Review

Viewpoint: Glucocorticoids in the treatment of rheumatoid arthritis: points to (re)consider

Maarten Boers 1, *
1Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands
*Correspondence to: Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit, PO Box 7057, Amsterdam 1007 MB, The Netherlands.
E-mail: eds@amsterdamumc.nl

Abstract
Glucocorticoids (prednisone) are essential in the treatment of RA and other autoimmune diseases. They are widely used, but treatment guidelines advise against. This viewpoint article explains why and suggests a way forward.

Keywords: arthritis, rheumatoid, glucocorticoids, clinical trial, observational study, bias, confounding, guideline

The discovery of glucocorticoids (GCs) and their ‘miracle-like’ effects on RA resulted in the first and only Nobel prize for a rheumatologist, Philip Hench, in 1951. However, the history of their application in RA and other diseases is troubled. Although Hench had warned of the wide spectrum of side effects with uncontrolled use, the drug was released for general use without proper safeguards, with tragic consequences and great disillusionment in the medical community [1]. In the context of RA, this has led to the paradox of continuing widespread use despite the availability of apparent alternatives and guidelines advising against such use.

Evidence of widespread use worldwide comes from a variety of sources. In 2007, a survey over many countries revealed 26–88% of RA patients currently on GCs [2]. More recent data from registries suggest similar or increasing rates, even in the presence of biologics [3]. Similar percentages emerge from RA trials [4]. In a large observational study of routine clinical practice, GC use was shown to be dynamic, with patients continuously cycling on and off treatment [5]. In view of the widespread awareness of the potential risks of GC therapy, it is remarkable to see recent evidence of patients inappropriately treated with GCs before MTX [6]. Another paradox is that so many patients on GCs are not assessed and treated for manageable risks such as cardiovascular disease and osteoporosis and counselled on ways to prevent weight gain at the start of therapy. For instance, in the recent GLORIA trial [7] discussed below, 58% of patients had a history of osteoporosis but only 29% were treated with antiresorptive agents (the trial protocol advised calcium and vitamin D in all patients, but further treatment was left to the treating rheumatologist).

The ACR issued its most recent update of its treatment guidelines for RA in 2021, and EULAR in 2022 [8, 9]. The ACR’s was novel in that it included detailed consideration of GCs in the treatment algorithm for the first time; however, very specific study review questions excluded almost all high-quality evidence (trials) on GCs. In the end, the guideline issued a ‘conditional recommendation’ against short-term use and a ‘strong recommendation’ against long-term use of GCs based on low-quality evidence of harm and the fear that once started, GCs could not be stopped. In contrast, the EULAR has recommended GCs as a bridge therapy for early RA since 2013, but the latest revision has tightened its application to a treatment period of <3 months.

In my opinion, both sets of recommendations are problematic for two reasons: for harm, the undue weighting of observational data compared with data from randomized controlled trials (RCTs), and the influence of expert opinion; and for both harm and benefit, the negatively slanted interpretation of evidence on GCs as compared with evidence on other treatments. In addition, the recommendations have not been (fully) informed by a batch of recent
studies discussed below that have strongly advanced our understanding.

Until recently, one of the problems with assessing the balance of benefit and harm of GCs lay in the quality of evidence. For many years, RCTs with their unbiased comparisons were relatively scarce and mostly assessed early RA, but this is no longer true. In contrast, the number of observational studies is very high and increasing every day. For benefit, most experts now accept the strong evidence that GCs are effective on inflammation and damage, i.e. are ‘disease-modifying’, although they are still named as a class separate from DMARDs. For harm, RCTs have shown no or scarce signals of harm, but in view of their modest sample size, the power to detect rare events is low. However, meta-analyses of trials have shown a similar picture [10, 11]. Observational studies that often show harm of GCs can be much larger, but these are difficult or impossible to interpret due to confounding (bias) by indication [12]. Briefly, this means that there is a strong tendency (reinforced by guidelines) to treat only severe cases with GCs. Severe RA itself causes many of the manifestations also associated with GCs, such as cardiovascular disease, osteoporosis and infections. So RA patients on GCs (also) do worse because they have severe disease, and it becomes impossible to disentangle what is due to disease and what is due to GCs. Many observational studies should simply be discarded because they make no attempt to address this confounding. Increasingly, researchers are trying to address the problem, by recording and correcting (adjusting) for prognostic factors, e.g. by a technique called propensity score matching. However, such techniques are highly dependent on the knowledge, selection and availability of data on the factors that actually drive the choice for GC treatment. In addition, all techniques fail in the face of strong confounding. For example, adjusting for disease activity means comparing the outcome of patients on and off GCs who have the same level of disease activity. But this ignores the fact that a patient on GCs would have higher disease activity if the GCs were stopped, i.e. has more active/severe disease than the patient not on GC; so in the end, the comparison of patients adjusted for disease activity is still confounded.

