Osteoarthritis and Cartilage (1997) **5**, 217–226

1063–4584/97/040217 + 217 $12.00/0

OSTEOARTHRITIS and CARTILAGE

REVIEW

Is conventional radiography suitable for evaluation of a disease-modifying drug in patients with knee osteoarthritis?

By Steven A. Mazzuca, Kenneth D. Brandt and Barry P. Katz

Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202-5103, U.S.A.

Introduction

Much interest currently exists in the possibility of pharmacologically modifying the disease process in osteoarthritis (OA). A number of pharmacologic agents have been shown to reduce proteolytic cartilage breakdown and/or stimulate matrix repair in animal models of OA. Such agents have been called 'chondroprotective' drugs, although it has been suggested recently that the preferable term is DMOAD (disease-modifying OA drug) [1]. Purported DMOADs range from empirical compounds, e.g., tissue extracts [2], to site-specific collagenase inhibitors designed, by structural analysis, to fit precisely into the catalytic site of the enzyme [3]. Excellent reviews of the subject by Howell *et al.* [4] and by Di Pasquale [5] have been published recently. Agents that have been reported to exhibit a DMOAD effect in animal models of OA include tribenoside, tamoxifen, diacerhein, chloroquin, hyaluronic acid, glucocorticoids, tranexamic acid, heparinoids, nonsteroidal anti-inflammatory drugs (NSAIDs), and doxycycline.

The evidence that progression of articular cartilage damage in animal models of OA may be modified pharmacologically heightens the need for better methods to detect early cartilage damage and assess progressive cartilage changes in humans with OA. When considering guidelines for testing DMOADs, committees of the International League against Rheumatism, Osteoarthritis Research Society and World Health Organization have recently emphasized that assessment of disease modification in OA requires measurement of changes in the anatomy of the joint [1], rather than merely in concentrations of biochemical or immunochemical 'markers' of joint damage or repair in serum, synovial fluid or urine [6]. Chondroscopy [7] and ultrasonography [8] hold potential for serial assessment of patients with OA. Although the resolution of magnetic resonance imaging (MRI) has steadily improved, surface coils and careful technique are required for detection of focal cartilage thinning and surface defects [9], and the ability of diagnostic MRI to provide the precise measurements of cartilage thickness needed to evaluate changes in this parameter over time in a large joint, such as the knee [10], has not yet been sufficiently validated; and the cost of MRI remains relatively high. The above committees considered that none of these alternative approaches has been sufficiently validated to permit a recommendation that it be used as an outcome measure at this time and that 'joint radiography, if standardized with respect to technique and views obtained, is the best technique available today for use in a large clinical trial of a DMOAD' [1]. In support of that view, double-contrast arthrography has recently confirmed the accuracy and precision of radiographic measurement of joint space width (JSW) for assessing cartilage thickness in the OA knee [11].

Prospects for identifying a DMOAD study population at risk for rapid progression

Research on the effects of DMOADs in humans has been hampered by uncertainty regarding the appropriate patient population for clinical trials of such agents. Clinical trials of DMOADs may be conducted in subjects with established OA, or alternatively, in subjects in whom the target joint is radiographically normal but at high risk for the

Submitted 25 June 1996; accepted 2 December 1996.
Supported in part by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (AR20582 and AR43348).
Address correspondence to: Steven A. Mazzuca, Ph.D., Rheumatology Division, Indiana University School of Medicine, Clinical Building, Room 492, 541 Clinical Drive, Indianapolis, IN 46202-5103, U.S.A.

rapid development of OA. The choice of a high-risk, but radiographically normal, knee as a target joint may be influenced also by the reasonable expectation that a DMOAD effect will be more readily demonstrated in an OA joint in which pathologic changes are mild than in a joint in which they are much more advanced [6, 7]. Although that contention is hypothetical, it is supported by the results of studies of doxycycline therapy in a canine cruciate-deficiency model of OA: when treatment was delayed until cartilage lesions were well-established, cartilage damage progressed more slowly than in the untreated OA controls—but damage was clearly present; when treatment commenced before gross or histologic cartilage changes had developed, the severity of OA was strikingly reduced, and in some cases, the cruciate-deficient knee remained grossly normal, whereas untreated controls showed extensive end-stage changes of OA [12, 13].

Recent epidemiologic data from Chingford, England, strongly suggest that a select population of obese, middle-aged women with radiographic evidence of unilateral knee OA may afford an opportunity to observe (and prevent) the rapid onset of OA in a joint at high risk for OA—the radiographically normal contralateral knee [14]. Forty-seven per cent of this group (15 of 32) progressed from unilateral to bilateral disease within 2 years. It should be noted, however, that OA progression in the Chingford Study was determined chiefly on the basis of osteophyte growth; only two of the 15 subjects, whose OA was considered to progress, showed narrowing of JSW in addition to osteophyte growth. However, radiographs in this study were obtained with knees in full extension, and strict attention was not paid to standardizing the position of the knee on repeated examinations. (This probably accounts for the fact that osteophytosis seen in one subject at baseline was not apparent in the repeat radiograph.) Furthermore, OA progression in the Chingford Study was adjudicated without quantitative measurement of JSW [14]. The degree to which the development of OA in the high-risk contralateral knee of this select group was accompanied by joint space narrowing (JSN) is, therefore, unknown.

With respect to the power of a DMOAD trial, the advantage of rapid OA progression in a high-risk knee which is radiographically normal at baseline may be multiplied by increased precision with which JSW can be measured in such a joint, in comparison with a knee with established bony changes of OA. Positioning—and reproducibility of repositioning—of the knee of a patient with severe OA, with constraining osteophytes and capsular fibrosis, may be more difficult than that of a knee in a subject with less-advanced changes. Furthermore, osteophytosis, subchondral sclerosis and cartilage calcification can obscure the margins of the joint space and make measurements of JSW less reproducible than in the radiographically normal knee. This advantage is theoretical, however; the degree to which the severity of OA impairs the reproducibility of JSW measurements is unknown.

