Seminars in Arthritis and Rheumatism ▮ (2018) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Treatments for giant cell arteritis: Meta-analysis and assessment of estimates reliability using the fragility index

Alvise Berti, MD[a,b], Divi Cornec, MD, PhD[b,c], Jose R. Medina Inojosa, MD, MSc[d], Eric L. Matteson, MD, MPH[e,f,*], M. Hassan Murad, MD, MPH[g]

[a] Immunology, Rheumatology, Allergy and Rare Diseases Department, San Raffaele Scientific Institute, Milan, and Santa Chiara Hospital, Trento, Italy
[b] Division of Pulmonary and Critical Care Medicine, Mayo Clinic College of Medicine and Science, Rochester, MN
[c] INSERM UMR1227, Lymphocytes B et Autoimmunité, Université de Bretagne Occidentale, CHU de Brest, Brest, France
[d] Division of Preventive Cardiology, Department of Cardiovascular Medicine, Mayo Clinic 200 First St S.W., Rochester, MN 55905
[e] Division of Rheumatology, Mayo Clinic College of Medicine and Science, Rochester, MN
[f] Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic College of Medicine and Science, Rochester, MN
[g] Evidence-based Practice Center, Mayo Clinic College of Medicine and Science, Rochester, MN

ARTICLE INFO

Keywords:
Fragility index
Giant cell arteritis
Tocilizumab
Randomized control trials
Meta-analysis

ABSTRACT

Background: To better communicate the results of randomized controlled trials (RCTs) of giant cell arteritis (GCA), we propose the use of the fragility index (FI), which is an intuitive measure defined as the minimum number of subjects whose status would have to change (e.g., from having the outcome to not) to render a statistically significant result nonsignificant, or vice-versa.
Methods: We conducted a systematic review and random-effects meta-analysis of RCTs of glucocorticoid (GC) sparing strategies for relapse-free maintenance in GCA, and used the FI to simplify the presentation of results.
Results: Ten RCTs (nine phase II and one phase III enrolling 645 subjects) were included. Tocilizumab, IV GC and methotrexate significantly improved the likelihood of being relapse free with relative risks and 95% confidence intervals of 3.54 (2.28, 5.51), 5.11 (1.39, 18.81) and 1.54 (1.02, 2.30); respectively. The median FI was 4.5 (range, 1–28), and was generally higher for negative RCTs ($n = 6$; median FI 4.5) than for positive RCTs ($n = 4$; median FI 3.5). The range of FI per treatment was (1–8) for methotrexate, (2–6) for anti-TNF agents, 4 for abatacept, 3 for IV GC pulses and (4–28) for tocilizumab.
Conclusion: Tocilizumab, IV GC and methotrexate improve the likelihood of being relapse-free in subjects with GCA. Assessment of GC sparing strategies in GCA has long depended on imprecise trials that would change significance if outcomes were reversed for a handful of subjects. FI may be used in rheumatology to simplify communication of statistical significance and overcome limitations of *p*-value.

Introduction

Giant cell arteritis (GCA) is the most common systemic idiopathic vasculitis, mainly involving the large- and medium-size vessels, particularly the extracranial branches of the carotid arteries [1–3]. Glucocorticoids (GC) are the cornerstone for GCA treatment. More than 50% of patients experience at least 1 severe disease flare, and GC therapy for GCA is often protracted [2,4–6]. The vast majority of these patients develop important side effects related to GC glucocorticoid treatment such as diabetes, hypertension, fractures, and cataracts [5,7,8]. In the last decades, several randomized controlled trials (RCTs) exploring new treatment strategies have been conducted to evaluate strategies for reducing the relapse rate and avoid the potential toxicity of GC [9–20].

To date, several RCTs investigating GC sparing strategies for relapse-free mainteance in GCA have been tested, with disparate and sometimes contradictory results [9–18]. The majority these trials are exploratory phase II trials. However, they have been used both by trialists and clinicians for decision making due to the lack of other evidence. In general, for preventing a new disease flare, trials have been considered as positive (as the anti-interleukin (IL)-6 agent tocilizumab [TCZ] [9] and the anti-cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) agent abatacept [ABA] [11]), negative [anti-tumor necrosis factor (TNF) agents [12–14]], or mixed-potentially positive [intravenous (IV) GC [18], methotrexate (MTX) [15–17]]. Recently, a phase III trial reporting positive efficacy results and favorable safety profile of TCZ has been completed [10].

* Corresponding author at: Divisions of Rheumatology and Epidemiology, Mayo Clinic, College of Medicine and Science, 200 First St S.W., Rochester, MN 55905.
E-mail address: matteson.eric@mayo.edu (E.L. Matteson).

https://doi.org/10.1016/j.semarthrit.2017.12.009
0049-0172/

In determining "positive" vs "negative", researchers and subsequently clinicians have depended on a threshold-based p-value to determine statistical significance, conventionally set at 0.05 [21]. However, this false and artificial dichotomy has been heavily criticized as misleading [22]. The GRADE approach (Grading of Recommendations Assessment, Development, and Evaluation) [23] for assessing certainty in evidence does not depend on p-values at all, and rather emphasizes the importance of several other domains, one of which is precision (i.e., would a decision about efficacy differ if the truth was one boundary of the confidence interval *versus* another) [24]. To simplify the issue of precision, one possible approach that may be more intuitive to stakeholders is the fragility index (FI). FI is defined as the minimum number of subjects in a randomized controlled trial (RCT) whose status would have to change (e.g., from having the outcome to not) to render a statistically significant result nonsignificant, or vice-versa [21]. To our knowledge, FI has never been used in rheumatology trials, and specifically in GCA.

