJ, Meenan RF. Time for changes in the design, analysis, and reporting of rheumatoid arthritis clinical trials. *Arthritis Rheum* 1990;**33**:140–9.
15 **Gotzsche PC**. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;**10**:31–56.
16 **Klippel JH**. Comment: Winning the battle, losing the war? Another editorial about RA. *J Rheumatol* 1990;**17**:1118–22.
17 **Peto R**, Collins R, Gray R. Large-scale randomized evidence: large, simple trials and overviews of trials. *J Clin Epidemiol* 1995;**48**:23–40.
18 **Concato J**, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92.
19 **Horwitz RI**. Clinical versus statistical considerations in the design and analysis of clinical research. *J Clin Epidemiol* 1998;**51**:305–7.
20 **Feinstein AR**, Horwitz RI. Problems in the "evidence" of "evidence-based medicine". *Am J Med* 1997;**103**:529–35.
21 **Kvien TK**, Mikkelsen K, Nordvag B. Results from controlled clinical trials: how relevant for clinical practice? *J Rheumatol* 2003;**30**:1135–7.
22 **Pincus T**. Rheumatoid arthritis: disappointing long-term outcomes despite successful short-term clinical trials. *J Clin Epidemiol* 1988;**41**:1037–41.
23 **Pincus T**, Wolfe F. Response to letter: Gold therapy for rheumatoid arthritis: Challenges to traditional paradigms. *Ann Intern Med* 1992;**117**:169–70.

24 **Pincus T**. Limitations of randomized controlled clinical trials to recognize possible advantages of combination therapies in rheumatic diseases. *Semin Arthritis Rheum* 1993;**23**(suppl 1):2–10.

25 **Pincus T**, Stein M. What is the best source of useful data on the treatment of rheumatoid arthritis: Clinical trials, clinical observations, or clinical protocols? *J Rheumatol* 1995;**22**:1611–17.

26 **Pincus T**. Analyzing long-term outcomes of clinical care without randomized controlled clinical trials: The consecutive patient questionnaire database. *Adv Mind Body Med* 1997;**13**:3–32.

27 **Pincus T**, Stein CM. Why randomized controlled clinical trials do not depict accurately long-term outcomes in rheumatoid arthritis: some explanations and suggestions for future studies. *Clin Exp Rheumatol* 1997;**15**(suppl 17): S27–S38.

28 **Felson DT**, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis: results of two metaanalyses. *Arthritis Rheum* 1990;**33**:1449–61.

29 **Pincus T**, Marcum SB, Callahan LF. Long-term drug therapy for rheumatoid arthritis in seven rheumatology private practices: II. Second-line drugs and prednisone. *J Rheumatol* 1992;**19**:1885–94.

30 **Sokka T**, Pincus T. Most patients receiving routine care for rheumatoid arthritis in 2001 did not meet inclusion criteria for most recent clinical trials or American College of Rheumatology criteria for remission. *J Rheumatol* 2003;**30**:1138–46.

31 **Sokka T**, Pincus T. Eligibility of patients in routine care for major clinical trials of anti-tumor necrosis factor alpha agents in rheumatoid arthritis. *Arthritis Rheum* 2003;**48**:313–18.

32 **Bathon JM**, Genovese MC. The Early Rheumatoid Arthritis (ERA) trial comparing the efficacy and safety of etanercept and methotrexate. *Clin Exp Rheumatol* 2003;**21**:S195–S197.

33 **Genovese MC**, Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, *et al*. Etanercept versus methotrexate in patients with early rheumatoid arthritis: two-year radiographic and clinical outcomes. *Arthritis Rheum* 2002;**46**:1443–50.

34 **Lipsky PE**, van der Heijde DMFM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, *et al*. Infliximab and methotrexate in the treatment of rheumatoid arthritis. *N Engl J Med* 2000;**343**:1594–602.

35 **Maini R**, St Clair EW, Breedveld F, Furst D, Kalden J, Weisman M, *et al*. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: A randomized phase III trial. *Lancet* 1999;**354**:1932–9.

36 **Stein CM**, Wood AJJ, Pincus T. Implementation of multiple outpatient formularies: undesirable effects. *Clin Pharmacol Ther* 1997;**61**:1–7.

37 **Wolfe F**, Michaud K. The clinical and research significance of the erythrocyte sedimentation rate. *J Rheumatol* 1994;**21**:1227–37.

38 **Felson DT**, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum* 1993;**36**:729–40.

