GUEPARD treat-to-target strategy is significantly more efficacious than ESPOIR routine care in early rheumatoid arthritis according to patient-reported outcomes and physician global estimate

Isabel Castrejón, Theodore Pincus, Martin Soubrier, Yih Chang Lin, Anne-Christine Rat, Bernard Combe and Maxime Dougados

Abstract

Objective. To analyse seven RA Core Data Set measures and three indices for their capacity to distinguish treatment results in early RA in the GUEPARD treat-to-target clinical trial vs ESPOIR routine care.

Methods. Post hoc analyses compared 65 GUEPARD and 130 matched control ESPOIR patients over 6 and 12 months for mean changes in measures, relative efficiencies and standardized response means (SRM). Three indices—28-joint disease activity score (DAS28), clinical disease activity index (CDAI) and routine assessment of patient index data (RAPID3)—were compared for mean changes and numbers of patients with high, moderate or low activity or remission using κ values.

Results. Greater improvement was seen for GUEPARD vs ESPOIR, statistically significant for physician and patient global estimates and pain and health assessment questionnaire physical function (HAQ-FN), but not joint counts and laboratory tests. Relative efficiencies with tender joint count as the referent measure indicated that pain (2.57) and global estimates by patient (3.13) and physician (2.31) were most efficient in distinguishing GUEPARD from ESPOIR. Mean improvements in GUEPARD vs ESPOIR were −3.4 vs −2.6 for DAS28 (0−10) (24%), −29.8 vs −23.1 for CDAI (0−76) (23%) and −13.0 vs −7.8 for RAPID3 (0−30) (40%) (all P < 0.01); agreement was moderate between CDAI vs DAS28 (κ = 0.56) and vs RAPID3 (κ = 0.48), and fair between DAS28 vs RAPID3 (κ = 0.26).

Conclusion. Patient and global measures indicate greater efficacy than joint counts or laboratory measures in detecting difference between GUEPARD treat-to-target and ESPOIR routine care. A RAPID3 of only patient measures may help guide treat-to-target in busy clinical settings.

Key words: treat-to-target, patient-reported outcomes, assessment, rheumatoid arthritis, patient questionnaires.

Introduction

RA is characterized by the absence of a single gold standard measure, such as blood pressure or serum haemoglobin A1c, that can be applied to all individual patients with a specific diagnosis [1]. Therefore a core data set of eight measures and indices of multiple measures have been developed for patient assessment [2]. The RA Core Data Set includes three measures from a health professional [tender joint count (TJC), swollen joint count (SJC), physician global estimate of status (DROCGL)]; three from a patient self-report questionnaire [physical function (FN), pain
and patient global estimate of status (PATGL)]; one laboratory test of an acute phase reactant, either ESR or CRP; and, in studies of 12 months or longer, a radiographic score.

Indices of multiple measures from the Core Data Set have been developed to assess and monitor RA patients in clinical trials and clinical care. The 28-joint DAS (DAS28) is the most specific index for RA [3], but it has limitations in routine care, including complex calculations (albeit available at an excellent website, http://www.4s-dawn.com/DAS28/DAS28.html); the need for a laboratory test, which often is not available at the time of the visit and in some situations is poorly informative [4], and the need for a formal joint count. The clinical disease activity index (CDAI) [5, 6] simplifies calculations and does not require a laboratory test, but does require formal joint counts, which have limitations in measurement [7] and usually are not performed at most visits of RA patients [8]. Routine assessment of patient index data (RAPID3), an index of only the three patient self-report Core Data Set measures, distinguishes active from control treatments in clinical trials of LEF [9], MTX [9], adalimumab [10], abatacept [11] and certolizumab [12] at levels similar to DAS28 and CDAI, and requires only about 5 s to score on a multidimensional health assessment questionnaire (MDHAQ) vs more than 100 s to score a DAS28 or CDAI [13].