Arguably, confounding becomes worse as the duration of exposure increases: patients who need GCs for a long period of time have much more serious disease than patients who don’t, so poor outcomes are much more likely in the former. Confounding is thus also a reasonable explanation for the oft-cited claim that GC-related adverse events may accrue slowly over time, also related to the cumulative dose. For example, in the widely quoted DESIR cohort, the increased risk of cardiovascular disease on GC treatment appeared only after lengthy treatment, with observation exceeding 7 years [13, 14]. Confounding by indication is not limited to studies of GCs. For example, a cohort study assessed the ‘real-world’ protective effect of H2-blocking drugs in patients starting NSAIDs. In the crude analysis, gastroprotection was associated with ‘more’ ulcers and bleeds, evidence of strong confounding. Adjustment through propensity score matching had limited effects, finally resulting in only a non-significant association of gastroprotection with fewer ulcers and bleeds [15].

I think for now we should consider most if not all current observational studies that compare adverse events in patients treated or not treated with GCs as (almost) impossible to interpret—in RA, but most likely in all other diseases rheumatologists treat. I do not doubt that high-dose GCs over a long period of time can cause great harm. But to date I have not seen convincing evidence to conclude that long-term, low-dose GCs do more harm than good in RA. On the contrary, the available evidence from RCTs strongly points the other way. Future observational studies in GCs may perform better if they only collect prospective, high-quality data on GC initiators and concurrent non-initiators. Ideally, non-initiators should be patients with an indication for GCs who do not start for reasons unrelated to the disease. Data should include detailed information on GCs and other antimicrobial therapy, including dosing over time and documentation of the motivation for starting and changing doses or drugs; such information should then be incorporated in the analysis to improve adjustment for confounding. The hurdles to achieving these standards are substantial.

Unfortunately, in the guideline committees, the sheer weight of the number of observational studies compared with the RCTs appears to overwhelm methodological reasoning [16]. This is strengthened by systematic reviews where both types of studies undergo quality assessment with different standards but the same words for assessment. Observational studies labelled ‘inconclusive’ should carry much lower weight than RCTs labelled ‘high risk of bias’. For example, in the systematic review that informed the EULAR recommendations [17], the GLORIA study discussed below was deemed to have a high risk of bias for the single fact that co-intervention was allowed. However, this was a key feature of the pragmatic design, to stay as close as possible to routine practice; in addition, the potential for bias was addressed in a predefined analysis. This analysis indeed suggested bias, but in an unexpected direction: benefit of GCs was probably underestimated. To the casual reader such reviews combining RCTs and observational studies suggest that the quality of evidence from both sources is poor, but a great number of observational studies are conclusive for the existence of harm. In discussions, ‘tunnel vision’ is apparent: expert confidence goes up and the caveats progressively disappear, so that even when the review stated that observational studies on a certain harm (e.g. cardiovascular disease in the EULAR systematic review on GCs) [17] show ‘conflicting results’ and generally a high risk of bias, its conclusion ‘confirmed the well-known safety risks’ and the summary of the discussion in the recommendation article stated such evidence was ‘well established’ [9].

Another potential bias lies in the selection of studies for the review, as evidenced by the ACR process. In another example, the EULAR review excluded the long-term follow-up studies [18–22] of the COBRA [23] and BeSt [24] trials, respectively. These studies had shown sustained benefit and no excess harm in the group initially treated with combined treatment compared with the original monotherapy group. The reason for exclusion was that post-trial, treatment was ‘uncontrolled’, i.e. both groups could receive GCs (Bergstra SA, personal communication). This is an astonishing decision, because ‘control’ would mean an observational comparison of GC exposure vs non-exposure, and thus confounding. In contrast, studying outcome in the groups as originally randomized regardless of subsequent exposure to GCs guaranteed an unconfounded comparison of the long-term effects of the original intervention. Because of this decision by the EULAR reviewers, key findings favourable for the 6-month ‘COBRA bridging scheme’ applied in both trials were withheld from the committee’s deliberations.
1. GCs are highly effective DMARDs in early and established RA. GCs decrease inflammation and slow joint damage progression, even at doses as low as prednisolone 5 mg/day.

2. The harm of judicious use of GCs is acceptable and manageable. Sufficient evidence from randomized trials is now available that supports the safety of short-term, high doses and prolonged use of low-moderate doses (oral prednisolone 5–10 mg/day, methylprednisolone intramuscularly 120 mg/month) for up to 2 years. Judicious use requires a high-quality rheumatology clinic setting like the one provided in trials, with proper monitoring and treatment of risk factors such as weight, cardiovascular disease and osteoporosis. To date, observational studies that suggest harm of longer-term use have been almost impossible to interpret due to strong confounding by indication.

