HAL open science

Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: A simulation study.

Gabriel Baron, Philippe Ravaud, Adeline Samson, Bruno Giraudeau

▶ **To cite this version:**

Gabriel Baron, Philippe Ravaud, Adeline Samson, Bruno Giraudeau. Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: A simulation study.. Arthritis and Rheumatism, 2008, 59 (1), pp.25-31. 10.1002/art.23253 . inserm-00217436

HAL Id: inserm-00217436

https://inserm.hal.science/inserm-00217436

Submitted on 7 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

HAL author manuscript

Arthritis and Rheumatism 2008;59(1):25-31

HAL author manuscript inserm-00217436, version 1

Title: Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: a simulation study

Authors: Gabriel Baron[1], Philippe Ravaud[1], Adeline Samson[2], Bruno Giraudeau[3]

[1]Gabriel Baron, MSc, Philippe Ravaud, MD, PhD: INSERM, U738, Paris, France; Université Paris 7 Denis Diderot, UFR de Médecine, Paris, France; AP-HP, Hôpital Bichat, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Paris, France

[2]Adeline Samson, PhD, UMR CNRS 8145, Laboratoire de Mathématiques Appliquées, Université Paris 5, Paris, France

[3]Bruno Giraudeau, PhD: INSERM, CIC 202, Tours, France ; CHRU de Tours, CIC 202, Tours, France

Address correspondence and reprint requests to

Gabriel Baron, Département d'Epidémiologie Biostatistique et Recherche Clinique, Groupe Hospitalier Bichat-Claude Bernard, 46 rue Henri Huchard, 75018 Paris, France.

Tel: +33 01 40 25 79 39

Fax: +33 01 40 25 67 73

E-mail: gabriel.baron@bch.aphp.fr

Word count: 3473

HAL author manuscript inserm-00217436, version 1

ABSTRACT

Objective. The purpose of this study was to assess the impact, in terms of statistical power and bias of treatment effect, of approaches to dealing with missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes.

Methods. A simulation study was performed. The missingness mechanisms we investigated copied the process of withdrawal from trials due to lack of efficacy. We compared three methods of managing missing data: (i) all available data (case complete), (ii) last observation carried forward (LOCF) and (iii) multiple imputation. Data were then analyzed by classical t-test (comparing the mean absolute change between baseline and final visit) or F test (estimation of treatment effect with repeated measurements by a linear mixed-effects model).

Results. With missing data rate close to 15%, the treatment effect was underestimated by 18% as estimated by a linear mixed-effects model with a multiple imputation approach to missing data. This bias was lower than that obtained with the case-complete (-25%) or LOCF (-35%) approaches. This statistical approach (combination of multiple imputation and mixed-effects analysis) was moreover associated with a 70% power (for a 90% nominal level), whereas LOCF was associated with a power of 55% and case complete a power of 58%. T-test analysis gave qualitatively equivalent but poorer quality results, except when multiple imputation was applied.

Conclusion. Our simulation study showed multiple imputation offering the smallest bias in treatment effect and the highest power. These results can help in planning trials, especially in choosing methods of imputation and data analysis.

HAL author manuscript inserm-00217436, version 1

Key words: rheumatoid arthritis, randomized trials, simulation study, radiographic outcome, missing data

HAL author manuscript inserm-00217436, version 1

Rheumatoid arthritis (RA) is the most common chronic inflammatory joint disease and is responsible for symptomatic manifestations (e.g., functional status, pain) and structural damage (i.e., damage of the articular cartilage and bone) (1). Effective disease-modifying anti-rheumatic drugs are increasingly available as therapy (2). Assessing such treatments requires the measurement of structural outcomes in randomized controlled trials (RCTs) to demonstrate a retardation of disease progression. Radiographic outcomes are often used as primary endpoints for assessing structural severity (3-6).

Because retardation of structural damage in RCTs requires observation over time, follow-up of patients often necessitates intermediate visits, requiring more than two sessions of radiography in most trials (7). Specific methods such as linear mixed-effects models, which exploit the richness of the dynamics obtained with such longitudinal, or repeated, measurements could be applied to estimate the treatment effect (8). Despite repeated measurements, calculating the mean change between baseline visit and end of the study in each group and comparing it by use of the classical t-test (or Mann-Whitney test for nonparametric comparisons) remains the standard analysis.

