



Parsons, C. M., Judge, A., Meyer, R., Bruyere, O., Petit-Dop, F., Chapurlat, R., Reginster, J-Y., Cooper, C., & Inskip, H. (2021). Determining individual trajectories of joint space loss: improved statistical methods for monitoring knee osteoarthritis disease progression. *Osteoarthritis and Cartilage*, 29(1), 59-67.
<https://doi.org/10.1016/j.joca.2020.09.009>

Peer reviewed version

License (if available):
CC BY-NC-ND

Link to published version (if available):
[10.1016/j.joca.2020.09.009](https://doi.org/10.1016/j.joca.2020.09.009)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://doi.org/10.1016/j.joca.2020.09.009>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Determining individual trajectories of joint space loss: improved statistical methods for monitoring knee osteoarthritis disease progression

Camille M Parsons¹, cp@mrc.soton.ac.uk

Andrew Judge², Andrew.judge@bristol.ac.uk

Renate Meyer³, renaite.meyer@auckland.ac.nz

Olivier Bruyere⁴, olivier.bruyere@uliege.be

Florence Petit-Dop⁵, Florence.petit-dop@servier.com

Roland Chapurlat⁶, roland.chapurlat@inserm.fr

Jean-Yves Reginster⁴, jyreginster@ulg.ac.be

Cyrus Cooper^{1,7,8}, cc@mrc.soton.ac.uk

Hazel Inskip¹, hmi@mrc.soton.ac.uk

¹ *MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, UK;*

² *Musculoskeletal Research Unit, University of Bristol, Bristol, UK;*

³ *Department of Statistics, University of Auckland, New Zealand;*

⁴ *Department of Public Health and Health Economics, University of Liege, 4020 Liege, Belgium;*

⁵ *Innovative Therapeutic Pole of Rheumatology, Servier, Suresnes, France;*

⁶ *INSERUM UMR 1033, Service de Rhumatologie et Pathologie Osseuse, Hopital Edouard Herriot, Universite de Lyon, Lyon, France;*

⁷ *National Institute for Health Research Musculoskeletal Biomedical Research Unit, University of Oxford, Oxford, UK;*

⁸ *National Institute for Health Research Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, UK;*

Correspondence to: Professor Cyrus Cooper, MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, SO16 6YD, UK.

Tel: +44 (0)23 8077 7624 Fax: +44 (0)23 8070 4021 Email: cc@mrc.soton.ac.uk

Running title: Statistical methods for monitoring OA progression

Keywords: longitudinal, osteoarthritis, progression

Abstract

Objectives

Knee osteoarthritis (KOA) progression is frequently monitored by calculating the change in knee joint space width (JSW) measurements. Such differences are small and sensitive to measurement error. We aimed to assess the utility of two alternative statistical modelling methods for monitoring KOA.

Material and Methods

We used JSW on radiographs from both the control arm of the Strontium Ranelate Efficacy in Knee Osteoarthritis trial (SEKOIA), a 3-year multicentre, double-blind, placebo-controlled phase 3 trial, and the Osteoarthritis Initiative (OAI), an open-access longitudinal dataset from the USA comprising participants followed over 8 years. Individual estimates of annualised change obtained from frequentist linear mixed effect (LME) and Bayesian hierarchical modelling, were compared with annualised crude change, and the association of these parameters with change in WOMAC pain was examined.

Results

Mean annualised JSW changes were comparable for all estimates, a reduction of around 0.14mm/y in SEKOIA and 0.07mm/y in OAI. The standard deviation (SD) of change estimates was lower with LME and Bayesian modelling than crude change (SEKOIA SD=0.12, 0.12 and 0.21 respectively; OAI SD=0.08, 0.08 and 0.11 respectively). Estimates from LME and Bayesian modelling were statistically significant predictors of change in pain in SEKOIA (LME p-value=0.04, Bayes p-value=0.04), while crude change did not predict change in pain (p-value=0.10).

Conclusions

Implementation of LME or Bayesian modelling in clinical trials and epidemiological studies, would reduce sample sizes by enabling all study participants to be included in analysis regardless of incomplete follow up, and precision of change estimates would improve. They provide increased power to detect associations with other measures.

Introduction

Knee joint space width (JSW) is a continuous measurement that is used as an outcome to monitor disease progression by determining cartilage loss within the joint (1). JSW measurements are commonly obtained from radiographic images, and are used in clinical trials assessing the potential of disease-modifying osteoarthritis drugs (DMOADs). JSW is currently the Food and Drug Administration's (FDA) only approved endpoint for such trials (2). Knee JSW measurements are small, being assessed in a standard metric scale of millimetres. In healthy individuals, maximum values are around 8mm (3), and it has been estimated that JSW measurements could be in error by up to 1mm (4). Previous studies have demonstrated that both the technique used to read the radiograph and the positioning of the knee during the radiograph can have a substantial influence on measured JSW (5, 6).

