The use of different methods for rapid determination of the ESR induces DAS28 misclassification in clinical practice

A. van der Maas¹, C.H.M. van den Ende¹, J. van Eerd², J. Fransen³, A.A. den Broeder¹

¹Department of Rheumatology, and ²Department of Clinical Chemistry, Sint Maartenskliniek, Nijmegen, The Netherlands; ³Department of Rheumatology, UMCN St. Radboud, Nijmegen, The Netherlands.

Abstract Objectives

Monitoring of disease activity using DAS28 is more effective than routine RA care, but the ESR measurement is time consuming. Alternative rapid ESR determination methods can be used but effects on DAS28 classification are unknown.

Methods

Alternative rapid ESR methods, including the Starrsed 30-minute mode and Alifax Roller Test-1TH, were compared to the Westergren method. Mean difference, limits of agreement (LoA) and intraclass correlation coefficients (ICC) were calculated. Based on these results, using a longitudinal design the percentage of DAS28 misclassification for the Alifax Roller Test-1TH was measured.

Results

The Alifax showed acceptable ICCs, but LoA were large. ICC was 0.67 (0.56-0.76), LoA -43;34. The longitudinal study on the Alifax (n=125) showed an ICC of 0.93, a kappa of 0.61, but disease activity was misclassified in 26% of the patients. Use of the ESR from the previous visit resulted in comparable levels of misclassification.

Conclusion

ESR measured by automated analysers like Alifax show acceptable ICC but LoA are large compared to the Westergren ESR. The Alifax Roller Test-1TH is very rapid but DAS28 misclassification is considerable and even as large as when using the ESR of the previous visit.

Key words

ESR, DAS28, disease activity, RA, monitoring

Aatke van der Maas, MD Cornelia H.M. van den Ende, PhD Juliëtte van Eerd, PhD Jaap Fransen, PhD Alfons A. den Broeder, MD, PhD

Please address correspondence and reprint requests to: Aatke van der Maas Sint Maartenskliniek, Department of Rheumatology, P.O. Box 9011, 6500 GM, Nijmegen, The Netherlands E-mail: a.vandermaas@maartenskliniek.nl Received on November 9, 2009; accepted in revised form on February 1, 2010.

© Copyright CLINICAL AND EXPERIMENTAL RHEUMATOLOGY 2010.

Introduction

Intensive monitoring of disease activity in patients with rheumatoid arthritis (RA) including settings goals and adjusting therapy accordingly has proved to be more effective than routine outpatients care (1-4). The Disease Activity Score 28 (DAS28) (5-6) is a validated and frequently used tool to monitor disease activity (6). It is a composite score based on the number of tender and swollen joints (of 28 joints), the erythrocyte sedimentation rate (ESR) and patient's global assessment of disease activity or general health on a 100 mm visual analogue scale (VAS) (5). On the basis of validated cut-off points DAS28 scores can be categorised into different levels of disease activity (ranging from remission, low, moderate and high disease activity) (5-8).

Originally the DAS28 has been validated using the Westergren method (9), which takes at least 60 minutes. This hampers the use of the DAS28 in daily practice. Several options can be considered to reduce the time needed to determine DAS28. First of all, automated rapid ESR analysers are available. These analysers provide an estimation of the Westergren ESR either by extrapolating from a shorter sedimentation time (for example ESR measured by Starrsed in 30-minute mode) or measure the ESR in a different method, for example by means of density measurement with infrared reading of the dynamics of erythrocytes (for example Alifax Roller Test-1 TH). Other options include the use of the ESR of the previous visit, which is frequently done when the current ESR is missing, or ask patients to come for ESR measurements a few days in advance, but this requires an extra visit. Also a DAS28 based on C-reactive protein (CRP) can be calculated, but the gain in time is only limited compared to using the ESR measured by Starrsed in 30-min mode. Furthermore Castrejón et al. showed that DAS28-ESR and DAS28-CRP are not fully equivalent (10). Finally, other disease activity scores not using ESR as one of their components can be used, like for example the CDAI, but since a considerable difference between the various scores was revealed further research seems warranted before using these indices in clinical practice (11-12). Feasible options at the moment are thus limited to the application of rapid ESR analysers or the use of the ESR of the previous visit.

