Advantages and Limitations of Quantitative Measures to Assess Rheumatoid Arthritis
Joint Counts, Radiographs, Laboratory Tests, and Patient Questionnaires

Theodore Pincus, M.D.

Abstract
Medical care is advanced by quantitative measures, all of which have advantages and limitations. No single “gold standard” measure, analogous to blood pressure, is available for diagnosis, prognosis, and management of rheumatoid arthritis (RA). Four types of measures have been used, including joint counts, radiographs, laboratory tests, and patient questionnaires. Joint counts are the most specific measure for RA but are poorly reproducible and not performed in most standard care. Radiographs provide an objective record of joint damage, but are scored quantitatively only in clinical research and have little prognostic value for long-term outcomes such as work disability and mortality. Laboratory tests are helpful when positive but frequently are “false negative”—for example, rheumatoid factor (RF), erythrocyte sedimentation rate, or C-reactive protein are normal in 30% to 45% of patients. “False positive” results are also seen; most people with RF or antinuclear antibody do not have a disease. Patient questionnaires are useful to assess and monitor patient status and provide the most significant predictive measures for long-term work disability and mortality. A multidimensional health assessment questionnaire is useful in all rheumatic diseases, with scoring templates and medical history information to save time for the rheumatologist and patient in standard care.

Modern medical care has been greatly advanced through quantitative measures to describe clinical problems as numerical data. Four types of quantitative measures, joint counts, radiographic scores, laboratory tests, and patient self-report questionnaires, have been widely used in quantitative assessment of patients with rheumatoid arthritis (RA). None of these measures can stand alone as a single “gold standard” to assess each individual patient in clinical trials and clinical care, in contrast to blood pressure or serum cholesterol, which serve as a single predictor of primary outcomes in clinical trials and targets in clinical care. Therefore, the measures generally are assembled into pooled indices, such as the American College of Rheumatology (ACR) Core Data Set and disease activity score (DAS).

The rheumatology literature concerning the different types of measures used in quantitative assessment of patients with RA has emphasized their validity (what is thought to be assessed by the measure) and their reliability (whether they appear to provide reproducible data). Appropriate validity and reliability are essential attributes of any clinical measure. However, other important considerations include prognostic values for long-term clinical outcomes, feasibility in clinical settings, costs in time and financial terms, acceptance by patients, etc. Many valid and reliable measures are not useful for clinical management.

All quantitative clinical measures have advantages and limitations for diagnosis, prognosis, and monitoring patient status. In this brief critical review, advantages and limitations of the four types of measures commonly used in clinical research and care of patients with RA, joint counts, radiographic scores, laboratory data and patient questionnaire measures, are discussed (Table 1).

Joint Counts
A quantitative joint count has been used in many formats. In 1965, a 66/68 joint count was described in the glossary of the (ACR). This joint count includes the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal joints of the hands, metatarsal phalangeal
(MTP) and distal interphalangeal joints of the feet, shoulder, elbow, wrist, hip, knee, ankle, tarsus, and temporomandibular, sternoclavicular, and acromioclavicular joints (Table 1). Five abnormalities were described: tenderness, swelling, pain on motion, limited motion and deformity, in graded scales—generally with 4 categories: none, mild, moderate, or severe.11

The original ACR joint count has undergone various proposed modifications, which have included fewer joints, fewer variables, and simple binary “yes/no,” which appears as effective as graded descriptions of abnormalities.15 One version, the Ritchie Index, involves description of tenderness in 52 joints, with MCP and PIP joints scored as a group.12 Further reduced joint counts that include 36,13,28,14 and 42 joints have been reported.8 In most clinical trials and clinical research, including long-term observational studies, only swelling and tenderness are assessed, both of which are manifestations of inflammation. Most clinical trials use a count of 66 to 68 joints or 28 joints (Table 1).