39 **Tugwell P**, Boers M. OMERACT Committee. Proceedings of the OMERACT Conferences on outcome measures in rheumatoid arthritis clinical trials, Maastricht, the Netherlands. *J Rheumatol* 1993;**20**:527–91.

40 **van Riel PLCM**. Provisional guidelines for measuring disease activity in clinical trials on rheumatoid arthritis [editorial]. *Br J Rheumatol* 1992;**31**:793–4.

41 **van der Heijde DMFM**, van't Hof MA, van Riel PLCM, Theunisse LM, Lubberts EW, van Leeuwen MA, *et al*. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;**49**:916–20.

42 **van der Heijde DMFM**, van't Hof M, van Riel PLCM, van de Putte LBA. Development of a disease activity score based on judgment in clinical practice by rheumatologists. *J Rheumatol* 1993;**20**:579–81.

43 **Goldsmith CH**, Smythe HA, Helewa A. Interpretation and power of pooled index. *J Rheumatol* 1993;**20**:575–8.

44 **Steinbrocker O**, Blazer A. A therapeutic score card for rheumatoid arthritis. *N Engl J Med* 1946;**14**:501–6.

45 **Lansbury J**. A method for summation of the systemic indices of rheumatoid activity. *Am J Med Sci* 1956;**232**:300–10.

46 **Paulus HE**, Egger MJ, Ward JR, Williams HJ, Cooperative Systematic Studies of Rheumatic Diseases Group. Analysis of improvement in individual rheumatoid arthritis patients treated with disease-modifying antirheumatic drugs, based on the findings in patients treated with placebo. *Arthritis Rheum* 1990;**33**:477–84.

47 **Smolen JS**, Breedveld FC, Schiff MH, Kalden JR, Emery P, Eberl G, *et al*. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. *Rheumatology (Oxford)* 2003;**42**:244–57.

48 **Pincus T**, Strand V, Koch G, Amara I, Crawford B, Wolfe F, *et al*. An index of the three core data set patient questionnaire measures distinguishes efficacy of active treatment from placebo as effectively as the American College of Rheumatology 20% response criteria (ACR20) or the disease activity score (DAS) in a rheumatoid arthritis clinical trial. *Arthritis Rheum* 2003;**48**:625–30.

49 **Tugwell P**, Wells G, Strand V, Maetzel A, Bombardier C, Crawford B, *et al*. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. *Arthritis Rheum* 2000;**43**:506–14.

50 **Strand V**, Cohen S, Schiff M, Weaver A, Fleischmann R, Cannon G, *et al*. Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. *Arch Intern Med* 1999;**159**:2542–50.

51 **Fries JF**, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;**23**:137–45.

52 **Meenan RF**, Gertman PM, Mason JH. Measuring health status in arthritis: the arthritis impact measurement scales. *Arthritis Rheum* 1980;**23**:146–52.

53 **Pincus T**, Callahan LF, Sale WG, Brooks AL, Payne LE, Vaughn WK. Severe functional declines, work disability, and increased mortality in seventy-five rheumatoid arthritis patients studied over nine years. *Arthritis Rheum* 1984;**27**:864–72.

54 **Yelin E**, Meenan R, Nevitt M, Epstein W. Work disability in rheumatoid arthritis: effects of disease, social, and work factors. *Ann Intern Med* 1980;**93**:551–6.

55 **Brook A**, Corbett M. Radiographic changes in early rheumatoid disease. *Ann Rheum Dis* 1977;**36**:71–3.

56 **Fuchs HA**, Kaye JJ, Callahan LF, Nance EP, Pincus T. Evidence of significant radiographic damage in rheumatoid arthritis within the first 2 years of disease. *J Rheumatol* 1989;**16**:585–91.

57 **Thould AK**, Simon G. Assessment of radiological changes in the hands and feet in rheumatoid arthritis: their correlation with prognosis. *Ann Rheum Dis* 1966;**25**:220–8.

58 **Scott DL**, Grindulis KA, Struthers GR, Coulton BL, Popert AJ, Bacon PA. Progression of radiological changes in rheumatoid arthritis. *Ann Rheum Dis* 1984;**43**:8–17.

59 **Isomäki HA**, Mutru O, Koota K. Death rate and causes of death in patients with rheumatoid arthritis. *Scand J Rheumatol* 1975;**4**:205–8.

60 **Fries JF**, Williams CA, Bloch DA, Michel BA. Nonsteroidal anti-inflammatory drug-associated gastropathy: incidence and risk factor models. *Am J Med* 1991;**91**:213–22.