Several automated ESR analysers are available nowadays, including the Alifax Roller Test-1TH (13-15) with an analysis time of 3 minutes. Automated rapid ESR analysers have shown to have good agreement with the Westergren ESR in terms of intraclass correlation coefficient >0.80 (13-15). However, two aspects remain poorly addressed. Firstly, limits of agreement (LoA) can be large even with adequate intraclass correlation coefficient (ICC) (16). Secondly, as the contribution of the ESR on the DAS28 is a result of a natural log transformation, data on agreement between various methods on the basis of non-transformed ESR scores are not adequate to study the impact on DAS28. Therefore, the influence of different ESR measurements on agreement and percentage of patients misclassified using the DAS28 is the most important variable to know before using these analysers, since this ultimately affects treatment decisions.

We therefore set out to determine the best alternative for determination of ESR to be used for DAS28 assessment in daily clinical practice, with emphasis on the proportion of patients being misclassified using the different methods.

Patients and methods

Study design and analysers

The study consists of three sub-studies. In the first two studies the ESR and LN(ESR) were measured in a crosssectional study design comparing ESR measured according to Westergren with ESR measured after 30 minutes and extrapolated and with a rapid ESR analyser, namely the Alifax Roller Test-1TH. After this, a third longitudinal study was done to determine the agreement on the level of DAS28 classification between the Alifax Roller Test-1TH and ESR according to Westergren.

The Alifax Roller Test-1TH measures the sedimentation and aggregation capacity of erythrocytes via optical density. During 20 seconds the density is

Competing interests: none declared.

measured a 1000 times, which then allows an approximation of the ESR. The Alifax Roller Test-1TH requires less than three minutes to assess the ESR (Alifax S.p.A., Polverara, Italy). ESR according to Westergren was measured by Starrsed, which is an automated Westergren method (Mechatronics B.V., Hoorn, the Netherlands). This analyser also has an option to extrapolate 30-minute sedimentation values to 60-minute Westergren values in a 30minute mode (30 minutes calculation method).

Study I: Starrsed in 30 minutes compared to Westergren by Starrsed in a cross-sectional study design

The ESR was measured in patients with rheumatologic diseases (RA, Psoriatic arthritis (PsA) and spondylarthropathies (SpA)) using the Westergren method via Starrsed (ESR_{WEST60}) and compared to the Starrsed 30-minute mode (ESR_{WEST30}). Mean difference with 95%CI, Bland and Altman analyses and ICC were calculated before and after log transformation. In Bland and Altman analyses the 95% limits of agreement, estimated by mean difference ± 1.96 standard deviation of the differences, provide an interval within which 95% of differences between measurements by the two methods are expected to lie (16).

Study II: Alifax Roller Test-ITH compared to Westergren by Starrsed

in a cross-sectional study design In a set of patients with rheumatologic diseases, the ESR was measured by the Alifax Roller Test-1 and compared to the 60-minute Westergren method measured by Starrsed. The same statistical analyses were done as in sub-study I.

Study III: Alifax Roller Test-1 compared to Westergren by Starrsed in longitudinal study design

The ESR was measured longitudinally in patients with RA who were treated with infliximab using the Alifax Roller Test-1TH and compared to the Starrsed according to the 60-minute Westergren method. A DAS28 was measured on several consecutive visits with time interval of 4-12 weeks when patients **Table I.** Mean difference, LoA and ICC of ESR measured by Starrsed in 30 minutes and Alifax Roller Test-1 compared to the ESR measured by the Westergren method (sub study I and II).