The major advantage of a joint count in RA is that it is the most specific measure to assess patient status.15 Therefore, the joint count has traditionally been regarded as the most important measure for clinical trials to distinguish active from control treatments.16 Rheumatologists regard the joint

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Advantages and Disadvantages of Quantitative Measures in Rheumatoid Arthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Joint Counts:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. Most specific measure to assess RA</td>
<td></td>
</tr>
<tr>
<td>2. Most important measure in clinical trials</td>
<td></td>
</tr>
<tr>
<td>3. 28-joint count as useful in clinical trials as 66-68 joint counts</td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. Formal studies indicated poor reproducibility</td>
<td></td>
</tr>
<tr>
<td>2. Tedious to perform—interrupt visit</td>
<td></td>
</tr>
<tr>
<td>3. Joint counts more likely to improve with placebo than patient questionnaire scores—poor relative efficiency</td>
<td></td>
</tr>
<tr>
<td>4. Swollen and tender joint count improves over 5-15 years while radiographic and functional scores indicate progressive damage</td>
<td></td>
</tr>
<tr>
<td>5. Most visits to rheumatologist do not include a formal joint count (but usually include a careful joint examination)</td>
<td></td>
</tr>
<tr>
<td><strong>Radiographs:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. Excellent quantitative scoring systems—Sharp, van der Heijde, Larsen, Genant</td>
<td></td>
</tr>
<tr>
<td>2. Erosions are closest to pathognomonic sign in RA</td>
<td></td>
</tr>
<tr>
<td>3. Reflect cumulative damage of disease</td>
<td></td>
</tr>
<tr>
<td>4. No work for rheumatologist</td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. Quantitative score tedious to perform</td>
<td></td>
</tr>
<tr>
<td>2. Treatments should be initiated prior to erosions—MRI, ultrasound are more sensitive</td>
<td></td>
</tr>
<tr>
<td>3. Radiographic damage has little prognostic value for work disability, death and even joint replacement</td>
<td></td>
</tr>
<tr>
<td><strong>Laboratory Tests:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. Most important measure in most clinical situations, e.g., cholesterol, hemoglobin, creatinine, glucose, etc.</td>
<td></td>
</tr>
<tr>
<td>2. Many tests may be of value—CBC, ESR, CRP, RF, anti-CCP</td>
<td></td>
</tr>
<tr>
<td>3. No work for rheumatologist</td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. ESR and CRP—normal in 40% at presentation</td>
<td></td>
</tr>
<tr>
<td>2. Anti-CCP and rheumatoid factor—negative in 20-50% of patients</td>
<td></td>
</tr>
<tr>
<td>3. May be unchanged in many patients, despite clinical improvement</td>
<td></td>
</tr>
<tr>
<td>4. Treatment decisions are based primarily on clinical criteria</td>
<td></td>
</tr>
<tr>
<td>5. Lab tests have good prognostic value for radiographic damage, but poor prognostic value for work disability or death</td>
<td></td>
</tr>
<tr>
<td><strong>Patient Questionnaires:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. Significant correlation with joint counts, ESR, x-ray scores, physical measures</td>
<td></td>
</tr>
<tr>
<td>2. More reproducible and less likely to improve with placebo than traditional joint counts, ESR, x-ray scores, physical measures</td>
<td></td>
</tr>
<tr>
<td>3. As informative as the ACR 20, 50, 70, or DAS in clinical trials</td>
<td></td>
</tr>
<tr>
<td>4. Predicts work disability, costs, joint replacement, and premature death better than traditional joint counts, radiographs, and laboratory tests</td>
<td></td>
</tr>
<tr>
<td>5. Patient may serve as own “control” over time</td>
<td></td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td></td>
</tr>
<tr>
<td>1. May improve due to non-specific developments unrelated to RA—improve with progressive damage</td>
<td></td>
</tr>
<tr>
<td>2. Subject to cultural and linguistic differences</td>
<td></td>
</tr>
<tr>
<td>3. May be subject to “gaming” by certain patients to give desired answers</td>
<td></td>
</tr>
</tbody>
</table>
count as the most important measure of clinical status.\textsuperscript{77}