A treat-to-target strategy guided by DAS28 results in better outcomes than traditional non-quantitative care [14]. A recent report compared results of treatment in two populations of early RA patients: the GUEPARD (Guérir la Polyarthrite Rhumatoïde Débutante) clinical trial with treat-to-target tight control based on DAS28 vs the ESPOIR (Etude et Suivi des Polyarthrites Indifférenciées Récentes) cohort with routine care [15]. Improvement appeared greater in GUEPARD vs ESPOIR patients, at higher levels according to patient and global RA Core Data Set measures than according to joint counts and laboratory tests. These observations suggested that further formal comparisons of each RA Core Data Set measure and DAS28, CDAI and RAPID3 to distinguish results of GUEPARD tight control vs ESPOIR routine care would be of value, as presented in this report.

Materials and methods

Databases

Post hoc analyses were performed on two databases of patients with early RA: 65 patients from the GUEPARD tight control trial and 130 matched patients from the ESPOIR usual care cohort [15]. GUEPARD includes patients with RA as defined by the 1987 criteria of the ACR with maximum disease duration of 6 months and DAS28-ESR >5.1. Patients who met eligibility criteria were randomized to receive MTX monotherapy or a combination of MTX and adalimumab (ADA), with adjustment of the treatment every 3 months to achieve a DAS28-ESR <3.2. If this target was not met, the following steps were taken sequentially: MTX and ADA (40 mg every other week), MTX and ADA (40 mg/week), MTX and etanercept (25 mg twice a week) and MTX combined with LEF.

ESPOIR is a large national multicentre, longitudinal and prospective cohort initiated by the French Society of Rheumatology [16]. Patients with undifferentiated arthritis or RA, with disease duration <6 months who were both DMARD- and glucocorticoid-naïve, were recruited. Patients were seen every 6 months during the 2 first years and then every year. The patients were treated with routine care, without a specific target, and followed by their rheumatologists, without a treatment protocol for a target. The approach was determined entirely by the treating rheumatologist, including a decision to initiate biologic agents. These prior studies and compilation of the databases for further analyses were conducted with the approval of the ESPOIR central ethics committee of Montpellier and the GUEPARD central ethics committee of Cochin, Paris. The analysis plan for this study was submitted to the Institutional Review Board (IRB) of the New York University School of Medicine, which deemed the planned analyses of deidentified data exempt from requirement of IRB approval.

A total of 130 patients in the ESPOIR cohort fulfilling the inclusion criteria of the GUEPARD trial and with no missing data at baseline were matched 1:2 using a propensity score. A propensity score is used in statistical analyses of non-randomized data to attempt to match patients to provide an unbiased estimation of group effects by adjusting for specific variables. The main advantage of a propensity score over random assignment is that it avoids the ethical considerations that arise when a potentially beneficial treatment or strategy is denied at random. Non-randomized data are less costly than clinical trials, and a propensity score allows the analysis of existing data. A propensity score is limited, as one controls only for recognized variables, while hidden bias may remain because of other non-recognized variables. Random assignment provides more confidence that both groups are similar on both observed and non-observed characteristics. In addition, matching can only estimate effects where there is overlap between groups, whereas random assignment ensures that there is common support across the whole sample.

The propensity score was computed using a multivariate logistic model. The following demographic and disease characteristics at baseline were used as covariates in the model: treatment centre; gender; age at inclusion; disease duration; BMI; RF and anti-CCP antibody status; ESR; CRP; tender and swollen 28 joint counts; patient’s assessments of pain, disease activity and fatigue; doctor’s assessment of disease activity on a 0–100 mm visual analogue scale (VAS); health assessment questionnaire physical function (HAQ-FN); physical and mental components of the 36-item Short Form Health Survey (SF-36); erosive disease, erosion score and narrowing score. A more detailed description of the matching methodology and these patient groups is found in a previous report [15].
Study RA Core Data Set measures

The RA Core Data Set measures include the number of tender joints (TJC) (0–28); number of swollen joints (SJC) (0–28); physician global estimate (DOCGL) (VAS 0–100); HAQ-FN (0–3); pain (VAS 0–100); patient global estimate (PATGL) (VAS 0–100); ESR (mm/h) or CRP (mg/dl); and, in studies of ≥12 months, radiographic scores according to the modified Sharp/van der Heijde score. Each of these variables was analysed at baseline, and at 6 and 12 months of follow-up. HAQ-FN was recalculated from 0 to 0 (multiplied by 3.3) in order to score RAPID3 (see below).