| | Mean difference (95% CI) | | LoA | | ICC | |
|------------------------------------------------------------|--------------------------|-----------------------|----------------|--------------|---------------------|---------------------|
| | Untransformed | 0.7*Ln | Untransformed | 0.7*Ln | Untransformed | 0.7*Ln |
| ESR _{WEST60} vs. ESR _{WEST30} (n=421) | 0.14 (-0.31– 0.60) (- | -0.01 -0.03 - 0.01 | -9.4 ; 9.7 | -0.47 ; 0.45 | 0.97 (0.96-0.98) | 0.93 (0.92-0.94) |
| ESR _{WEST60} vs. Alifax (n=125) | -4.4 (-7.81.0) (- | -0.01 0.09 – 0.07 | -43 ; 34 7) | -0.98 ; 0.87 | 0.67 (0.56-0.76) | 0.74 (0.65-0.81) |

LoA: limits of agreement; ICC: Intraclass Correlation Coefficient; ESR: Erythrocyte sedimentation rate; ESR_{WEST60}: ESR measured by Westergren method in 60 minutes; ESR_{WEST60}: ESR measured by Westergren method in 30 minutes and extrapolated to 60 minutes; 95% CI: 95% confidence interval.

received infliximab. Disease activity was classified - using validated cut-off points - as remission when the DAS28 was less than 2.6, as low when between 2.6 and 3.2, moderate between 3.2 and 5.1 and high when the DAS28 exceeded 5.1 (7-8). The different ESR measurements were used to calculate the DAS28 and afterwards the mean difference in DAS28, LoA, ICC, unweighed kappa (DAS28 classification in a 4x4 table) and percentage of misclassification were calculated. Finally the same calculations were done comparing DAS28 using the Westergren ESR measured on the last visit to the DAS28 using current Westergren ESR.

It was also investigated what the agreement was in classification of DAS28 change of more than 1.2 between two consecutive visits. Hereto, patients were classified as improved >1.2, not changed, and worsened >1.2 using the two ESR measurement methods (Alifax Roller Test-1TH and Westergren 60 minutes) and a 3x3 table was made and kappa analyses were done.

Results

Study I: Westergren by Starrsed in 30 minutes compared to Westergren by Starrsed in a cross-sectional study design

In 421 patients with various inflammatory (RA, AS, SpA) rheumatologic diseases ESR was measured with Starrsed in 60 and 30 minutes. There was good agreement in ESR values measured after 60 minutes and the extrapolated value after 30 minutes, shown by a mean difference in ESR of only 0.14 mm/h (95%CI -0.31–0.59 mm/h). LoA were -9.4 and 9.7 mm/h. Furthermore, the ICC was 0.97 (95% CI 0.96–0.98). After 0.7*log transformation the mean difference was even less, -0.01 (95%CI -0.03–0.02) and the LoA were -0.47 and 0.45. The ICC was 0.93 (95%CI 0.92–0.94) (Table I).

Study II: Alifax Roller Test-1TH compared to Westergren by Starrsed in a cross-sectional study design

In 125 patients with various inflammatory rheumatologic diseases the mean difference in ESR between Westergren and Alifax Roller Test-1TH was -4.4 mm/h (95% CI -7.8-1.0 mm/h). The ICC between Alifax Roller Test-1TH and Westergren was 0.67 (95% CI 0.56–0.76). The LoA between the different methods were -42.6 to 33.8. After 0.7*log transformation the mean difference between Westergren and Alifax Roller Test-1TH was -0.01 (95% CI -0.12-0.10). The ICC between Westergren and Alifax Roller Test-1TH was 0.74 (0.65–0.81). The LoA were -1.27 to 1.25 after log transformation (Table I).

Study III: Alifax Roller Test-ITH compared to Westergren by Starrsed in longitudinal study design

One hundred and twenty-five patients with RA treated with infliximab were included. The group consisted of 86 women and 39 men, and their average age was 59 years (SD 12). DAS28 at inclusion was 3.6 (SD 1.1) and ESR values ranged from 2 to 120 with both analysers. The mean difference in ESR was 15.1 and LoA were between -18.2 and 48.4 (see Fig. 1). The mean difference in DAS28 between the two methods



Fig. 1. Bland and Altman analysis of ESR measured by Starrsed and Alifax Roller Test-1.



was 0.29 (SD 0.28), with LoA between -0.27 and 0.85 (see Fig. 2). The ICC in DAS28 was 0.93 (95% CI 0.71-0.97). The kappa was 0.61 and of all patients 26% were misclassified, with a slight tendency towards overestimation of the disease activity (Table II).