Key variables in knee radiography

Obtaining a satisfactory plain radiograph of the knee is not a trivial matter. As discussed recently by Buckland-Wright [15, 16], several steps in the production of the conventional plain knee radiograph make quality control difficult: for example, the technician performing the examination may have developed his or her own preference for positioning of patients, especially when faced with someone who is markedly obese or who otherwise has difficulty standing or walking. Idiosyncratic variation in technique can lead to unintended variation in the degree of knee flexion, misalignment of the X-ray beam, and magnification of the radiographic image of the joint.

DEGREE OF KNEE FLEXION

Since publication of the classic monograph by Ahlback [17] and the supporting paper by Leach *et al.* [18], the standard knee radiograph has typically been obtained with the patient standing and the joint fully extended. However, in patients with advanced OA, radiographs obtained with the knee fully extended tend to overestimate the amount of cartilage remaining on the articular surface. Exaggeration of the magnitude and variability of JSW in radiographs taken with the knee in the extended position (see below) is caused by the femur 'riding up' on cartilage at the anterior margin of the tibia [19]. In contrast, the semiflexed position more closely approximates the normal anatomic standing position of the tibiofemoral joint than the fully-extended view [20, 21]. Arthroscopic findings confirm that the semiflexed view is more likely than a full-extension view to display the region of the tibiofemoral compartment in which cartilage damage in OA is most prevalent [19].

Various specific degrees of flexion have been suggested as providing increased sensitivity for demonstrating cartilage loss in the standing knee

radiograph. For example, Resnick and Vint [22] described six patients in whom the standing anteroposterior (AP) view underestimated the degree of joint space loss evident on a 'tunnel' view, i.e., knee flexed to 60° or 70°. Messieh *et al.* [19] compared the standing AP view with a posteroanterior view with 30° flexion and found 10 patients who had normal joint space on the fully-extended view, but marked narrowing on the flexion view. Alternatively, Rosenberg *et al.* [23] have suggested that 45° of knee flexion is superior to either 10° or 30° of flexion for detecting JSN. While the optimal degree of flexion for knee radiographs may be debated, the need for standardization with respect to flexion was demonstrated recently by Ravaud *et al.* [24], who found 10–12.5% variation in estimates of tibiofemoral JSW from repeated radiographs of normal knees in which knee flexion was manipulated by as little as 5°.

Buckland-Wright has recently reported [25] that the precision (i.e., reproducibility of repeated measures) of JSW can be improved markedly by use of fluoroscopy to standardize knee flexion. Unlike the techniques in the above studies, which required fixed degrees of flexion, Buckland-Wright's protocol specifies the identification by fluoroscopy of the degree of flexion for each subject that results in superimposition of the anterior and posterior lips of the tibial plateau. This degree of flexion results in the positioning of the plateau in a plane parallel to the floor. This radiographic technique is further standardized by having the subject, while still under fluoroscopy, rotate the foot (with the heel fixed) until the tibial spines are centered below the femoral notch. The outline of the foot is traced on a film jacket to facilitate repositioning of the knee during repeat examinations [25]. Deviations in foot rotation of as little as 15° may result in significant variation of estimates of JSW in repeated radiographs [24].

X-RAY BEAM ALIGNMENT

The position of the central ray of the X-ray beam relative to the center of the joint, i.e., the joint space, is another important variable. The X-ray beam is tangent to the plane of the joint at a single point on the radiograph. All other points in the image are distorted because the X-ray beam diverges in a cone-shaped manner around the central tangent ray. Therefore, a change in the angle of the beam will result in distortion of the relationships of the articular margins on the radiograph projection. This distortion increases with increasing angulation (i.e., increasing distance from the central ray). The degree of misalignment of the beam necessary to alter results is not large; Fife *et al.* [26] found a 17% decrease in JSW when the X-ray beam was displaced by 1 cm below its original alignment centered at the mid-point of the patella. Ravaud *et al.* [24] also detected significant reductions of JSW in repeated measurements of JSW in which the angle of inclination varied 5° from a line parallel to the joint space.

RADIOGRAPHIC MAGNIFICATION

Although radiographic magnification is not generally taken into account, the distance between the center of a joint and the X-ray film will affect the degree of magnification of the radiographic image. The distance between the center of the joint and the X-ray film can be large, and is influenced by factors such as obesity (common in subjects with knee OA) and restriction of joint movement because of pain, osteophytosis or soft tissue contracture. In an assessment of standard radiographs of the knee obtained in the standing extended view, Buckland-Wright *et al.* [25] found magnification of JSW ranging from 9–35% relative to a fixed magnification marker. Buckland-Wright's protocol for standardized knee radiography requires that a magnification marker (i.e., a 5 mm ball encased in plexiglass or another semi-radiolucent material) be affixed with tape to the skin overlying the head of the fibula [25]. Any measurement of JSW from that image can be corrected for the degree of magnification apparent in the image of the marker.

Reproducibility of quantitative radiographic measurements

The validity of measurements of the radiographic features of OA is dependent not only upon image quality, but also upon the reproducibility of the mensural procedure [27, 28]. Some investigators have reported estimates of JSW without describing their methods in detail [29, 30]. Others have used a ruler [31, 32] or calipers [1, 33] and/or a magnifying lens with a fitted graticule [34]. Although regarded as more precise than semi-quantitative scoring systems [35], such as the Kellgren and Lawrence scale [36], the reproducibility of these quantitative methods (i.e., the degree to which repeated examinations of the same joint by the same observer, or by different observers, yield the same estimate of JSW) is subject to observer error.