61 **Griffin MR**, Ray WA, Schaffner W. Nonsteroidal anti-inflammatory drug use and death from peptic ulcer in elderly persons. *Ann Intern Med* 1988;**109**:359–63.

62 **Wolfe F**, Pincus T. Standard self-report questionnaires in routine clinical and research practice—an opportunity for patients and rheumatologists. *J Rheumatol* 1991;**18**:643–6.

63 **Wolfe F**, Pincus T. Data collection in the clinic. *Rheum Dis Clin North Am* 1995;**21**:321–58.

64 **Callahan LF**, Bloch DA, Pincus T. Identification of work disability in rheumatoid arthritis: physical, radiographic and laboratory variables do not add explanatory power to demographic and functional variables. *J Clin Epidemiol* 1992;**45**:127–38.

65 **Pincus T**, Brooks RH, Callahan LF. Prediction of long-term mortality in patients with rheumatoid arthritis according to simple questionnaire and joint count measures. *Ann Intern Med* 1994;**120**:26–34.

66 **Callahan LF**, Smith WJ, Pincus T. Self-report questionnaires in five rheumatic diseases: comparisons of health status constructs and associations with formal education level. *Arthritis Care Res* 1989;**2**:122–31.

67 **Pincus T**, Callahan LF, Brooks RH, Fuchs HA, Olsen NJ, Kaye JJ. Self-report questionnaire scores in rheumatoid arthritis compared with traditional physical, radiographic, and laboratory measures. *Ann Intern Med* 1989;**110**:259–66.

68 **Pincus T**, Wolfe F. An infrastructure of patient questionnaires at each rheumatology visit: improving efficiency and documenting care. *J Rheumatol* 2000;**27**:2727–30.

69 **Pincus T**. Documenting quality management in rheumatic disease: are patient questionnaires the best (and only) method? *Arthritis Care Res* 1996;**9**:339–48.

70 **Wolfe F**, Pincus T. Listening to the patient: a practical guide to self-report questionnaires in clinical care. *Arthritis Rheum* 1999;**42**:1797–808.

71 **Pincus T**, Brooks RH, Callahan LF. A proposed standard protocol to evaluate rheumatoid arthritis (SPERA) that includes measures of inflammatory activity, joint damage, and longterm outcomes. *J Rheumatol* 1999;**26**:473–80.

72 **Wolfe F**, Lassere M, van der Heijde D, Stucki G, Suarez-Almazor M, Pincus T, *et al*. Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. OMERACT IV: Outcome measures in rheumatology. Cancun, Mexico, April 16–20, 1998. *J Rheumatol* 1999;**26**:484–9.

73 **Fries JF**. Time-oriented patient records and a computer databank. *JAMA* 1972;**222**:1536–42.

74 **Fries JF**. A data bank for the clinician? [editorial]. *N Engl J Med* 1976;**294**:1400–2.

75 **Fuchs HA**, Brooks RH, Callahan LF, Pincus T. A simplified twenty-eight joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum* 1989;**32**:531–7.

76 **Fuchs HA**, Callahan LF, Kaye JJ, Brooks RH, Nance EP, Pincus T. Radiographic and joint count findings of the hand in rheumatoid arthritis: related and unrelated findings. *Arthritis Rheum* 1988;**31**:44–51.

77 **Callahan LF**, Pincus T, Huston JW III, Brooks RH, Nance EP Jr, Kaye JJ. Measures of activity and damage in rheumatoid arthritis: depiction of changes and prediction of mortality over five years. *Arthritis Care Res* 1997;**10**:381–94.

78 **Pincus T**, Swearingen C, Wolfe F. Toward a multidimensional health assessment questionnaire (MDHAQ): Assessment of advanced activities of daily living and psychological status in the patient friendly health assessment questionnaire format. *Arthritis Rheum* 1999;**42**:2220–30.

79 **Pincus T**, Callahan LF, Fuchs HA, Larsen A, Kaye J. Quantitative analysis of hand radiographs in rheumatoid arthritis: Time course of radiographic changes, relation to joint examination measures, and comparison of different scoring methods. *J Rheumatol* 1995;**22**:1983–9.

80 **Molenaar ET**, Edmonds J, Boers M, van der Heijde DM, Lassere M. A practical exercise in reading RA radiographs by the Larsen and Sharp methods. *J Rheumatol* 1999;**26**:746–8.