Change in JSW is slow in the general population, often over decades, however, in some individuals disease progression occurs rapidly over a short period (7). Such wide variation in progression between individuals, and the presence of measurement error, make it extremely difficult to distinguish those individuals who have experienced real deterioration, and thus a narrowing of their knee JSW, from those with an apparent change that is simply due to error within the measurement.

Longitudinal JSW measurements are increasingly collected in both clinical and research settings. Traditionally, epidemiological studies and clinical trials use statistical techniques such as paired t-tests (8) or non-parametric rank comparisons (9) to assess group change in JSW between two time points, usually the first and last measurement. Such techniques provide summary statistics at the population level, but the potential for an individual's change in the observed difference between knee JSW measurements to be dominated by measurement error is obscured and rarely considered. Not only do such techniques provide no information about disease progression at the individual level, but they

only use two time points rather than all available repeated measurements. It may also mean that large numbers of participants are excluded from study analysis due to the participant dropping out before the final visit, despite data being available at other study follow-up visits.

Several risk factors have been identified as being associated with JSW narrowing, and thus disease progression, such as obesity (10), increasing age (11) and gender (12, 13). Yet, there is still debate in other areas as to whether certain modifiable risks are linked to increased JSW narrowing. This may in part be due to real change over time being obscured by measurement error and therefore weakening the possibility of finding associations. We have previously demonstrated the value of the reliable change index (RCI) as an analysis methodology that enables change in JSW to be identified, removing many of the apparent changes that are likely due to measurement error (14). Although the RCI allows for identification of individuals who had statistically significant reliable change in knee JSW across differing time frames, the method only uses two measurements. Increasingly in both clinical and research environments multiple JSW measurements are obtained over time and these should be used to the full.

Linear mixed effect (LME) modelling is an established frequentist statistical technique that can be used to model longitudinal change in JSW, using all JSW measurements available, while accounting for the potential of measurement error within individual measurements. Bayesian hierarchical modelling is an alternative statistical method that also uses all repeated JSW measurements obtained.

Thus, this study aimed to assess the utility of both the frequentist LME and Bayesian modelling methods for monitoring of change in knee JSW by comparing individual estimates of change in JSW obtained from these alternative statistical modelling methods with crude change in JSW.

Methods

Study Design

Joint space width measurements from two datasets were used in this study; the Osteoarthritis Initiative (OAI), and the placebo arm of the Strontium ranelate Efficacy in Knee Osteoarthritis trial (SEKOIA), both described in detail previously (15, 16).

In brief, the OAI is a multicentre, longitudinal, prospective observational study following study participants in the United States of America, with the overarching aim of improving public health through prevention, or alleviation, of pain and disability from OA. To be eligible for entry into the OAI study, participants had to be aged between 45 and 79 years, and have established radiographic knee OA as defined using the OARSI atlas (17) or be identified as at risk of developing knee OA when they entered the study. The OAI study participants were recruited between February 2004 and May 2006. The OAI study data is an open access source, and all OAI data used in this study were downloaded between October and December 2016. The data used within this study cover 96 months of follow up.

The SEKOIA study was a 3-year international, multicentre, double-blind, randomised placebo-controlled phase 3 trial that was established to assess the structure-modifying effect of a drug treatment, strontium ranelate, on radiological and clinical progression of OA in the knee joint. Study participants were recruited into the trial from 98 study centres across 18 different countries between 2006 and 2008, and were randomised to either a drug regime of strontium ranelate 1 g/day, strontium ranelate 2g/day, or a placebo treatment. To be eligible for entry into the SEKOIA study, participants had to be Caucasian men or women aged over 50 years with a primary diagnosis of knee OA as defined by the clinical criteria of the American College of Rheumatology (ACR) (18), on radiograph have a

Kellgren and Lawrence (K&L) grade of 2 or 3 (19), JSW between 2.5mm and 5mm at an inclusion screen, and predominant OA of the medial tibiofemoral compartment. The SEKOIA study conformed to the principles of the Declaration of Helsinki (ISRCTN41323372). The data used within this study are from the placebo arm of the trial.

Longitudinal joint space width measurement

In the OAI study, knee radiographs were performed at baseline, and at 12, 24, 36, 48, 72 and 96 months using the 'fixed flexion' knee radiograph protocol in all study centres (20). All OAI radiographic images were read centrally, and were examined in pairs of study participants' radiographic images but blinded to chronological order of the image; the minimum distance within the medial compartment of the knee was measured using a customized software tool. If both knees met the inclusion criteria for the OAI study, both knees were entered into the observation study. However for the purpose of this study, only one knee per study participant was included in the analysis. The knee with the highest K&L grade and the smallest JSW at baseline was chosen for inclusion.