The reproducibility (or precision) of a mensural procedure for JSW measurement can be expressed as the standard deviation (S.D.) of repeated measurements of JSW within the same joint(s)—also referred to as the standard error of measurement (SE_m). To standardize the scale of precision estimates, the reproducibility of repeated measurements of the same individuals is frequently quantified as a coefficient of variation (CV), the ratio of the S.D. to the mean of repeated measurements. Lack of attention to standardization of the technique in routine clinical knee radiography can result in a CV as high as 20% for repeated manual measurements (i.e., with a ruler) of JSW made directly on fully-extended, AP views of the same subjects [37]. In contrast, Lequesne [1, 33] has described a highly-standardized method of manual JSW measurement in which the points of a pair of calipers are used to measure the interbone distance on a radiograph. The points are then used to prick a sheet of paper on which the distance between the pinpricks is measured with a 10× magnifying lens fitted with a 10 mm graticule with 0.1 mm divisions. The intra-observer CV for repeated measures with this technique was 3.8% [25].

The magnitude of human error that may be present even in highly-standardized manual measurement of JSW was illustrated recently in a study of 25 patients with knee OA and 10 normal controls [25], in which four repeated JSW measurements, made manually with a calipers on conventional plain radiographs of the knee (i.e., fully-extended, AP view, no correction for magnification) were compared with JSW measurements made with specialized edge-detection computer software [38] in digitized radiographic images of the same subjects, obtained during four examinations with standardized positioning in semiflexion and correction for image magnification [25]. For the 10 radiographically normal knees, the variability (i.e., S.D.) of repeated manual measurements from conventional extended view radiographs [median S.D. = 0.31 mm, 95% confidence interval (CI) = 0.14–0.48 mm], in relation to mean JSW, yielded a CV of 6.4%. Computerized measurement alone did not reduce the error variation in JSW measurement caused by lack of standardization of the position of the knee (median S.D. = 0.29 mm, CV = 6.2%), while addition of magnification correction resulted in only modest further reduction of the S.D. of repeated measures (median S.D. = 0.21 mm, CV = 5.4%). However, when the automated, magnification-corrected measurement system was applied to JSW measurement on radiographs of normal knees in the semiflexed position, the S.D. was decreased by nearly half (median S.D. = 0.11 mm, CV = 3.2%, $P < 0.01$).

In the above analysis, application of automated JSW measurement and magnification correction to images of 25 OA knees in optimal semiflexed position significantly ($P < 0.01$) reduced error variance across four repeated measurements (median S.D. = 0.19 mm, CV = 5.5%) relative to manual measurement in conventional radiographs (median S.D. = 0.30 mm, CV = 6.4%). Although the improvement was not as great as that observed in examinations of radiographically normal knees, automated, magnification-corrected measurement in studies of the semiflexed OA knee reduced error variation by about one third, compared with manual measurement of JSW without these refinements. Even greater improvement in the reproducibility of automated, magnification-corrected JSW measures was reported with the use of high-definition 5×macroradiographic images (median S.D. = 0.06, 95% CI = 0.04–0.08, CV = 1.6%) [25]. However, microfocal radiography of joints is not widely available in clinical radiology departments in the United States, and therefore, could not be readily employed today in a multicenter clinical trial of a DMOAD in this country.

Quantitative radiographic studies of knee OA

Because disease modification in OA should preserve articular cartilage and slow the rate of JSN, two questions are key to the design of a clinical trial of a DMOAD in which the primary outcome is based on serial measures of JSW: what is the mean and variability of JSW in the target population? At what overall rate does JSW narrow over time? Here, too, interpretation of published estimates of these population parameters must be tempered by knowledge of the subjects studied and the radiographic and mensural procedures employed.

JOINT SPACE WIDTH

Table I summarizes published estimates of minimum JSW (mean ± S.D.) in the medial tibiofemoral compartment in normal subjects and patients with knee OA [31, 33, 39–41]. Although variability in knee positioning confounds efforts to compare the results of these studies, the largest estimates of JSW were reported when the radiograph was obtained with the knee in extended position. Indeed, the mean JSW in OA knees in extended position estimated by Kirwan *et al.* [39] (4.45 mm) was larger than that found by Buckland-Wright [40] for normal knees in the standardized,

semiflexed position (4.06 mm). Relatively-large mean values for OA knees in the extended position also have been reported by others (3.56 mm by Lequesne [33], 4.89–5.06 mm by Mazières [31]).

Even more striking than variance in the mean JSW—and more relevant to sample size requirements in DMOAD trials—the S.D. of JSW measurements of OA knees in the extended view (2.01 mm) found by Kirwan *et al.* [39] was substantially larger than those reported by Buckland-Wright for the semiflexed view of either normal or OA knees (Table I). In the latter study, the S.D. for computerized measures of tibiofemoral JSW in the normal semiflexed knee was 0.79 mm, while that for OA knees with a JSW greater than 3.0 mm was 0.50 mm, and for knees with JSW of 1.5–3.0 mm, S.D. was 0.58 mm [40]. Similarly, S.D.s of 0.74–0.81 mm were reported for 29 OA knees of participants in the placebo group of a clinical trial of NSAID, in which baseline JSW in all patients was greater than 2 mm (Table I) [41].

In summary, conventional radiography and manual measurement of JSW appear to offer larger, more variable estimates of JSW than does the combination of computerized measurement of JSW in semiflexed views of the knees. The implication of the data presented in Table I is that, as designers of DMOAD trials decide upon the magnitude of the between-group difference in JSW to be detected, that difference will be a smaller proportion of within-group variability for conventionally derived outcome measures than for highly-standardized and computerized measurements. In theory, therefore, conventional radiography and manual measurement of JSW will reduce the power of a DMOAD trial.