Composite RA indices

Three composite indices—DAS28, CDAI and RAPID3—are calculated according to the formula: DAS28 = (0.56×√(TJC) + (0.28×√(SJC) + (0.7×ln(ESR)) + [0.014 × PATGL] (in mm)), for a total score of 0–10 [3]. DAS28 categories include remission (≤2.6), low activity (2.6–3.2), moderate activity (3.2–5.1) and high activity (>5.1) (Supplementary data Table S1, available at Rheumatology Online).

CDAI is a simple sum of TJC (0–28) + SJC (0–28) + DOCGL (0–10) + PATGL (0–10), for a total score range of 0–76 [6]. CDAI categories include remission (≤2.8), low activity (2.9–10.0), moderate activity (10.1–22.0) and high activity (>22) (Supplementary data Table S1, available at Rheumatology Online).

RAPID3 is a composite index of three 0–10 scales for HAQ-FN (recalculated from 0–3 on the HAQ, as noted above), pain and PATGL, for a total score range of 0–30 [12]. RAPID3 categories include remission (≤3), low activity (3.1–6.0), moderate activity (6.1–12.0) and high activity (>12) [13] (Supplementary data Table S1, available at Rheumatology Online).

Statistical analysis

All statistical analyses were performed using Stata 12.0 for Windows (StataCorp LP, College Station, TX, USA). Baseline demographic and clinical measures were compared in GUEPARD tight control vs ESPOIR routine care using the Student’s t test for variables normally distributed and Wilcoxon rank-sum test for variables with non-normal distribution.

Differences between GUEPARD (tight control) vs ESPOIR (routine care) in the responsiveness of each measure to document clinical changes between baseline and 12 months were calculated according to the method of Tugwell et al. [17]. Initially, standardized response means (SRMs) were calculated for each measure, as described by Wells et al. [18] (see supplementary data Appendix for formulas, available at Rheumatology Online). Relative efficiencies are expressed in relation to the TJC as the referent measure [17].

A relative efficiency >1 indicates that the measure is more efficient than TJC to detect a difference between results of two treatment strategies. These analyses were performed without adjustment for mean daily dosage of corticosteroids or the duration of biological agent therapy, as in the previous report [15]. Results were similar at 6 and 12 months, and only 12-month data are presented for analyses of individual measures.

Each of the three composite indices—DAS28, CDAI and RAPID3—was calculated at baseline, 6 months and 12 months. Mean changes from baseline to endpoints at 6 months and 12 months were compared in GUEPARD vs ESPOIR using Student’s t test, with statistical significance adjusted for eight comparisons for individual measures and three comparisons for indices. The numbers and proportions of patients in four categories—high, moderate and low activity and remission—were computed at base line, 6 and 12 months according to each index. The agreement of the disease activity categories was measured using k statistics [19]; strength of agreement was interpreted as k > 0.81 = very good agreement, 0.61–0.80 = substantial agreement, 0.41–0.60 = moderate agreement, 0.21–0.40 = fair agreement and ≤ 0.20 = poor agreement [20].

Results

Individual RA Core Data Set measures

Baseline demographic and disease variables were similar in GUEPARD and ESPOIR (Table 1). The only significant differences were that patients in the GUEPARD tight control clinical trial had longer mean disease duration vs the ESPOIR routine care patients (5.4 vs 3.5 months), higher likelihood of treatment with biologic therapies (82% vs 13%), a lower percentage taking glucocorticoids (50% vs 75%) and a lower mean dose of glucocorticoids (4.7 vs 8.1) (all P < 0.001) (Table 1).

Substantial improvement was seen in both GUEPARD and ESPOIR after 6 and 12 months of follow-up (Table 2). Differences in improvement were statistically significantly greater in the GUEPARD tight control group for HAQ-FN, pain, PATGL and TJC unadjusted, and all but TJC adjusted for eight comparisons. However, differences were not statistically significantly greater in GUEPARD vs ESPOIR for SJC, ESR, CRP and radiographic progression (Table 2).