Several important limitations are seen in performance of the joint count. First, formal studies indicate that joint counts are poorly reproducible.\textsuperscript{18-22} Although reproducibility can be improved with training,\textsuperscript{21} joint counts remain less reproducible than, say, patient questionnaires. Among the ACR Core Data Set measures, improvement of patients in clinical trials who receive placebo or control treatment generally is greater for swollen and tender joint counts than for patient measures and laboratory tests,\textsuperscript{23} a phenomenon consistent with relatively poor reliability. The data suggest a possible unconscious bias on the part of an assessor to identify more involved joints at screening and fewer involved joints after 4 to 12 months of observation in a clinical trial.

Joint tenderness and swelling are sensitive to changes in inflammatory activity, and therefore are included in randomized controlled clinical trials. However, many studies now indicate that the number of swollen and tender joints often are improved over 5 to 15 years, while patients experience joint damage, functional declines, work disability, and premature death.\textsuperscript{24-33} Therefore, improvement in a tender and/or swollen joint count, even at a 20% level, may nonetheless be associated with further joint damage over time, manifested as joint deformity, radiographic progression, and functional declines. Recognition of this phenomenon has led to new emphasis on low disease activity or remission as goals of therapy for RA,\textsuperscript{34-38} rather than an improvement at a 20% (or even 50%) level, which may leave a patient vulnerable to long-term disease progression.

Joint counts are rather tedious to perform, requiring 1 to 4 minutes, during which time normal conversation between the patient and rheumatologist is interrupted. A joint count may occupy 15% to 25% of a 15- to 20-minute visit, during which time it might be preferable for the rheumatologist and patient to discuss drug therapy, psychosocial matters, work issues, and other patient concerns, rather than to obtain a detailed count of joints, which may have poor accuracy and not necessarily be reproducible. Although rheumatologists are taught that a joint count should be a component of each visit of a patient with RA and suggest that they are regarded as a desirable practice,\textsuperscript{39} joint counts are not performed at most visits of patients with RA.\textsuperscript{40}

Joint counts are needed to calculate a DAS, which may be a requirement for new therapies, including anti-tumor necrosis factor alpha (TNF) biologic therapies. However, very few rheumatologists have adequate joint count data in their records which could be used in a manner analogous to that seen in controlled clinical trials or for erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP), to describe clinical status over long periods. Even if tender and swollen joints counts were available at every visit, improvement in these measures would not necessarily indicate that the patient has not experienced progressive joint destruction.

In recent years, a number of self-report joint counts in different formats have been reported.\textsuperscript{31-50} The findings on self-report joint counts are correlated at levels of $r=0.44$ to 0.87, indicating statistically significant correlations, which are not, however, identical.\textsuperscript{44} These levels of correlation are in the range of, say, correlations of joint count with ESR or CRP. The degree of correlation between self-reported versus observer-reported joint count data is considerably higher for tender joints than for swollen joints. The widely used rapid assessment of disease activity in rheumatology (RADAR),\textsuperscript{41} adapted as a rheumatoid arthritis disease activity index (RADAI)\textsuperscript{42} includes only painful or tender joints. Self-report joint counts may be more prominently used by rheumatologists in clinical trials and clinical care over the coming decades.

**Radiographs**

Radiographs have several desirable features in assessment of RA. First, characteristic erosions or symmetric joint space narrowing may provide a finding that is closest to a “pathognomonic” sign in RA, with few “false positive” results when present. The radiograph reflects cumulative damage resulting from uncontrolled inflammation over time. Several excellent quantitative scoring systems are available, including the classical scoring systems developed by Larsen\textsuperscript{52,53} and Sharp,\textsuperscript{55,56} and modifications by van der Heijde,\textsuperscript{55,56} Rau,\textsuperscript{57} and others. At the same time, radiographs present several important limitations as a measure of clinical status in patients with RA. Radiographs change slowly in most people, so that at least 6 months to a year may be required to assess changes in an individual patient, although it is possible to ascertain change in a large group of patients over 12 weeks. Quantitative radiographic scores are tedious to perform and rarely are assigned outside of formal clinical studies. Indeed, most rheumatologists (and radiologists) have no experience in quantitative assessment of radiographs.