The intention-to-treat (ITT) principle is the cornerstone of RCTs (9-11) and is widely recommended to demonstrate the superiority of one treatment over another (12, 13). However, few trialists use this principle in analyzing their data (14, 15), particularly in trials evaluating radiographic outcomes in RA (16). The ITT principle requires that all patients, whether their data is complete or incomplete, be included in the statistical analysis. Approximately two-thirds of RCTs of RA have a missing data rate greater than 10% for radiographic outcomes (16), and trialists must use methods for dealing with missing data to apply the ITT principle.

In trials involving longitudinal measurements of radiographic outcome, missing data can result from lack of efficacy or adverse events, for example. When data are incomplete,

HAL author manuscript inserm-00217436, version 1

results of the trial can be affected in two major ways. First, missing data can result in a bias of treatment effect estimates. For example, patients experiencing greater deterioration in structural damage may be less likely to complete the visits. If missing data are ignored and analyses are based on only the data of patients who are doing well, then the disease progression could be underestimated (17). Second, missing data can result in loss of statistical power (i.e., the ability of the trial to detect a difference between groups) if data for some subjects are excluded from the analysis (17).

Several methods exist to adjust for missing data (18). However, conclusions of trials (i.e., superiority of one treatment over another or not ?) and treatment effect may be affected by the method used to handle missing data. We aimed to compare the impact of different approaches chosen to deal with missing data under a scenario that mimics trials of RA with a radiographic outcome. We performed a simulation study. Such studies, increasingly common in the medical literature, are used to assess the performance of statistical methods in relation to the known truth (19). We compared approaches to handle missing data in terms of statistical power and magnitude of bias introduced by missing data on treatment effect.

HAL author manuscript inserm-00217436, version 1

METHODS

The underlying clinical trial

We considered a simulation trial based on a 2-armed RCT resembling RA trials with a control group and an experimental group. The trial had a 2-year duration with 3 time points of measurement, 1 year apart, including that at baseline (7).

The primary endpoint was the Sharp-Van der Heijde score (20, 21), a semi-quantitative radiologic measure recommended as one of the two possible primary endpoints in evaluating structural damage (22). This score, ranging from 0 to 448, assesses erosions and joint-space narrowing separately in the hands and feet. Thirty-two joints in the hands and 12 in the feet are scored for erosions, with a maximum of 5 erosions per joint in the hands and 10 in the feet. Joint-space narrowing was graded from 0 to 4 in 30 joints in the hands and in 12 joints in the feet. The Sharp-Van der Heijde score is the sum of the erosion and joint-space narrowing scores.

Simulations of longitudinal measurements involved use of a linear mixed-effects model with a random intercept and slope. The individual intercept and slope were simulated for each patient, and radiographic data were simulated with a linear model from these individual parameters. All simulation values were chosen according to published data of the TEMPO study (23-25). We assumed that the baseline distribution of the radiographic score could be approximated by a log-normal distribution (mean=45, standard deviation=45). The mean progression can be assumed to be linear (26), although the evolution of data for individual patients shows high variability (27). The slope was simulated by a normal distribution.

HAL author manuscript inserm-00217436, version 1

Simulations were done under the alternative and the null hypothesis. Under the alternative hypothesis, the slope and its standard deviation were assumed to be greater in the control group than in the experimental group (mean change over 2 years=3, standard deviation=10, vs. mean=0, standard deviation 5, respectively), which reflects fewer benefits from treatment (i.e., greater deterioration of structural damage) in the control group. A sample size estimation for a two-sided test of efficacy (t-test) resulted in a sample size of 150 patients per group to achieve a type I error of 5% and a power of 90%. Under the null hypothesis, the mean change over 2 years in each group was the same (also with 150 patients per group).

Missingness mechanism

After the complete data sets were simulated, patients' data were deleted according to a predefined missingness mechanism. We considered only missing data with a monotone pattern (i.e., data for a patient up to a certain time). We assumed that all baseline data were observed. Further to our previous literature review (16), we assumed a dropoutrate of 15% at 2 years. Patients with disease progression greater or lower than a defined limit between 2 occasions dropped out of the trial. The probability of dropout was (i) 2.5% if the slope between 2 successive visits was negative (i.e., improvement), (ii) 5% if the slope was between 0 and 5 points (i.e., slight deterioration) and (iii) 20% if the slope was greater then 5 points (i.e., substantial deterioration) (scenario A). The limit of 5 points was chosen in accordance with published estimations of the minimal clinically important difference and the smallest detectable difference of the Sharp-Van der Heijde score, which are very close (approximately 5 points) (21). In scenario A, the probability of a missing value was arbitrarily chosen to ensure a dropout rate of approximately 15% at the final visit. In scenario B, the probability of a missing value was divided by 2 to achieve a dropout rate of approximately 7.5%.

