Joint Bone Spine 82 (2015) 172–176

Available online at ScienceDirect www.sciencedirect.com

Elsevier Masson France EM|consulte www.em-consulte.com/en

Original article

Impact of repeated measures of joint space width on the sample size calculation: An application to hip osteoarthritis

Paul Ornetti[a,b,c,*], Laure Gossec[d,e], Davy Laroche[f], Christophe Combescure[g], Maxime Dougados[c], Jean-Francis Maillefert[a,b]

[a] Rheumatology department, Medicine Faculty, Dijon University Hospital, 21000 Dijon, France
[b] Inserm U1093, Burgundy University, 21000 Dijon, France
[c] Paris-Descartes University, Medicine Faculty, UPRES-EA 4058, Cochin Hospital, Rheumatology B Department, 75014 Paris, France
[d] CIC-P Inserm 1432, Plateforme d'Investigation Technologique, Dijon University Hospital, Dijon, France
[e] Rheumatology department, Pitié-Salpétrière Hospital, AP–HP, 75013 Paris, France
[f] CIC-P Inserm 1432, Plateforme d'Investigation Technologique, Dijon University Hospital, 21000 Dijon, France
[g] Division of clinical epidemiology, Geneva University Hospital and University of Geneva, 1211 Geneva, Switzerland

ARTICLE INFO

Article history:
Accepted 22 October 2014
Available online 6 January 2015

Keywords:
Sample size
Randomized clinical trial
Joint space
Hip
Osteoarthritis
Smallest detectable difference

ABSTRACT

Objective: The aims of the study were to investigate the impact of aggregation of repeated readings (a) on minimising variability of joint space width (JSW) measurements based on calculation of the smallest detectable difference (SDD) and (b) on sample size calculation in a hip OA randomized controlled trial (RCT).

Methods: (a) Post-hoc analysis of 50 radiographs from a hip OA RCT (ECHODIAH). JSW reliability was calculated by the SDD through the aggregation of repeated readings of hip OA radiographs by an experienced rheumatologist. (b) The observed SDDs were applied to the real ECHODIAH data, to determine the post-hoc theoretical sample sizes.

Results: (a) Although the intra-reader reliability did not improve, SDDs decreased from 0.75 mm to 0.27 mm when aggregating two or more consecutive readings. (b) A significant decrease of sample size was noticed until three consecutive paired JSW measurements, with a sample size ranging from 6588 patients per group (SDD = 0.75 mm) to 377 patients (SDD = 0.45 mm). However, an increase of the sample size was observed for smaller SDDs.

Conclusion: In hip OA, the aggregation of repeated radiograph readings leads to a significant decrease in SDD, although the intra-reader reliability of the experienced reader remains stable. The decrease in SDD allowed to reduce significantly the post-hoc calculation of sample size until a SDD value of 0.45 mm. For smaller cut-offs of radiological progression, the sample size increased again indicating that gain in sensitivity does not automatically lead to gain in sample size, particularly if the treatment effect is limited.

In recent years, interest has grown among the scientific community, drug companies, and regulatory agencies for the development of drugs that might influence the natural history of rheumatic diseases by preventing, retarding, or even reversing structural degradation. These disease-modifying drugs have to be evaluated using primary outcome measures that reflect structural modification, i.e., radiographs in osteoarthritis (OA), rheumatoid arthritis (RA) and ankylosing spondylitis, ultrasonography or magnetic resonance imaging. At the present time, in OA, structural variables, particularly minimal joint space width (JSW) on plain radiographs are considered as the most appropriate primary outcome measure [1,2]. However, change in JSW, expressed in millimeters, is a continuous variable that does not provide information on the proportion of patients with favorable evolution, i.e., does not make it possible to classify patients as "progressors" or "non-progressors". The percentage of patients who have structural progression during a study is easier to understand than the average change in JSW. A task force was created under the auspices of Osteoarthritis Research Society International (OARSI) and Outcome Measures in Rheumatology Clinical Trials (OMERACT) in order to propose a definition of clinically relevant radiological progression in hip/knee OA [1]. Based on the data extracted from the literature, the expert committee failed to propose a data-driven definition of clinically relevant change, so proposed to define a change as an absolute change of JSW in millimeters above the measurement error, i.e., above the

* Corresponding author. Department of Rheumatology, Medicine Faculty, Dijon University Hospital, 3, boulevard Delattre-de-Tassigny, 21000 Dijon, France.
E-mail address: paul.ornetti@chu-dijon.fr (P. Ornetti).

http://dx.doi.org/10.1016/j.jbspin.2014.10.006
1297-319X/

smallest detectable difference (SDD) calculated using the Bland and Altman reliability method [3] in each RCT. This method allows to distinguish in a clinical trial between progressors (with a change greater than measurement error, i.e., SDD) and patients who remain stable (with a change possibly explained by the measurement error alone) [4]. Therefore, reliability will impact on the outcomes of a trial through the SDD, since a variation in the SDD leads to a variation in the proportion of progressors with consequences on the calculation of sample sizes.