RATE OF JSN

The magnitude of the difference in JSW between treatment and placebo groups to be detected in a DMOAD trial will be predicated on assumptions regarding the overall rate of JSN expected in the placebo group and the degree to which JSN will be slowed in the treatment group by the DMOAD. An assumption about the overall linear rate of JSN for the placebo group cannot be derived easily. The course of cartilage loss in individuals is likely to be marked by intermittent intervals of progression, of varying duration. It also is reasonable to expect biological variability among subjects in the rate of JSN. JSN in the general elderly population or in people with untreated, early knee OA (who are likely to be identified by community-based recruiting efforts) may be slower than that in a clinic population of OA patients with established disease. Therefore, designers of DMOAD trials will need to consider both the overall rate and variability of JSN in potential populations.

Table II summarizes published estimates of the rate of medial tibiofemoral JSN in patients with knee OA. Unfortunately for comparative purposes, previous studies of JSN in OA knees have varied markedly from one another in many important respects: sample size, source of the sample (clinic vs community), severity of OA, heterogeneity of patient characteristics that may relate to OA progression (e.g., age, sex, obesity), radiographic technique, mensural procedures and duration of observation. Because of these differences in methodology, estimates of the annual rate of JSN in OA knees in the seven studies listed in Table II range from 0.06 mm/year [42] to 0.60 mm/year [39, 43]—10-fold variation—far too large to be

Table I
Published estimates of medial tibiofemoral joint space width (JSW) at the narrowest point

Study (first author)	Knee position/ method of measurement	Number of knees	OA severity	Mean JSW, mm (±S.D. when reported)
Kirwan [39]	Extended/ruler	150	Variable	4.54 ± 2.01
Lequesne [33]	Extended/calipers	22	Variable	3.56
Mazières [31]	Extended/ruler	167	Variable	right: 4.89 left: 5.06
Buckland-Wright [40]	Semiflexed/computer*	14	Normals	4.06 ± 0.79
		90 { 56	Minimal†	3.77 ± 0.50
		90 { 19	Moderate†	2.04 ± 0.58
		90 { 15	Marked†	1.64 ± 0.79
Buckland-Wright [41]	Semiflexed/computer*	29	JSW > 2 mm	3.39 ± 0.74 (baseline) 3.36 ± 0.77 (6 months) 3.35 ± 0.73 (12 months) 3.27 ± 0.81 (18 months)

*Computerized measurement of digitized 5×macroradiographic images. †Minimal OA severity = JSW > 3 mm; moderate OA = JSW 1.5–3.0 mm; marked OA = JSW < 1.5 mm. NA = not available.

Table II
Published studies of medial tibiofemoral joint space narrowing (JSN) in knees with bony changes of OA

Study (First author)	Number of knees	Average follow-up (years)	Observed JSN over follow-up period, mean ± s.d. (mm)	Annual rate of JSN (mm/year)	Point of measurement
Ravaud [43]	55	1	*Reader 1* *Reader 2*		
			0.42 ± 1.1 0.47 ± 0.9	0.42–0.47	Narrowest
			0.45 ± 1.2 0.60 ± 1.1	0.45–0.60	Midpoint
			0.26 ± 1.2 0.37 ± 1.1	0.26–0.37	10 mm from medial extremity
Mazières [31]	167	1	Left knee: 0.26 ± 1.6	0.26	Not reported
			Right knee: 0.13 ± 1.3	0.13	Not reported
Buckland-Wright [41]	17	1.5	0.275 ± 0.268‡	0.183	Narrowest
Kirwan [39]	150	3	1.85 ± 1.88	0.60	Narrowest
Lethbridge-Çejku [42]	36†	4	Women: 0.24 ± 0.56*	0.06	Not reported
			Men: 0.36 ± 0.68*	0.09	
Lequesne [33]	22	3.9	1.01 ± 0.78*	0.26	Midpoint
	24	7.7	1.69 ± 1.62*	0.22	Midpoint
Neuhauser [44]	40†	8.1	0.81 ± 1.05*	0.10	Narrowest

*Cumulative measurements were not reported. Observed JSN is extrapolated from reported annual averages of JSN.
†A subset of a population-based sample; all other studies reflect JSN in clinic-based samples of OA patients.
‡Signal knees: the symptomatic OA knee of each subjects for which JSW was nearest to (but still greater than) 2 mm at baseline [personal communication from the authors].

conclusive for the planning of a placebo-controlled DMOAD trial without consideration of the patient characteristics and methodologic features that may, in part, be responsible for such variability. It is noteworthy, however, that both Mazières [31] and Lequesne [33], in studies that differed markedly with respect to sample size and duration, obtained estimates of 0.26 mm/year in patients with established knee OA. In fact, 0.26 mm/year is the median of the estimates of annual rate of JSN presented in Table II.

Table II lends support to the hypothesis that subjects with knee OA recruited from the community will have a slower overall rate of OA progression than those from clinic populations. Estimates of JSN in the population-based Baltimore Longitudinal Study of Aging [42] and the Framingham Study [44] yielded rates at the lower end of the spectrum (0.06–0.10 mm/year). The remaining studies in Table II, all of which used clinical samples, had consistently higher values for annual JSN (i.e., generally more than 0.20 mm/year).