The DAS28 based on the Westergren

ESR of the preceding visit had a mean difference of -0.07 (SD 0.31) compared to the current Westergren. LoA were - 0.69 to 0.55. ICC, kappa and percentage of misclassification were 0.96, 0.68 and 22% respectively, and results were comparable to values found using the Alifax Roller Test-1TH (Table II).

When the patients were categorised according to change in DAS28 between two consecutive visits, change was misclassified in 9.6% of the patients with a kappa of 0.63. The relatively low percentage of misclassification compared to the modest kappa was caused by a low incidence of DAS28 change: 81% of the patients did not have a change in DAS28 exceeding -1.2 or 1.2 according to both ESR measurement methods (Table III).

Discussion

The ESR measured by analysers like Alifax Roller Test-1TH compared to Westergren indeed demonstrated an acceptable intra-class correlation coefficient (13-15). Limits of agreement, however, were very large and a considerable percentage of patients were misclassified in DAS28 disease activity level, thus not even outperforming the use of the ESR of the previous visit. To see if the initial misclassification of the ESR measured by Alifax Roller Test-1TH could be eliminated two correction algorithms were used (data not shown). First the ESR was corrected for the ratio between ESR measured on the preceding visit by Alifax Roller Test-1 and Westergren. The second correction included the use of individual regression analyses to correct the Alifax Roller Test-1 ESR for the individual preceding measurements of two previous visits. Our data did not support the use of these correction algorithms since the proportion of misclassification did not improve. Furthermore significant change in DAS28 was also misclassified in nearly 10% of the patients. Misclassification in change is lower than misclassification in absolute disease activity, but this overestimates the agreement since the proportion of patients with change in DAS28 exceeding 1.2 was low (81% of patients had no change in DAS28 of more than 1.2 between the two visits). This is illustrated by a moderate kappa 0.6 despite low percentage of misclassification. The Starrsed 30-minute ESR however appears to be a good alternative to the 60-minute measurement with small limits of agreement and a very good ICC.

Internal validity of our study seems ade-

Table II. Mean difference, LoA, ICC, kappa and DAS28 misclassification in DAS28 with ESR measured by Alifax Roller Test-1 and with the previous ESR compared to the ESR measured by the Westergren method.

| | Mean difference (SD) | e LoA | ICC (95% CI) | Kappa | % misclassification |
|---------------------------------------------------------|-------------------------|--------------|------------------|-------|---------------------|
| DAS28 _{Alifax} vs. DAS28 _{WEST60} | 0.29 (0.28) | -0.27 ; 0.85 | 0.93 (0.71-0.97) | 0.61 | 26% |
| DAS28 ESR previous visit vs. DAS28 _{WEST60} | 0.07 (0.31) | -0.69 ; 0.55 | 0.96 (0.94-0.97) | 0.68 | 22% |

DAS28: Disease Activity Score 28; SD: standard deviation; 95% CI: 95% confidence interval; DAS_{WEST60} : DAS28 calculated with ESR measured by Westergren method in 60 minutes.

Table III. 3x3 Table of change in DAS28 where ESR is measured by Westergren and then compared to change in DAS28 when ESR is measured by Alifax Roller Test-1.