The statistical methodology for calculating sample sizes in clinical trials has been extensively developed over the years [5]. One of the first steps in calculating sample size in a randomized controlled trial (RCT) is to consider the expected difference between groups, which depends on the inherent efficacy of the treatment but also depends on the variability of the measurement process. This may include standardization of radiographic acquisition (radiographic view [1], joint positioning [6], centralization of acquisition, film exposure [7], use of digitized image or not, computer based measurement or not and experience of the reader in radiographic measurement) [8,9]. In RCTs evaluating a potential disease-modifying OA drug (DMOAD), one feels intuitively that a reduction in the variability of JSW measurements that minimize the measurement error should allow to detect more radiological progressors, therefore to improve responsiveness and to reduce the sample size. However, minimizing the variability might allow to detect a greater percentage of progressors but potentially in both groups (active group and placebo group). Thus, as the difference between groups strongly affects the calculation of the sample size, it is necessary to investigate the relation between variability of measurement and sample size. Repeated measures may impact variability and therefore, might reduce the required sample size in a two-armed trial as suggested in RA, for Disease Activity Score [10]. To our knowledge, this has not been assessed in OA to date.

Thus, the aims of the present study were to investigate the impact of aggregation of repeated readings:

- on minimising variability of JSW measurements based on calculation of the smallest detectable difference (SDD);
- on sample size calculation in a hip OA randomized controlled trial (RCT).

1. Methods

1.1. Data extraction

Fifty pelvic radiographs were randomly extracted from the ECHODIAH database, a multicenter, prospective, longitudinal, placebo-controlled RCT [11] evaluating the effect of diacerein in reducing the progression rate of hip OA JSW over 3 years.

1.2. Study design

The study designs are:

- blinded radiographic analysis of JSW of hip OA radiographs performed 12 times by a senior rheumatologist (JFM) with extensive experience in examining radiographs of hip OA [12,13];
- assessment of the SDD with different numbers of repeated readings from the radiographic sample (from 1 to 6 repeated readings) by obtaining mean measures for groups of readings;
- sample size calculation from the observed SDDs applied to the database of the same RCT.

Table 1
Patient's characteristics (ECHODIAH completers and random sample). Results are presented as mean ± standard deviation unless otherwise mentioned.

	ECHODIAH study ($n = 50$)	Pilot study ($n = 50$)	P
Age (years)	62.8 (6)	63.4 (8)	0.39
Sex (% female)	67.8	64.9	0.32
Body mass index (kg/m^2)	27.1 (4.2)	26.4 (4.9)	0.14
OA disease duration (years)	5.1 (4.8)	5.3 (4.7)	0.57
Pain (VAS 0–100)	49 (21)	51 (23)	0.46
Lequesne index (0–24)	8.1 (2.4)	8.4 (3.1)	0.22
Femoral head migration			
Supero-lateral	63%	56%	0.11
Supero-medial	30%	36%	0.17
Concentric	7%	8%	0.32
Kellgren & Lawrence grading			
II	65%	67%	0.37
III	34%	33%	0.56
IV	1%	0%	NA

1.2.1. Patients' characteristics

The inclusion criteria have been described elsewhere [11]. Briefly, patients between 50 and 75 years fulfilled ACR criteria for the diagnosis of hip OA [14] and were symptomatic. Baseline characteristics of hip OA completers population are detailed in Table 1. All the patients signed an informed consent. The study protocol was approved by the local ethics committee.

1.3. Analysis of radiographs

Radiographs were obtained following previously published guidelines [15]. An antero-posterior pelvic weight-bearing radiograph of the pelvis with the lower limbs in 15° internal rotation was obtained with a size ratio of 1:1. The interbone distance at the narrowest point was measured in millimetres, using a 0.1 mm graduated magnifying glass laid directly over the radiograph.

1.4. Statistical analysis

1.4.1. Calculation of SDD

In the first step, mean and standard deviation of JSW in millimeters was calculated for each reading [reading 1 (R1) to reading 12 (R12) and for the following reading sequence: R1 + R2, R3 + R4, R1 + R2 + R3, R4 + R5 + R6, R1 + R2 + R3 + R4, R5 + R6 + R7 + R8, R1 + R2 + R3 + R4 + R5, R6 + R7 + R8 + R9 + R10, R1 + R2 + R3 + R4 + R5 + R6, R7 + R8 + R9 + R10 + R11 + R12]. For example, for R1 + R2, the mean (SD) values of (R1 + R2/2) were assessed. In the second step, the SDD was assessed using the Bland and Altman method [3] for the different numbers of repeated readings, i.e., for consecutive couples of radiograph readings (R1 vs R2, R3 vs R4, etc.) and for consecutive reading sequence (S1 to S6, with S1 = R1 vs R2, S2 = R1 + R2 vs R3 + R4, S3 = R1 + R2 + R3 vs R4 + R5 + R6, S4 = R1 + R2 + R3 + R4 vs R5 + R6 + R7 + R8, S5 = R1 + R2 + R3 + R4 + R5 vs R6 + R7 + R8 + R9 + R10, S6 = R1 + R2 + R3 + R4 + R5 + R6 vs R7 + R8 + R9 + R10 + R11 + R12) For example, in the sequence 4, the SDD was assessed between the mean value of the 4 first readings (R1 + R2 + R3 + R4) and the mean value of the 4 next readings (R5 + R6 + R7 + R8). The SDD is then defined as 1.96 SD of the difference between measurements and provides an absolute estimate of measurement error. Ninety-five percent confidence intervals (CI) of SDDs were estimated with bootstrapping methods [16]. In the third step, the difference between the SDD values in the consecutive couples of radiographs (one global P value) and in