Additional support of this hypothesis can be found in the study by Buckland-Wright *et al.* [41], who measured, on average, reasonably rapid progression (0.183 mm/year) in the index or 'signal' knees (i.e., the symptomatic knee with JSW nearest to, but still greater than 2 mm at baseline) of 17 OA patients in the placebo group of a clinical trial of an NSAID (J. C. Buckland-Wright, personal communication). However, in separate analyses of all 34 knees of the participants in the placebo group, the mean rate of narrowing was approximately 0.08 mm/year among knees with greater than 50% of JSW at baseline, but about 0.25 mm/year in knees with less than 50% of JSW at baseline [41]. Although sample size limitations preclude firm conclusions, the results of analyses of subgroups by Buckland-Wright [41] are consistent with the hypothesis that JSN accelerates with disease severity.

Consequently, the radiographic inclusion criteria employed by a DMOAD trial (e.g., to require that subjects have, e.g., at least 2 mm of tibiofemoral JSW at baseline) and the recruitment strategy (i.e., from clinic and/or community sources) dictated by sample size requirements should be taken into account when developing an expectation for the overall rate of JSN in the placebo group. Furthermore, these data suggest that randomization of participants to treatment groups should be stratified by recruitment source to balance the effect of possibly differing base rates of JSN.

As noted above, the variance of JSN over a given interval will have two components: error variation (which can be minimized by use of mensural procedures with superior reproducibility) and true biological variation. Measurement error will be distributed randomly and will be of a magnitude dictated by the reproducibility of repeated measures [45]. With regard to biological variation, there is no question that in most patients, joint damage is ultimately progressive. In sufficient time, the disease will progress; and mean joint space will diminish (i.e., cumulative JSN will

increase) in the vast majority of patients. On theoretical grounds, therefore, one might hypothesize that the longer the period of observation, the smaller the variability of JSN will be in relation to the mean.

Indeed, among the studies presented in Table II, the two of shortest duration (i.e., 1-year) [31, 43] estimated the S.D. of JSN to be 200–1000% of the mean. In contrast, in four of five studies of longer duration (i.e., ≥ 3 years) [33, 39, 42, 44], the S.D. was similar to, or even smaller than, the mean value for JSN (i.e., $\pm$ 30%). This apparent inverse relationship between the duration of observation and ratio of S.D. to mean JSN suggests that, given sufficient time, true JSN will become sufficiently great as to overshadow measurement error and biological variability.

However, pragmatic considerations will require that DMOAD studies be as short as possible. Time is money, and, the longer the study the greater the problems with subject retention are likely to be. Therefore, any clue as to how to shorten the time required for reduction of the within-group variability in JSN in relation to the magnitude of mean JSN (and between-groups difference in mean JSN) is worth noting. In the only study in Table II to employ computerized measurement of JSW from digitized (albeit macro-) radiographic images of OA knees in standardized semiflexed position [41], the S.D. of JSN (0.268 mm) was comparable with the mean cumulative JSN (0.275 mm) after only 18 months of observation (J. C. Buckland-Wright, personal communication). Although, again, a caveat must be offered because of the small number of observations on which these parameters are estimated. This study, nevertheless, illustrates the potential benefit of increased precision in measurement of radiographic outcomes in a DMOAD trial.

Implications for the design of a DMOAD trial

EFFECT OF MEASUREMENT PRECISION ON SAMPLE SIZE

Because the SE_m (i.e., the S.D. of repeated measurements) for any outcome variable in a clinical trial is a component of within-group variability (i.e., the 'error term' used in statistical tests of study hypotheses), the precision of mesurement can have a large impact on the power to detect a significant DMOAD effect. If the outcome at the end of the trial were a single estimate of JSW, rather than the magnitude of JSN, the effect of radiographic or mensural imprecision could be readily calculated. Power calculations for JSW would be based upon its variance, which is the sum of two components: the biological variation among subjects and the variation associated with measurement error. As demonstrated by Bloch [45], a decrease in measurement error relative to biological variation will directly reduce the error term used in the statistical analysis of the outcome variable and will increase the power of the trial accordingly.

In comparison with JSW, the variance of JSN and its relationship to SE_m is more complicated, because JSN is the numerical difference between two measurements of JSW—both of which will contain measurement error. As with JSW, imprecision of measurement inflates the error term directly for statistical tests of JSN. Moreover, measurement error reduces the correlation between repeated estimates of JSW, further increasing unaccountable within-group variation in JSN.

To illustrate the effect of measurement precision on power, we have performed sample size calculations for a hypothetical DMOAD trial using respective SE_m estimates appropriate for manual measurement of JSW in conventional extended view radiographs and for computerized, magnification-corrected measurement of JSW in semiflexed views of normal and OA knees [25]. In this illustration, we have assumed a conservative annual rate of JSN (0.20 mm/year), study duration of 2 years, and a value for the S.D. of JSN equal to 100% of the mean JSN in the placebo group (i.e., 0.40 mm). To simplify this exercise, we have used a single value (0.80) for the correlation between the baseline and end-of-trial estimates of JSW measured without error. Sample size estimates were obtained using PC-Size software [46].

The remarkable potential for improved measurement precision to increase the power of a placebo-controlled DMOAD trial to detect, e.g., a 30% reduction of the rate of JSN is shown in Fig. 1.

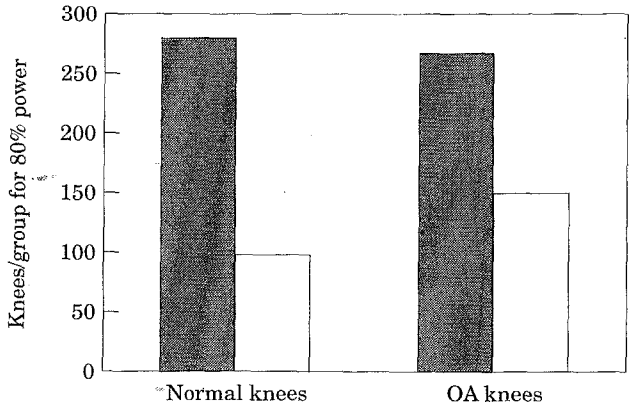

FIG. 1. The effect of measurement precision on sample size in a DMOAD trial designed to detect a 30% drug effect in which the rate of JSN is 0.2 mm/year in the placebo group. (▩) Extended view, manual measurement; (□) semiflexed view, automated measurement.