| | | Alifax | | | | |
|------------|--------------|--------|-------------|------|-------|--|
| | DAS28 change | <1.2 | -1.2 to 1.2 | >1.2 | Total | |
| Westergren | <-1.2 | 5 | 2 | 0 | 7 | |
| | -1.2 to 1.2 | 2 | 101 | 2 | 105 | |
| | >1.2 | 0 | 6 | 7 | 13 | |
| | Total | 7 | 109 | 9 | 125 | |

quate. The groups were adequate in size, as witnessed by the small confidence intervals on all measures. The third substudy was performed in a longitudinal cohort of 125 RA patients, representative for RA patients in daily clinical practice. Ideally all ESR measurement techniques could have been used in the last experiment, however this was not possible due to practical reasons, such as conversion and calibrating issues when using simultaneously the Starrsed in 30 and 60-minute mode. However, the use of several different ESR values (Westergren 60-minute and 30-minute mode, Alifax and previous visit Westergren) and different analyses techniques (ICC, LoA, DAS28 misclassification, DAS28 change misclassification) remain strong points of our study.

In addition to showing that the Alifax is not superior for daily clinical use to determine the DAS28 than use of the ESR of the previous visit, our data underscore the notion that to investigate the influence of the method used for ESR determination on the DAS28, it is not enough to only use correlation and kappa statistics. The percentage of misclassification can be substantial even when ICC and kappa seem adequate, as demonstrated by the LoA. These differences between the methods used to determine the test characteristics of a test strongly support the need to investigate a test in the same way as it is used in daily practice. Another point of interest in the assessment of a new ESR measurement method is that any alternative should at least outperform the misclassification based on the ESR of the previous visit. Our results showed no difference between the performance of the Alifax Roller Test-1TH and the use of the ESR of the previous visit. Some points of discussion remain. For example, the percentage of misclassification would be smaller if the DAS is calculated instead of the DAS28, since the relative contribution of the ESR is smaller in this calculation. However, the DAS is more extensive and therefore more time consuming. A second argument that can be made in favour for the use of rapid ESR analysers could be that it is only relevant to classify patients in low and non-low (moderate and high) disease activity. Although this also would reduce misclassification (from 26% to 15% in our study), this would hamper the measurement of remission, which is the future goal of RA therapy. Thirdly, all validation work on the DAS(28) has originally been done using the ESR according to the Westergren, either by hand or automated using the Starrsed (personal communication). This supports our choice for the Starrsed as gold standard, but DAS28 validation can of course be repeated using other methods as well. Finally, we've focused on the influence of different ESR measurement in DAS28 misclassification, but random or systemic errors in measuring the other components of the DAS28 can also cause misclassification and the magnitude of these errors in clinical practice are thus far not fully known. Also, other rapid ESR measurement methods might become available in the future, and their influence on DAS28 misclassification may be more acceptable.

Only one other study concerning this subject has been published. Levitus et al also investigated the influence of ESR measurement by Alifax Roller Test-1TH on disease activity classification in patients with rheumatoid arthritis (17). Comparable - though somewhat smaller - percentages of misclassification were found (11% vs. 15%) when disease activity was classified in two levels. Their study, however, did not - amongst other things - include the use of other ESR methods like Starrsed 30-minute mode and previous ESR, did not use longitudinal data and also did not assess misclassification of change in DAS28 (17).

In conclusion, it is important to realise according to which method ESR is calculated in clinical practice since it substantially affects the validity of the DAS28, the use of which is advocated by guidelines (18-20). Alifax Roller Test-1TH is – though used in many laboratories – has shown to be not suitable for this purpose.

References:

- 1. GRIGOR C, CAPELL H, STIRLING A *et al.*: Effect of a treatment strategy for tight control of rheumatoid arthritis (the TICORA study): a single blind randomised controlled trial. *Lancet* 2004; 364: 263-9.
- FRANSEN J, MOENS HB, SPEYER I, VAN RIEL PLCM: Effectiveness of systematic monitoring of rheumatoid arthritis disease activity in daily practice: a multicentre, cluster randomised controlled trial. *Ann Rheum Dis* 2005; 64: 1294-8.
- GOEKOOP-RUITERMAN YP, DE VRIES-BOU-WSTRA JK, ALLAART CF et al.: Clinical and radiographic outcomes of four different treatment strategies in patients with early

Rapid ESR determination and DAS28 misclassification / A. van den Maas et al.

rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum* 2005; 52: 3381-90.