3. GCs remain essential as add-on treatment for RA, applied as bridge and as prolonged therapy. To date, RA treatment guidelines have mostly ignored or tried to strongly limit GC use. The field does not adhere to these guidelines because, in practice, no rheumatologist can do without the strong benefits of GCs. This is true regardless of access to biologics.

4. GCs are not difficult to taper and stop. Most patients are able to taper and stop as part of a treatment protocol. Clinically relevant adrenal insufficiency is rare with 3-month tapering.

5. The benefits of GCs decrease if bridging treatment is ≤3 months. As a general rule, drugs (including GCs) only work when taken. However, there is evidence for a ‘window of opportunity’ in early RA, where bridging schedules with at least 6 months of treatment have shown long-term beneficial effects even after GCs are stopped. Three months is probably too short to obtain optimum benefit in early RA.

6. Further high-quality evidence on the balance of benefit and harm of GCs is not likely to be forthcoming soon. Most GC trials have been publicly funded. The low cost of GC therapy makes it an unlikely target for improvement through pharmaceutical innovation. Such innovation is currently the only major driver for large-scale trials. Observational studies that more fully address confounding have yet to appear.

7. Not just for GCs, but for any effective treatment, both the decision to initiate and the decision to stop involves considerations of benefits and harms. For stopping, the benefit is avoidance of future harm, but the harm is the immediate risk of disease recurrence (flare).

Finally, GCs hold a special place in the assessment and interpretation of evidence compared with other antirheumatic drugs. For instance, many physicians strongly overestimate the frequency of GC-associated adverse events, to levels even above the already biased estimates from observational studies [25]. Another example is the recent handling of the evidence from the ORAL surveillance trial on tofacitinib [26]. This large, randomized trial in 4362 high-risk RA patients was set up as a non-inferiority trial to prove tofacitinib was not more harmful than the active comparator, a TNF inhibitor. Unfortunately, tofacitinib proved inferior, i.e. more harmful, with relative risks of 1.33 (95% CI 0.91, 1.94) for major cardiovascular events and 1.48 (95% CI 1.04, 2.09) for malignancy, thus even failing the arguably very wide non-inferiority margin of 1.8. These findings did not prompt guideline committees to recommend that Janus kinase inhibitors should be given ‘only as bridging therapy, for a period no longer than 3 months’ or to ‘conditionally (or strongly) advise against’ their use.

Recent studies on GCs have strongly advanced our understanding of bridging and chronic add-on GC therapy, including the pros and cons of stopping. The double-blind SEMIRA trial showed that RA patients with stable low disease on tocilizumab and prednisone 5 mg/day can taper and stop GCs in 16 weeks, with substantial risk of flare but no signs of adrenal insufficiency [27]. The double-blind GLORIA trial in senior patients with established RA demonstrated that add-on prednisolone 5 mg/day for 2 years is effective on inflammation and damage without undue harm [7]. After successful completion, blinded tapering of GCs was feasible in 3 months, associated with trends of increasing disease activity and flares but no signs of adrenal insufficiency [28]. Further, even the EULAR recommendations review group reported that in RCTs with a bridging schedule, the large majority of patients were able to taper and stop GCs [29, 30]. This suggests that judicious use is possible if such protocols are properly implemented in the clinic. Finally, the recent CORRA trial showed that GC bridging with an initial dose of 60 mg/day, tapered to zero in 3 months, briefly suppressed disease activity but had no effect on damage progression at 1 year compared with MTX monotherapy [31]. In other words, bridging protocols with a maximum period of 3 months as recommended by the EULAR are suboptimal [32].

To sum up, I offer some points for (re)consideration (Table 1). Regardless of the importance one places on the various sources of information on GCs, it is clear that there is often a trade-off between immediate benefit (on disease activity and damage) and the risk of harm, where a potentially increased risk of major clinical events in the long term is most feared. In view of the most recent evidence, I hope guideline committees can turn the page, (re)consider all available evidence and improve guidance on bridging and long-term GC treatment of RA with both benefits and harms in mind.

**Data availability**

No new data are available for this article.

**Authors’ contributions**

M.B. contributed everything.

**Funding**

No specific funding was received from any bodies in the public, commercial or not-for-profit sectors to carry out the work described in this article.

Disclosure statement: The author has received speaker fees from Novartis and Pfizer, provides expert testimony for Celltrion and was the principal investigator of two trials on glucocorticoids: COBRA and GLORIA.
Glucocorticoids in the treatment of RA

References

12. Jacobs JWG, Pereira D. Glucocorticoids are always under suspicion - is the perception of their risks unbiased? J Rheumatol 2018;45:293-6.