HAL author manuscript inserm-00217436, version 1

Data analysis strategies

The 2-sided t-test was used to test the absolute change between the 2 groups (progression estimated by the change between the baseline and 2-year visit).

A linear model with mixed effects for repeated measurements with random intercepts and slopes was also considered. This model takes into account intermediate measurements (not just 2 measurements). With missing data, this model uses all available measurements. Restricted maximum likelihood estimation was performed. In the model, the interaction of group by time of visit (i.e., the difference in slopes) was a fixed effect. A F-test on this fixed effect was used to compare slopes.

Methods of managing missing data

We considered 3 methods of managing missing data. The two first methods are traditionally used (16), and the third one is promising. Of the first two methods, the case-complete analysis ignores the problem of missing data. When considering absolute change, the case-complete analysis affects only patients with complete data (i.e., complete data at baseline and final visits). In a linear mixed-effects model, the analysis refers to all available measurements. The second and most popular method of single imputation was the last observation carried forward (LOCF) method, whereby the last observation is carried forward and used for all missing observations at the remaining time points. The third method was multiple imputation (28, 29). Instead of filling in a single value for each missing value, this technique replaces each missing value of an incomplete data set with a set of plausible values (n=5 in our study) that represent the uncertainty in the correct value to impute. The data augmentation Monte Carlo Markov

HAL author manuscript inserm-00217436, version 1

chain replacement method was used. Then, each completed data set is analyzed by the analysis of choice, and results of imputed datasets are combined in a single analysis yielding point estimates and standard errors.

Methods of managing missing data and data analysis strategies were combined (Table 1) and applied to the 2 scenarios, A and B, of missing data.

Type I error, power and bias

The type I error and power of the t-test and F test under different approaches were computed. To estimate the empirical type I error of each approach, the entire trial simulation was repeated 1000 times under the null hypothesis. The empirical type I error was calculated as the proportion of P values < 0.05 from testing the null hypothesis of no difference in each simulated trial. Similarly, to estimate the empirical power of each approach, the entire trial simulation was repeated 1000 times under the alternative hypothesis. The estimated power for each approach was the proportion of these 1000 simulated trials showing statistically significant results ($P < 0.05$).

Under the alternative hypothesis (i.e., the hypothesis of a treatment effect), estimators of treatment effect were computed. Treatment effect was the difference in the absolute mean change between the 2 groups by t-test analysis or estimation of difference in slopes between the 2 groups by linear mixed-effects analysis. Averaging the estimates derived from the 1000 simulated trials allowed for estimating the expected mean of the treatment effect. Then comparing this expected mean to the "true" value (used to simulate data) allowed for calculating the bias. Bias was expressed as relative bias (in percentage) and as absolute bias (in Sharp-van der Heidje units).

HAL author manuscript inserm-00217436, version 1

Type I error, power and bias were also computed before the missingness mechanism was applied (i.e., on complete data sets for which an "ideal" analysis could be performed) as a reference and to check the validity of simulations.

Simulations of trials involved use of R 2.2.0. Management of missing data and data analysis involved SAS 9.1.

HAL author manuscript inserm-00217436, version 1

RESULTS

Under the alternative hypothesis, the missingness mechanism provided a missing data rate at 2 years of 15.3% for scenario A and 7.9% for scenario B (Table 2). Proportions of dropout in scenario A for conditions i, ii and iii were 2.5%, 5.0% and 20.0%, respectively. In scenario B, these proportions also approximated expected values (1.2%, 2.5% and 10.1%, respectively).