Compared with a trial in which JSW is measured manually in conventional extended views of radiographically normal knees, computerized measurement would reduce the sample size needed to achieve 80% power from 280 to 97 participants/group. For knees with established OA, in which computerized measurement is less precise than in normals, the required sample size would decrease to a lesser extent (i.e., from 267 to 149 participants/group). Nevertheless, a proportional (44%) reduction in recruitment, screening and post-randomization costs would be substantial.

We emphasize that this exercise is intended for illustrative purposes and that these sample size estimates are not meant to be used in the planning of DMOAD trials. Actual sample size calculations will be very sensitive to many study-specific parameters, including characteristics of the target population, sampling strategy, radiographic methods, outcome variables and anticipated effect size.

EFFECT OF THE RATE OF JSN ON SAMPLE SIZE AND STUDY DURATION

The same caveats apply to the following illustration of the effect of the rate of JSN on the design of a DMOAD trial. Fig. 2 illustrates the large effect (and the potential for underpowering a DMOAD trial) that small differences in assumption about the rate of JSN in the placebo group would have on the size and duration of a hypothetical trial designed to detect a 30% reduction in the rate of JSN. For the sake of this illustration, we have taken a range of rates of JSN in the placebo group (0.10–0.25 mm/year) from the low end of the spectrum presented in Table II. The illustration also presumes that the SD of JSN is 0.35 mm after 24 months (i.e., 70–175% of the expected mean JSN in the placebo group across the range of annual rates).

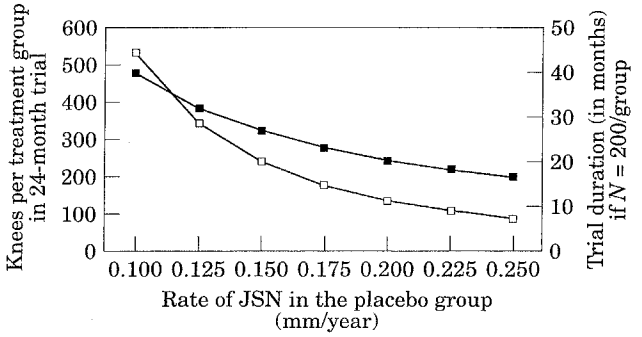

FIG. 2. The effect of the rate of JSN in the placebo group on sample size and duration of a DMOAD trial designed to detect a 30% drug effect (□) sample size; (■) duration.

In a 24-month DMOAD trial, if the rate of JSN in the placebo group were 0.25 mm/year and treatment with the drug reduced the rate by 30%, 87 participants would be required to complete the trial in both the treatment group and placebo-control groups to have 80% power to detect a statistically significant ($P < 0.05$) drug effect. However, for every decrease of 0.025 mm/year in the base rate of JSN on the placebo group, sample size requirements grow 23–56%. At the extreme, if the mean rate of JSN were 0.10 mm/year, the required sample size would be 535 participants/group sixfold greater than if the rate of JSN were 0.25 mm/year. These sample size estimates are about 25% larger than those offered recently by Buckland-Wright [16] for a trial of similar duration, power, and assumed rate of JSN in the placebo group. The reason for this discrepancy is that our assumed S.D. for JSN is larger, in absolute terms and in relation to the mean, than the value of 0.245 mm used by Buckland-Wright [16].

Alternatively, if the effect of the annual rate of JSN on study duration, rather than on sample size, is considered, Fig. 2 illustrated that in a trial with 200 knees/group, if the overall rate of JSN in the placebo group is 0.25 mm/year, a 30% drug effect should be detectable with 80% power after 16 months of treatment. If the size of the sample is kept constant, each successive 0.025 mm/year reduction in JSN in the placebo group would extend the minimum length of the trial by 2–8 months. For a rate of JSN of only 0.1 m/year in the placebo group, the projected duration of the trial would rise to 40 months.

Conclusion

Is conventional radiography suitable for evaluation of a DMOAD in patients with knee OA? Based on our knowledge of (1) the reproducibility of measurement of JSW, (2) the nature and rate of OA progression in potential target populations for a DMOAD trial and (3) the pragmatics and costs of clinical trials, it is apparent that conventional radiographic technique, in concert with manual measurement of JSW, cannot be endorsed as methods for quantitating primary outcomes in DMOAD trials. Failure to standardize crucial elements of radioanatomic positioning and to automate measurement of JSW has been shown to introduce significant and probably insurmountable error variation to estimates of JSW [24, 25]. The lack of standardization with respect to radiographic methods in previous studies of OA progression are, in all likelihood, chiefly responsible for the highly-variable estimates of popu-

lation parameters (i.e., mean and variance of JSW) and the rate of OA progression that these studies offer [31, 33, 39–44].

Descriptions of highly-standardized protocols for knee radiography [16, 25] and for measurement of JSW [33, 38] are available. We have illustrated how the level of precision achievable with fluoroscopically assisted flexion of the knee and rotation of the foot, with computerized, magnification-corrected measurement of JSW can, in theory, decrease the sample size required to detect a DMOAD effect. Quantitation of JSW notwithstanding, a higher standard of reproducibility of radioanatomic positioning in knee radiography also will increase our sensitivity to detect progression of bony features of OA (e.g., osteophytes).