High SRMs >4 were observed for PATGL, pain and DOCGL, compared with 2–3 for TJC and HAQ-FN, and <1 for SJC, ESR, CRP and radiographic score (Table 3). Relative efficiencies with TJC as the referent measure indicated that PATGL (3.13), pain (2.57) and DOCGL (2.31) were more efficient than TJC to distinguish GUEPARD from ESPOIR (Table 3), while HAQ-FN (0.43), SJC (0.41), ESR (0.01), CRP (0.12) and radiographic score (0.31) were less efficient than TJC to recognize differences between the two patient groups (Table 3).

Three RA indices: DAS28, CDAI and RAPID

Statistically significant improvements from mean baseline values over 6 and 12 months were seen in both GUEPARD and ESPOIR, according to DAS28, CDAI and RAPID3 (Table 4). Mean improvement over 6 and 12 months was
TABLE 1 Baseline demographic and clinical characteristics of patients in GUEPARD (tight control) and ESPOIR (routine care)

<table>
<thead>
<tr>
<th>Demographic variables</th>
<th>GUEPARD (n = 65)</th>
<th>ESPOIR (n = 130)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic variables</strong></td>
<td><strong>Age, mean (s.d.), years</strong></td>
<td>47.9 (15.7)</td>
<td>48.1 (11.5)</td>
</tr>
<tr>
<td></td>
<td><strong>Women, n (%)</strong></td>
<td>52 (80.0)</td>
<td>98 (75.4)</td>
</tr>
<tr>
<td><strong>Core Data Set measures</strong></td>
<td><strong>TJC (0–28), mean (s.d.)</strong></td>
<td>14.0 (6.7)</td>
<td>14.1 (6.8)</td>
</tr>
<tr>
<td></td>
<td><strong>SJC (0–28), mean (s.d.)</strong></td>
<td>10.1 (4.9)</td>
<td>10.7 (5.7)</td>
</tr>
<tr>
<td></td>
<td><strong>Physician global assessment (VAS 0–100), mean (s.d.)</strong></td>
<td>67.6 (17.2)</td>
<td>65.9 (18.1)</td>
</tr>
<tr>
<td></td>
<td><strong>HAQ-FN score adjusted (0–3 scale), mean (s.d.)</strong></td>
<td>1.5 (0.6)</td>
<td>1.5 (0.7)</td>
</tr>
<tr>
<td></td>
<td><strong>Patient global assessment (VAS 0–100), mean (s.d.)</strong></td>
<td>59.5 (22.0)</td>
<td>55.7 (20.7)</td>
</tr>
<tr>
<td></td>
<td><strong>Patient global assessment (VAS 0–100), mean (s.d.)</strong></td>
<td>68.1 (18.7)</td>
<td>67.9 (21.9)</td>
</tr>
<tr>
<td></td>
<td><strong>ESR, mean (s.d.), mm/h</strong></td>
<td>37.3 (22.4)</td>
<td>39.5 (25.7)</td>
</tr>
<tr>
<td></td>
<td><strong>CRP, mean (s.d.), mg/l</strong></td>
<td>28.7 (33.4)</td>
<td>29.8 (32.3)</td>
</tr>
<tr>
<td></td>
<td><strong>Modified Sharp–van der Heijde (0–448), mean (s.d.)</strong></td>
<td>7.7 (13.4)</td>
<td>5.8 (8.0)</td>
</tr>
<tr>
<td></td>
<td><strong>Erosion score (0–280), mean (s.d.)</strong></td>
<td>2.4 (4.9)</td>
<td>2.4 (3.7)</td>
</tr>
<tr>
<td></td>
<td><strong>JSN score (0–168), mean (s.d.)</strong></td>
<td>5.3 (9.5)</td>
<td>3.5 (5.4)</td>
</tr>
<tr>
<td><strong>Other disease measure</strong></td>
<td><strong>Symptom duration, mean (s.d.), months</strong></td>
<td>5.4 (4.7)</td>
<td>3.5 (2.0)</td>
</tr>
<tr>
<td><strong>Therapy variables</strong></td>
<td><strong>Patients receiving biologic therapy during the study, n (%)</strong></td>
<td>53 (82)</td>
<td>17 (13)</td>
</tr>
<tr>
<td></td>
<td><strong>Duration of biologic therapy, mean (s.d.), months</strong></td>
<td>6.8 (3.6)</td>
<td>4.9 (3.6)</td>
</tr>
<tr>
<td></td>
<td><strong>Patients receiving glucocorticoid treatment, n (%)</strong></td>
<td>33 (50)</td>
<td>97 (75)</td>
</tr>
<tr>
<td></td>
<td><strong>Daily dose of prednisolone, mean (s.d.), mg</strong></td>
<td>4.7 (3.3)</td>
<td>8.1 (5.2)</td>
</tr>
</tbody>
</table>