For scenario A under the null hypothesis, type I errors were maintained (i.e., approximately 5%) except for the following strategies: the case-complete and LOCF approaches by t-test (i.e., when comparing absolute change), and the LOCF approach by F test (i.e., when comparing slopes by linear mixed effect model) (Table 3).. As expected, the power for complete data sets (i.e., without missing data) was close to 90% whatever the data analysis used (absolute change or linear mixed-effects model). By scenario of missing data and when data were analyzed after using imputation techniques (i.e., LOCF or multiple imputation) or not (i.e., case complete), power ranged from 51% (t-test performed on case-complete data without missing-data handling methods) to 74% (t-test on missing data handled by multiple imputation). Whatever the data analysis, the all-available data and LOCF methods gave equivalent power. By multiple imputation, the power was always higher than that with other imputation strategies (Table 3).

As expected, the bias of treatment effect with complete data sets (i.e., without missing data) was very low (i.e., close to 0%) whatever the data analysis (Table 3). For scenario A, with 15.3% missing data at 2 years, the treatment effect estimated by linear mixed-effect model with multiple imputation approach was underestimated by 17.7% (equivalent to an underestimation of 0.53 points in Sharp-van der Heidje units) (Table 3). This bias was lower than that obtained with all-available-data (-25.0%) or LOCF (-35.4%) methods. When considering absolute change, again, multiple imputation gave the most precise treatment-

effect estimates (underestimation of 17.2% vs. –34.0% and –35.8% for case-complete and LOCF, respectively).

For scenario B, type I error was maintained for all strategies under the null hypothesis. Results under the alternative hypothesis with 7.9% missing data at 2 years showed less bias than with scenario A but followed approximately the same pattern (Table 3). Bias was minimal with multiple imputation (-8.6% for absolute change and –8.2% for linear mixed-effects model) and power with multiple imputation was better than with other imputation strategies.

HAL author manuscript inserm-00217436, version 1

HAL author manuscript inserm-00217436, version 1

DISCUSSION

This simulation study evaluated the impact of various approaches dealing with missing data in RCTs of RA with radiographic outcomes. With rate of missing data approximately 15% for scenario A, our simulation results showed multiple imputation with better power and more precise estimations of treatment effect, on average, than other approaches, whatever the data analysis, absolute change or linear mixed-effects model. With a linear mixed-effect model, treatment effect was underestimated by 17.7%. The power was lower than the nominal power (70% vs. 88%), but the ability to detect a difference was superior than with other methods of managing missing data. When using this statistical approach, type I error was maintained.

The problem of dealing with missing data is tackled extensively in methodological work involving radiographic endpoints in RA (30-32). The advice is generally to perform sensitivity analyses (i.e., a set of analyses showing the influence of different methods of handling missing data on the study results (17)) to ensure that the qualitative conclusion of a randomized trial provided by the main analysis is not affected by how missing data are handled. Sensitivity analyses allow for assessing the robustness of the results and are used as an additional support to the main analysis. Recently, 2 sensitivity analyses, evaluating different methods of handling radiographic missing data, were performed to confirm the robustness of radiographic results in published reports of RA trials (24, 33). However, although sensitivity studies should always be included in the statistical analysis plan of an RCT, they do not allow for drawing general conclusions (i.e., conclusions applicable to different trials) regarding the most appropriate method to use in dealing with missing data for the main analysis. Moreover, we thought it also of interest to assess whether the treatment effect is affected by the method use to handle missing data. This situation cannot be assessed in the framework of a sensitivity analysis (since the “true” effect is unknown) but can be

HAL author manuscript inserm-00217436, version 1

studied in the framework of a simulation study. To our knowledge, this is the first time a simulation approach has been used to investigate consequences of methods for dealing with missing data on power and bias of treatment effect in RA. Such an approach has been used for osteoporosis (34), another progressively deteriorating disease. In this study, the authors found no strategy adequate for universal use, particularly with a high missing data rate.

In RA trials involving longitudinal measurements of radiographic outcomes, 2 main sources of missing data are identified: lack of efficacy and adverse events. Unexpected selective dropout (preferentially in one group) because of lack of clinical efficacy may bias the trial results. In general, patients with a worse prognosis (greater disease activity, greater radiographic evidence of disease progression) have a higher prior probability of premature discontinuation in any clinical trial, and patients completing the entire trial have a more favourable prognosis, either by nature or by treatment (35). In our missingness scenario, we particularly focused on missing data due to lack of efficacy by our excluding data for patients having deteriorated disease with a high probability of dropout. We did not deal with missing data related to adverse events. Adverse events can lead patients to leave the trial regardless of good or bad treatment results. Because we compared two active treatments we assumed that missing data due to adverse events would have been similar in both groups and that they would have equally affected treatment results in each group (36), so we did not consider missing data related to adverse events. However, patients experiencing adverse events might have more comorbidities than others, which leads them to leave the trials.