This study aims to systematically review the current evidence on relapse free maintenance for GC sparing strategies in GCA and to evaluate its robustness (precision) using FI. We conjecture that this concept is easy for clinicians (and perhaps also for patients) to understand, and may improve communication and subsequently shared decision making.

Methods

Systematic review

The literature search was defined according to the PICO format (patients, intervention, comparator, outcomes; further detailed in the Supplementary Material Table 1). We used the PRISMA guidelines for conducting and reporting this systematic review [25]. MEDLINE (*Ovid MEDLINE®*), *EMBASE*, *Scopus*, and *Web of Science* were searched from inception through September 1, 2017, without age restrictions. The search strategy included different terms for GCA and treatments, including disease modifying antirheumatic drugs (DMARDs such as MTX, azathioprine), and the following biological agents: adalimumab [ADA], etanercept [ETN], infliximab [IFX], TCZ, ABA, and was restricted to English language articles. Only randomized clinical trials (RCTs) with adequate blinding procedures were included. A hand search through Pubmed for RCTs in GCA was performed and confirmed the abstract selection. Further details of the search strategy are provided in Supplemental Table 1.

Two reviewers independently and in duplicate screened titles and abstracts using broad inclusion criteria. The full-text of all potentially relevant trials was assessed using predefined eligibility criteria. Discrepancies between authors were resolved by discussion. Because remission was defined using slightly different parameters in each study, the outcomes definitions of the authors of the cited RCTs were utilized for each RCT.

Synthesis of results

Meta-analysis was performed if more than one RCT for a given outcome was available and data from these RCTs were sufficiently homogeneous regarding clinical, methodological and statistical characteristics. A random effects model was used. The relative effect of MTX was derived from an individual patient meta-analysis that reported time-to-event outcomes because this was considered a more accurate estimate of relative effect [26]. Certainty about the evidence was assessed using the GRADE approach (Grading of Recommendations Assessment, Development, and Evaluation) [23].

Fragility index

The FI was calculated for each RCT reporting the outcome of being relapse free at the end of the follow-up, which for most of the studies was 52 weeks. When the follow-up was shorter than 52 weeks, or the data could not be extracted at 52 weeks with sufficient certainty, the original study follow-up length was used. The results of each trial were represented in a two-by-two contingency table using the subject sample that the authors employed in their original analysis [21]. For trials with a time to-event-outcome, the number of events in each group for the entire follow-up period was used as events to construct a two-by-two table. FI was calculated by adding an event from the group with the smaller number of events (and subtracting a non-event from the same group to keep the total number of subjects constant) and by recalculating the two-sided *P* value for Fisher's exact test or χ^2, according to the test originally used in each study. Events were iteratively added until the calculated *p*-value became equal to or greater than 0.05 for the studies originally reported as positive, or smaller than 0.05 for the studies initially reported as negative. The number of additional events required was considered the FI for that trial result [21]. All calculations were performed using JMP-SAS version 9.4 (SAS Institute, Cary, NC, USA). The concept of FI using data from one of the included trials [17] is illustrated in Figure 1.

Results

Systematic review

Of the 1645 identified articles, 10 fulfilled the selection criteria (Fig. 2, Supplementary Table 1) [9–18], and all but one were phase II RCTs [10]. In 8 of these 10 trials [9,10,12–14,16–18], the primary study endpoint was the outcome of relapse-free status at the end of a definite period. The number of subjects who relapsed and those who did not at the end of the follow-up could be analyzed for all these 10 RTCs.

A total of 645 subjects with GCA were enrolled in these 10 trials. The majority of the RCTs assessing a GC sparing strategy for relapse-free mainteance in GCA had similar inclusion criteria (mostly American College of Rheumatology 1990 classification criteria [20] and/or biopsy proven GCA) (Supplementary Table 2). The main features of the subject groups and trials included in FI calculations are shown in the Table.

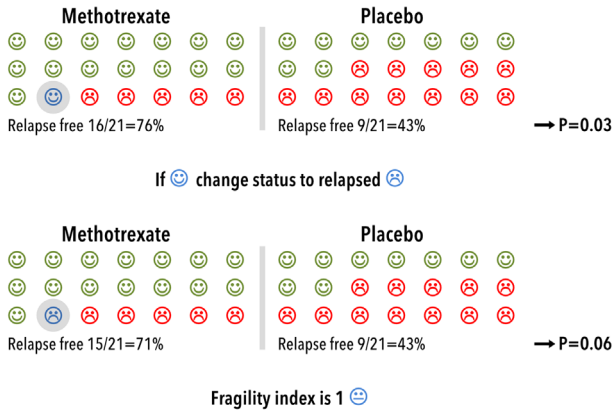

Fig. 1. Illustration of the concept of fragility index using data from one of the included trials. Happy and sad face symbol represents subjects who are relapse free or have relapsed; respectively, at the end of trial follow up. *P* value is 2-sided and calculated using the chi-square test. Statistical significance and FI are calculated following the intention to treat principle.