We also have presented limited, but encouraging, data suggesting that the level of precision achievable today can reduce to 18 months the time required before power calculations for hypothesis testing can be predicated on the assumption that within-group variability of JSN is near the mean of JSN in the placebo group. In contrast, among studies using conventional radiography and manual measurement, the minimum interval required for this level of statistical power may be twice as long.

The precision associated with this level of standardization comes at a cost: special training of technologists in radioanatomic positioning of the subject (and assurance that the technologists maintain the acquired skills), use of fluoroscopy and image digitization, computer software and hardware. Nevertheless, if additional studies establish high inter-technologist reproducibility of radioanatomic positioning of the knee—a necessity for a multicenter trial—a powerful argument can be made for the use of this system in future DMOAD studies.

Finally, the design of an efficient DMOAD trial would be assisted by the identification of inclusion criteria that would increase the likelihood, if not the rate, of OA progression among randomized participants. Considerable interests exists, therefore, in biochemical and immunochemical analyses of synovial fluid or serum [47] and in imaging procedures [48] which may identify individuals who are at greater risk for the progression of OA than others. To permit accurate evaluation of reports of the use of such surrogates, if radiography is used to determine the progression of OA, it is important that detailed information is provided concerning the mensural procedures employed and their precision. In the absence of such information, considerable risk exists that the potential use of such surrogates may be misjudged.

References

1. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes JC, Pelletier JP, Altman R. Guidelines for testing slow-acting and disease-modifying drugs in osteoarthritis. J Rheumatol 1994;21(Suppl 41):65–71.
2. Burkhart D, Ghosh P. Laboratory evaluation of antirheumatic drugs as potential chondroprotective agents. Sem Arthritis Rheum 1987;17:3–34.
3. Lovejoy B, Cleasby A, Hassell AM, Luther MA, Weigl D, McGeehan G, Lambert MH, Jordan SR. Structural analysis of the catalytic domain of human fibroblast collagenase. In Greenwald RA, Golub LM, Eds. Inhibition of Matrix Metalloproteinases: Therapeutic potential. New York, NY: New York Acad Sci. 1994:375–8.
4. Howell DS, Altman RD, Pelletier, J-P, Martel-Pelletier J, Dean DD. Disease modifying antirheumatic drugs: current status of their application in animal models of osteoarthritis. In Kuettner K, Goldberg V, Eds. Osteoarthritic Disorders. Rosemont, IL: American Academy of Orthopedic Surgeons 1995:365–77.
5. Di Pasquale G. Pharmacologic control of cartilage degeneration in osteoarthritis. In Woessner JF, Howell DS, Eds. Joint Cartilage Degradation: Basic and Clinical Aspects. New York: NY: Marcel Dekker 1993:475–501.
6. Lohmander S. Molecular markers of cartilage turnover. In Kuettner K, Schleyerbach R, Peyron JG, Hascall VC. Articular Cartilage and Osteoarthritis. New York, NY: Raven Press 1992:653–7.
7. Ayral X, Dougados M, Listrat V, Bonvarlet J-P, Simmonet J, Poiraudeau S, Amor B. Chondroscopy: a new method for scoring chondropathy. Sem Arthritis Rheum 1993;22:289–97.
8. Myers SL, Dines K, Brandt DA, Brandt KD, Albrecht ME. Assessment by high frequency ultrasound of articular cartilage thickness and osteoarthritic changes. J Rheumatol 1995;22:109–16.
9. Buckwalter KA. Imaging in OA and crystal deposition diseases. Curr Opin Rheumatol 1993; 5:503–9.
10. Hodler J, Trudell D, Pathria MN, Resnick D. Width of the articular cartilage of the hip: quantification by using fat-suppression spin-echo MR imaging in cadavers. Am J Roentgenol 1992;159:351–5.
11. Buckland-Wright JC, Macfarlane DG, Lunch JA, Jasani MK, Bradshaw CR. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. Ann Rheum Dis 1995;54:263–8.
12. Yu LP Jr, Smith GN, Brandt KD, Myers SL, O'Connor BL, Brandt DA. Reduction of the severity of canine osteoarthritis by prophylactic treatment with oral doxycycline. Arthritis Rheum 1992;35:1150–9.
13. Brandt K, Yu L, Smith G, O'Connor B, Myers S. Therapeutic effect of doxycycline in canine osteoarthritis. Osteoarthritis Cartil 1993;1:14.
14. Spector TD, Hart DJ, Doyle DV. Incidence and progression of osteoarthritis in women with unilateral knee disease in the general population: the effect of obesity. Ann Rheum Dis 1994;53:565–8.
15. Buckland-Wright C. Protocols for precise radioanatomical positioning of the tibiofemoral and