Modern treatment of RA suggests that many patients at this time are (appropriately) treated prior to radiographic damage, with normal radiographs that do not show any changes other than possible osteopenia. Magnetic resonance imaging (MRI) and ultrasound are more sensitive techniques,\textsuperscript{58,59} but are expensive, are not performed at many sites, and methods for quantitation remain under study.

Perhaps a most important limitation of radiographs is that scores have far less prognostic capacity for long-term outcomes in RA, including work disability, costs, and premature death, than patient questionnaires or joint count data.\textsuperscript{1,60-61} This observation remains relatively poorly recognized among general physicians and even rheumatologists, who may regard radiographic structural change as a most “objective” measure and therefore of great prognostic importance.

One explanation for the poor prognostic capability of radiographs in RA clinical research may be that most research studies of radiographs involve the hands and feet, which
appear less important in the prognosis of work disability or survival than large joints such as knees, hips, and shoulders.\textsuperscript{17} Radiographs of all joints are correlated significantly with one another, and with other measures of clinical status in patients with RA. Perhaps with data from large joints, the radiograph will provide greater prognostic value for long-term outcomes in RA. However, at this time, radiographic data has relatively limited value in the prognosis of outcomes of RA.

**Laboratory Tests**

Laboratory tests present the obvious advantage that most physicians and patients believe that data from the laboratory, such as serum cholesterol, creatinine, glucose, hemoglobin A1C, etc, provide the most definitive information in diagnosis and management of many diseases, according to a “biomedical model.”\textsuperscript{62} Furthermore, reductions in ESR and CRP are seen in all successful clinical trials of therapies of RA, which indicate efficacy of an active treatment compared to a control treatment. An ESR greater than 28 or abnormal CRP is often an inclusion criterion in clinical trials.\textsuperscript{63} An ESR less than 30 in a woman and less than 20 in a man is required to meet ACR remission criteria.\textsuperscript{64} A serologic test for anti–cyclic citrullinated proteins (anti-CCP) and rheumatoid factor (RF) can provide apparent reassurance that a diagnosis of RA is correct.\textsuperscript{65}

At the same time, laboratory testing in RA includes several important limitations. A primary limitation is that ESR and CRP are normal in about 40% of patients with RA, reported initially by Wolfe and Michaud \textsuperscript{66} and confirmed at two other sites.\textsuperscript{67} The data are taken from rheumatology treatment centers to which patients are referred by family practitioners and general internists. Such referral may be less likely if these tests are normal. Therefore, the conclusion that 40% of patients with RA have normal ESR and CRP may be an underestimate.

A further limitation is that while a decline in ESR or CRP is seen with clinical improvement in groups of patients, ESR and CRP tend to be stable in many patients, even with clinical improvement.\textsuperscript{68} Another limitation of laboratory tests is that they are often not available at the time when a clinical decision is made, leading to a derivative index known as a clinical disease activity index (CDAI), which includes swollen and tender joint counts, and physician/assessor and patient global assessments, but no laboratory tests.\textsuperscript{69} Of course, the clinician may arrange for the test in advance or contact the patient at a later date, but this is generally inconvenient and time-consuming.