aWilcoxon rank-sum test for non-normally distributed variables. *Statistically significant results at P < 0.05. Student’s t-test for normally distributed variables.

TABLE 2 Changes in variables over 12 months in GUEPARD (tight control) vs ESPOIR (routine care)

<table>
<thead>
<tr>
<th>GUEPARD (n = 65)</th>
<th>ESPOIR (n = 130)</th>
<th>P-valuea</th>
<th>Adjusted P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TJC (0–28)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>P-value</strong></td>
</tr>
<tr>
<td><strong>Physician global assessment (VAS 0–100)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>P-value</strong></td>
</tr>
<tr>
<td><strong>CRP, mg/l</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>P-value</strong></td>
</tr>
<tr>
<td><strong>Modified Sharp–van der Heijde score (0–448 scale)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>Mean change (95% CI)</strong></td>
<td><strong>P-value</strong></td>
</tr>
</tbody>
</table>

NP: not performed (adjusted P-value not calculated when unadjusted P-value >0.05). aStudent’s t-test. *Statistically significant results at P < 0.005 (Bonferroni adjustment for nine comparisons).

Differences at 6 months were −3.3 vs −2.1 (33.0%) for DAS28 (0–10) and −3.1 vs −1.8 (32.2%) for CDAI (0–76), and −1.2 vs 0.2 (45.0%) for RAPID3 (0–30) (all P < 0.001). Differences at 12 months were −3.4 vs −2.6 (23.5%) for DAS28 (0–10), −29.8 vs −23.1 (22.4%) for CDAI (0–76) and −13.0 vs −7.8 (40.0%) for RAPID3 (0–30) (all P < 0.01) (Table 4).
The percentage of patients in remission after 6 months was significantly higher in the GUEPARD tight control group and similar for the three indices (Table 5). At 6 months, moderate agreement was seen between CDAI and DAS28 ($k = 0.56$) and between RAPID3 and CDAI ($k = 0.48$); fair agreement was seen between DAS28 and RAPID3 ($k = 0.26$). At 12 months, substantial agreement was seen between CDAI and DAS28 ($k = 0.62$); fair agreement was seen between RAPID3 and CDAI ($k = 0.37$) and between DAS28 and RAPID3 ($k = 0.31$).

**Discussion**

This report presents data consistent with extensive evidence that a treat-to-target strategy has greater efficacy in the treatment of RA than routine care [14]. To the best knowledge of the authors, this is the first report to compare relative efficiencies of RA Core Data Set measures in two independently recruited patient groups rather than two arms of a clinical trial. The greater efficacy of the GUEPARD tight control treat-to-target strategy vs ESPOIR routine care is seen at higher levels according to patient self-report and global measures compared with joint counts and laboratory tests.