patellofemoral compartments of the knee. Osteoarthritis Cartil 1995;3(Suppl A):71–80.
16. Buckland-Wright C. Quantitation of radiographic changes. In Brandt KD, Doherty M, Lohmander S, Eds. Textbook of Osteoarthritis. London: Oxford University Press. In Press.
17. Ahlback S. Osteoarthritis of the knee: a radiographic investigation. Acta Radiol 1968;1(Suppl.):277.
18. Leach RE, Gregg T, Siber FJ. Weight bearing radiography in osteoarthritis of the knee. Radiology 1970;97:265–8.
19. Messieh SS, Fowler PJ, Munro T. Anteroposterior radiographs of the osteoarthritic knee. J Bone Joint Surg 1990;72B:639–40.
20. Maquet P. Biomethanics of the Knee. Berlin: Springer-Verlag 1976.
21. Buckland-Wright JC, Macfarlane DG. Radioanatomic assessment of therapeutic outcome in osteoarthritis. In Kuettner K, Goldberg V, Eds. Osteoarthritic Disorders. Rosemont, IL: American Academy of Orthopedic Surgeons 1995:51–66.
22. Resnick D, Vint V. The "tunnel" view in assessment of cartilage loss in osteoarthritis of the knee. Radiology 1980;137:547–8.
23. Rosenberg TD, Paulos LE, Parker RD, Coward DB, Scott SM. The forty-five-degree posteroanterior flexion weight-bearing radiograph of the knee. J Bone Joint Surg 1988;70A:1479–83.
24. Ravaud P, Auleley G-R, Chastang C, Rousselin B, Paolozzi L, Amor B, Dougados M. Knee joint space width measurement: an experimental study of the influence of radiographic procedure and joint positioning. Br J Rheumatol 1996;35:761–6.
25. Buckland-Wright JC, Macfarlane DG, Williams SA, Ward RJ. Accuracy and precision of joint space width measurements in standard and macroradiographs of osteoarthritic knees. Ann Rheum Dis 1995;54:872–80.
26. Fife RS, Brandt KD, Braunstein EM, Katz BP, Shelbourne KD, Kalasinski LA, Ryan S. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. Arthritis Rheum 1991;34:377–82.
27. Buckland-Wright JC. Quantitative radiography of osteoarthritis. Ann Rheum Dis 1994;53:268–75.
28. Dieppe PA, Brandt K, Lohmander S, Felson D. Detecting and measuring modification in osteoarthritis: the need for standardized methodology. J Rheumatol 1995;22:201–3.
29. Dieppe P, Cushnaghan J, Jasani MK, McRae F, Wall I. A two-year placebo-controlled trial of non-steroidal anti-inflammatory therapy in osteoarthritis of the knee joint. Br J Rheumatol 1993;32:595–600.
30. Dougados M, Gueguen A, Nguyen M, Thiesce A, Listrat V, Jacob L, Naqkashe JP, Gabriel KR, Lequesne M, Amor B. Longitudinal radiologic evaluation of osteoarthritis of the knee. J Rheumatol 1992;19:378–84.
31. Mazières B, Levourc'h P, Bunouf P. Mesures des interlignes du genou arthrosique: variations en un an. Rev Rheum 1990;57:680.
32. Jonsson K, Buckwalter K, Helvie M, Niklason L, Martel W. Precision of hyaline cartilage thickness measurements. Acta Radiol 1992;33:234–9.
33. Lequesne M. Quantitative measurements of joint space during progression of osteoarthritis: "chondrometry." In Kuettner K, Goldberg V, Eds. Osteoarthritic Disorders. Rosemont, IL: American Academy of Orthopedic Surgeons 1995:427–44.
34. Laoussadi S, Menkes CJ. Amélioration de la précision de la mesure visuelle de la hauteur de l'interligne articulaire du genou et de la hanche à l'aide d'une loupe graduée. Rev Rhum Maladies Osteoarticulaires 1991;58:678.
35. Altman RD, Fries JF, Bloch DA, Carstens J, Cooke TD, Genant H, Gofton P, Groth H, McShane DJ, Murphy WA, Sharp JT, Spitz P, Williams CA, Wolfe F. Radiographic assessment of progression in osteoarthritis. Arthritis Rheum 1987;30:1214–25.
36. Kellgren JH, Lawrence, JS. Radiographic assessment of osteoarthritis. Ann Rheum Dis 1957;16:494–502.
37. Mazières B, Levourc'h P, Boicimt D. Mesures des interlignes du genou arthrosique: reproducibilite, coefficients de variation. Rev Rheum 1990;57: 680.
38. Lynch JA, Buckland-Wright JC, Macfarlane DG. Precision of joint space width measurement in knee osteoarthritis from digital analysis of high definition macroradiographs. Osteoarthritis 1993; 1:209–18.
39. Kirwan JR, Cushnaghan J, Dacre J, McAlindon T, Dieppe PA, Rogers J. Progression of joint space narrowing in knee osteoarthritis. Arthritis Rheum 1992;35(9 suppl.):S134.
40. Buckland-Wright JC, Macfarlane DG, Jasani MK, Lynch JA. Quantitative microfocal radiographic assessment of osteoarthritis of the knee from weight bearing tunnel and semiflexed standing views. J Rheumatol 1994;21:1734–41.
41. Buckland-Wright JC, Macfarlane DG, Lynch JA, Jasani MK. Quantitative microfocal radiography detects changes in OA knee joint space width in patients in placebo-controlled trial of NSAID therapy. J Rheumatol 1995;22:937–43.
42. Lethbridge-Çejku M, Hochberg MC, Scott WW Jr, Plato CC, Tobin JD. Longitudinal change in joint space of the knee: data from the Baltimore Longitudinal Study of Aging. Arthritis Rheum 1995;38(9 suppl.):S262.
43. Ravaud P, Giraudeau B, Auleley G-R, Chastang C, Poiraudeau S, Ayral X, Dougados M. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol. 1996;23:1756–64.
44. Neuhauser KB, Anderson JJ, Felson DT. Rate of joint space narrowing in normal knees and knees with osteoarthritis. Arthritis Rheum 1994;37(9 Suppl.):S423.
45. Bloch DA. Sample size requirements and the cost of a randomized clinical trial with repeated measurements. Stat Med 1986;5:663–7.
46. Dallal GE. PC-Size: a program for sample size determinations. Amer Stat 1986;40:52.
47. Heinegard D, Petersson I, Boegard T, Dahlstrom J, Svensson B, Poole AR, Ionescu M, Saxne T. Changes in serum levels of cartilage and bone markers appear to predict early osteoarthritis of the knee. Trans Orthop Res Soc 1996;21:219.
48. Dieppe P, Cushnahan J, Young P, Kirwan J. Prediction of the progression of joint space narrowing in osteoarthritis of the knee with bone scintigraphy. Ann Rheum Dis 1993;52:557–63.