Tests for RF and anti-CCP also have important limitations. In recent years, it has been recognized that anti-CCP may add sensitivity, enabling the clinician to make a diagnosis of RA in a patient with early undifferentiated arthritis.\textsuperscript{70} Indeed, some patients who have negative tests for RF have positive tests for anti-CCP. Nonetheless, more than one third and up to 40% of patients with RA do not have anti-CCP or RF.\textsuperscript{65} Although these antibodies may be of great value in further understanding of pathogenesis, they may be interpreted as indicating the absence of RA in many people who may need treatment.\textsuperscript{65}

In addition, “false negative” tests are seen for RF and anti-CCP. Indeed, if 1% of the normal population has a positive test for RF or anti-CCP, as has been found in population surveys, these tests are positive in people who do not have RA as often (or more often) than in people who have this disease.\textsuperscript{65} Obviously, tests are not ordered in all people, but the prevalence of musculoskeletal symptoms in the population is 15% (fibromyalgia is 5%). Therefore, many people with anti-CCP or RF have fibromyalgia—probably at least as many as have RA based on population data. Overall, although clinicians and patients often attribute disproportionate importance to laboratory tests in medical management of RA, their use has significant limitations.

**Patient Self-Report Questionnaires**

Patient self-report questionnaires, in which data are derived from patients rather than from imaging or laboratory tests, have become prominent in assessment and monitoring of patients with rheumatic diseases over the past two decades. The health assessment questionnaire (HAQ)\textsuperscript{71} published in 1980 was a major milestone in rheumatology and is the most widely used questionnaire at this time. The HAQ includes a scale of 20 activities of daily living (ADL) in eight categories to assess functional disability, with four patient response options: “without any difficulty” = 0, “with some difficulty” = 1, “with much difficulty” = 2 and “unable to do” = 3. The eight categories, each of which includes two or three ADL, address dressing, arising, eating, walking, bathing, reaching, gripping, and performing errands. The score for each category is the highest score among the two or three ADL within the category; 1 is added to the score if the patient uses aids or devices for that category. The total HAQ score is the mean score derived from eight scores with a range of 0 to 3 for each category. The HAQ also includes 10-cm visual analog scales to assess pain and global status, the other two patient-reported outcomes (PRO) measures in the Core Data Set.

The HAQ has been widely used in clinical trials and clinical research, but not in standard clinical care. Although the HAQ can be completed in 5 minutes in a waiting room, it is not easily scored by a busy clinician, based on a need to examine two sides of one page to identify an appropriate score among two or three for each category, including whether help or aid or device might be used in raising the score. Therefore, several modifications of the HAQ have been developed for ease of scoring in standard clinical care. A modified HAQ (MHAQ)\textsuperscript{72} and a multidimensional HAQ (MDHAQ)\textsuperscript{73} include all ADL on one side of one page, and can be scored rapidly based on the mean of 8 of 10 scores (rather than 20 without aids and devices) as with the HAQ. The MDHAQ includes two complex ADL—“Are you able to
walk 2 miles?” and “Are you able to participate in recreation and sports as you would like?” These ADL were added based on higher standards for rheumatology care at this time than in the 1970s when the HAQ was developed.

The MDHAQ also includes three psychological items concerning sleep, anxiety, and depression, queried in the standard patient-friendly HAQ format. Overall, 13 items are included in the MDHAQ. 10 concerning physical function and three concerning sleep, anxiety and depression. The MDHAQ includes visual analog scales for pain, global status, and fatigue, as well as length of morning stiffness, and change in status, on one side of one page. The questionnaire can be scanned by the physician (“eyeballed”) in less than 5 seconds. Visual analog scales are available as 21 circles rather than a 10-cm line, which facilitates scoring by not requiring a ruler. The 21 circles present a further advantage over a 10-cm line, in that photocopies or faxes may distort the 10-cm line and a score can be assigned on the basis of intervals of 0.5 and 1.0 from 0 to 10. Scoring templates are included at the right on the questionnaire so that physicians or their associates can score the 3 major items on the questionnaire in 15 seconds or less.

Despite extensive evidence of the value of patient questionnaires in standard clinical care, most rheumatologists do not use questionnaires in their own clinical patient care. The primary reasons cited included “takes too much staff time” and “the staff will not cooperate.” These impressions concerning patient questionnaires are generally derived from lengthy questionnaires designed for clinical research.