In SEKOIA, radiographs were performed at the time of selection and then annually on the target knee, giving up to four measurements per person, using a standardised technique across all centres, as described elsewhere (21). All the SEKOIA radiographic images were measured centrally (INSERM UMR 1033, Lyon, France) by a single reader blinded to the study participant and treatment allocation. Each blinded post-baseline radiographic image was measured in comparison with the study participant's baseline image to optimise reproducibility and sensitivity, and minimal JSW (mm) at the tibiofemoral compartment was measured using a standardised computer-assisted method (22).

Linear mixed effect modelling

The frequentist technique of LME modelling is an established statistical method for modelling longitudinal data (23-25). LME modelling allows multiple JSW measurements for each individual to be included within a single statistical model, while allowing each individual to have a different baseline JSW measurement (random intercept) and trajectory of change (random slope). The results of LME modelling give a study population average overall estimate, and individual estimates of change (Best Linear Unbiased Predictors (BLUPS)) can also be calculated during the modelling process. A major assumption of LME modelling that was made in these analyses is that the relationship between the outcome and predictor is linear. There is little evidence in previous epidemiological research about the form of longitudinal trajectories of knee JSW across the lifecourse, and so it seemed appropriate in the context of this study to assume that trajectories were linear. *Also, most analyses implicitly assume a linear relationship by taking the difference between baseline and final measurements, and we aimed to compare the two modelling methods with this form of assessment.*

Bayesian longitudinal modelling

The overriding principle of Bayesian analysis, based on the Bayes' theorem, (26) is a way of calculating conditional probabilities, i.e. inference is made about what is not known given previous knowledge and observations.

The practical application of conditional probability occurs considerably more often than may be generally thought. For example, in a clinical setting, a rheumatologist may suspect knee OA given the description of joint pain, stiffness and restricted movement that a patient gives. This is conditional probability, as the probability of having a disease is dependent on the probability of having a set of symptoms.

Bayesian hierarchical modelling allows use of all JSW measurements from all study participants by allowing different initial JSW measurements (random intercept) and trajectories of change (random slope). The assumption of exchangeable observations was used (27), and, as little is known about change in knee JSW over time, the parameters in the Bayesian hierarchical model used within this study were assigned non-informative priors (28).

Association with pain progression

To further assess the performance of the individual trajectories of change, change in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain over the study duration of SEKOIA and OAI was considered as a 'gold' standard of disease progression. Conditional change in WOMAC pain was characterised by the residuals obtained from linear regression of WOMAC pain at follow-up on WOMAC pain at baseline; this measure of change is independent of baseline pain level. The association between the change in WOMAC pain score and each of the three estimates of JSW changes was assessed using linear regression.

Statistical analysis

Study participants' continuous characteristics were summarised using means and standard deviations (SD), after checking for normality. For comparison, two definitions of crude annualised change were used. First, in line with the majority of clinical studies, we considered only those participants included in the intention-to-treat (ITT) analysis, i.e. only those participants with baseline and end of study JSW measurements. For each study participant in both datasets with baseline and end of study JSW measurements, crude annualised change (ITT) was calculated, using the Stata software (release 14.0 STATA Corp, College Station, TX, USA), by dividing change in JSW by study follow-up (29). To use as much data as possible, a second definition of annualised crude change was calculated using each study

participants last measured JSW, which might have been before the end of the study, and dividing by the number of years the study participant remained within the study.

To directly compare the crude change (ITT) with estimates for LME and Bayesian hierarchical models, the study populations were restricted to those 326 study participants in SEKOIA and 1918 study participants in the OAI with a baseline and end of study JSW measurement. All Bayesian analysis in this project was undertaken using WinBUGS 14 (27) implemented through the statistical package RStudio (30). LME modelling and analysis of individual posterior estimates obtained from Bayesian analysis were undertaken in Stata, release 14.0 (STATA Corp, College Station, TX, USA) (29). The random slope parameters, the estimates of individual annual change in knee joint space obtained from LME and Bayesian modelling, and crude annualised changes in JSW were compared using means and SDs, and using the Bland-Altman method for limits of agreement (31).

Results

A total of 3469 study participants were enrolled into the OAI, and 559 study participants were randomised to the placebo arm of the SEKOIA study and included in this study. The participants' characteristics from both datasets are given in table 1. Just under 60% of study participants in the OAI were women, and just over 70% in the SEKOIA study. The mean (SD) age at baseline of study participants in the OAI was 61.6 (9.1) years and in SEKOIA was 62.8 (7.5) years. Study participants in both datasets had similar mean (SD) BMI at baseline, 29.1 (4.8) kg/m² in the OAI and 29.8 (5.1) kg/m² in SEKOIA.