In trials designed with longitudinal measurements, the use of appropriate statistical methods for repeated measurements is now recommended (24). Use of a linear mixed-effects models does not handle missing values; estimates take into account all available data whether longitudinal data are complete or not. If data are missing completely at random (i.e., when the missingness is not related to the observed or potential outcome of patients) or missing at

HAL author manuscript inserm-00217436, version 1

random (i.e., when the missingness can be explained entirely by the observed but not unobserved outcomes) (37), then the estimates will be unbiased. However, this case is no longer true when the missingness depends on the unobserved data (i.e., when the missing data are missing not at random) (37). This case cannot be excluded with missing radiographic data (e.g., selective dropout).

With linear mixed-effects analysis, tests theoretically tend to be more powerful than with absolute change analysis. In fact, the linear mixed-effects model takes into account how the disease and treatment affects each patient over time and how the radiographic data of the same patient are correlated. To exploit the richness of all measurements, this strategy could be particularly interesting when the number of intermediate visits is high (38). In this study, we considered only a 2-year trial duration and one intermediate visit. So our results do not totally confirm this improvement in power. However, in trials involving more visits, this method could be promising.

The case-complete approach conflicts with the ITT principle and its use is decreasing. However, the use of this approach gave us a reference for quantifying bias in treatment effect introduced by missing data. This bias cannot be neglected (i.e., approximately 34% with 15% missing data rate).

With the LOCF approach, the missing radiographic value is replaced by the last available value, assuming no change in radiographic score after the patient drops out. In RA trials, this concern can lead to considering dropouts as showing less radiographic evidence of deterioration as compared with completers. This approach is widely criticized and, not surprisingly, introduced the highest bias in our estimates of treatment effect.

As compared with the LOCF approach, multiple imputation has theoretically good statistical properties (e.g., unbiased estimates), provided that data are missing at random. In our study, whatever the data analysis strategy, absolute change or linear mixed-effects model,

HAL author manuscript inserm-00217436, version 1

multiple imputation gave substantially improved values for power and minimized the bias in estimates of treatment effect as compared with other approaches. However, this result could be improved by including in the imputation model selected covariates such as baseline predictors of dropout or correlated variables of the radiographic score (37).

A well-designed study should always consider the problem of minimizing missing data (e.g., ensuring appropriate follow-up for all randomized patients by scheduling patients for radiography even if they drop out of the study) (39). Statistical methods, however well designed, cannot address missing values of high proportions. In our work, the proportion of missing values was sufficiently low (i.e., 7.9% or 15.3%) so as to be reasonably considered with statistical methods. However, some statistical approaches dealing with missing data can minimize bias with reasonably low rates of missing data but cannot avoid it completely, particularly with selective dropout.

This study has caveats and limitations. First, these results may not be generalizable to all radiographic scores used in RA. Even if similar results in relative bias could be expected, extrapolation to another radiographic score would require a new simulation study. Second, the random simulations carried out may not reflect all the patterns of missing data that could occur in real situations. Furthermore, we did not explore all methods of dealing with missing data. Particularly, Cook and al. have proposed a specific method to deal with data missing not at random, which uses measurements obtained from out-of-study sources to estimate values for missing study data in a random effects model (40). In the same way, some other sophisticated methods, such as selection models or pattern mixture models, were not considered (41). However, because these models rely on many assumptions, their use suggests many assumptions, so they cannot be used as a main strategy when analyzing the results of an RCT. These models can be used in the framework of a sensitivity analysis.

In this simulation study, we showed the influence of the choice of analysis and methods for handling missing data when analyzing results of RCTs with radiographic outcomes. Our results, especially those obtained with multiple imputation, can help investigators in planning clinical trials, especially in choosing methods of imputation and data analysis. Regardless of the method of handling missing data chosen for the main analysis, sensitivity analysis is essential to confirm the robustness of the results.

HAL author manuscript inserm-00217436, version 1

HAL author manuscript inserm-00217436, version 1

References

1. Symmons DP. Disease assessment indices: activity, damage and severity. Baillieres Clin Rheumatol 1995;9:267-85.

2. Lee DM, Weinblatt ME. Rheumatoid arthritis. Lancet 2001;358:903-11.

3. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum 1993;36:729-40.