Patient questionnaires designed for research differ substantially from patient questionnaires designed for standard care. Research questionnaires may require 15 to 30 minutes for a patient to complete and generally cannot be reviewed (“eyeballed”) or scored easily during standard clinical care, adding a burden without adding to care. Patient questionnaires designed for standard care can save time for the clinician and improve the quality of patient visits, with the following features:

- Completed by most patients in 5 to 10 minutes;
- Scanned (“eyeballed”) by a clinician in 5 to 10 seconds;
- Designed to facilitate scoring, often with scoring templates on the questionnaire;
- Scored and available to enter into a flow sheet in 10 to 20 seconds; and
- Informed in patients with all rheumatic diseases.

Almost all the work is done by the patient, not the physician or the staff, and the physician spends only a few valuable seconds reviewing the data.

Disadvantages are also seen in the use of patient questionnaires. First, patient questionnaires are not as specific as joint counts to assess RA, particularly degree of inflammation. A patient’s pain or global score may improve or may be considerably higher or lower because of developments other than direct changes of RA. For example, severe back pain can lead to a much higher pain score, while an important psychosocial development, including good news about a desired goal or relative success may improve (lower) pain and global scores nonspecifically. Secondly, patient questionnaires may be subject to cultural and linguistic differences. For example, scores for pain were found to be higher in Hispanic patients compared to Caucasians compared to Asian patients. These cultural influences could hypothetically be overcome by measurement of physical function using performance measures such as grip strength and walking time. However, these tests currently are rarely used in clinical care despite their documented value in monitoring patient prognosis. Thirdly, patient questionnaires may hypothetically be subject to “gaming” on the part of the patient, who might regard certain responses as likely to engender a desired result. For example, a patient seeking work disability may indicate severe problems with physical function, pain, and global status, which may exaggerate the problems, while a patient who is trying to gain employment may underestimate such problems. Nonetheless, patient questionnaire data are quite effective in clinical trials, with relative efficiencies in the range of joint counts. Furthermore, patient questionnaires—not a joint count, radiographic score, or laboratory test—provide the most significant predictors of all severe long-term outcomes in patients with RA, including functional status, work disability, costs, joint replacement surgery, and premature death.

If quantitative data are recorded, an opportunity for documentation and more rational monitoring is gained, along with enhanced efficiency of patient care. If no data are recorded, this opportunity is lost and can never be replaced. It has been advocated that all rheumatologists would find it valuable to ask each patient to complete a questionnaire, such as an MDHAQ, at each visit in standard clinical care.

**Concluding Thoughts**

Most rheumatic diseases are characterized by the absence of a single quantitative measure that can serve as a pathognomonic diagnostic test and to assess and monitor clinical status in individual patients. Therefore, an extensive array of disease-specific quantitative measures and indices of these measures have been developed to quantitate patient status for clinical trials, clinical research, and clinical care. However, most of these measures remain research tools and are not applied to assess and monitor patient status in standard clinical care. From a pragmatic perspective, a simple patient questionnaire such as the MDHAQ, which has been found useful in patients with all rheumatic diseases, may provide a promising approach to introducing quantitative measurement into standard clinical rheumatology care.

**References**


15. Pincus T: The DAS is the most specific measure, but a patient questionnaire is the most informative measure to assess rheumatoid arthritis. J Rheumatol 2006;33:834-7.


40. Pincus T, Segurado OG: Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. Ann Rheum Dis 2006;23.
57. Pincus T, Sokka T: Prevalence of normal erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) on presentation of patients with rheumatoid arthritis (RA) at two rheumatology settings, one in the US and the other in Finland: Is a patient questionnaire a better quantitative measure of clinical severity? Arthritis Rheum 2005;52(9):S127.
60. van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, et al:


78. Pincus T, Wolfe F: Patient questionnaires for clinical research and improved standard patient care: is it better to have 80% of the information in 100% of patients or 100% of the information in 5% of patients? J Rheumatol 2005;32:575-7.