These findings also are consistent with evidence of high relative efficiencies of patient report measures compared with joint counts and laboratory tests to distinguish active from control treatments in clinical trials of LEF [9], MTX [9], adalimumab [10], abatacept [11] and certolizumab [12]. In these trials, variation in the most efficient measure is seen, although DOGGL, PATGL, pain and HAQ-FN generally....
show greater efficiency than TJC, SJC, ESR and CRP. However, each measure is most efficient in a certain trial, underscoring a need for an RA Core Data Set of seven measures. Each RA Core Data Set measure has a similar order of magnitude to detect differences in clinical trials and clinical care. Patient self-report and global measures have greater sensitivity, while joint counts and laboratory tests have greater specificity.

RAPID3 is an index that includes only self-report data provided by patients, without a need for formal joint counts from health professionals, scored in 5 s on a multidimensional health assessment questionnaire (MDHAQ) vs ≥ 100 s for DAS28 or CDAI [13]. RAPID3 yielded similar, but not identical, results compared to DAS28 or CDAI for categories of high, moderate and low activity, and remission in this study and others [13, 21]. All RAPID3 data are from the same observer—the patient—overcoming the need for the same health professional observer to perform a joint count at each visit. Indeed, patient self-report measures appear as reproducible as physician-performed formal joint counts [7], in addition to their similar or greater relative efficiencies. Moreover, completion of a self-report questionnaire by a patient in the waiting area provides relevant information to the clinician before seeing the patient.

An international expert committee recommended that a treat-to-target strategy in daily clinical practice should include a validated composite measure of disease activity to guide treatment decisions in routine clinical practice [22]. Collection of RAPID3 scores in the waiting area at each visit ensures that some quantitative data will be available in addition to narrative descriptions of patient status. A careful joint examination, but not necessarily a formal joint count, must also be performed.

Patient completion of an MDHAQ prior to seeing the doctor in no way prevents collection of a formal joint count, DAS28, CDAI or any other measure. As laboratory tests and SJCs are more prognostic of structural damage than patient self-report measures, these measures and a DAS28 or CDAI should be obtained to guide clinical care. Nonetheless, physical function on a patient questionnaire is more significant than radiographs or laboratory tests in the prognosis of most severe outcomes of RA, including work disability, mortality, costs and even joint replacement surgery [23]. A RAPID3 score and DAS28 or CDAI would appear valid for a treat-to-target strategy for optimal care of patients with RA.

Rheumatology key messages

- Efficacy of treat-to-target in patients with RA appears greater for patient-reported measures than joint counts or laboratory tests.
- RAPID3 detects differences between RA treatment strategies as effectively as DAS28 and CDAI.
- RAPID3 may be of value to assess the clinical status of patients with rheumatic diseases in busy clinical settings.

Acknowledgements

The authors thank Nathalie Rincheval for expert monitoring and data management, and all the investigators who recruited and followed the patients (F. Berenbaum, Paris-Saint Antoine; M.C. Boissier, Paris-Bobigny; A. Cantagrel, Toulouse; B. Combe, Montpellier; M. Dougados, Paris-Cochin; P. Fardelonne and P. Boumier, Amiens; B. Fautrel and P. Bourgeois, Paris-La Pitié; R.M. Filpo, Lille;
An unrestricted grant from Merck Sharp and Dohme (MSD) was allocated for the first 5 years of the GUEPARD clinical trial and the ESPOIR routine care cohort study that collected the data analysed herein. Two additional grants from INSERM were obtained to support part of the GUEPARD database. The French Society of Rheumatology, Abbott, Pfizer and Roche-Chugal also supported the ESPOIR cohort study.

**Funding:** The analyses reported in this manuscript were supported by internal funding of Dr Pincus.

**Disclosure statement:** M.D. has participated at symposia and advisory boards organized by Pfizer, Abbott, UCB, Roche, BMS and Novartis and his department has received research grants from Pfizer, Abbott, UCB, Roche, BMS and Novartis. All other authors have declared no conflicts of interest.

### Supplementary data

Supplementary data are available at *Rheumatology* Online.

### References

8. Pincus T, Segurado OG. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. *Ann Rheum Dis* 2006;65:820–2.
21. Pincus T, Swearingen CJ, Bergman M *et al*. RAPID3 (routine assessment of patient index data 3), a rheumatoid arthritis index without formal joint counts for routine