At baseline, study participants in the OAI and SEKOIA had on average a knee JSW of 3.99mm, and 3.51mm, respectively. Over the 96 months of the OAI study crude mean JSW reduced to 3.74mm, and over the 3-year duration of the SEKOIA study, the crude average JSW reduced to 3.15mm.

A total of 19491 knee JSW measurements across the 3385 study participants in the OAI, and 1765 knee JSW measurements across 558 study participants in SEKOIA and were included within the LME and Bayesian modelling. Estimates from LME modelling indicated that, on average, knee JSW decreased by 0.08 mm (95% confidence interval: -0.083, -0.077) per 12 months in the OAI study and 0.14mm (95% confidence interval: -0.144, -0.127) for each successive year in the SEKOIA study (table 2). Posterior estimates of average annual knee JSW obtained from Bayesian modelling provided very similar population level estimates, with, on average, knee JSW having decreased by 0.08mm (95% credible interval: -0.276, 0.057) per year in the OAI study and 0.14mm (95% credible interval: -0.393, 0.073) per year in SEKOIA (table 2).

The estimates of mean crude change when using as much data as possible were -0.146mm per year in SEKOIA and -0.081mm in OAI, whereas mean crude change (ITT) was -0.066mm per year in the OAI and -0.137mm per year in SEKOIA (table 2). These are comparable to estimates of average annual knee JSW from LME and Bayesian modelling, but the standard deviation (SD) of change estimates was lower with LME and Bayesian modelling than crude change (SEKOIA SD=0.12, 0.12 and 0.21 respectively; OAI SD=0.08, 0.08 and 0.11 respectively).

Figure 1 contains the Bland-Altman plot assessing the level of agreement between annual change in knee JSW in the OAI study obtained from LME and Bayesian modelling. The mean difference between the two estimates is 0.002mm, and the limits of agreement were -0.023mm and 0.027mm. There is

no systematic trend in the magnitude of the differences between the two estimates. Figure 2 and 3 contain comparisons of LME and Bayesian modelling estimates with crude annualised change. Although the mean difference between estimates is small, 0.002mm between LME and crude estimates and 0.003mm between Bayes and crude estimates, there is a systematic trend in the magnitude of the differences between the two modelling techniques and crude change. The larger the magnitude in average change in JSW obtained from either modelling technique, the larger the difference between the estimate and crude change. Similar patterns between estimates are seen in SEKOIA, see supplementary material.

Table 3 contains the results of comparison of the three estimates of JSW change with conditional change in WOMAC pain. In both studies, estimates of annualised change in JSW obtained from LME and Bayesian modelling provided stronger associations with change in pain than did the crude changes, indicating that the modelling methods provide greater power to detect associations with other measures of disease. When adding age, BMI and gender as confounders into the model these relationships remained unchanged.

Discussion

Key results

In the OAI and SEKOIA data, when estimates of annual change in knee JSW obtained from LME and Bayesian modelling were compared, mean differences between the two estimates were small and no systematic magnitude of difference was observed. Therefore LME models and Bayesian hierarchical modelling methodologies proved comparable when applied in these longitudinal analyses.

In those study participants with baseline and end of study JSW measurements, mean annualised crude change estimates were comparable to estimates obtained from LME model and Bayesian modelling in both the OAI and SEKOIA. However, both statistical modelling techniques provide estimates with greater precision, as demonstrated by the smaller standard deviations observed for LME and Bayesian change estimates when compared with crude change. This is also demonstrated by the larger effect sizes and smaller p-values found when the two modelling JSW estimates were related to WOMAC pain than when the crude changes were used.

Implication of results

The majority of previous epidemiological studies assessing change in knee JSW over time only utilise the first and last study visit measurements. This means that all the longitudinal JSW measurements collected throughout a study are rarely used. For example, in a study by Fukui et al (32), radiographic images were obtained at baseline and every 6 months during the 3 year follow-up period, giving the possibility of 7 knee JSW measurements. However, within this study, change was reported using JSW narrowing rate (mm/year), which appears to have been calculated using only the baseline and 3-year JSW measurements and thus not making full use of the repeated measures. One of the strengths of Bayesian and LME modelling is that both methods allow for the use of all repeated JSW measurements and are also flexible in that a balanced study design, with complete data on all participant's is not required. Therefore, implementing either of these methods in future studies would reduce the sample sizes required because data from all study participants can be used regardless of incomplete study follow up.