- 4. GOEKOOP-RUITERMAN YP, DE VRIES-BOU-WSTRA JK, KERSTENS PJ *et al.*: DAS-driven therapy versus routine care in patients with recent-onset active Rheumatoid Arthritis. *Ann Rheum Dis* 2010; 69: 65-9.
- PREVOO ML, VAN 'T HOF MA, KUPER HH, VAN LEEUWEN MA, VAN DE PUTTE LBA, VAN RIEL PLCM: Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995; 38: 44-8.
- PREVOO ML, VAN RIEL PLCM, VAN 'T HOF MA et al.: Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. Br J Rheumatol 1993; 32: 589-94.
- 7. VAN GESTEL AM, PREVOO ML, VAN 'T HOF MA, VAN RIJSWIJK MH, VAN DE PUTTE LBA, VAN RIEL PLCM: Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. Arthritis Rheum 1996 Jan; 39: 34-40.
- 8. FRANSEN J, CREEMERS MC, VAN RIEL PLCM: Remission in rheumatoid arthritis: agree-

ment of the disease activity score (DAS28) with the ARA preliminary remission criteria. *Rheumatology* (Oxford) 2004; 43: 1252-5.

- 9. WESTERGREN A: Studies of the suspension stability of the blood pulmonary tuberculosis. *Acta Med Scand* 1921; 54: 247-82.
- CASTREJÓN I, ORTIZ AM, GARCÍA-VICUÑA R *et al.*: Are the C-reactive protein values and erythrocyte sedimentation rate equivalent when estimating the 28-joint disease activity score in rheumatoid arthritis? *Clin Exp Rheumatol* 2008; 26: 769-75.
- 11. SALAFFI F, CIMMINO MA, LEARDINI G et al.: Disease activity assessment of rheumatoid arthritis in daily practice: validity, internal consistency, reliability and congruency of the Disease Activity Score including 28 joints (DAS28) compared with the Clinical Disease Activity Index (CDAI). Clin Exp Rheumatol 2009; 24: 552-9.
- MICHAUD K, MIKULS TR, CALL SE *et al.*: Poor to modest agreement between rheumatoid arthritis response measures in clinical practice. *Clin Exp Rheumatol* 2009; 27: 633-40.
- PLEBANI M, DE TONI S, SANZARI MC, BER-NARDI D, STOCKREITER E: The TEST 1 automated system: a new method for measuring the erythrocyte sedimentation rate. *Am J Clin Pathol* 1998; 110: 334-40.
- 14. GIAVARINA D, CAPUZZO S, CAUDURO F, CARTA M, SOFFIATI G: Internal quality control for erythrocyte sedimentation rate meas-

ured by TEST-1 Analyzer. Clin Lab 2002; 48: 459-62.

- DE JONG N, SEWKARANSING I, SLINGER J, RIJSDIJK JJM: Erythrocyte sedimentation rate by the Test-1 analyzer. *Clin Chem* 2000; 46: 881-2.
- BLAND JM, ALTMAN DG: Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135-60.
- 17. LEVITUS M, PELLICIA A, VAN DE STADT RJ, VOSKUYL AE, BOUMAN AA: Is the Alifax Test-1TH useful to determine the Disease Activity Score (DAS28) in rheumatoid arthritis patients? *Clin Rheumatol* 2009; 28: 469-74.
- 18. COMBE B, LANDEWE R, LUKAS C et al.: EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2007; 66: 34-45.
- KENNEDY T, MCCABE C, STRUTHERS G et al.: BSR guidelines on standards of care for persons with rheumatoid arthritis. *Rheuma*tology 2005; 44: 553-6.
- 20. VAN RIEL PLCM, BARENDSEN BC, CROONEN-BORG JJ, FOLMER H, FORMSMA SA, HAZES JMW: Richtlijn Diagnostiek en behandeling van reumatoide artritis. 2009 Nederlandse vereniging Reumatologie. Van Zuiden communications. *ISBN* 978-90-8523- 178-3.