4. Recommendations for the registration of drugs used in the treatment of rheumatoid arthritis. Group for the Respect of Ethics and Excellence in Science (GREES): rheumatoid arthritis section. Br J Rheumatol 1998;37:211-5.

5. Boers M, van der Heijde DM. Prevention or retardation of joint damage in rheumatoid arthritis: issues of definition, evaluation and interpretation of plain radiographs. Drugs 2002;62:1717-24.

6. Soubrier M, Dougados M. How to assess early rheumatoid arthritis in daily clinical practice. Best Pract Res Clin Rheumatol 2005;19:73-89.

7. Baron G, Boutron I, Giraudeau B, Ravaud P. Reporting of radiographic methods in randomized controlled trials assessing structural outcomes in rheumatoid arthritis. Ann Rheum Dis 2006.

8. Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer; 2000.

9. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. Stat Med 1999;18:1903-42.

HAL author manuscript inserm-00217436, version 1

10. Lewis JA, Machin D. Intention to treat--who should use ITT? Br J Cancer 1993;68:647-50.

11. Lachin JM. Statistical considerations in the intent-to-treat principle. Control Clin Trials 2000;21:167-89.

12. CPMP CfPMP. ICH Topic E9 Statistical principles for clinical trials. The European Agency for the Evaluation of Medicinal Products 1998;CPMP/ICH/363/96.

13. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med 2001;134:663-94.

14. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. Jama 2003;290:1062-70.

15. Thabut G, Estellat C, Boutron I, Samama CM, Ravaud P. Methodological issues in trials assessing primary prophylaxis of venous thrombo-embolism. Eur Heart J 2006;27:227-36.

16. Baron G, Boutron I, Giraudeau B, Ravaud P. Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. Arthritis Rheum 2005;52:1858-1865.

17. CPMP CfPMP. Points to consider on missing data. The European Agency for the Evaluation of Medicinal Products 2001;CPMP/EWP/1776/99.

18. Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, et al. Analyzing incomplete longitudinal clinical trial data. Biostatistics 2004;5:445-64.

19. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Stat Med 2006;25:4279-92.

HAL author manuscript inserm-00217436, version 1

20. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27:261-3.

21. Bruynesteyn K, van der Heijde D, Boers M, van der Linden S, Lassere M, van der Vleuten C. The Sharp/van der Heijde method out-performed the Larsen/Scott method on the individual patient level in assessing radiographs in early rheumatoid arthritis. J Clin Epidemiol 2004;57:502-12.

22. van der Heijde D, Simon L, Smolen J, Strand V, Sharp J, Boers M, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. Arthritis Rheum 2002;47:215-8.

23. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet 2004;363:675-81.

24. van der Heijde D, Landewe R, Klareskog L, Rodriguez-Valverde V, Settas L, Pedersen R, et al. Presentation and analysis of data on radiographic outcome in clinical trials: experience from the TEMPO study. Arthritis Rheum 2005;52:49-60.

25. van der Heijde D, Klareskog L, Rodriguez-Valverde V, Codreanu C, Bolosiu H, Melo-Gomes J, et al. Comparison of etanercept and methotrexate, alone and combined, in the treatment of rheumatoid arthritis: two-year clinical and radiographic results from the TEMPO study, a double-blind, randomized trial. Arthritis Rheum 2006;54:1063-74.

26. Hulsmans HM, Jacobs JW, van der Heijde DM, van Albada-Kuipers GA, Schenk Y, Bijlsma JW. The course of radiologic damage during the first six years of rheumatoid arthritis. Arthritis Rheum 2000;43:1927-40.

HAL author manuscript inserm-00217436, version 1

27. Plant MJ, Jones PW, Saklatvala J, Ollier WE, Dawes PT. Patterns of radiological progression in early rheumatoid arthritis: results of an 8 year prospective study. J Rheumatol 1998;25:417-26.

28. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: J. Wiley & Sons; 1987.

29. Schafer J. Analysis of Incomplete Multivariate Data: New York: Chapman & Hall; 1997.

30. Symmons DP. Methodological issues in conducting and analyzing longitudinal observational studies in rheumatoid arthritis. J Rheumatol Suppl 2004;69:30-4.

31. Johnson K. Evidence from rheumatoid arthritis trials for approval: what does it mean? With special reference to using radiographic endpoints. J Rheumatol 2000;27:549-51.