Few studies mentioned the issue of measurement error, and, as highlighted by Ravaud et al (33), without accounting for measurement error, the differences observed may not be 'true organic

change'. The application of Bayesian and LME modelling would account for measurement error during the modelling process by 'smoothing' estimates, as both models use individual trajectories to provide estimates at the population level. So application of these techniques in the research environment, in particular in DMOADs, would allow for the use of all available measurements for all study participants diluting the effect of measurement error to enable provision of robust estimates of OA disease progression, and increasing precision of estimates. Given that the standard deviations of the estimates of change from the modelling methods are smaller than those for the crude changes, the power of studies increases, and use of the modelling methods leads to fewer trial participants required to determine treatment efficacy.

A further important consideration for implementation of LME or Bayesian modelling is the ethical obligation for researcher to use as much data as possible. If a study team asks individuals to join a study that requires them to attend for study visits and radiate them to obtain radiographic images, then researchers should ensure that as much of the data collected as possible is used within the study analyses.

Strengths of study

Bayesian modelling is an established statistical modelling technique that is widely used nowadays in many fields, with software being available in which to conduct the analyses. However, to date, no previous epidemiological studies have been identified that use Bayesian analysis to monitor change in knee JSW, and only a handful use LME modelling.

This is also the first study to fully explore the performance of different statistical methodologies in monitoring change in different datasets with different study time frames containing study participants

with varying disease severities. Results from the Bayesian and LME modelling were comparable in both SEKOIA and the OAI, indicating that even in this simple application of modelling longitudinal JSW data both techniques are robust to the number of JSW measurements, study durations and disease severity.

Limitations

This study has presented two methodologies that would prove beneficial in calculating a robust OA progression estimate in different settings. Bayesian modelling is not yet a routine approach and so this is unlikely to be useful in a clinical setting without further development. However, in a research setting, with skilled statisticians in the team, the approach could readily be implemented and allows all measurements to be used from a longitudinal dataset, adding depth to the analysis. In contrast, LME models are an established methodology and many software packages have built-in commands to handle the analysis, such as the 'mixed' command in the statistical software Stata (29).

There is currently no gold standard method for assessing change in knee JSW measurements. So there is no official gold standard comparator for the statistical methods presented in this study. However pain progression is one of the most widely used methodologies to monitor symptomatic disease status, with previous studies demonstrating an association between pain and structural progression of knee OA (34).

In all analyses presented in this study, the assumption was made that change in JSW over time is linear. Some studies have demonstrated JSW progression may be greater in more severe disease. For example Halilaj and colleagues used least absolute shrinkage and selection (LASSO) regression models to predict whether study participants belonged to stable, improving or worsening clusters, while

allowing different trajectories of progression for each cluster (35). This study used different statistical methodology and assumptions, further highlighting the complexity of defining disease progression. Although the LME and Bayesian models used within this study allowed for individual trajectories, potential clustering of these individual trajectories, interactions with time, and quadratic relationships should be explored in further studies.

The assumption was also made that the mechanism of missingness was 'missing at random', which means when outcome measures with a primary endpoint combined across multiple time points, models estimates are produced using full information maximum likelihood (FIML) techniques.

The statistical methodologies used within this study have been applied to two large (>300 study participants) datasets. These studies were used to ensure a large number of study participants with differing JSW ranges were included. However further exploration of these methods in smaller research studies (<300 study participants) should help further understanding of how these approaches can be used in estimating individual OA progression.

Conclusions

Implementation of LME or Bayesian modelling in clinical trials and epidemiological studies, would reduce sample sizes required by maximising the use of data and enabling all study participants to be included in analysis regardless of incomplete study follow up. The estimates would be more robust to measurement error by 'smoothing' the estimates, both at the population level and for individual estimates. As the precision of change estimates would improve, the power of these methods to detect associations with other measures would also increase. The power would also be increased in trials,

thus reducing the number of participants who need to be recruited, providing a reduction in research costs and participant burden.

Acknowledgements

We acknowledge the directors and personnel of the 98 investigating centres, all members of the SEKOIA management team (appendix 1) and the study participants. We would also like to express our gratitude to the participants of the OAI. The OAI is a public-private partnership comprised of five contracts (NO1-AR-2-2258, NO1-AR-2-2259, NO1-AR-2-2260, NO1-AR-2-2261, NO1-AR-2-2262) funded by the National Institutes of Health, a branch of the U.S. Department of Health and Human Services, in four clinical sites and conducted by the OAI study investigators.

Author Contributions

CP, AJ, RM, CC and HI were involved in the design of the study, the statistical analysis and interpretation of the results. OB, FPD, RC and J-YR all assisted in the production of the manuscript. All authors read and approved the manuscript.

Conflict of interest

CP and RM has no conflicts of interest to report. Outside of the submitted work, AJ has received consultancy fees from Freshfields Bruckhaus Derringer and Anthera Pharmaceuticals LTD. Outside the submitted work, OB reports grants from Biophytis, IBSA, MEDA, Servier, and SMB, personal fees from Amgen, Aptissen, Biophytis, IBSA, MEDA, Sanofi, Servier, SMB, and UCB. Outside of the submitted work, J-YR reports consultancy fees from IBSA Genevrier, Mylan, Radius Health, and Pierre Fabre, grants from IBSA Genevrier, Mylan, CNIEL, and Radius Health, and payment for lectures from IBSA Genevrier, Mylan, CNIEL, Dairy Research Council (DRC). HMI salary is provided from the UK Medical

Research Council, and outside of the submitted work HMI's institution has received grants from NIHR and the British Heart Foundation. FP-D salary if provided by Servier. Outside of the submitted work, CC reports personal fees from Alliance for Better Bone Health, Amgen, Eli Lilly, GSK, Medtronic, Merck, Novartis, Pfizer, Roche, Servier, Takeda and UCB. Outside of the submitted work, RC reports board membership and consultancy fees from Pfizer.

References

1. Buckland-Wright JC, Macfarlane DG, Lynch 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(4):263-8. PubMed PMID: PMC1005571.
2. Conaghan PG, Hunter DJ, Maillefert JF, Reichmann WM, Losina E. Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group. *Osteoarthritis Cartilage*. 2011;19(5):606-10. doi: 10.1016/j.joca.2011.02.018. PubMed PMID: PMC3260466.
3. Deep K, Norris M, Smart C, Senior C. Radiographic measurement of joint space height in non-osteoarthritic tibiofemoral joints. A comparison of weight-bearing extension and 30 degrees flexion views. *J Bone Joint Surg Br*. 2003;85(7):980-2. Epub 2003/10/01. PubMed PMID: 14516031.
4. Ravaud P, Giraudeau B, Auleley GR, Drape JL, Rousselin B, Paolozzi L, et al. Variability in knee radiographing: Implication for definition of radiological progression in medial knee osteoarthritis. *Ann Rheum Dis*. 1998;57(10):624-9. PubMed PMID: 1998380183.
5. Nevitt MC, Peterfy C, Guermazi A, Felson DT, Duryea J, Woodworth T, et al. Longitudinal performance evaluation and validation of fixed-flexion radiography of the knee for detection of joint space loss. *Arthritis Rheum*. 2007;56(5):1512-20. PubMed PMID: 2007242128.
6. Brandt KD, Mazuca SA, Conrozier T, Dacre JE, Peterfy CG, Provvedini D, et al. Which is the best radiographic protocol for a clinical trial of a structure modifying drug in patients with knee osteoarthritis? *J Rheumatol*. 2002;29(6):1308-20. Epub 2002/06/18. PubMed PMID: 12064851.
7. Hunter DJ. Risk stratification for knee osteoarthritis progression: a narrative review. *Osteoarthritis Cartilage*. 2009;17(11):1402-7. doi: <https://doi.org/10.1016/j.joca.2009.04.014>.
8. Benichou OD, Hunter DJ, Nelson DR, Guermazi A, Eckstein F, Kwok K, et al. One-Year Change in Radiographic Joint Space Width in Patients With Unilateral Joint Space Narrowing: Data From The Osteoarthritis Initiative. *Arthritis Care Res (Hoboken)*. 2010;62(7):924-31. doi: 10.1002/acr.20149. PubMed PMID: PMC2927100.
9. Theiler R, Stucki G, Schütz R, Hofer H, Seifert B, Tyndall A, et al. Parametric and non-parametric measures in the assessment of knee and hip osteoarthritis: interobserver reliability and correlation with radiology. *Osteoarthritis Cartilage*. 1996;4(1):35-42. doi: [http://dx.doi.org/10.1016/S1063-4584\(96\)80005-7](http://dx.doi.org/10.1016/S1063-4584(96)80005-7).
10. Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. *BMC Musculoskelet Disord*. 2008;9:132. Epub 2008/10/04. doi: 10.1186/1471-2474-9-132. PubMed PMID: 18831740; PubMed Central PMCID: PMC2573886.