32. Landewe R, van der Heijde D. Presentation and analysis of radiographic data in clinical trials and observational studies. Ann Rheum Dis 2005;64 Suppl 4:iv48-51.

33. Leung H, Hurley F, Strand V. Issues involved in a metaanalysis of rheumatoid arthritis radiographic progression. Analysis issues. J Rheumatol 2000;27:544-8; discussion 552.

34. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. Stat Med 2001;20:3931-46.

35. Landewe RB, Boers M, van der Heijde DM. How to interpret radiological progression in randomized clinical trials? Rheumatology (Oxford) 2003;42:2-5.

36. Mallinckrodt CH, Sanger TM, Dube S, DeBrota DJ, Molenberghs G, Carroll RJ, et al. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. Biol Psychiatry 2003;53:754-60.

37. Liu M, Wei L, Zhang J. Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. Pharm Stat 2006;5:7-18.

HAL author manuscript inserm-00217436, version 1

38. Diggle PJ, Heagarty P, Liang K, Zeger SL. Analysis of longitudinal data: Oxford University Press; 2002.

39. Auleley GR, Giraudeau B, Baron G, Maillefert JF, Dougados M, Ravaud P. The methods for handling missing data in clinical trials influence sample size requirements. J Clin Epidemiol 2004;57:447-53.

40. Cook NR. Imputation strategies for blood pressure data nonignorably missing due to medication use. Clin Trials 2006;3:411-20.

41. Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. Stat Med 2004;23:1455-97.

HAL author manuscript inserm-00217436, version 1

Table 1. Possible combinations of methods of managing missing data and data analysis strategies

Approach	Management of missing data	Data analysis strategies	Test
1	Case-complete analysis	Absolute change	t-test
2	Case-complete analysis*	Linear mixed-effects model	F test
3	LOCF	Absolute change	t-test
4	LOCF	Linear mixed-effects model	F test
5	Multiple imputation	Absolute change	t-test
6	Multiple imputation	Linear mixed-effects model	F test

*All available data in the context of linear mixed effect model. LOCF: last observation carried forward.

HAL author manuscript inserm-00217436, version 1

Table 2. Simulated rate of missing data in each arm of the trial at each radiology visit for scenario A, with 15.3% missing data, and scenario B, with 7.9% missing data under the alternative hypothesis.

		Rate of missing data (%)		
Scenario	**Radiology visit**	**Experimental group**	**Control group**	**Global**
A	**One year**	6.1	10.5	8.3
	Two year	11.1	19.5	15.3
B	**One year**	3.1	5.1	4.1
	Two year	6.0	9.9	7.9

HAL author manuscript inserm-00217436, version 1

Table 3. Results of simulation study (type I error, power, relative bias expressed as percentage; absolute bias expressed in terms of Sharp-van der Heijde units) for a dropout rate of 15.3% (scenario A) and 7.9% (scenario B) at 2 years.

Scenario/ Data analysis	Management of missing data	*Null hypothesis* Type I error	*Alternative hypothesis* Power	Treatment effect Relative Bias	Treatment effect Absolute Bias
A					
	Complete data set*	4.9	91.1	+0.1	+0.00
Absolute Change	Case complete	12.6	50.6	-34.0	-1.02
	LOCF	8.2	58.1	-35.8	-1.07
	Multiple imputation	6.2	74.3	-17.2	-0.52
	Complete data set*	4.2	88.4	+0.4	+0.01
Linear mixed-effects model	All available data	5.7	57.9	-25.0	-0.75
	LOCF	7.0	55.0	-35.4	-1.06
	Multiple imputation	4.1	69.5	-17.7	-0.53
B					
	Complete data set*	4.6	90.4	-0.8	-0.02
Absolute Change	Case-complete	5.9	75.4	-16.8	-0.51
	LOCF	5.5	77.8	-18.7	-0.56
	Multiple imputation	4.4	84.0	-8.6	-0.26
	Complete data set*	3.7	87.4	-0.8	-0.02
Linear mixed-effects model	All available data	5.4	72.0	-12.9	-0.38
	LOCF	4.7	72.4	-14.4	-0.43
	Multiple imputation	4.2	81.2	-8.2	-0.25

* Before applying missingness mechanism (i.e., results provided are obtained on complete datasets). LOCF: last observation carried forward.