11. Garstang SV, Stitik TP. Osteoarthritis: epidemiology, risk factors, and pathophysiology. *Am J Phys Med Rehabil.* 2006;85(11 Suppl):S2-11; quiz S2-4. Epub 2006/11/03. doi: 10.1097/01.phm.0000245568.69434.1a. PubMed PMID: 17079976.
12. Silverwood V, Blagojevic-Bucknall M, Jinks C, Jordan JL, Protheroe J, Jordan KP. Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. *Osteoarthritis Cartilage.* 23(4):507-15. doi: 10.1016/j.joca.2014.11.019.
13. Felson DT, Naimark A, Anderson J, Kazis L, Castelli W, Meenan RF. The prevalence of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. *Arthritis Rheum.* 1987;30(8):914-8. Epub 1987/08/01. PubMed PMID: 3632732.
14. Parsons CM, Judge A, Leyland K, Bruyere O, Petit Dop F, Chapurlat R, et al. Novel Approach to Estimate Osteoarthritis Progression: Use of the Reliable Change Index in the Evaluation of Joint Space Loss. *Arthritis Care Res (Hoboken).* 2019;71(2):300-7. Epub 2018/05/10. doi: 10.1002/acr.23596. PubMed PMID: 29741284; PubMed Central PMCID: PMC6129388.
15. Nevitt M. The Osteoarthritis Initiative: Protocol for the Cohort Study [cited 2016 3rd April]. Available from: <https://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf>.
16. Reginster JY, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X, et al. Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: Results of a double-blind, randomised placebo-controlled trial. *Ann Rheum Dis.* 2013;72(2):179-86. PubMed PMID: 2013022574.
17. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage.* 2007;15 Suppl A:A1-56. Epub 2007/02/27. doi: 10.1016/j.joca.2006.11.009. PubMed PMID: 17320422.
18. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis Rheum.* 1986;29(8):1039-49.
19. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis.* 1957;16(4):494-502. doi: 10.1136/ard.16.4.494. PubMed PMID: MEDLINE:13498604.
20. Peterfy C, Li J, Zaim S, Duryea J, Lynch J, Miaux Y, et al. Comparison of fixed-flexion positioning with fluoroscopic semi-flexed positioning for quantifying radiographic joint-space width in the knee: test-retest reproducibility. *Skeletal Radiol.* 2003;32(3):128-32. Epub 2003/02/28. doi: 10.1007/s00256-002-0603-z. PubMed PMID: 12605275.
21. Cooper C, Reginster JY, Chapurlat R, Christiansen C, Genant H, Bellamy N, et al. Efficacy and safety of oral strontium ranelate for the treatment of knee osteoarthritis: rationale and design of randomised, double-blind, placebo-controlled trial. *Curr Med Res Opin.* 2012;28(2):231-9. Epub 2011/12/14. doi: 10.1185/03007995.2011.648758. PubMed PMID: 22148897.
22. Gensburger D, Arlot M, Sornay-Rendu E, Roux JP, Delmas P. Radiologic assessment of age-related knee joint space changes in women: a 4-year longitudinal study. *Arthritis Rheum.* 2009;61(3):336-43. Epub 2009/02/28. doi: 10.1002/art.24342. PubMed PMID: 19248124.
23. West BT. Linear mixed models.
24. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986;42(1):121-30. Epub 1986/03/01. PubMed PMID: 3719049.
25. Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods: Sage; 2002.
26. Bayes M, Price M. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions.* 1763;53:370-418. doi: 10.1098/rstl.1763.0053.
27. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing.* 2000;10(4):325-37.
28. Kass RE, Wasserman L. The selection of prior distributions by formal rules. *Journal of the American Statistical Association.* 1996;91(435):1343-70.
29. 2015 S. Stata Statistical Software. Texas 2015.

30. (2015) RT. RStudio: Integrated Development for R. Boston, MA2015.
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-10. Epub 1986/02/08. PubMed PMID: 2868172.
32. Fukui N, Yamane S, Ishida S, Tanaka K, Masuda R, Tanaka N, et al. Relationship between radiographic changes and symptoms or physical examination findings in subjects with symptomatic medial knee osteoarthritis: A three-year prospective study. *BMC Musculoskelet Disord*. 2010;11 (no pagination)(269). PubMed PMID: 2010702568.
33. Ravaud P, Giraudeau B, Auleley GR, Edouard-Noel R, Dougados M, Chastang C. Assessing smallest detectable change over time in continuous structural outcome measures: Application to radiological change in knee osteoarthritis. *J Clin Epidemiol*. 1999;52(12):1225-30. PubMed PMID: 1999397744.
34. Wang Y, Teichtahl AJ, Abram F, Hussain SM, Pelletier J-P, Cicuttini FM, et al. Knee pain as a predictor of structural progression over 4 years: data from the Osteoarthritis Initiative, a prospective cohort study. *Arthritis Res Ther*. 2018;20(1):250. doi: 10.1186/s13075-018-1751-4.
35. Halilaj E, Le Y, Hicks JL, Hastie TJ, Delp SL. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. *Osteoarthritis Cartilage*. 2018;26(12):1643-50. Epub 2018/08/22. doi: 10.1016/j.joca.2018.08.003. PubMed PMID: 30130590; PubMed Central PMCID: PMC6469859.

Figure Legends

Figure 1: Bland-Altman plot for estimates of annual change (mm) in the OAI

Figure 2: Bland-Altman for estimates of change in the OAI from crude annualised change and Bayes estimates

Figure 3: Bland-Altman for estimates of change in the OAI from crude annualised change and LME estimates

Supplementary figure 1: Bland-Altman plot for estimates of annual change (mm) in SEKOIA

Supplementary figure 2: Bland-Altman for estimates of change in SEKOIA from crude annualised change and Bayes estimates

Supplementary figure 3: Bland-Altman for estimates of change in SEKOIA from crude annualised change and LME estimates

Table 1 Participants characteristics

	SEKOIA (n=559)		OAI (n=3469)	
	Mean	SD	Mean	SD
Age (years)	62.8	7.5	61.6	9.1
BMI (Kg/m ²)	29.8	5.1	29.1	4.8
	n	%	n	%
Female	392	70.1	2042	58.9
Severity of knee osteoarthritis	Mean	SD	Mean	SD
Joint space at baseline (mm)	3.51	0.83	3.99	1.34
Joint space at end of study duration (mm)	3.15	1	3.74	1.35
Joint space narrowing over study duration (mm)*	-0.41	0.63	-0.52	0.88
	Minimum	Maximum	Minimum	Maximum
Joint space at baseline (mm)	0.65	6.11	0.61	8.87
Joint space at end of study duration (mm)	0.38	5.5	0.70	8.49
Joint space narrowing over study duration (mm)*	-3.34	1.59	-3.35	2.63
Kellgren and Lawrence Grade	n	%	n	%
0	-		545	15.9
1	-		374	10.9
2	350	62.6	1343	39.1
3	209	37.4	883	25.7
4	-	-	292	8.5

*36 month in SEKOIA and 96 months in the OAI

1

2

Table 2: Comparison of estimates from LME and Bayesian modelling

	LME model				Bayesian hierarchical model				Crude change			
	Estimate	95% Confidence interval	SD	N	Estimate	95% Credible interval	SD	N	Estimate	95% Confidence interval	SD	N
SEKOIA	-0.136	-0.144, -0.127	0.099	558	-0.135	-0.393, 0.073	0.102	558	-0.146	-0.169, -0.123	0.257	472
OAI	-0.08	-0.083, -0.077	0.082	3,385	-0.078	-0.276, 0.057	0.081	3469	-0.081	-0.089, -0.074	0.213	3301
	Restricting estimates only to those directly comparable											
	LME model			Bayesian hierarchical model			Crude change (ITT sample)					
	Estimate	95% Confidence interval	SD	Estimate	95% Credible interval	SD	Estimate	95% Confidence interval	SD			
SEKOIA (N=336)	-0.132	-0.144, -0.120	0.115	-0.132	-0.436, 0.079	0.121	-0.137	-0.160, -0.115	0.209			
OAI (N=1918)	-0.068	-0.071, 0.064	0.083	-0.068	-0.273, 0.070	0.084	-0.066	-0.071, -0.061	0.111			

3

4

5

6

7

8

9

10

Table 3: Comparison of estimates of change in joint space width with change in WOMAC pain

	LME model			Bayesian hierarchical model			Crude change		
	Beta	95% Confidence interval	p-value	Beta	95% Credible interval	p-value	Beta	95% Confidence interval	p-value
SEKOIA	-0.966	-1.900, -0.324	0.043	-0.93	-1.819, -0.423	0.04	-0.480	-0.987, 0.027	0.064
OAI	-1.128	-1.550, -0.706	<0.001	-1.097	-1.522, -0.671	<0.001	-0.353	-0.540, -0.166	<0.001
Restricting estimates only to those directly comparable									
	LME model			Bayesian hierarchical model			Crude change (ITT sample)		
	Beta	95% Confidence interval	p-value	Beta	95% Credible interval	p-value	Beta	95% Confidence interval	p-value
SEKOIA (N=326)	-0.894	-1.829, 0.041	0.061	-0.859	-1.748, 0.031	0.058	-0.432	-0.947, 0.083	0.100
OAI (N=1918)	-1.264	-1.766, -0.762	<0.001	-1.244	-1.734, -0.751	<0.001	-0.894	-1.269, -0.519	<0.001
Restricting estimates only to those directly comparable, adjusting for confounders									
	LME model			Bayesian hierarchical model			Crude change (ITT sample)		
	Beta	95% Confidence interval	p-value	Beta	95% Credible interval	p-value	Beta	95% Confidence interval	p-value
SEKOIA (N=326)	-0.852	-1.785, 0.081	0.073	-0.851	-1.737, 0.035	0.060	-0.399	-0.911, 0.114	0.127
OAI (N=1918)	-0.934	-1.360, -0.508	<0.001	-0.908	-1.336, -0.480	<0.001	-0.753	-1.13, -0.378	<0.001

